

Universidad de Sevilla

*Modelo mixto de Credit Scoring construido con
Análisis discriminante y el Algoritmo de
Kohonen. Valoración de las componentes de
riesgo según Basilea II*

Memoria presentada por Dña Elena M^a Esteve López para alcanzar el grado de doctor
y dirigida por la catedrática de la Universidad de Sevilla Dña M^a José Vázquez Cueto

Febrero 2005

ÍNDICE

AGRADECIMIENTOS	4
INTRODUCCIÓN	5
Bibliografía	16

CAPÍTULO I **19**

EL NUEVO ACUERDO DE BASILEA **19**

1.1 HISTORIA DEL CREDIT SCORING	20
1.2 EL NUEVO ACUERDO DE CAPITAL DE BASILEA	22
1.3 MÉTODOS DE CÁLCULO DE RIESGO PARA EXPOSICIONES AL DETALLE	26
1.3.1 EL MÉTODO IRB	26
1.3.2 COMPONENTES DE RIESGO: TASA DE MOROSIDAD, SEVERIDAD Y PÉRDIDA ANTICIPADA	28
1.3.3 FÓRMULA PARA LA DERIVACIÓN DE PONDERACIONES DEL RIESGO	31
1.3.4 LOS REQUISITOS MÍNIMOS PARA LAS EXPOSICIONES AL DETALLE	35
1.3.5 EL MÉTODO ESTÁNDAR	59
Bibliografía	61

CAPÍTULO II **64**

UNA REVISIÓN DE LOS MÉTODOS ESTADÍSTICOS EMPLEADOS EN LA CONSTRUCCIÓN DE CLASIFICACIONES DE CRÉDITO **64**

2.1 INTRODUCCIÓN	64
2.2 ANÁLISIS DISCRIMINANTE: UNA PERSPECTIVA DESDE LA TEORÍA DE LA DECISIÓN	66
2.2.1 EL CASO DE UNA NORMAL UNIVARIANTE	73

2.2.2 EL CASO DE UNA NORMAL MULTIVARIANTE CON COVARIANZA COMÚN	74
2.2.3 EL CASO DE UNA NORMAL MULTIVARIANTE CON DIFERENTES MATRICES DE COVARIANZAS	76
2.3 ANÁLISIS DISCRIMINANTE: SEPARACIÓN EN DOS GRUPOS	77
2.4 ANÁLISIS DISCRIMINANTE: UNA FORMA DE REGRESIÓN LINEAL	79
2.5 REGRESIÓN LOGÍSTICA	84
2.6 OTROS MÉTODOS DE REGRESIÓN NO LINEAL	88
2.7 CLASIFICACIÓN EN ÁRBOL (MÉTODO DE PARTICIÓN RECURSIVA)	90
2.7.1 MEDIDAS DE DIVISIÓN DEL MÉTODO DEL ÁRBOL	93
2.8 EL MÉTODO DE VECINDAD	98
2.9 EL MÉTODO DE LA DISCRIMINACIÓN MÚLTIPLE	105
Bibliografía	111

CAPÍTULO III **113**

UNA REVISIÓN SOBRE ALGUNOS MÉTODOS NO ESTADÍSTICOS EMPLEADOS EN LA CONSTRUCCIÓN DE CLASIFICACIONES DE CRÉDITO **113**

3.1 INTRODUCCIÓN	113
3.2 PROGRAMACIÓN LINEAL	115
3.3 PROGRAMACIÓN ENTERA	122
3.4 REDES NEURONALES	124
3.4.1 ELEMENTOS Y ARQUITECTURA DE UNA RED NEURONAL ARTIFICIAL	125
3.4.2 MECANISMO DE APRENDIZAJE	130
3.4.3 REDES CON APRENDIZAJE SUPERVISADO	132
3.4.4 REDES CON APRENDIZAJE NO SUPERVISADO	137
3.5 ALGORITMOS GENÉTICOS	140
3.5.1 INTRODUCCIÓN	140
3.5.2 EL ALGORITMO GENÉTICO SIMPLE	143

3.5.3 CODIFICACIÓN	145
3.5.4 PATRONES	149
Bibliografía	153

CAPÍTULO IV **159**

CONSTRUCCIÓN DE UN MODELO MIXTO DE CREDIT SCORING **159**

4.1 INTRODUCCIÓN	159
4.2 MODELO MIXTO DE CREDIT SCORING	162
4.2.1 MODELO DE ANÁLISIS DISCRIMINANTE	162
4.2.2 ELABORACIÓN DEL MODELO	168
4.2.3 ALGORITMO DE KOHONEN	172
4.3 CÁLCULO DE LAS COMPONENTES DE RIESGO	182
4.4 CONCLUSIONES	185
Bibliografía	187

CAPÍTULO V **191**

APLICACIÓN PRÁCTICA DE CREDIT SCORING MIXTO **191**

5.1 INTRODUCCIÓN	191
5.2 ANÁLISIS DE LOS RESULTADOS EMPÍRICOS	193
5.3 CONCLUSIONES	215
Bibliografía	219

AGRADECIMIENTOS

A través de estas líneas quiero manifestar mi más sincero agradecimiento a todas aquellas personas que han colaborado para que esta tesis llegase a su término.

En primer lugar, a mi directora la Dra. Dña. M^a Josefa Vázquez Cueto que me ha brindado tanto sus excelentes conocimientos como su apoyo académico en todo momento, pero sobre todo su ánimo y confianza en este trabajo y ante todo quiero reconocer en estas líneas que sin ella esta tesis nunca hubiera visto la luz. ¿Cómo se puede reflejar en palabras y sobre papel mi gratitud de corazón...?

Agradecer a todos los compañeros que confiaron en mí y en la posibilidad de que esta tesis saliera adelante en momentos de desfallecimiento y desconfianza por mi parte, en especial al Dr. D. Miguel Ors Montenegro, por horas y horas de fructíferas conversaciones y al director de la Universidad Cardenal Herrera-CEU, D. Francisco Sánchez Martínez por hacerme sentir su mano en mi hombro en los peores momentos.

Pero sobre todo quiero agradecer la infinita paciencia y el apoyo sin medida que he recibido de mi marido, y la fuerza de mis hijos, que como motor de mi vida que son, han conseguido que esta tesis esté finalizada.

Mamá, gracias por todo.

INTRODUCCIÓN

Cuando un prestamista concede un crédito asume un riesgo que consiste en la probabilidad de que el prestatario incurra en un impago de la deuda ya sea total o parcial.

Este hecho es el que nos conduce a preguntarnos cómo podemos valorar este riesgo de impago para, de esta manera, conocido el valor del riesgo, poder establecer una cobertura del capital implícito en estas circunstancias.

Podemos encontrarnos en dos situaciones que a nuestro entender deben ser diferenciadas:

1. Los Prestatarios son grandes empresas o instituciones y los prestamistas pueden ser grandes empresas o agentes económicos individualizados (pequeña y mediana empresa y consumidores).
2. Los Prestatarios son agentes económicos individualizados y los prestamistas generalmente son entidades bancarias.

En el primer caso las empresas o entidades con necesidad de liquidez emiten activos financieros y en esta emisión se comprometen a devolver el capital más intereses a quienes han adquirido estos activos.

Estos prestamistas pueden incurrir en un riesgo de impago ya que dependen de la solvencia de la empresa emisora durante la vida del préstamo.

Introducción

Para estudiar la valoración del riesgo de estas emisiones el primer paso es elaborar un rating¹ de estas grandes empresas o instituciones, aún sin la necesidad de que todas las empresas incluidas en el rating sean emisoras de deuda en ese momento.

Normalmente se utiliza el vocablo inglés rating en lugar de su vocablo castellano clasificación puesto que las primeras empresas que decidieron realizar una clasificación de entidades financieras fueron empresas estadounidenses, y por lo tanto utilizaban el término rating que después se fue incorporando en las empresas europeas y del resto del mundo.

La razón por la que surgen las empresas de clasificación la encontramos en la emisión de bonos que realizaron las empresas de ferrocarriles en Estados Unidos para financiar su proyecto [10]. En ese momento las empresas de clasificación se dedican a estudiar la mayor o menor solvencia de las diferentes empresas ferroviarias para ver si estas serían capaces o no de responder después de su deuda.

A partir de aquí, las empresas clasificadoras extienden su cometido a todo tipo de empresa siendo una garantía de solvencia la emisión de deuda por parte de aquellas entidades que ocupan buen lugar en el rating elaborado por estas empresas de clasificación.

Las instituciones más importantes que se dedican a la realización de ratings son Moody's y Standard & Poor.

Ambas cuentan con más de cien años de experiencia en este terreno, son estadounidenses pero han extendido su mercado por el resto del mundo, por ejemplo, en España, Moody's se ha establecido con su propio nombre, sin embargo S&P se ha establecido con la empresa Iberating.

Como decíamos al inicio de esta sección nos podemos encontrar con un segundo caso donde los prestamistas son entidades bancarias y los

¹ Clasificación

prestatarios son normalmente pequeñas y medianas empresas y consumidores.

Aquí, las entidades bancarias también confeccionan una clasificación para determinar el valor del riesgo en el que incurren cuando conceden un crédito. La clasificación elaborada por las entidades bancarias se denomina Credit Scoring.

Un Credit Scoring es un método estadístico usado para predecir la probabilidad de impago de un préstamo.

Utilizando datos históricos y técnicas estadísticas, el credit scoring trata de aislar los efectos de varias características de clientes en el impago de los créditos. El método produce una "calificación" que el banco puede utilizar para clasificar un préstamo en términos de riesgo.

Mientras la historia de los créditos se remonta a 5000 años atrás en la antigua Babilonia, la historia de los credit scoring tiene solamente 50 años de antigüedad. Un credit scoring es esencialmente un método para diferenciar grupos de población cuando no se pueden ver claramente las características que definen los diferentes grupos. Si el lector tiene interés por conocer la evolución histórica del credit scoring puede encontrarla en la sección 1 del capítulo primero de este trabajo.

En la actualidad, el énfasis en la elaboración de un credit scoring radica en el cambio de objetivos para tratar de minimizar el riesgo de impago de un cliente en un producto en particular y al mismo tiempo buscar para la entidad bancaria el máximo beneficio que puede conseguir con este cliente.

Como ya hemos indicado un credit scoring es un método de evaluación del riesgo cuando se solicita un préstamo.

Para construir un modelo de medición se tienen que analizar los datos históricos correspondientes a las características de los clientes a los que

se les concedió un préstamo y determinar las que se utilizan para indicar que el préstamo fue bien concedido.

Un modelo de credit scoring bien construido debería dar un alto porcentaje de altas calificaciones para los clientes que pagaron su préstamo y un alto porcentaje de bajas calificaciones para los clientes que no pagaron su préstamo o viceversa.

Pero el modelo así planteado no es perfecto, y algunos malos clientes reciben calificaciones tan altas (bajas) como los buenos clientes.

Los datos que se suelen tener en cuenta, según [10] son:

- 1.Renta mensual
- 2.Deudas
- 3.Activos financieros
- 4.Tiempo de permanencia en el trabajo
- 5.Comportamiento de pago o impago en otros préstamos
- 6.Cuantía destinada al pago de vivienda (alquiler o hipoteca)

y son estos datos con los que el banco puede proceder a dar una calificación a su cliente.

Nuestro trabajo estará enfocado hacia la creación de Credit Scoring para entidades bancarias, no somos nosotros quienes concedemos importancia a este tema, es el propio Comité de Supervisión Bancaria quien atribuye la relevancia que le corresponde al realizar un Nuevo Acuerdo de Basilea.

El Comité realiza el llamado Nuevo Acuerdo de Capital de Basilea que reemplaza al realizado en 1988, el nuevo acuerdo es más extenso y

elaborado pero sigue reafirmando los tres pilares en los que se basaba su antecesor, estos son:

1. Requisitos de capital mínimo
2. Proceso de examen supervisor
3. Disciplina de mercado

El Pilar 1 abarca los requisitos de capital por riesgo de mercado, riesgo crediticio y riesgo operativo. Para mejorar la sensibilidad al riesgo, el Comité propone una serie de opciones para tratar tanto el riesgo crediticio como operativo. El Comité ha decidido tratar el riesgo de tipo de interés en la cartera de inversión en el Pilar 2. Dada la cantidad de supuestos subyacentes necesarios, el Comité opina que el proceso de examen supervisor dará lugar a un tratamiento mejor y más sensible al riesgo que los requisitos de capital mínimo.

En el capítulo 1 de esta tesis se desarrollan los requisitos sobre valoración y cobertura de riesgo que deben cumplir los bancos para poder aplicar el método IRB (Método basado en ratios internos). Cuando un banco tenga la potestad de poder aplicar este método significará que el comité supervisor ha llegado a la conclusión de que dicho banco cumple todos los requisitos de cobertura de riesgo ante posibles impagos.

De forma general, para que un modelo interno sea considerado IRB, debe reunir unos requisitos imprescindibles que son:

1. Los modelos deben ofrecer:
 - El incentivo de perfeccionar sistemas y los resultados obtenidos de ellos a partir de los datos de partida.
 - Una mejor información en conjunto de límites y reservas
 - Más precisión en la valoración del riesgo y la estructura

utilizada para ello, lo que debe contribuir a un proceso más transparente de la toma de decisiones; y

- Unas bases más consistentes para la asignación del capital económico

2. Desde una perspectiva supervisora, el desarrollo de la metodología del modelo debe llevar, consecuentemente, el perfeccionamiento en el rigor y consistencia del proceso de control del riesgo.

El Nuevo Acuerdo se plantea la implantación del método IRB para subsanar una serie de deficiencias presentadas por los modelos que los bancos utilizan en la actualidad, como por ejemplo:

1. Los bancos han indicado el uso de grandes intervalos de confianza en la medida del riesgo de crédito. No está claro si tales intervalos de confianza pueden ser estimados con una precisión razonable, y todavía no se comprende bien la utilización de supuestos en modelos que fallan en los extremos de la distribución arrastrando en el error al montante de capital necesario para soportar el riesgo obtenido. Consecuentemente, es susceptible de ser cuestionado el hecho de que el uso de grandes intervalos de confianza producirían requerimientos de capital que son altamente dependientes del modelo planteado, o no son compatibles con los de otras instituciones. De ahí que se siga dando importancia a la validación tanto interna como externa.
2. En la actualidad, no es comúnmente aceptada una estructura de verificación periódica de la precisión de los modelos de riesgo de crédito; yendo más lejos, los métodos que verifican una cierta sensibilidad son verosímiles para jugar un papel importante en

este proceso. Finalmente, es importante advertir que el ambiente interno en el que opera el modelo (la cantidad de fallos en el control, la calidad de los controles internos, el rigor de las pruebas, etc.) continuará siendo una parte clave en la evaluación de la estructura de control de riesgo de los bancos.

En el segundo capítulo se elabora un resumen de métodos estadísticos empleados hasta la actualidad para la construcción de credit scoring.

La aplicación de análisis discriminante a todas las variables que hemos indicado anteriormente como importantes en la elaboración de estos modelos nos permite decidir que combinación de estos factores o datos es la mejor para predecir el impago del préstamo y cuál es el peso que tiene cada característica en el impago del crédito.

Según Fair, [5] se pueden considerar 50 o 60 variables para construir un modelo típico de credit scoring pero defiende que la combinación que mejor predice la calificación del cliente oscila entre las ocho a doce variables. Por otro lado, Saunders [16] defiende el uso de 48 factores para evaluar la probabilidad de impago de un crédito.

En la mayoría de los sistemas de calificación (no en todos), una alta calificación indica poco riesgo, y por lo tanto habrá una calificación de corte que determina un conjunto de clientes a los que se concederá el crédito porque el riesgo de la concesión es menor.

Aplicando estrictamente el modelo pertinente, el banco debería aprobar aquellos préstamos cuya calificación quedara por encima de la calificación de corte y denegar aquellos que quedaran por debajo de la misma. Habría que estudiar con detenimiento aquellos que no superan o no alcanzan la nota de corte de forma holgada.

Incluso un buen sistema de credit scoring no predice con total certidumbre lo que va a ocurrir con cualquier préstamo individual, pero sí

proporciona una predicción bastante precisa de la probabilidad de que la concesión de un crédito bajo unas características pueda resultar impagado.

Para construir un buen modelo de credit scoring se necesitan suficientes datos históricos, que reflejen el desarrollo de los créditos tanto en periodos en los que las condiciones económicas han sido buenas como en los periodos en los cuales las condiciones económicas han sido malas².

En el artículo de Roszbach [15] podemos ver la aplicación de un modelo Tobit para la obtención de un credit scoring, donde la autora demuestra que la aplicación de este modelo permite encontrar una resolución mejor para el problema que empleando un modelo probit.

También se emplean métodos no estadísticos en la construcción de credit scoring, entre los que destacamos las redes neuronales, y desarrollamos un resumen de los mismos en el capítulo 3.

Los métodos no estadísticos son menos exigentes con las características que deben cumplir las variables del modelo que los métodos estadísticos, por lo tanto, tienen el potencial para ser utilizados en el desarrollo de modelos que son más heterogéneos como préstamos a empresas.

Las redes neuronales son algoritmos de inteligencia artificial que permiten desde el aprendizaje a través de la experiencia discernir la relación entre las características del cliente y la probabilidad de impago y determinar qué características son más importantes en la predicción del impago [12] y [1] (ver artículos de Malhotra y coautores y de Edward Altman).

² Patrick McAllister y John Mingo estiman que para desarrollar un modelo de predicción para préstamos comerciales, se necesitan de 20000 a 30000 aplicaciones.

Este método es más flexible que las técnicas estadísticas ya que no establece ningún supuesto sobre la forma de la función de la relación entre las características del cliente y la probabilidad de impago o sobre la distribución de las variables o los errores del modelo, ni sobre las correlaciones entre las características.

Algunos argumentan que las redes neuronales se muestran mucho más prometedoras para un credit scoring para préstamos comerciales, pero otros han argumentado que el modelo es más ad hoc que los métodos estadísticos tradicionales.

En el artículo de Altman y Saunders [1,16] concluyen que las redes neuronales no son claramente mejores que los métodos estándar, pero sugieren usar ambos tipos de métodos en ciertas aplicaciones, especialmente en las complejas en las que la flexibilidad de las redes neuronales tendrían un valor particular.

Pero conociendo el funcionamiento de las redes neuronales, es importante tener presente que la aplicación de este método no se reduce únicamente a su entrenamiento con los datos disponibles: uno de los principales problemas es escoger previamente las variables más relevantes, ya que un exceso de variables puede introducir ruido que oculte las más significativas, mientras que su defecto puede provocar una falta de información [4].

La estructura característica de una red neuronal es que la neuronas están agrupadas en capas. Existe una primera capa por la que entra la información, una o varias capas ocultas que la procesan, y una capa de salida que proporciona los resultados de la red.

Para nuestro trabajo nos encontramos con el problema de no poder utilizar una estructura de capas por varias razones:

1. Si utilizamos todas las características del prestatario como datos de información de entrada, la red neuronal concede a las mismas igual importancia, y este es un supuesto inicial que no podemos asumir en la construcción de nuestro credit scoring, y como ya hemos indicado antes hay que proceder a una selección de las más significativas.
2. Los resultados de salida de la red utilizando como entrada todas las características no tienen solución única, significa esto, que de antemano ya sabemos que incurrimos en un error, mínimo pero no medible ya que cada vez que diéramos solución al problema obtendríamos una solución diferente, siempre cercana a la primera, pero nunca podemos decir que los datos de entrada convergen a una única solución de salida.

Sabiendo las limitaciones del modelo de redes neuronales, y siguiendo la sugerencia de Saunder y Altman [1] que llegan a la conclusión de que los resultados más fieles en modelos de credit scoring se obtienen mediante la combinación de redes neuronales con un modelo estadístico tradicional decidimos construir un modelo que combine ambos métodos.

En la literatura consultada nos llama la atención que la mayoría de las aplicaciones realizadas para el estudio de solvencia económica de empresas lo que hacen los autores es establecer la comparación de resultados entre un modelo estadístico y un modelo de red neuronal, siendo normalmente superior el resultado obtenido por la red neuronal como el trabajo realizado por Surkan y Singleton (1990) [18] donde comparan la solución al problema proporcionada por un perceptrón multicapa y el análisis discriminante, o el de Martínez de Lejarda (1996) [11] que utiliza los mismos modelos aplicados al sector de seguros español.

Con estos argumentos la propuesta de este trabajo, que se desarrolla en el capítulo 4, consiste en la creación de un modelo de credit scoring utilizando las dos técnicas conjuntamente, redes neuronales y métodos estadísticos, concretamente, el Algoritmo de Kohonen y el análisis discriminante.

En cuanto a la aplicación del Algoritmo de Kohonen en este tema encontramos que ya ha sido empleado en alguna ocasión, lo que nos reafirma en la decisión de su empleo en este trabajo, dos ejemplos a mencionar son: el trabajo de Serrano Cinca y Martín del Brio (1993) [17] que analizan la crisis bancaria española durante los años 80 y el de López González y Flórez López (1999) [9] que caracterizan la solvencia empresarial.

El objetivo de este capítulo es desarrollar un modelo de Credit Scoring que, además de cumplir los objetivos tradicionales de calificación y clasificación de los clientes para poder tomar una decisión de conceder o no un préstamo solicitado, incorpore todos aquellos requisitos necesarios para ser considerado un modelo interno para la valoración de riesgo, como es requerido por Basilea II.

En el capítulo 5 se elabora una aplicación práctica del modelo propuesto que conjunta la elaboración de un credit scoring para préstamos de pequeña cuantía, con el cumplimiento de la normativa del Nuevo acuerdo de Basilea II referido a las exposiciones al detalle. Para su desarrollo ha sido necesario el empleo de tres programas informáticos, Matlab, SPSS, y Excel.

La mayor dificultad para llevar a cabo este capítulo no ha sido tanto la necesidad del empleo del software informático como el disponer de los datos necesarios para la elaboración del Credit Scoring debido a la

negativa de las entidades bancarias a conceder datos anónimos para poder trabajar con ellos.

BIBLIOGRAFÍA

[1] : Altman, Edward I., Giancarlo Marco y Franco Varetto. Corporate Distress Diagnosis: Comparisons Using Linear Discriminant Analysis and Neural Networks (The Italian Experience). *Journal of Banking and Finance* 18, pp.505-29 (1994).

[2] : Altman, Edward I , Anthony Saunders. Credit Risk Measurement: Developments Over the last 20 Years. *Journal of Banking and Finance*, forthcoming (New York University Salomon Center Working Paper S-96-40) (1997).

[3] : BBVA. Basilea II: Impacto en las entidades financieras. Madrid (2001).

[4] : Bosch, J., Garrido, L., Gómez, S., Predicción de índices futuros financieros mediante redes neuronales. Universidad de Barcelona. (1997).

[5] : Fair, Isaac. Low to Moderate Income and High Minority Area Case Studies Fair, Isaac and Company, INC. Discussion Paper (October 4, 1996).

[6] : Guillén M. y Arís M., Count Data Models for a credit scoring system. Universidad de Barcelona (1992).

Bibliografía

[7] : Jost, A. Data Mining" "Credit Risk Modeling: Design and Application Elizabeth Mays Editor. (1998).

[8] : Killian M.T., Credit Scoring. Forum on Credit Scoring. (2000).

[9] : López González, E., Flórez López, R. El análisis de solvencia empresarial utilizando redes neuronales autoasociativas Proceedings of the IV International Meeting on Advances in Computational Management, Reus (1999).

[10] : López Pascual J., El rating y las agencias de calificación. Dykinson. (1996).

[11] : Martínez de Lejarza, I. Forecasting company failure: neural approach versus discriminant analysis. an application to spanish insurance companies of the 80's. Intelligent systems in Accounting and finance, Huelva, pp. 169-186.(1996).

[12] : Malhotra, D.K., Rashmi Malhotra, y Robert W. McLeod. Artificial Neural Systems in Commercial Lending The Bankers Magazine, pp.40-44 (noviembre/ diciembre 1994).

[13] : McAllister, Patrick H.,y John J. Mingo. Commercial Loan Risk Management, Credit-Scoring, and Pricing: The Need for a New Shared Database. Journal of Commercial Lending, pp. 6-22 (mayo 1994).

[14] : Mester, L.J., What's the point of Credit Scoring?, Business Review, Septiembre/Octubre 1997, Federal Reserve Bank oh Philadelphia (1997).

Bibliografía

[15] : Roszbach, K., Bank lending policy, credit scoring and the survival of loans, Escuela de Economía de Estocolmo, (1998).

[16] : Saunders, Anthony, Financial Institutions Management: A Modern Perspective, 2nd edition, Chapter 10, Boston: Irwin: Boston, (1997).

[17] : Serrano Cinca, C., Martín del Brio, B. Predicción de la quiebra bancaria mediante el empleo de redes neuronales artificiales Revista española de Financiación y Contabilidad Vol.22, nº 74, pp.153-176.(1993).

[18] : Surkan, A., Singleton, J. Neural networks for bond rating improved by multiple hidden layers Neural networks in Finance and Investing, Probus Publishing, Chicago (1992).

Capítulo I

EL NUEVO ACUERDO DE BASILEA

En este primer capítulo, en su primera sección hacemos una incursión por la historia del Credit Scoring hasta nuestros días que nos conduce a enlazar con la siguiente sección donde desarrollamos las principales propuestas de Basilea II sobre requisitos de capital mínimo y las normas generales para la medición de riesgo de las operaciones realizadas por las entidades bancarias, entre las que destaca la incorporación de un método interno para la medición de dicho riesgo. Después, en la sección tres del capítulo nos centraremos en las exposiciones al detalle o pequeños préstamos, y será aquí donde se definan todos y cada uno de los elementos necesarios para que una entidad bancaria pueda desarrollar un método interno que permita minimizar el riesgo de capital en estas exposiciones al detalle.

1.1 Historia del Credit Scoring

Mientras la historia de los créditos se remonta a 5000 años atrás en la antigua Babilonia, la historia de los credit scoring tiene solamente 50 años de antigüedad. Un credit scoring es esencialmente un método para diferenciar grupos de población cuando no se pueden ver claramente las características que definen los diferentes grupos. El primer método para resolver este problema de identificación de grupos en una población fue introducido en estadística por Fisher [6](1936). Lo utilizó para resolver un problema de biología. En 1941, Durand [5] fue el primero en reconocer que se pueden utilizar las mismas técnicas para discriminar entre clientes solventes e insolventes frente a un préstamo, este trabajo lo desarrolló como un proyecto de investigación para el U.S. National Bureau of Economic Research pero no estaba enfocado con una intención predictiva, solo de clasificación.

En la década de los 30, algunas empresas de ventas por correo contrataron analistas de créditos que introdujeron sistemas de calificación numérica para tratar de corregir las inconsistencias en las decisiones tomadas, esto lo podemos encontrar en Weingartner [15] (1966) y Smalley y Sturdivant [12] (1973). Con el comienzo de la Segunda Guerra Mundial, estas empresas de ventas por correo y otras dedicadas a las pequeñas finanzas comenzaron a experimentar dificultades con los préstamos concedidos ya que los analistas crediticios se alistaban en el servicio militar, y por lo tanto era difícil encontrar personas con estos conocimientos. En esta época los análisis para decidir a quien se concedía un crédito se realizaban a mano y consistían por un lado en la introducción de sistemas de calificación numérica y por otro lado en describir un conjunto de condiciones que el cliente debía cumplir.

Historia del Credit Scoring

No mucho tiempo después de finalizada la guerra se comenzaron a ver los beneficios de la utilización de modelos estadísticos en decisiones de préstamos. La primera empresa en elaborar un credit scoring fue creada por Bill Fair y Earl Isaac en San Francisco a principios de la década de los 50, y sus clientes eran principalmente empresas de ventas por correo y las denominadas casas de finanzas.

La llegada de las tarjetas de crédito a finales de los años 60, hace que las entidades bancarias y otras empresas emisoras de tarjetas se planteen la utilidad de los credit scoring, pero el número de personas a las que diariamente se les adjudicaba una tarjeta de crédito era tan alto que era imposible en términos económicos y de mano de obra aplicar un método de decisión de crédito, y es solo el desarrollo de los ordenadores lo que hace factible esta aplicación. Las entidades que deciden aplicar credit scoring buscan que estos sean el mejor método de predicción y que los ratios de impago disminuyan al menos en un 50% ó más (Myers y Forgy [10] (1963) y Churchill [3] (1977)). Sin embargo, Capon [2] (1982) argumentaba que la fuerza empírica bruta del credit scoring viola las tradiciones de la sociedad americana, este autor pensaba que debería haber más dependencia de la historia del crédito y esto posibilitaría explicar porqué ciertas características son necesarias en un sistema de clasificación y otras no. El evento que reafirmó la completa aceptación del credit scoring fue el artículo Actos de igual oportunidad de crédito y su enmienda a la Constitución de Estados Unidos en 1975 y 1976 que de él se derivó. Esto implica que la discriminación en la concesión de un crédito es ilegal a menos que esta esté empíricamente demostrada y sea estadísticamente válida.

En la década de los 80, el éxito de la aplicación de credit scoring en la concesión de tarjetas de crédito hizo que los bancos comenzaran a utilizarlos para otros productos, como préstamos personales y en los

Capítulo 1. El Nuevo Acuerdo de Basilea

últimos años de esta década para préstamos hipotecarios y pequeñas empresas. En los 90, se comenzó a aplicar en marketing para la clasificación del ratio de respuestas a las campañas publicitarias. De hecho, este fue uno de los primeros usos en los años 50, cuando Sears utilizó un scoring para decidir a quien enviaba sus catálogos [9].

En la actualidad, el énfasis en la elaboración de credit scoring radica en el cambio de objetivos para tratar de minimizar el riesgo de impago de un cliente en un producto en particular y al mismo tiempo buscar para la entidad bancaria el máximo beneficio que puede conseguir con este cliente.

1.2 El Nuevo Acuerdo de Capital de Basilea

El Comité de Supervisión Bancaria de Basilea está compuesto por los gobernadores de los bancos centrales del Grupo de los Diez en 1975. Sus miembros son altos representantes de la supervisión bancaria y bancos centrales de Bélgica, Canadá, Francia, Alemania, Italia, Japón, Luxemburgo, Holanda, Suecia, Suiza, el Reino Unido y Estados Unidos.

En 1999 se comenzó a trabajar en el llamado Nuevo Acuerdo de Capital de Basilea [7] que reemplazaría al realizado en 1988, el objetivo inicial era que estuviera vigente en el año 2004, pero las más de 200 rectificaciones recibidas ha provocado que se dilate en el tiempo su puesta en funcionamiento, siendo la última fecha propuesta finales del 2006.

El Nuevo Acuerdo es más extenso y elaborado, siendo su aportación fundamental la incorporación de los métodos internos para poder calcular el mínimo riesgo de sus exposiciones de capital aunque sigue reafirmando los tres pilares en los que se basaba su antecesor, estos son:

El Nuevo Acuerdo de Capital de Basilea

- 1.-Requisitos de capital mínimo.
- 2.-Proceso de examen supervisor.
- 3.-Disciplina de mercado.

Este primer pilar abarca los requisitos de capital por riesgo de mercado, riesgo crediticio y riesgo operativo. Para mejorar la sensibilidad al riesgo, se propone una serie de opciones para tratar tanto el riesgo crediticio como operativo. Así mismo, decide tratar el riesgo de tipo de interés en la cartera de inversión en el punto dos. Dada la cantidad de supuestos subyacentes necesarios, el proceso de examen supervisor dará lugar a un tratamiento mejor y más sensible al riesgo que los requisitos de capital mínimo.

Es en la primera sección del Nuevo Acuerdo cuando se refiere al riesgo crediticio concretamente, donde el Comité presenta los métodos y los requisitos asociados a ellos, para calcular los activos ponderados por riesgo con arreglo a los métodos estándar y al método basado en ratios internos (Internal ratings-Based) que puede ser de dos clases, básico y avanzado, que llamaremos IRB a partir de ahora.

En su propuesta original, se planteó que solo algunos bancos más desarrollados utilizarían sus evaluaciones internas del riesgo crediticio para fijar exigencias de capital, sin embargo, después de un estudio más profundo, se cree que las normas mínimas de elegibilidad del método IRB están en realidad al alcance de una gama más amplia de bancos. En general, Basilea II propone que los requisitos mínimos de capital crecerán si todos los bancos adoptan el enfoque estandarizado y disminuirán si eligen el avanzado, pero las divergencias sobre esta creencia son muchas, nos encontramos con países como Estados Unidos que se ha opuesto a la implantación del Nuevo Acuerdo, y países como

Capítulo 1. El Nuevo Acuerdo de Basilea

Alemania también se muestran reticentes ya que piensan que hay un sesgo hacia lo grande frente a lo pequeño y esto afecta directamente a las pymes, importantísimas para este país. Pensamos que aún cumpliéndose lo que pronostica Basilea II, algunas entidades se iban a ver forzadas a reducir el volumen de créditos concedidos y esto, sin duda trasciende de los bancos a la sociedad, de ahí las reticencias de algunos países, incluso del Parlamento Europeo, a su incorporación.

Si una entidad bancaria decidiera aplicar métodos internos de cálculo de riesgo, las exposiciones empresariales, bancarias y soberanas³ recibirían un tratamiento similar, sin embargo, su utilización para las exposiciones al detalle sería diferente. Para cada categoría de exposición, el tratamiento de implantación de métodos internos se basaría en tres elementos:

1. Componentes del riesgo. La entidad financiera puede utilizar sus propias estimaciones o las estimaciones supervisoras estándar.
2. Utilización de una función de ponderación del riesgo. Esta convierte los componentes del riesgo en ponderaciones de riesgo a ser utilizadas por los bancos para calcular el capital mínimo.
3. Establecimiento de los requisitos mínimos que un banco debe cumplir para poder aplicar el método IRB.

Sin estas condiciones, no sería posible confiar en las estimaciones internas de los bancos. El cumplimiento de estos requisitos es una inversión que los bancos deberían hacer para aprovechar la mayor

³ El vocablo "soberano" incluye estados soberanos y bancos centrales tratados como estados soberanos por el supervisor nacional

El Nuevo Acuerdo de Capital de Basilea

sensibilidad al riesgo que ofrece el método IRB, y poder minimizar así el capital de exposición al riesgo. Según la consultora Mercer Oliver Wyman la entrada en vigor de Basilea II en 2006 le va a suponer a la banca española entre 500 y 600 millones de euros para poder cumplir todos los requisitos que impone este acuerdo, estas exigencias van a suponer importantes inversiones, principalmente, en herramientas informáticas (entre el 35% y 45%) y cambios en contabilidad y gestión. Para el resto del mundo, el mismo informe afirma que la banca tendrá que invertir 21.350 millones.

También podría ocurrir que la implementación del método interno constituya un desafío para algunos supervisores en vista de la importancia que cobran en él la especificidad de la validación y el examen supervisor, esto significa que no es suficiente que la entidad desarrolle internamente, o adquiriera un sistema de medición de riesgos, por muy adecuado que este sea, tenga los datos necesarios y vigile su funcionamiento. Será necesario que pase la prueba de uso de cada entidad, y esto significa que el método debe formar parte de la arquitectura y de la cultura de la entidad. Así mismo, se alienta a las autoridades nacionales a tomar medidas para implementar estos mecanismos, teniendo en cuenta que el diálogo e intercambio de información entre supervisores, tanto bilaterales como multilaterales, serían parte integral de este proceso.

El trabajo del Comité en el campo de las exposiciones empresariales, bancarias y soberanas es el más elaborado pero también presenta propuestas específicas para las exposiciones al detalle y es en estas últimas donde vamos a hacer hincapié.

1.3 Métodos de cálculo de riesgo para exposiciones al detalle

1.3.1 El método IRB

El método IRB propuesto en el Nuevo Acuerdo para las exposiciones al detalle es distinto en cuanto a insumos, estructura de la ponderación del riesgo y requisitos mínimos al que se propone para ser aplicado a otro tipo de exposiciones cuyo riesgo inicial se prevé mayor, como pueden ser las exposiciones empresariales a las que se les puede aplicar tanto el método avanzado como el básico. En este sentido, se hace necesaria una definición objetiva de las exposiciones al detalle a las que solo se les puede aplicar el método avanzado. La definición propuesta se basa en ciertos criterios cuyo objetivo es captar carteras homogéneas con un gran número de préstamos de bajo valor orientados ya sea al consumidor o a la empresa y donde el riesgo adicional de cualquier exposición individual es pequeño.

Teniendo en cuenta que las exposiciones al detalle engloban una amplia gama de productos, cada uno de los cuales presenta su particular trayectoria de pérdidas históricas, el Comité clasifica las exposiciones al detalle en tres grandes categorías:

- 1) Exposiciones garantizadas mediante hipotecas para adquisición de vivienda
- 2) Exposiciones al detalle autorrenovables admisibles (QRRE), y
- 3) Otras exposiciones no respaldadas por hipotecas también conocidas por otras minoristas.

Métodos de cálculo de riesgo para exposiciones al detalle

En términos generales, la categoría QRRE incluye créditos autorrenovables no asegurados que presentan características de pérdidas apropiadas, donde se incluirían muchas exposiciones derivadas de la financiación mediante tarjetas de crédito. El resto de créditos al consumo distintos de hipotecas, incluidas las exposiciones frente a pequeñas empresas, forman parte de la categoría otras minoristas.

Una de las diferencias más marcadas entre las carteras al detalle y empresariales es la forma en que los bancos diferencian el riesgo. Para las exposiciones al detalle, la utilización de una escala fija de calificación y la asignación de calificaciones a los prestatarios es mucho menos común. Más bien lo que hacen los bancos es dividir la cartera en "segmentos" conformados por exposiciones con características de riesgo parecidas, sobre la base de características de los prestatarios, transacción/producto y otras.

Por lo tanto, para aplicar el método IRB, los bancos tendrán que agrupar las exposiciones al detalle en segmentos determinados internamente, de acuerdo con un conjunto de requisitos mínimos. La evaluación de los componentes de riesgo se realizaría a nivel de segmento y no a nivel de grado de calificación, como en el caso de las exposiciones empresariales.

Vemos entonces, que los requisitos que deben cumplirse son los siguientes:

1. El banco debería establecer una clasificación de las exposiciones de capital.
2. Para cada clase de exposición, debería proporcionar los componentes de riesgo ciertos, utilizando parámetros estandarizados o sus estimaciones internas.

Capítulo 1. El Nuevo Acuerdo de Basilea

3. Se debería configurar una función de medición de riesgo la cual proporcione medidas (y por lo tanto requerimientos de capital) para el conjunto dado de los componentes.
4. Cada una de las componentes de la clasificación de las exposiciones debe cumplir los puntos anteriormente expuestos.

Como vemos en el primer requisito indicado anteriormente, se debe establecer una clasificación de las exposiciones, dentro de esta clasificación tenemos las exposiciones al detalle, y estas son las que suscitan nuestro interés.

1.3.2 Componentes de riesgo: Tasa de morosidad, Severidad y Pérdida Anticipada

En este apartado nos limitamos a explicar cuales son los componentes de riesgo que define Basilea II, en la sección 1.3.4. se desarrolla de forma más exhaustiva cuales son los requisitos mínimos que estos componentes deben cumplir para ser considerados como tales.

El riesgo de crédito da lugar a la pérdida crediticia (dinero no reembolsado por los prestatarios y gastos incurridos por la entidad al intentar recuperarlo). Esta pérdida es una parte inevitable de las operaciones crediticias, y por lo tanto, afecta a casi todas las unidades de negocio de una entidad financiera.

Aunque el nivel de pérdida varía de mes en mes o de año en año, se puede calcular estadísticamente el nivel medio de pérdida crediticia. La media matemática de la pérdida crediticia se denomina Pérdida

Anticipada, denominada según BIS II, Pérdida esperada, y se debe considerar como un coste de negocio.

Métodos de cálculo de riesgo para exposiciones al detalle

La Pérdida Anticipada o Pérdida Esperada (EL)⁴ viene dada por tres parámetros: Tasa de morosidad anticipada o Probabilidad de impago (PD)⁵, Exposición (EAD)⁶, y Severidad o Pérdida dada de impago (LGD)⁷. La Pérdida Anticipada depende de la calidad crediticia del cliente, el tipo de operación y de las garantías⁸.

La tasa de morosidad anticipada o Probabilidad de impago (PD) se define como la probabilidad de que el cliente entre en mora en un determinado periodo de tiempo. La morosidad se define, típicamente, como un impago tanto del principal como de los intereses o una reestructuración de las condiciones del préstamo para evitar dichos impagos. Esta definición es consistente con la utilizada por la mayoría de las agencias externas de rating, como S&P y Moody's.

La Probabilidad de impago está completamente separada de los otros componentes de la Pérdida Anticipada, en línea con las recomendaciones de Basilea II.

La Exposición en una operación de activo puede ser definida como la cantidad de dinero que el banco arriesga en el caso de que el cliente entre en mora. En otras palabras, la cantidad que podría ser perdida, asumiendo que no hubiera ningún tipo de recuperación, en caso de préstamos sería el capital pendiente.

En cuanto a la Severidad o Pérdida dada de impago (LGD) podemos decir que cuando un prestatario entra en mora, el banco no perderá necesariamente el importe total del préstamo. La LGD representa el ratio

⁴Expected Loss

⁵Probability of default

⁶Exposure of default

⁷Loss Given default

⁸ La Pérdida Anticipada puede ser calculada como un valor absoluto (en euros) o, más frecuentemente, como porcentaje del tamaño de la operación.

Capítulo 1. El Nuevo Acuerdo de Basilea

entre las pérdidas efectivas incurridas, como consecuencia de la entrada en mora y la Exposición. Para su cálculo es necesario pasar todos los flujos asociados a la recuperación, tanto positivos (pagos/venta de activos) como negativos (costes internos/judiciales), al valor presente en el día de entrada en mora.

$$\% \text{ Recuperación} = \frac{\text{Valor presente recuperaciones} - \text{Costes de recuperación}}{\text{Importe de la deuda al momento de entrar en mora}}$$
$$\% \text{ LGD} = 1 - \% \text{ Recuperación}$$

El Comité propone, en general, que los estimadores de pérdidas sean cuantificados en el mismo horizonte temporal. Por otra parte el vencimiento de una exposición no se reconoce como un componente de riesgo para la aplicación del IRB en las exposiciones al detalle, así, se reconoce que la restricción del mismo horizonte temporal puede no ser apropiado para los préstamos al detalle y abre diferentes alternativas.

Hay dos clases de componentes de riesgo para las exposiciones al detalle, y por lo tanto, dos métodos para implementar el sistema IRB:

1. Estimaciones separadas de la probabilidad de impago, PD y la pérdida dada de impago, LGD: En esta opción, el banco proporciona estimaciones internas del valor medio de PD y LGD para cada segmento de riesgo.
2. Estimación de la Pérdida Esperada, EL: En esta opción, lo que se requiere es una estimación de la pérdida media esperada para cada segmento de riesgo. Para las exposiciones individuales, EL está definida como:

Métodos de cálculo de riesgo para exposiciones al detalle

$$EL=PD \times EAD \times LGD$$

Mientras el banco debe proporcionar estimaciones internas de EL, en esta opción, no necesita identificar por separado los valores de PD y LGD.

La decisión del Nuevo Acuerdo de sacar adelante estas dos opciones refleja la diversidad en la práctica bancaria en lo que se refiere a los préstamos al detalle.

Las estimaciones que se obtengan en el empleo de cualquiera de los dos métodos dependen de la estimación de la exposición al incumplimiento, EAD, proporcionada por los bancos, y esta a su vez, depende del volumen y riqueza de los datos que los bancos posean sobre el comportamiento del riesgo y de los prestatarios en sus carteras al detalle.

Algunos bancos prefieren estimar los componentes PD y LGD por separado porque creen que la separación de la estimación proporciona un beneficio consistente en una información adicional. Las entidades que optan por este método reconocen el valor del incremento de información como una mejora significativa en el análisis de la calidad de su cartera y su riesgo.

Por otro lado, el argumento más importante para la aplicación de la estimación de EL hace referencia a que las estimaciones de los componentes PD y LGD no proporcionan el suficiente incremento de beneficios que garanticen el gasto y el esfuerzo asociados con el cambio de métodos para pronosticar la pérdida.

1.3.3 Fórmula para la derivación de ponderaciones del riesgo

Una vez conocidas los componentes de riesgo que establece Basilea II,

Capítulo 1. El Nuevo Acuerdo de Basilea

pasamos a explicar como debe ponderarse el riesgo de las exposiciones al detalle.

Las exposiciones al detalle recibirán una ponderación del riesgo que dependerá ya sea de PD , LGD y EAD o de la pérdida prevista (EL) de la exposición. La fórmula de ponderación de riesgo solo considera como componentes de riesgo PD , LGD y EAD pero también se planea diseñar un mecanismo análogo para las entidades bancarias que deseen considerar como componente de riesgo la pérdida esperada (EL).

Hay que considerar que la ponderación de riesgo para una exposición al detalle no depende del vencimiento (M) de la exposición⁹.

La fórmula de ponderación del riesgo para las exposiciones al detalle [7] es la siguiente:

$$\text{Activos ponderados por su nivel de riesgo} = K \times 12.5 \times \text{EAD}$$

Ahora bien, el cálculo de K (requerimiento de capital) difiere dependiendo del tipo de exposición al detalle que estemos considerando, obteniéndolo de la siguiente manera:

- Para exposiciones garantizadas mediante hipotecas sobre viviendas:

$$\text{Requerimiento de capital(K)} = \text{LGD} \times N[(1 - R)^{-0.5} \times G(\text{PD}) + (R/(1 - R))^{0.5} \times G(0.999)]$$

⁹ Los valores de PD, LGD y EAD serán expresados en números enteros y no en decimales, salvo anotación explícita en contrario, por ejemplo, un valor de 100% para LGD sería introducido como 100. La excepción se presenta en el contexto del requerimiento de capital (K). En esta ecuación, PD se expresa como un decimal (p.ej. una probabilidad de incumplimiento de 1% se expresaría como 0,01)

Métodos de cálculo de riesgo para exposiciones al detalle

$$\text{Correlación}(R)=0.15$$

- Para exposiciones minoristas autorrenovables admisibles

$$\begin{aligned} \text{Requerimiento de capital}(K) &= \text{LGD} \times \\ &\times N[(1-R)^{-0.5} \times G(\text{PD}) + (R/(1-R))^{0.5} \times G(0.999)] - \\ &0.75 \times \text{PD} \times \text{LGD} \end{aligned}$$

$$\begin{aligned} \text{Correlación}(R) &= 0.02 \times \\ &(1 - e^{(-50 \times \text{PD})}) / (1 - e^{-50}) + 0.11 \times [1 - (1 - e^{(-50 \times \text{PD})}) / (1 - e^{-50})] \end{aligned}$$

- Para otras exposiciones minoristas

$$\begin{aligned} \text{Requerimiento de capital}(K) &= \text{LGD} \times \\ &N[(1-R)^{-0.5} \times G(\text{PD}) + (R/(1-R))^{0.5} \times G(0.999)] \end{aligned}$$

$$\begin{aligned} \text{Correlación}(R) &= 0.02 \times (1 - e^{(-35 \times \text{PD})}) / (1 - e^{-35}) + \\ &0.17 \times [1 - (1 - e^{(-35 \times \text{PD})}) / (1 - e^{-35})] \end{aligned}$$

Capítulo 1. El Nuevo Acuerdo de Basilea

donde $N(x)$ es la función de distribución de una variable aleatoria normal estándar, sabiendo que $G(PD)=z$, donde $P(X \leq z)=PD$.

Las ponderaciones del riesgo para las exposiciones al detalle se basan en evaluaciones separadas de PD y LGD como insumos de la función de ponderación del riesgo, se requiere un mecanismo para trasladar esa estimación a la estructura de ponderaciones del riesgo PD-LGD, es decir para (EL), como ya indicábamos anteriormente.

Cuando esto se consiga, las ponderaciones del riesgo para todas las exposiciones al detalle serían determinadas mediante una fórmula común que relaciona las características de riesgo de una exposición (PD y LGD o EL) a una ponderación del riesgo correspondiente. Esta ponderación de riesgo sería aplicable a todos los tipos de productos incluidos en la categoría de exposición al detalle, aunque se está estudiando la conveniencia de tener fórmulas de ponderación de riesgo diferentes para cada uno de los productos.

En este apartado, hemos hecho una incursión en los requisitos de carácter más cuantitativo exigibles para la incorporación del método IRB en una entidad bancaria, y pensamos que son numerosos y complejos y requieren además que el modelo sea robusto, datos suficientes y que los directivos bancarios lo aprueben, lo que supone que el modelo que elaboren debe ser claro en su explicación y aplicación.

Una entidad puede desarrollar internamente un sistema de scoring y cuantificación suficiente, puede adquirirlo de un tercero, o desarrollarlo conjuntamente con otras entidades, pero además, tiene que demostrar que tiene datos propios, internos, adecuados para calcular las variables relevantes relativas a sus propios clientes, y este es un problema

Métodos de cálculo de riesgo para exposiciones al detalle

importante con el que creemos se pueden encontrar en España las entidades bancarias para poder aplicar un método interno¹⁰.

1.3.4 Los requisitos mínimos para las exposiciones al detalle

Queremos aclarar llegados a este punto, que el supervisor puede aceptar que un sistema determinado cumple todos los requisitos cuantitativos, que los datos de la entidad cumplen estos requisitos, y que el banco tiene un sistema interno de validación adecuado, pero para darlo por adecuado debe asegurarse que se ajusta a las necesidades y realidad de la entidad.

En este apartado vamos a desarrollar el último de los requerimientos necesarios para que una entidad bancaria pueda implantar un método interno de medición de riesgo. Las entidades bancarias deberían cumplir unos requisitos mínimos para poder incorporar un método IRB, y estos se deberán mantener en el tiempo para poder seguir empleando este modelo. Los requisitos mínimos los enumeramos a continuación:

- 1.Segmentación del producto y el riesgo.
- 2.Exhaustividad e integridad de las asignaciones de calificación de los clientes.
- 3.Vigilancia sobre el sistema y los procesos de calificación.
- 4.Criterios sobre la orientación del sistema de calificación.
- 5.Estimación de la Tasa de Morosidad (PD), Severidad (LGD) y Pérdida

¹⁰ En el Club de Riesgos celebrado en Marzo de 2002, era uno de los principales problemas expuestos por las entidades bancarias para la elaboración de un método interno de medición de riesgo. La solución que apuntaban era el comienzo desde esa fecha para la recopilación de datos, aunque esto no les permitiría llegar a tiempo a muchas de ellas para incorporar un método interno en la fecha de entrada del Nuevo Acuerdo, prevista para el año 2004.

Capítulo 1. El Nuevo Acuerdo de Basilea

Anticipada (EL).

6.Documentación y recopilación de datos.

7.Utilización de las calificaciones internas.

8.Validación interna.

9.Requisitos de divulgación.

Muchos de ellos son idénticos a los que sustenta el método IRB para las exposiciones empresariales, de ahí que se hagan referencias a estas. Sin embargo existen diferencias importantes en algunos aspectos, las que reflejan características particulares de las carteras al detalle. Estas diferencias y el desarrollo de los requisitos mínimos antes enumerados se expresan a continuación.

Los requisitos mínimos de la segmentación

Los sistemas de calificación para las exposiciones al detalle deben estar orientados al riesgo del prestatario y al riesgo del producto, y a recoger todas las características pertinentes de ambos. Este requisito es diferente al de las exposiciones empresariales y refleja la práctica que actualmente predomina en el mercado para las exposiciones al detalle y que consiste en combinar características del prestatario y del producto que nos determinan la estimación del riesgo dentro de un mismo segmento.

Toda exposición comprendida dentro de la definición al detalle para fines IRB, será asignada a un segmento de riesgo particular por el banco. Las entidades bancarias deberían demostrar que el nivel de segmentación adoptado internamente, permite una diferenciación significativa del riesgo y da lugar a una agrupación de conjuntos de préstamos suficientemente homogéneos además de asegurar que las características de riesgo del conjunto de préstamos subyacente, permanezcan

Métodos de cálculo de riesgo para exposiciones al detalle

relativamente estables a lo largo del tiempo y pueden ser seguidas por separado. La segmentación debe orientarse hacia el riesgo, tanto del prestatario, como de la transacción. Una vez identificado un segmento de riesgo, los bancos deben tratar de la misma manera a todos los prestatarios y transacciones del segmento en lo que concierne a la suscripción y estructuración de los préstamos, asignación de capital económico, fijación de precios y otras condiciones del acuerdo de préstamo, control y notificación interna. Esto servirá para demostrar la homogeneidad del riesgo de las exposiciones dentro de cada segmento.

Una entidad bancaria debería segmentar sus exposiciones al detalle sobre la base de las siguientes cuatro técnicas. Las dos primeras serán obligatorias para todos los bancos. Las dos siguientes también deberían realizarse, salvo que el banco logre demostrar a su supervisor, que un nivel tal de segmentación no sería apropiado, en vista del tipo de exposiciones al detalle o de la envergadura de sus operaciones.

1. Segmentación por tipo de producto

En primer lugar se debería segmentar las exposiciones al detalle según los siguientes tipos de productos.

- a. Tarjetas de crédito.
- b. Préstamos reembolsables a plazos (préstamos personales, compra financiada de automóviles, etc.)
- c. Créditos renovables.
- d. Crédito inmobiliario.
- e. Créditos para pequeños negocios.

2. Segmentación por riesgo de prestatario

Capítulo 1. El Nuevo Acuerdo de Basilea

El banco debería segmentar por puntuaciones de crédito o medida equivalente. Esto incluye la segmentación basada en la puntuación de solicitudes (puntuación atribuida en función de la información completa proporcionada en la solicitud de crédito). La puntuación continua o de conducta (basada en los datos de la agencia de informes comerciales o en los datos internos del banco) debe utilizarse como una base para volver a evaluar las estimaciones de la pérdida asociada a cada segmento, más que como base de la segmentación.

3. Segmentación por estado de morosidad.

Los bancos deberían normalmente separar los préstamos morosos de los que no lo son. Para ello, debería existir por lo menos dos categorías separadas e identificables para los conjuntos de préstamos en mora. Los bancos que no presenten este nivel de segmentación tendrían que demostrar a su supervisor que la morosidad no constituye un impulsor o pronosticador importante del riesgo en sus carteras al detalle. Además, tendrán que recopilar datos sobre este impulsor de riesgo para realizar evaluaciones periódicas y así determinar si la morosidad es lo suficientemente importante como para justificar una segmentación basada en ella.

4. Segmentación por madurez

Para captar los efectos de la maduración, las entidades financieras deberían segmentar sobre la base de la madurez de las exposiciones (momento en el que la transacción fue registrada en documentos). El periodo de maduración no debería pasar de un año. Los bancos que no segmenten por lo menos parte de sus exposiciones al detalle por

Métodos de cálculo de riesgo para exposiciones al detalle

madurez, tendrán que demostrar a su supervisor que la madurez no constituye un impulsor o pronosticador importante del riesgo de sus carteras al detalle. Además, estos bancos tendrían que recopilar datos sobre este impulsor del riesgo para realizar evaluaciones periódicas y así determinar si la madurez es lo suficientemente importante como para justificar una segmentación basada en ella.

Además de estas cuatro clases de segmentación, se proponen otras alternativas que enumeramos a continuación.

- a) Diferentes medidas de la relación préstamo a valor para los préstamos garantizados.
- b) Técnicas de mercadeo y distribución (tarjetas para grupos afines en los mercados objetivo, tarjetas oro, etc.).
- c) Prestatario tipo/perfil demográfico (ocupación, edad, etc.)
- d) Tamaño del préstamo.
- e) Vencimiento (hipotecas de 10 años, de 30 años, etc.)

Los bancos que segmenten sus exposiciones al detalle sobre la base de estas características de riesgo, tendrán que convencer a sus supervisores de que el tipo de segmentación que realizan permite una diferenciación significativa del riesgo.

Otro punto importante que hay que considerar es que para cada segmento identificado, el banco deberá proporcionar una medida cuantificable de las características de pérdida (PD y LGD, o EL) de ese segmento. Por lo tanto, el nivel de segmentación para el modelo interno utilizado, debe asegurar que el número de préstamos de un segmento dado, es el adecuado para que las pruebas estadísticas utilizadas para cuantificar conceptos de pérdida basados en el segmento tengan una fuerza razonable.

La distribución de prestatarios y exposiciones entre los segmentos al detalle deberá ser significativa. Ningún segmento de riesgo individual

Capítulo 1. El Nuevo Acuerdo de Basilea

deberá incluir una concentración indebida de la exposición al detalle total del banco.

El criterio utilizado por la entidad bancaria para asignar exposiciones de capital a los segmentos determinados no puede ser aleatorio, así, un banco debe contar con criterios para encajar una exposición dentro de un segmento. El banco debe demostrar que sus criterios cubren todos los factores que influyen en el análisis de riesgo. Estos factores deben demostrar habilidad para diferenciar el riesgo, tener poder de pronóstico y discriminación y ser tanto verosímiles como intuitivos y mostrar estabilidad dentro del esquema de calificación contemplado.

Al asignar exposiciones a un segmento, los bancos deben tener en cuenta toda la información pertinente. Las metodologías y datos utilizados en la asignación de las exposiciones a un segmento deben estar claramente documentados.

La exhaustividad e integridad de las asignaciones de calificación

Por lo expuesto en el apartado anterior, cada prestatario quedará asignado a un segmento de riesgo que se caracterizará por tener una calificación determinada por las características de pérdida que en él concurran.

Esto significa que la entidad bancaria deberá seguir y controlar continuamente su cartera para determinar si una exposición está en el segmento correcto y si las características de pérdida del segmento han cambiado. El propósito de este tipo de control es identificar tendencias emergentes o señales de "alerta anticipada".

Como mínimo, un banco debería examinar anualmente la actuación (características de pérdida) y estado de morosidad de cada segmento de riesgo identificado. También debería controlar la situación de los prestatarios individuales dentro de cada segmento de riesgo, para

Métodos de cálculo de riesgo para exposiciones al detalle

confirmar que las exposiciones siguen siendo asignadas a la calificación correcta. El examen de una muestra representativa de las exposiciones del segmento bastaría para satisfacer este último requisito.

Cuando los bancos cuentan con metodologías de puntuación o criterios de evaluación de riesgo, los casos de invalidación de estos criterios deben ser excepcionales.

Cuando se proceda excepcionalmente a la invalidación, las exposiciones interesadas estarán sujetas a un examen cuidadoso continuo, independientemente del proceso del muestreo.

La vigilancia sobre el sistema y los procesos de calificación

Todos los aspectos importantes del proceso de calificación y estimación de PD deben ser aprobados internamente por la directiva, comité de gestión y personal directivo superior¹¹ de la entidad bancaria. Todos ellos deberían poseer un conocimiento general de los métodos descritos en la documentación del sistema y proceso de calificación y aprobar cualquier discrepancia importante entre el procedimiento establecido y lo que ocurre en la práctica.

La calificación interna debe ser parte esencial de los informes presentados a todos los gestores anteriormente mencionados. Los informes deberán ser mensuales y cubrir el perfil de riesgo, la migración,

¹¹ Esta norma se refiere a una estructura administrativa compuesta por una directiva y personal directivo superior. El Comité entiende que los marcos legislativos y reguladores de los países difieren mucho en cuanto a las funciones de las directivas y del personal directivo superior. En algunos países, la directiva tiene la función principal, y no exclusiva, de supervisar al órgano ejecutivo (administrador jefe, gerente general) con el fin de asegurar que el mismo cumpla con sus tareas. Por esta razón, en algunos casos se le conoce como junta de supervisión. Esto quiere decir, que la directiva no posee funciones ejecutivas. En otros países en cambio, la directiva tiene competencias más amplias en el sentido que establece el marco general para la gestión del banco. En vista de estas diferencias, las nociones de directiva y personal directivo superior son utilizadas en el Nuevo Acuerdo para ponerle nombre a dos funciones decisorias del banco y no para identificar construcciones legales.

Capítulo 1. El Nuevo Acuerdo de Basilea

y cuantificación de las estimaciones de pérdida para todos y cada uno de los segmentos, así como la comparación entre índices de incumplimiento efectivo y expectativas.

El personal directivo debería asegurarse de que el proceso, los criterios y el resultado de la calificación estén ampliamente documentados en papel o en forma electrónica. La documentación debería ser lo suficientemente específica como para permitir una evaluación por terceros de las calificaciones asignadas y la calibración de una PD media por segmento. La documentación debe ser accesible a todas las personas interesadas en el proceso de calificación.

Cuando se utilicen modelos estadísticos en el proceso de calificación, el personal directivo deberá asegurarse que el banco cuente con un documento de metodología para el modelo utilizado. El documento de metodología deberá:

- a) Proporcionar un esbozo detallado de la teoría, supuestos y/o base matemática y empírica de la asignación de PD estimadas a los segmentos o deudores individuales, y la(s) fuente(s) de datos utilizada(s) para estimar el modelo;
- b) Establecer un proceso estadístico riguroso (con pruebas de desempeño fuera de tiempo y fuera de muestra) para validar la selección de variables explicativas; y
- c) Explicar las circunstancias bajo las cuales el modelo no funciona eficazmente para que el banco esté informado de sus limitaciones.

El personal directivo deberá velar constantemente por el correcto funcionamiento del sistema de calificación del riesgo. Para ello, será necesario una interacción estructurada entre el personal directivo y las funciones de control del banco, particularmente el departamento de

Métodos de cálculo de riesgo para exposiciones al detalle

control del riesgo crediticio y la auditoría interna. Esta interacción debe estar relacionada con la eficiencia del sistema y la suficiencia de los recursos, lo que implica una acción correctiva con respecto a deficiencias observadas en el sistema de calificación del riesgo.

Por otra parte, los auditores internos deberán examinar anualmente el sistema de calificación del banco, incluyendo en su examen a la calificación interna. El cumplimiento de todos los requisitos mínimos aplicables será asimismo examinado por los auditores, quienes deberán documentar todo el trabajo realizado.

Algunos supervisores nacionales podrán además solicitar una auditoría externa del proceso de asignación de calificaciones del banco, así como de su estimación de las pérdidas.

El banco deberá contar con una o más unidades independientes de control del riesgo crediticio, encargadas del diseño, puesta en práctica y desempeño de su sistema interno de calificación. Las unidades deben ser operativamente independientes del personal y funciones administrativas responsables de originar las exposiciones. Sus áreas de responsabilidad deben incluir lo siguiente:

- a) Asignación y/o examen y seguimiento de la calificación interna;
- b) Elaboración y análisis de informes sobre los resultados del sistema interno de calificación, datos históricos sobre el comportamiento de exposiciones crediticias en el pasado por segmento interno, análisis de la migración, comparación de los segmentos asignados con las calificaciones externas o modelos de pronóstico del incumplimiento y seguimiento global de créditos en cada segmento por criterio principal de calificación;
- c) Verificación de la operabilidad de los procedimientos destinados a controlar la asignación de calificaciones de acuerdo a políticas y criterios

Capítulo 1. El Nuevo Acuerdo de Basilea

establecidos. Toda incompatibilidad entre la asignación efectiva y las reglas establecidas deberá ser rápidamente identificada y corregida; y

d) Examen y documentación de cualquier cambio en el proceso de calificación y de las razones que lo motivaron.

La unidad de control del riesgo crediticio asumirá la responsabilidad de los modelos utilizados en el proceso de calificación y los controlará. Será en última instancia responsable del examen constante del modelo y de los cambios que pudieran introducirse en el futuro. Esta unidad deberá ser operativamente independiente del personal con funciones administrativas responsables de originar las exposiciones, y de cualquier otro personal que tenga control sobre el modelo. Será necesario difundir información y conocimiento del modelo y su metodología fuera de este grupo directivo.

Los miembros del personal responsable de cualquier aspecto del proceso de calificación deberán poseer las calificaciones y formación necesarias para asumir este papel. La administración deberá destinar recursos cualificados y competentes suficientes a estas funciones de control. Las personas responsables de asignar o examinar las calificaciones de riesgo deberán recibir una formación adecuada para promover una asignación consecuente y correcta de la calificación.

Los criterios sobre la orientación del sistema de calificación

Los bancos deberían tener un proceso específico y criterios para asignar una exposición a un segmento de riesgo. Esto se hace generalmente sobre la base de criterios uniformes o aplicando una ficha de puntuación a todos los prestatarios de una cartera o a un segmento homogéneo de la misma. Estos criterios deben ser lo suficientemente específicos como para que un tercero pueda evaluar la asignación de una

Métodos de cálculo de riesgo para exposiciones al detalle

exposición a un segmento de riesgo particular, deben demostrar habilidad para diferenciar el riesgo y deben ser tanto verosímiles como intuitivos.

La evaluación de riesgos del banco debería ser conservadora, sobre todo en las áreas en las que el perfil del prestatario da lugar a dudas. Toda decisión sobre la calificación de riesgo debe tener en cuenta la calidad de la información financiera y de otro tipo; extendiéndose más allá de la información contable cuando corresponda. Será necesario realizar un análisis más profundo del crédito a medida que la condición financiera del prestatario se deteriora y aumenta la probabilidad de incumplimiento.

Al asignar un prestatario a un segmento, el banco deberá evaluar factores de riesgo del horizonte futuro, basándose sobre información actual y su experiencia con el prestatario. Dada la dificultad de pronosticar acontecimientos lejanos y su efecto sobre la condición financiera del prestatario, el banco deberá adoptar una visión conservadora de la información proyectada. Además cuando los datos disponibles son limitados, el banco deberá darle un sesgo conservador a su análisis.

Para la cuantificación del riesgo (el proceso de asignar probabilidades de incumplimiento (PD) a los segmentos), se utiliza un horizonte de un año.

Para cada segmento de su calificación interna, el banco estimará una probabilidad de incumplimiento de un año. Cada estimación de PD deberá representar una visión conservadora de una PD media de largo alcance para el segmento del prestatario en cuestión, y estar, por lo tanto, apoyada en la experiencia histórica y evidencia empírica. Al mismo tiempo, la estimación deberá ser de visión hacia el futuro. Al cumplir estos requisitos, los bancos podrán incorporar ajustes basados en varios

Capítulo 1. El Nuevo Acuerdo de Basilea

factores. Estos ajustes deberán realizarse mediante un proceso racional y un análisis bien elaborado y documentado. Además, deberán basarse en la evidencia empírica y otra información histórica como por ejemplo, un cambio importante en los índices de incumplimiento o en los impulsores principales del incumplimiento futuro. Todo ajuste realizado deberá ser aplicado de manera conservadora y constante.

En lo que concierne a la durabilidad de la evaluación del riesgo, el requisito es que el banco examine al prestatario y el segmento al que es asignado, por lo menos anualmente y con mayor frecuencia en el caso de prestatarios de más riesgo. Ciertos créditos, especialmente los de mayor riesgo o los problemáticos, estarían sujetos a exámenes más frecuentes. Además, los bancos deberían hacer una nueva calificación cada vez que obtengan información pertinente sobre el prestatario.

Estimación de la Tasa de Morosidad (PD), Severidad (LGD) y Pérdida Anticipada

Los bancos deberían proporcionar una estimación explícita tanto de PD y LGD, identificadas por separado, o EL, para cada segmento. En lo que concierne a la noción de LGD o EL, la pérdida será entendida como pérdida económica. Con ello se pretende incluir los efectos del descuento, los costos de financiación y los costos directos e indirectos asociados con los cobros sobre el préstamo al determinarse la pérdida. Los bancos no deben medir simplemente la pérdida registrada en la contabilidad, aunque si deberían poder comparar las dos cosas. Además, un banco debe proporcionar una estimación explícita del monto de exposición de cada transacción (comúnmente llamada Exposición en el Momento del Incumplimiento (EAD) en los sistemas internos del banco). Todas estas estimaciones de la pérdida deben procurar captar todos los riesgos de una exposición subyacente.

Métodos de cálculo de riesgo para exposiciones al detalle

En el caso de productos al detalle con exposiciones futuras dudosas, como las tarjetas de crédito, los bancos deben tener en cuenta su evolución y/o la expectativa de retiros adicionales previos al incumplimiento, al calibrar las pérdidas estimadas (EL o LGD). En concreto, si el banco no refleja factores de conversión para líneas no retiradas en sus estimaciones de EAD, debería reflejar en sus LGD estimadas la probabilidad de retiros adicionales previos al incumplimiento.

Las estimaciones se basarán en un promedio de varios años, pero incluirán también un elemento de visión hacia el futuro.

Los bancos deberían utilizar la anterior definición de incumplimiento como pérdida económica para estimar estas medidas de pérdida (EL o LGD) y recopilar datos sobre el incumplimiento de su propia experiencia. De todas formas, podrán utilizar definiciones diferentes para los distintos productos al detalle, aunque todas las definiciones internas deberán ser consecuentes con la definición de referencia, como también lo tendrá que ser cualquier conjunto de datos externos utilizado para estimar estas medidas. Esta definición de incumplimiento no pretende de ninguna manera afectar los derechos y recursos legales de los bancos en el caso que el prestatario deje de cumplir sus obligaciones en virtud del contrato de préstamo, como tampoco pretende establecer o modificar normas contables aceptadas. Su único objetivo es abordar los problemas relacionados con la estimación uniforme de las características de pérdida del método IRB en todos los bancos y fuentes de datos, para uso posterior en los cálculos del capital regulador.

Se considerará que un deudor incurre en incumplimiento cuando tiene lugar uno o más de los siguientes casos:

a) Se determina que el deudor probablemente no pagará sus obligaciones (capital, intereses o cargos) enteras;

Capítulo 1. El Nuevo Acuerdo de Basilea

b) Una situación de crédito asociada con cualquier obligación del deudor, tales como, anulación de la deuda, reserva específica o reestructuración forzosa con condonación o aplazamiento del capital, intereses o cargos; cualquier reprogramación del vencimiento de un servicio financiero (extender la duración de una hipoteca para reducir los pagos mensuales) será considerada un evento de incumplimiento, siempre que se emprenda tal reprogramación en circunstancias difíciles para mitigar una situación de incumplimiento.

c) El deudor tiene una obligación en mora por más de 90 días; o

d) El deudor ha solicitado una declaración de quiebra u otra protección similar contra sus acreedores.

Los bancos deberán documentar la definición específica del incumplimiento utilizada internamente y demostrar su compatibilidad con la definición de referencia ya mencionada.

Los bancos deberán estudiar toda la información disponible para estimar el valor medio de PD/LGD o EL (las características de pérdida) para cada segmento, teniendo en cuenta las tres técnicas que se plantean en los requisitos para la estimación de PD que son:

a) Experiencia interna de incumplimiento

· El banco podrá utilizar datos sobre la experiencia interna de incumplimiento para estimar la PD.

b) Asociación a datos externos

· Se reconocerá asimismo el uso de técnicas asociativas. Para asociar sus calificaciones internas a la escala utilizada por una institución externa de evaluación de crédito o institución similar, los bancos podrán atribuir

Métodos de cálculo de riesgo para exposiciones al detalle

una PD a cada segmento interno y luego atribuir la característica de incumplimiento observada para los segmentos de la institución que se ha tomado como referencia a los segmentos internos del banco.

La asociación al conjunto de datos utilizado deberá ser significativa, evitando posibles inclinaciones o incongruencias en el enfoque o en los datos implícitos. En este sentido, el banco deberá demostrar que sus criterios de calificación interna pueden compararse con los utilizados para crear o diferenciar las frecuencias de incumplimiento incorporadas en la fuente de datos utilizada. Los criterios deberán estar orientados hacia el riesgo del prestatario y no reflejar características de las transacciones. En el análisis habrá que incluir también una comparación de la definición de incumplimiento utilizada.

c) Modelos estadísticos de incumplimiento

- Sujeto a los requisitos mínimos formulados, el banco podrá utilizar un promedio de estimaciones individuales de la probabilidad de incumplimiento para prestatarios de un segmento determinado usando modelos estadísticos de pronóstico de incumplimiento

El banco deberá contar con un proceso de revisión de la entrada de datos al modelo estadístico de pronóstico del incumplimiento que incluya la evaluación de la exactitud, integridad y propiedad de los datos específicos de la asignación de una calificación aprobada.

El banco debería demostrar que el universo de prestatarios representados en los datos es representativo del universo de los prestatarios reales del banco.

En vista de que la segmentación es específica para cada banco, los datos internos serán la fuente principal de información para estimar las

Capítulo 1. El Nuevo Acuerdo de Basilea

características de pérdida. Los bancos podrán utilizar datos externos o modelos estadísticos para la cuantificación siempre que sea posible demostrar la existencia de un vínculo estrecho entre la base de segmentación y el perfil de riesgo del banco. En todos los casos, los bancos utilizarán las fuentes de datos como puntos de comparación.

Los bancos deberían reconocer la importancia de las consideraciones de juicio en este proceso, particularmente en el sentido de asegurar una estimación con visión hacia el futuro de las características de pérdida. Juicios de este tipo serán aplicados con un sesgo conservador, manteniendo el grado de conservadurismo constante a lo largo del tiempo.

Visto todo lo anterior, se puede resumir que fuera cual fuese el método elegido, todos los métodos utilizados para estimar las características de pérdida deben cumplir los siguientes requisitos:

- a) El universo de exposiciones representado en el conjunto de datos es igual o por lo menos semejante al del segmento contemplado.
- b) Las normas de concesión de préstamos o suscripción utilizadas para generar las exposiciones en la fuente de datos son muy parecidas a las utilizadas por el banco para sus segmentos actuales;
- c) Las condiciones económicas o de mercado reinantes en el momento de la experiencia están relacionadas con las condiciones actuales y las condiciones predecibles; y
- d) El número de préstamos en la muestra y el periodo de datos utilizado para la cuantificación, hacen que el banco se sienta seguro de la precisión y solidez de las características de pérdida y el análisis estadístico subyacente.

Las entidades bancarias deberían contar continuamente con estimaciones de las características de pérdida que estén bien calibradas y que además logren incorporar rápidamente toda la información nueva que

Métodos de cálculo de riesgo para exposiciones al detalle

se presente. Como mínimo, los bancos deberán examinar anualmente las características de pérdida estimadas.

Sin tener en cuenta las fuentes de información utilizadas por el banco, sean éstas externas, internas, o comunes, o una combinación de las tres, para estimar las características de pérdida, el periodo de observación histórica subyacente utilizado debe durar por lo menos cinco años. Si el periodo de observación dura más tiempo, se utilizará este periodo más largo.

La documentación y recopilación de datos

Para las carteras de servicios bancarios al detalle, habrá que almacenar una historia completa de la evaluación del riesgo para cada prestatario o cada segmento de prestatarios. Específicamente, los bancos deben recopilar y almacenar datos sobre:

a) Las características del segmento, incluso las características de los productos utilizados para la segmentación, las características del prestatario utilizadas para la segmentación, madurez, y estado de morosidad; y

b) Las características de riesgo cuantificadas, asociadas con cada segmento (las probabilidades de incumplimiento, o pérdidas previstas asociadas con los segmentos). Para cada uno de estos conceptos de pérdida, el banco debe recopilar y almacenar las medidas pronosticadas y efectivas.

La utilización de las calificaciones internas

Las calificaciones internas asignadas y la información cuantitativa que de ellas se deriva deben ser parte integral del proceso diario de medición y gestión del riesgo crediticio.

Capítulo 1. El Nuevo Acuerdo de Basilea

Las calificaciones internas deberían jugar un papel preponderante en el proceso de aprobación de créditos.

Las probabilidades de incumplimiento asociadas con las calificaciones internas deberán emplearse en la valoración del riesgo crediticio. El costo del crédito debe reflejar información tanto del prestatario como de las calificaciones del servicio. Esta información deberá, a su vez, ser utilizada como un factor en la valoración de la exposición.

La distribución de las exposiciones entre los diferentes segmentos de calificación interna y la probabilidad de incumplimiento correspondiente deben estar incorporadas en el sistema de presentación de informes al personal directivo superior.

Así mismo, las calificaciones internas deben estar explícitamente vinculadas a la evaluación interna de la suficiencia de capital del banco.

Las calificaciones internas del banco y las PD estimadas deben ser consideradas en el proceso de dotación de reservas. El banco deberá contar con políticas claras con respecto al tratamiento de la pérdida prevista. La PD asociada con un segmento interno debe utilizarse como insumo del análisis de la rentabilidad bancaria la cual podrá, a su vez, ser utilizada como un elemento de los procesos administrativos del banco, tales como las decisiones estratégicas de asignación de recursos o los planes de remuneración por rendimiento.

Si un banco cuenta con un modelo de riesgo crediticio que forma parte del análisis de rentabilidad y/o asignación interna de capital, las características de incumplimiento estimadas deben ser también un insumo importante de este modelo.

Un banco debe contar con procesos sólidos de pruebas en la evaluación de la suficiencia de capital. En las pruebas de evaluación se identificarán los acontecimientos y posibles futuros cambios en las condiciones económicas que podrían afectar desfavorablemente a las

Métodos de cálculo de riesgo para exposiciones al detalle

exposiciones crediticias del banco y se evaluará la habilidad del banco para aguantar dichos cambios. Son tres las áreas que deberían ser examinadas por el banco para tenerlas en cuenta en su evaluación:

- 1) Desaceleración de la actividad económica o industrial
- 2) Situaciones de riesgo de mercado
- 3) Condiciones de liquidez

Como parte de las pruebas de evaluación, será necesario recrear escenarios específicamente diseñados para calificar cuantitativamente el efecto producido por la migración de las exposiciones hacia segmentos de calificación menores. Un análisis de este tipo deberá además examinar el efecto de índices de incumplimiento más elevados e índices de recuperación más bajos que los pronósticos de PD, LGD y la medida de exposición de capital del banco.

Los resultados de las pruebas serán enviados periódicamente a la administración superior, cualquiera que sea el método de prueba empleado, tomándose la acción apropiada en los casos en que los resultados sobrepasen los límites de tolerancia.

Un departamento o unidad independiente deberá realizar la prueba de evaluación, por lo menos semestralmente. Toda prueba deberá ser documentada adecuadamente.

Un banco debería poseer un historial convincente de utilización de información proveniente de la calificación interna. Por lo tanto, tendrá que demostrar que ha estado utilizando un sistema de calificación en consonancia general con los requisitos mínimos consignados en este documento durante los últimos tres años, por lo menos. La intención de este requisito no es imponer una moratoria sobre la modificación y el mejoramiento de los sistemas de calificación de los bancos, lo que se pretende es que la incorporación del método IRB se haga con todas las

Capítulo 1. El Nuevo Acuerdo de Basilea

garantías posibles de éxito, lo que implica una buena medición del riesgo de crédito.

Estas pruebas de utilización de las calificaciones internas descritas se aplicarán a las exposiciones al detalle con la siguiente modificación: Las estimaciones de pérdida asociadas con calificaciones internas, deben utilizarse para la valoración del riesgo crediticio, teniendo en cuenta las limitaciones impuestas por leyes o reglamentos antidiscriminatorios vigentes en cada país.

La validación interna

Otro de los requisitos mínimos a considerar para la implementación de un método interno que garantice el cálculo del capital mínimo expuesto al riesgo consiste en la validación interna de este método.

Los bancos deberán contar con un sistema eficiente para validar la precisión y coherencia de los sistemas de calificación, procesos y estimación de PD. Una entidad bancaria deberá demostrar a sus supervisores que el proceso de validación interna le permite evaluar, de manera significativa y constante, el rendimiento de los sistemas de calificación interna y cuantificación de riesgos.

El banco deberá contar con un proceso de revisión del aporte de datos, a fin de evaluar la exactitud, integridad y propiedad de los datos específicos que determinan la asignación de una calificación aprobada en particular.

Será necesario mantener documentación detallada de las excepciones que no se ajusten a los parámetros de introducción de datos, como parte del ciclo de procedimiento de la validación modelo.

Métodos de cálculo de riesgo para exposiciones al detalle

Además de lo expuesto, el ciclo de procedimiento de la validación incluirá:

- a) Control periódico permanente del rendimiento del modelo, con inclusión de la evaluación y prueba estadística rigurosa de la estabilidad dinámica del modelo y sus coeficientes principales;
- b) Identificación y documentación de las relaciones fijas individuales que ya no se adecuan al modelo;
- c) Pruebas periódicas del resultado del modelo comparado con los resultados anuales, como mínimo; y
- d) Un proceso de control riguroso de los cambios que estipule los procedimientos a seguir antes de introducir cambios en el modelo como respuesta a los resultados de la validación.

Los bancos deberán comparar regularmente el incumplimiento efectivo con la probabilidad de incumplimiento estimada para cada segmento y demostrar que el incumplimiento efectivo por segmento concuerda con las expectativas. Estas comparaciones deberán utilizar periodos de observación de datos históricos tan largos como sea posible. Las entidades bancarias deberán entender y documentar claramente los métodos y datos utilizados en las comparaciones, las que se llevarán a cabo con frecuencia (por lo menos una vez al año).

También se podrán emplear otras herramientas de validación cuantitativa. El análisis deberá basarse en datos regularmente actualizados que sean apropiados para la cartera y cubran un abanico de entornos económicos e, idealmente, un ciclo coyuntural completo.

Los bancos deberán demostrar que los métodos de prueba y datos cuantitativos son constantes a lo largo del tiempo. Los cambios efectuados en los métodos y datos (tanto fuentes de datos como periodos cubiertos) deben ser documentados con mucha precisión.

Los requisitos de divulgación

El Comité ha establecido, dentro de los requisitos mínimos, los denominados requisitos de divulgación, para que los bancos puedan emplear el método IRB. Así, los bancos deben divulgar:

Información general sobre metodología e insumos claves, también denominadas divulgaciones cualitativas, y estas son:

- a) La aceptación del método por el supervisor;
- b) Para cada cartera, si se utiliza una estimación propia o un vector de supervisión para LGD y/o EAD;
- c) Para cada cartera, métodos de estimación y validación de PD, así como LGD y EAD;
- d) Datos necesarios para la estimación del modelo, utilización interna por los bancos de las estimaciones, excluyendo para fines de capital IRB, responsabilidad e independencia del proceso de calificación;
- e) Relación entre calificaciones internas y externas;
- f) El proceso para administrar y reconocer la cobertura del riesgo crediticio;
- g) Para cada cartera, definiciones de incumplimiento utilizadas, así como LGD y EAD internamente para cada cartera en el marco IRB, y asociación de las definiciones internas y de referencia de incumplimiento, así como LGD y EAD, incluyendo la metodología utilizada por el banco, si la definición empleada se aleja de la definición de referencia; y
- h) Los bancos en transición, aprobada por los supervisores, entre métodos fundados en la calificación interna, deben divulgar: los requisitos mínimos específicos a los que se aplica la transición, las áreas y el grado de cumplimiento faltante, y el avance realizado hacia el

Métodos de cálculo de riesgo para exposiciones al detalle

cumplimiento de la serie completa de requisitos mínimos.

Información requerida para la evaluación del riesgo o divulgaciones cuantitativas, que son:

- a) El porcentaje de exposición nominal cubierta por el método IRB;
- b) En lo que respecta a montos nominales, valores de PD y LGD o EL para cada segmento de riesgo;
- c) Para créditos con exposición variable, supuestos para EAD, utilizados para la estimación, montos de la exposición nominal y estimaciones de EAD antes y después de la cobertura de riesgo crediticio reconocida;
- d) Para créditos con exposición variable, montos de la exposición nominal y valores para PD, LGD y EAD o EL para cada segmento de riesgo;
- e) La distribución de los deudores con calificación externa entre las diferentes categorías de calificación interna de PD.

Rendimiento ex post como indicación de la calidad y confiabilidad que también se consideran divulgaciones cuantitativas:

- a) Estadísticas sumarias de la distribución de LGD real, tales como, la desviación típica y el 10^o, 50^o y 90^o percentil, también ponderadas con la exposición. Para la cartera al detalle, valores para PD, LGD y EAD o EL para cada segmento de riesgo;
- b) Valores para el número de servicios con incumplimiento y servicios y monto girado para cada segmento de riesgo. Para cada cartera, en la medida que corresponda, estadísticas sumarias de la distribución de EAD, también ponderadas con la exposición, junto con el número de prestatarios;

Capítulo 1. El Nuevo Acuerdo de Basilea

c) La distribución de las migraciones de calificación ponderadas con la exposición nominal y EAD, respectivamente, en ambos casos después de 1, 2 y 3 años;

d) Para los bancos que utilizan sus propias estimaciones de LGD, una comparación entre capital económico, capital real mantenido y requerimiento de capital mínimo e indicadores sumarios del capital económico atribuido a las líneas comerciales más importantes.

Presencia de algunos problemas

Creemos que una de las cuestiones que surgen de la necesidad de evaluación que el supervisor debe hacer del cumplimiento de los repetidos requisitos es la igualdad en el trato a bancos de distintos países. Los requisitos son numerosos y complejos, pero, a pesar de ello, están definidos de una forma general. Por lo tanto, surgen problemas de interpretación, e incluso de definiciones, en el momento de la implantación. Esto podría llevar a que supervisores de distintos países traten de manera distinta a bancos similares.

Este problema no solo surge en el enfoque IRB, sino en cualquiera de los procedimientos de evaluación para la aplicación de metodologías avanzadas en el Acuerdo, donde el supervisor necesariamente tiene que aplicar una dosis de discrecionalidad, y es más complejo en la revisión de aquellas cuestiones que sean más cualitativas. Pensamos que para que el Acuerdo funcione racionalmente no es deseable que las reglas se establezcan hasta el más mínimo detalle, es necesaria una gran dosis de flexibilidad para que el supervisor pueda tratar caso a caso a las entidades en todo aquello que no puede incluir ni prever la norma, y al mismo tiempo dotar de esa misma flexibilidad a las entidades para poder desarrollar los mejores sistemas para su situación particular. Es más,

Métodos de cálculo de riesgo para exposiciones al detalle

creemos que sería hasta saludable para el buen funcionamiento del Acuerdo cierta divergencia en su aplicación en los distintos países, puesto que estos son objetivamente diferentes, y estas diferencias se refieren tanto a la forma en que cada uno supervisa como a cuestiones más profundas y estructurales referidas a las características de los sistemas legales, económicos y financieros.

1.3.5 El método estándar

Una de las grandes diferencias entre el Acuerdo de Capital de 1988 y el Nuevo Acuerdo (1999) radica en que el último hace hincapié en la utilización de métodos internos de cálculo de riesgo por parte de las entidades bancarias, como hemos visto en el apartado (1.3.1), donde se han desarrollado todos los requerimientos necesarios para su implantación.

Sin embargo, Basilea II mantiene la utilización del método estándar para aquellos bancos que no puedan o no quieran usar un método interno, que consiste en que las entidades bancarias deben mantener un porcentaje fijo de capital como previsión de riesgo. Según el banco de España los porcentajes a aplicar al riesgo crediticio son los recogidos en la siguiente tabla: [11]

Capítulo 1. El Nuevo Acuerdo de Basilea

Sin riesgo	0%	Sector público y garantías específicas
Riesgo bajo	0.1%	Hipot. vivienda y empresas con rating A o superior
Riesgo medio-bajo	0.4%	Leasing y demás garantías reales
Riesgo medio	0.6%	Riesgos no mencionados en otras categorías
Riesgo medio-alto	1%	Riesgos personales por consumo
Riesgo alto	1.5%	Saldos tarjetas de crédito y activos sin cobertura obligatoria

Como podemos observar hay una cierta segmentación del riesgo según el producto crediticio sometido a exposición.

Creemos que la utilización de un método interno construido con todas las garantías requeridas beneficiaría a las entidades bancarias puesto que supone una mayor información y un cálculo del capital de riesgo expuesto más acorde con las operaciones crediticias realizadas.

De todo lo expuesto anteriormente se infiere que si queremos construir un Credit Scoring enfocado a préstamos de bajo valor orientados ya sea al consumidor o a la empresa y donde el riesgo adicional de cualquier

Métodos de cálculo de riesgo para exposiciones al detalle

exposición individual es pequeño, su elaboración deberá cumplir la normativa destinada a exposiciones al detalle y además ser aceptado como método interno de medición de riesgo por la entidad supervisora que proceda, en nuestro caso el banco de España.

El Credit Scoring que elaboremos deberá permitir calcular los siguientes conceptos:

- a) Separación de préstamos en segmentos que reflejen un mismo nivel de riesgo
- b) Tasa de morosidad (PD) y Severidad (LGD) por separado, para cada segmento de riesgo ó
- c) La pérdida anticipada (EL), para cada segmento.

El cálculo de estos conceptos nos permitirá obtener la ponderación de riesgo derivado de la concesión de créditos.

Una vez conocidos los requisitos normativos pasaremos a estudiar los modelos matemáticos que nos permitan realizar un Credit Scoring con las características antes expuestas.

Bibliografía

[1] : Bosch Morro, A. Cómo llevar a cabo la integración de riesgos y la gestión de recursos. Recoletos Medios digitales.(2003).

[2] : Capon, N. Credit scoring systems: A critical analysis. J. Marketing, 46, 82-91. (1982).

[3] : Churchill, G., Nevin, J., Watson, R. The role of credit scoring in the loan decisión. Credit World, March, 6-10. (1977).

Bibliografía

[4] : www.ciss.noticias.es (2003).

[5] : Durand, D. Risk Elements in Consumer Instalment Financing. National Bureau of Economic Research. New York. (1941).

[6] : Fisher, R. The use of multiple measurements in taxonomic problems. Ann. Eugenics. 7, 179-188. (1936).

[7] : www.bis.org/publ/bcbsca.htm. (2001).

[8] : www.ceu.es/clubriesgos/seccionesRiesgoCredito.htm (2002).

[9] : Lewis, E. An Introduction to Credit Scoring. Athena Press, San Rafael, CA. (1992).

[10] : Myers, J., Forgy, E. The development of numerical credit evaluation systems. J.Amer.Statist. Assoc.,58, 799-806. (1963).

[11] : Orsikowsky, B. Los modelos internos en la provisión estadística de la Circular 9/99. (Delegado del banco de España). Conferencia del Club de Riesgos celebrada en Madrid en marzo de 2002.

[12] : Smalley, O., Sturdivant, F. The Credit Merchants: A History of Spiegel. Southern Illinois University Press, Carbondale, IL. (1973).

[13] : Trias, R. Basilea II puede obligar a reducir el volumen de créditos. La gaceta de los Negocios (2003).

Bibliografía

[14] : Vargas, F. El marco general de la validación de procedimientos internos en Basilea II: El enfoque IRB. Circular del Banco de España. (2002).

[15] : Weingartner, H. Concepts and utilization of credit scoring techniques. *Banking*, 58, 51-53. (1966).

Capítulo II

UNA REVISIÓN DE LOS MÉTODOS ESTADÍSTICOS EMPLEADOS EN LA CONSTRUCCIÓN DE CLASIFICACIONES DE CRÉDITO

2.1 Introducción

Cuando los primeros credit scoring fueron desarrollados en los años 50 y 60, el método utilizado fue la discriminación estadística. Aún hoy, dichos métodos son los más comúnmente utilizados para construir clasificaciones de crédito. La ventaja de estos es que permiten utilizar el conocimiento de las propiedades de los estimadores de la muestra, el uso de los intervalos de confianza y el contraste de hipótesis en este contexto. Estas técnicas estadísticas permiten identificar y eliminar características no relevantes y asegurar aquellas que son relevantes para la clasificación.

Aunque los métodos estadísticos fueron los primeros utilizados para construir sistemas de clasificación y aún permanecen entre los más

importantes, ha habido cambios en los últimos cuarenta años. Inicialmente, los métodos se basaron en métodos de discriminación sugeridos por Fisher (1936) para problemas de clasificación general. Esto condujo a seguir una línea de scoring basada en la función lineal discriminante de Fisher. Los supuestos que se necesitaban para asegurar que esta era la mejor manera para discriminar, entre buenos y malos clientes potenciales, eran extremadamente restrictivos y claramente no se daban en la práctica, aunque los scoring obtenidos eran muy robustos. El método de Fisher se puede entender como una forma de regresión lineal, y esto conlleva a una investigación de otras formas de regresión que tienen supuestos menos restrictivos para garantizar su optimalidad y poder conseguir una reglas de clasificación. El método con más éxito de todos los métodos de análisis de regresión lineal discriminante es la regresión logística. Otro método que también ha sido utilizado en los últimos 20 años es el de clasificación en árbol o método de partición recursivo. Este consiste en la división del conjunto de aplicaciones en un número de subgrupos determinado dependiendo de sus características. Una vez realizada esta operación, clasificamos cada subgrupo como satisfactorio o insatisfactorio. Aunque este método no proporciona un valor a cada una de las características como lo hacen los scoring lineales, el resultado es el mismo. Un método para decidir en todo caso cuando una nueva concesión de crédito será clasificada satisfactoria o insatisfactoriamente. Todos estos mecanismos han sido utilizados para proyectar scoring en organizaciones comerciales.

En este capítulo, revisamos cada uno de estos métodos y el fondo estadístico que lo sostiene. Comenzaremos con el análisis discriminante. Discutiremos como la función lineal discriminante se considera como un clasificador al que se llega por tres modelos diferentes. En el primero lo que se busca es minimizar el coste esperado en la decisión de aceptar a

Capítulo 2. Una revisión de los métodos estadísticos empleados en ...

un nuevo cliente. El segundo modelo está motivado por el trabajo original de Fisher, se trata de encontrar la mejor función que diferencie entre los grupos de buenos (satisfactorios) y malos (insatisfactorios) clientes y el tercero considera una ecuación de regresión que trata de encontrar el mejor estimador de la probabilidad de que un cliente sea bueno. Seguiremos con la construcción de un modelo de regresión donde la variable dependiente es una función no lineal de la probabilidad de que un cliente sea bueno. Veremos la regresión logística y otras funciones encabezadas por el análisis probit y tobit. Veremos después la clasificación mediante el método del árbol para finalizar con el modelo de la vecindad, que es bastante diferente de las formulaciones estadísticas de los problemas de clasificación original.

2.2 Análisis discriminante: Una perspectiva desde la teoría de la Decisión

El proceso de concesión de un crédito consiste en la elección entre dos acciones, una que trata de conceder un nuevo crédito y otra que implica la no concesión del crédito. El credit scoring trata de ayudar en la toma de esta decisión, buscando cual debería ser la mejor regla a aplicar a partir de una muestra de concesiones de crédito previas. La ventaja de esta es el conocimiento previo de la evolución de dichos préstamos. Si hay solamente dos acciones posibles, aceptar o rechazar, entonces no hay ninguna ventaja en clasificar el desarrollo de estas concesiones de crédito conocidas en más de dos clases distintas de las de buenos y malos. Se considera a un cliente bueno cuando el desarrollo de la misma es aceptable para la entidad prestamista, y es malo cuando ante el conocimiento de la evolución posterior al préstamo la entidad desearía haberla rechazado. Algunas entidades consideran como malo un cliente

Análisis discriminante: Una perspectiva desde la teoría de la decisión

cuando se dan por perdidos un número de pagos consecutivos, mientras que otras entidades lo consideran como malo cuando se da por perdido el número total de pagos.

Hay un sesgo inherente en este modelo en el que la muestra se toma de asignaciones previas de créditos que fueron concedidos y no hay información del desarrollo de aquellos créditos que fueron rechazados en el pasado. Así, la muestra es solo representativa de aquellas concesiones de crédito que fueron aceptadas y no es representativa de todas aquellas que se rechazaron.

Hay un argumento que sostiene que hay más opciones que elegir y rechazar en un proceso de concesión de crédito. Por ejemplo, se puede decidir solicitar información adicional o que el estudio de la asignación sea considerada por un analista crediticio para una decisión manual. De todas formas, la decisión final en cada caso será aceptar o rechazar la petición de crédito. No valdría la pena tratar de clasificar las concesiones en un grupo en particular cuando la decisión en todo caso de que pertenezcan a ese grupo dependa enteramente del prestamista. Así, incluso en este proceso de múltiples acciones es posible clasificar las aplicaciones en solo dos grupos, buenos y malos, de donde la decisión final resultará una de las dos opciones.

Definimos $\mathbf{X}=(X_1, X_2, \dots, X_p)'$ como el conjunto de p variables aleatorias que describen la información disponible en una valoración para la asignación de crédito. Usaremos tanto la palabra variable como característica para referirnos al mismo concepto, es decir para describir cualquier aplicación X_i . La primera la utilizamos cuando queremos enfatizar la naturaleza aleatoria de esta información entre las aplicaciones y la segunda cuando queremos recalcar que nos proporciona una información. El valor actual de las variables para una aplicación

Capítulo 2. Una revisión de los métodos estadísticos empleados en ...

particular se denota como $x=(x_1, x_2, \dots, x_p)'$. En la terminología de credit scoring, los diferentes valores posibles o respuestas, x_i , de la variable X_i se denominan atributos de una característica. Así, si una característica típica es el estatus residencial, entonces sus atributos serían, en propiedad, alquiler sin muebles, alquiler con muebles, viviendo con los padres, u otros. Diferentes entidades pueden tener distintos grupos de atributos de la misma característica. Así, otro prestamista puede decidir clasificar el estatus residencial en propiedad con hipoteca, propiedad sin hipoteca, en alquiler con muebles propios, en alquiler con muebles no propios, vivienda móvil, vivir con los padres, vivir con otros parientes, u otros. Para distinguir claramente entre característica y atributo diremos que la característica es la pregunta que se realiza para conceder o no el préstamo y los atributos son las posibles respuestas a esa pregunta.

Volviendo a la decisión que la entidad prestamista debe tomar, denotaremos por A al conjunto de todos los valores posibles que la variable

$$X=(X_1, X_2, \dots, X_p)'$$

puede tomar, es decir, el conjunto de todas las posibles respuestas que se pueden obtener. El objetivo es encontrar una regla que divida el conjunto A en dos subconjuntos A_G y A_B que clasifica las aplicaciones cuyas respuestas están en A_G como buenos y aceptarlas mientras que aquellas cuyas respuestas están en A_B son malos y rechazarlas minimizando el coste esperado para el prestamista. Los dos tipos de coste corresponden con los dos tipos de error que pueden surgir de esta decisión. Se puede clasificar a alguien que es bueno como malo y rechazar a esta persona.

Análisis discriminante: Una perspectiva desde la teoría de la decisión

Por lo tanto, incurrimos en un coste por rechazar la concesión de préstamos a clientes que hubieran sido solventes. Asumimos por ahora, que el coste de no conceder estos préstamos es el mismo para cada uno de ellos y lo denotaremos por L . El segundo error es clasificar un malo como bueno y aceptar la aplicación. En este caso se incurriría en un débito cuando el cliente impagara el crédito. Planteamos pues el asumir las siguientes hipótesis de trabajo

1. La pérdida esperada por préstamo concedido es la misma para todos los clientes y la denotaremos por D .
2. Denotaremos por $p(G)$ (respectivamente, $p(B)$) la proporción de asignaciones que no producen impago (respectivamente, que producen impago).
3. Las características de una asignación de crédito tiene un número finito y discreto de atributos, así A es finito y solo tiene un número finito de atributos diferentes x .

Notaremos por $p(x|G)$ a la probabilidad de que tenga los atributos x teniendo una buena asignación de crédito, de forma similar, definiremos $p(x|B)$ como la probabilidad de que tenga los atributos x sabiendo que tiene una mala asignación de crédito. Si $p(G|x)$ denota la probabilidad de que tenga una buena asignación de crédito teniendo como atributos x , con $p(G)=p(\text{asignación buena})$ y $p(B)=p(\text{asignación mala})$, el Teorema de Bayes nos dice

$$p(G|x)=p(x|G) p(G)/p(x). \quad (2.1)$$

Capítulo 2. Una revisión de los métodos estadísticos empleados en...

Un resultado similar lo encontramos para $p(x|B)$, la probabilidad de que un individuo tenga los atributos x sabiendo que tiene una mala asignación de crédito, y lo expresamos mediante

$$p(B|x) = p(x|B) p(B)/p(x). \quad (2.2)$$

Notamos que (2.1) y (2.2) se pueden expresar de la siguiente manera:

$$p(G|x)/p(B|x) = p(x|G) p(G)/p(x|B) p(B). \quad (2.3)$$

El coste de equivocarnos en la concesión si aceptamos las asignaciones con atributos en A_G y rechazamos aquellas con atributos en A_B es:

$$\begin{aligned} & L \sum_{x \in A_B} p(x|G)p(G) + D \sum_{x \in A_G} p(x|B)p(B) = \\ & = L \sum_{x \in A_B} p(G|x)p(x) + D \sum_{x \in A_G} p(B|x)p(x). \end{aligned} \quad (2.4)$$

La regla que minimiza este coste esperado es sencilla. Consideramos que L y D están dados, si tomamos un particular $x = (x_1, x_2, \dots, x_p)'$, o bien está en A_G o bien en A_B . Si está en A_G , entonces existe un coste si este incurre en impago, en cuyo caso el coste esperado es $Dp(x|B) p(B)$. Si x está clasificado en A_B existe un coste si este es solvente, y el coste esperado es $Lp(x|G) p(G)$. Así, se caracteriza la pertenencia de x en A_G mediante la desigualdad si

Análisis discriminante: Una perspectiva desde la teoría de la decisión

$$Dp(x|B)p(B) \leq Lp(x|G)p(G)$$

De esta manera la regla de decisión que minimiza el coste esperado viene dada por:

$$\begin{aligned} A_G &= \{x: Dp(x|B)p(B) \leq Lp(x|G)p(G)\} = \{x: (D/L) \leq p(x|G)p(G) / p(x|B)p(B)\} \\ &= \{x: (D/L) \leq p(G|x)/p(B|x)\}, \end{aligned} \quad (2.5)$$

donde la última expresión la obtenemos a partir de (2.3).

Una crítica al criterio anterior es que el coste esperado depende de los valores de L y D, los cuales pueden no ser conocidos. Así, en lugar de minimizar el coste esperado, se podría intentar minimizar la probabilidad de cometer un tipo de error mientras mantenemos la probabilidad de cometer el otro tipo de error en un valor dado. En el contexto de la concesión de créditos, lo obvio es minimizar el nivel de impago mientras mantenemos el porcentaje de aplicaciones aceptadas en un valor dado. El último requerimiento es equivalente a mantener la probabilidad de rechazar buenas aplicaciones en algún nivel fijado.

Suponemos que queremos que el porcentaje de aplicaciones aceptadas (el ratio de aceptación) sea a. Entonces A_G debe satisfacer:

$$\sum_{x \in A_G} p(x) = \sum_{x \in A_G} p(x|G) p(G) + \sum_{x \in A_G} p(x|B) p(B) = a, \quad (2.6)$$

mientras que al mismo tiempo minimizamos el ratio de impago

$$\sum_{x \in A_G} p(x|B) p(B).$$

Capítulo 2. Una revisión de los métodos estadísticos empleados en...

Si definimos para cada $x \in A$,

$$b(x) = p(x|B)p(B)$$

entonces nuestro objetivo es encontrar el conjunto A_G tal que podamos resolver el programa

$$\min \sum_{x \in A_G} b(x) = \sum_{x \in A_G} (b(x)/(p(x)))p(x) \quad (2.7)$$

$$\text{sujeto a } \sum_{x \in A_G} p(x) = a.$$

Empleando el método de los multiplicadores de Lagrange (o el sentido común usando el principio de codicia), se puede ver que este debe ser el conjunto de atributos x , que satisfacen

$$b(x)/(p(x)) \leq c,$$

donde c es elegido tal que la suma de las $p(x)$ satisface la restricción de ser igual a a . Entonces

$$A_G = \{x : b(x)/(p(x)) \leq c\} = \{x : p(B|x) \leq c\} = \quad (2.8)$$

$$= \{x : (1-c)/c \leq p(x|G)p(G)/(p(x|B)p(B))\},$$

y donde la segunda desigualdad se obtiene a partir de las definiciones de $p(x)$ y $b(x)$.

Así la forma de la regla de la decisión bajo este criterio es la misma que en (2.5), para $L=c$ y $D=1-c$.

Análisis discriminante: Una perspectiva desde la teoría de la decisión

El mismo análisis se podría plantear asumiendo que las características de la asignación son continuas y no variables aleatorias discretas. La única diferencia sería que las funciones de distribución condicional $p(x|G), p(x|B)$ son reemplazadas por funciones de densidad condicionales $f(x|G)$ y $f(x|B)$ y los sumatorios son reemplazados por integrales. Así el coste esperado si se divide el conjunto A en los conjuntos A_G y A_B y se aceptan solo aquellos que estén en A_G sería

$$L \int_{x \in A_B} f(x|G)p(G)dx + D \int_{x \in A_G} f(x|B)p(B)dx, \quad (2.9)$$

y la regla de decisión que minimiza es análoga a (2.5), y la expresamos como:

$$\begin{aligned} A_G &= \{x : D f(x|B)p(B) \leq L f(x|G)p(G)\} = \\ &= \{x : D p(B)/(Lp(G) \leq f(x|G)/(f(x|B))\}. \end{aligned} \quad (2.10)$$

2.2.1 El caso de una Normal Univariante

Consideramos el caso posible más simple donde solo hay una variable característica continua X y su distribución condicional respecto a los buenos $f(x|G)$ es una normal con media μ_G y varianza σ^2 , mientras la distribución respecto a los malos es una normal de media μ_B y varianza σ^2 . Entonces

$$f(x|G)=(2\pi)^{-1/2} \exp(-(x-\mu_G)^2/2\sigma^2),$$

y así la regla de (2.10) se transforma en:

$$\begin{aligned} f(x|G)/f(x|B) &= \exp(-(x-\mu_G)^2/2\sigma^2)/\exp(-(x-\mu_B)^2/2\sigma^2)= \\ &= \exp(-(x-\mu_G)^2 + (x-\mu_B)^2/2\sigma^2) \geq D p(B)/L p(G). \quad (2.11) \end{aligned}$$

Entonces

$$x(\mu_G - \mu_B) \geq (\mu_G^2 - \mu_B^2)/2 + \sigma^2 \ln(D p(B)/(L p(G))) \Leftrightarrow x \in A_G$$

2.2.2 El caso de una Normal Multivariante con covarianza común

Un ejemplo más realista lo tenemos cuando hay p características (variables) en la información obtenida para la aplicación y los resultados tanto de los buenos clientes como de los malos forman una distribución normal multivariante. Asumimos que la media de la distribución de los buenos es μ_G y la de los malos μ_B con una matriz de covarianzas común Σ . La correspondiente función de densidad en este caso es

$$f(\mathbf{x}|G)=(2\pi)^{-p/2} (\det\Sigma)^{-1/2} \exp(-(\mathbf{x}-\mu_G)'\Sigma^{-1}(\mathbf{x}-\mu_G)/2) \quad (2.12)$$

Análisis discriminante: Una perspectiva desde la teoría de la decisión

donde $(\mathbf{x} - \boldsymbol{\mu}_G)$ es un vector columna con p filas y $(\mathbf{x} - \boldsymbol{\mu}_G)'$ denota su traspuesta. Siguiendo lo obtenido en (2.11), tenemos:

$$f(\mathbf{x}|G)/f(\mathbf{x}|B) \geq D p(B)/L p(G).$$

En consecuencia,

$$\mathbf{x}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_G - \boldsymbol{\mu}_B) \geq (\boldsymbol{\mu}_G' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_G - \boldsymbol{\mu}_B' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_B)/2 + \sigma^2 \text{Ln}(Dp(B)/Lp(G)) \quad (2.13)$$

La parte izquierda de la expresión (2.13) es una suma ponderada de los valores de las variables, que expresamos como, $x_1 w_1 + x_2 w_2 + \dots + x_p w_p$, mientras la parte derecha de la expresión es una constante. Así, la expresión (2.13) nos lleva hacia una regla lineal de scoring, la cual es conocida como una función lineal discriminante.

En el ejemplo anterior se asume que las medias y las covarianzas de la distribución son conocidas. Este caso se da raramente, y es más normal reemplazarlas por los estimadores, conocidos como medias muestrales, y que denotaremos por \mathbf{m}_G y \mathbf{m}_B y la matriz de covarianzas muestrales que denotaremos por \mathbf{S} . Entonces la regla de decisión (2.13) quedaría expresada por

$$\mathbf{x}'\mathbf{S}^{-1}(\mathbf{m}_G - \mathbf{m}_B) \geq (\mathbf{m}_G' \mathbf{S}^{-1} \mathbf{m}_G - \mathbf{m}_B' \mathbf{S}^{-1} \mathbf{m}_B)/2 + \sigma^2 \text{Ln}(Dp(B)/Lp(G)) \quad (2.14)$$

2.2.3 El caso de una Normal Multivariante con diferentes matrices de covarianzas

Otra restricción obvia en el caso anterior es que las matrices de covarianzas son iguales para la población de buenos y malos clientes. Suponemos que la matriz en la población de buenos clientes es Σ_G y para los malos Σ_B . En este caso (2.14) se convierte en

$$f(x|G)/f(x|B) \geq Dp(B)/Lp(G) \quad (2.15)$$

Entonces,

$$\exp\{-1/2((x - \mu_G)' \Sigma_G^{-1} (x - \mu_G) - (x - \mu_B)' \Sigma_B^{-1} (x - \mu_B))\} \geq Dp(B)/Lp(G) \quad (2.16)$$

es decir,

$$\begin{aligned} (x' (-\Sigma_G^{-1} + \Sigma_B^{-1})x + 2x' (\Sigma_G^{-1} \mu_G - \Sigma_B^{-1} \mu_B)) &\geq (\mu_G' \Sigma_G^{-1} \mu_G - \mu_B' \Sigma_B^{-1} \mu_B) \\ &+ 2 \text{Ln}(Dp(B)/Lp(G)). \end{aligned} \quad (2.17)$$

Nótese que el lado izquierdo de la anterior desigualdad es una forma cuadrática en las variables x_1, x_2, \dots, x_p . Esta parece ser una regla de decisión más general y se podría esperar que este desarrollo fuese mejor que la regla lineal. En la práctica, sin embargo, se tiene que estimar el doble de parámetros, esto es, Σ_G y Σ_B . La incertidumbre extra que envuelve a estos estimadores hace que la regla de decisión cuadrática sea menos robusta que la lineal. Esto se confirma en los trabajos de Reichert, Cho y G.M. Wagner [12].

Análisis discriminante: Una perspectiva desde la teoría de la decisión

2.3 Análisis discriminante: Separación en dos grupos

En el trabajo original de Fisher (1936) [4], en la que se introducía la función lineal discriminante, la clave era encontrar la combinación de variables que mejor separaran dos grupos cuyas características eran observables. Estos dos grupos podían ser subespecies de una planta y las características son las medidas físicas, o podía ser quien sobrevive o sucumbe a algún trato traumático y las características son las respuestas a varias pruebas. En el contexto del credit scoring, los dos grupos son clasificados por el prestamista en buenos y malos clientes y las características son la información de la oficina de crédito y la aplicación de otros detalles.

Consideramos $Y = w_1 X_1 + w_2 X_2 + \dots + w_p X_p$ una combinación lineal de las características $\mathbf{X} = (X_1, X_2, \dots, X_p)'$. Parece obvio que medir la separación es como medir cuánto de diferentes son los valores de las medias de Y para los dos diferentes grupos de buenos y malos en la muestra. Así, buscamos elegir los valores de w_i , de forma que $\sum_i w_i = 1$, y maximicen la diferencia entre $E[Y|G]$ y $E[Y|B]$. El valor que obtenemos es similar si los datos de cada grupo están dispersos, que si están reagrupados uniendo los buenos y los malos clientes previamente cuando queremos discutir su separación y este segundo método facilita el cálculo y por lo tanto se procede de esta forma.

Fisher sugirió que si asumimos que los dos grupos tienen varianzas comunes, entonces una medida sensible de separación es

$M = (\text{distancia entre las medias muestrales de ambos grupos} / \text{varianza muestral de ambos grupos})^{1/2}$

Capítulo 2. Una revisión de los métodos estadísticos empleados en ...

Se divide entre la raíz cuadrada de la varianza de la muestra para hacer la medida escalar independiente, esto es, si se transforma la variable Y en cY , entonces la medida de M no cambia.

Al igual que en la anterior sección denotaremos las medias muestrales por \mathbf{m}_G y \mathbf{m}_B para los clientes buenos y malos, respectivamente, y S denotará la varianza muestral común. Entonces la distancia de separación correspondiente M sería

$$M = \mathbf{w}(\mathbf{m}'_G - \mathbf{m}'_B / \mathbf{w}'\mathbf{S}\mathbf{w})^{1/2} \quad (2.18)$$

Diferenciando M con respecto a \mathbf{w} y tomando la derivada igual a cero, obtenemos que el valor de M es mínimo cuando

$$(\mathbf{m}'_G - \mathbf{m}'_B / \mathbf{w}'\mathbf{S}\mathbf{w})^{1/2} - ((\mathbf{w}'(\mathbf{m}_G - \mathbf{m}_B))(\mathbf{S}\mathbf{w})) / \mathbf{w}'\mathbf{S}\mathbf{w} = 0,$$

es decir,

$$(\mathbf{m}_G - \mathbf{m}_B)(\mathbf{w}'\mathbf{S}\mathbf{w}) = (\mathbf{w}'(\mathbf{m}_G - \mathbf{m}_B))\mathbf{S}\mathbf{w} \quad (2.19)$$

El hecho de que la segunda derivada de M con respecto a \mathbf{w} sea una matriz definida positiva garantiza que el valor obtenido es efectivamente un mínimo.

$$(\mathbf{w}'\mathbf{S}\mathbf{w}) / (\mathbf{w}'(\mathbf{m}_G - \mathbf{m}_B))$$

es un escalar λ de forma que

Análisis discriminante: Separación en dos grupos

$$\mathbf{w} = \lambda (\mathbf{S}^{-1} (\mathbf{m}_G - \mathbf{m}_B)) \quad (2.20)$$

Estos valores son los mismos que hemos obtenido en la expresión (2.14), solo que esta vez no hemos asumido una distribución normal. Este es el mejor separador de los buenos y malos clientes bajo el criterio de que no conocemos sus distribuciones. Este resultado se obtiene para todas las distribuciones de varianza finita, porque la medida de la distancia M considera solamente la media y la varianza de las distribuciones, y esta da el mismo resultado para todas las distribuciones con la misma media y varianza.

2.4 Análisis discriminante: Una forma de regresión lineal

Es otro modelo de credit scoring, que también llega a una función lineal discriminante, es la regresión lineal. En este modelo se trata de encontrar la mejor combinación lineal de las características

$$w_0 + w_1 X_1 + w_2 X_2 + \dots + w_p X_p = (\mathbf{w}^*)' \mathbf{X}^*$$

que explique la probabilidad de impago. Siendo

$$\mathbf{w}^* = (w_0, w_1, w_2, \dots, w_p)', \quad \mathbf{X}^* = (1, X_1, X_2, \dots, X_p)'$$

Capítulo 2. Una revisión de los métodos estadísticos empleados en ...

Así, si p_i es la probabilidad de que la aplicación i en la muestra tenga impago, se quiere encontrar la mejor aproximación de \mathbf{w}^* tal que

$$p_i = w_0 + x_{i1} w_1 + x_{i2} w_2 + \dots + x_{ip} w_p \quad \text{para todo } i. \quad (2.21)$$

Suponemos una muestra con n elementos tal que $n_G + n_B = n$, donde n_B son malos, para facilitar la notación asumimos que son los primeros de la muestra y así $p_i = 1$ para $i = 1, \dots, n_B$. Los restantes n_G de la muestra con $i = n_B + 1, \dots, n_G + n_B$ son buenos, así para ellos $p_i = 0$. Buscamos los coeficientes w tales que minimicen el error cuadrático medio. Este viene dado aproximadamente por:

$$\sum_{i=1}^{n_B} \left(1 - \sum_{j=0}^p w_j x_{ij}\right)^2 + \sum_{i=n_B+1}^{n_B+n_G} \left(\sum_{j=0}^p w_j x_{ij}\right)^2. \quad (2.22)$$

En notación matricial, (2.21) se puede escribir

$$\begin{pmatrix} \mathbf{1}_B & \mathbf{X}_B \\ \mathbf{1}_G & \mathbf{X}_G \end{pmatrix} \begin{pmatrix} \omega_0 \\ \mathbf{w} \end{pmatrix} = \begin{pmatrix} \mathbf{1}_B \\ \mathbf{0} \end{pmatrix} \quad \text{ó} \quad \mathbf{Y}\mathbf{w} = \mathbf{b}, \quad (2.23)$$

donde

Análisis discriminante: Una forma de regresión lineal

$$\mathbf{Y} = \begin{pmatrix} 1_B & X_B \\ 1_G & X_G \end{pmatrix},$$

es una matriz de $n \times (p+1)$, y

$$\mathbf{X}_B = \begin{pmatrix} x_{1,1} & \cdots & \cdots & x_{1,p} \\ x_{2,1} & \cdots & \cdots & x_{2,p} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n_B,1} & \cdots & \cdots & x_{n_B,p} \end{pmatrix}$$

es una matriz $n_B \times p$, y

$$\mathbf{X}_G = \begin{pmatrix} x_{n_B+1,1} & \cdots & \cdots & x_{n_B+1,p} \\ x_{2,1} & \cdots & \cdots & x_{2,p} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n_B+n_G,1} & \cdots & \cdots & x_{n_B+n_G,p} \end{pmatrix},$$

es una matriz $n_G \times p$, y

$$\mathbf{b} = \begin{pmatrix} 1_B \\ 0 \end{pmatrix},$$

donde $\mathbf{1}_B$ es el vector $1 \times n_B$ con todas las entradas igual a 1.

Con esta nueva notación el problema se reduce a

$$\min (\mathbf{Y}\mathbf{w}-\mathbf{b})'(\mathbf{Y}\mathbf{w}-\mathbf{b}). \quad (2.24)$$

Calculamos las derivadas con respecto a \mathbf{w} e igualamos a cero, esto es,

$$\mathbf{Y}'(\mathbf{Y}\mathbf{w}-\mathbf{b})=0 \quad \text{ó} \quad \mathbf{Y}'\mathbf{Y}\mathbf{w}=\mathbf{Y}'\mathbf{b}$$

donde

$$\mathbf{Y}'\mathbf{b}=\begin{pmatrix} 1 & 1 \\ X'_B & X'_G \end{pmatrix} \begin{pmatrix} \mathbf{1}_B \\ \mathbf{0} \end{pmatrix} = \begin{pmatrix} n_B \\ n_G m_G \end{pmatrix}$$

e

$$\mathbf{Y}'\mathbf{Y}=\begin{pmatrix} 1 & 1 \\ X'_B & X'_G \end{pmatrix} \begin{pmatrix} 1 & X_B \\ 1 & X_G \end{pmatrix} = \begin{pmatrix} n & n_B m'_B + n_G m'_G \\ n_B m_B + n_G m_G & X'_B X_B + X'_G X_G \end{pmatrix} \quad (2.25)$$

Análisis discriminante: Una forma de regresión lineal

Si para obtener resoluciones explicativas consideramos los valores esperados muestrales iguales a los valores poblacionales y denotamos por \mathbf{S} a la matriz de covarianzas de la muestra, se tendrá que

$$\mathbf{X}'_G \mathbf{X}_G + \mathbf{X}'_B \mathbf{X}_B = n\mathbf{S} + n_G \mathbf{m}_G \mathbf{m}'_G + n_B \mathbf{m}_B \mathbf{m}'_B \quad (2.26)$$

Con (2.25) y utilizando (2.26) tenemos

$$\begin{aligned} n\mathbf{w}_0 + (n_G \mathbf{m}_G + n_B \mathbf{m}_B)' \mathbf{w} &= n_B \\ (n_G \mathbf{m}_G + n_B \mathbf{m}_B) \mathbf{w}_0 + (n\mathbf{S} + n_G \mathbf{m}_G \mathbf{m}'_G + n_B \mathbf{m}_B \mathbf{m}'_B) \mathbf{w} &= n_G \mathbf{m}_G \end{aligned} \quad (2.27)$$

Sustituyendo la primera ecuación de (2.27) en la segunda obtenemos

$$\begin{aligned} n_G \mathbf{m}_G &= (n_G \mathbf{m}_G + n_B \mathbf{m}_B) (1/n) (n_B - (n_G \mathbf{m}_G + n_B \mathbf{m}_B)' \mathbf{w}) \\ &+ (n_G \mathbf{m}_G \mathbf{m}'_G + n_B \mathbf{m}_B \mathbf{m}'_B) \mathbf{w} + n\mathbf{S}\mathbf{w}, \end{aligned}$$

Finalmente, podemos escribir que

$$\mathbf{S}\mathbf{w} = \mathbf{c}(\mathbf{m}_G - \mathbf{m}_B) \quad (2.28)$$

donde

$$\mathbf{c} = n_G n_B (1 - (\mathbf{m}_G - \mathbf{m}_B)' \mathbf{w}).$$

Así, (2.28) da la mejor elección de $\mathbf{w}=(w_1, w_2, \dots, w_p)$ para los coeficientes de la regresión lineal. Este es el mismo \mathbf{w} que obteníamos en (2.20) con la función lineal discriminante. Este método muestra, que se pueden obtener los coeficientes del credit scoring por el método de los mínimos cuadrados de la regresión lineal.

Hemos considerado en la ecuación de la regresión (2.21), que en el lado izquierdo de esta, los buenos clientes tienen valor 1 y los malos valor 0. Esto da un conjunto de constantes, que calificamos como $\mathbf{w}(\mathbf{1}, \mathbf{0})^*$. Si tomamos otros valores tales que los buenos clientes quedan a la izquierda de \mathbf{g} y los malos a la izquierda de \mathbf{b} , entonces los coeficientes en la regresión $\mathbf{w}(\mathbf{g}, \mathbf{b})^*$ difieren solo en el término constante w_0 tal que

$$\mathbf{w}(\mathbf{g}, \mathbf{b})^* = \mathbf{b} + (\mathbf{g} - \mathbf{b})\mathbf{w}(\mathbf{1}, \mathbf{0})^* \quad (2.29)$$

2.5 Regresión Logística

El método de regresión para discriminación lineal tiene un defecto obvio. En la ecuación (2.21), el lado derecho podría tomar valores desde $-\infty$ hasta $+\infty$, pero el lado izquierdo es una probabilidad y solo puede tomar valores entre 0 y 1. Esto mejoraría si en este lado tuviéramos una función de p_i , la cual pudiera tomar un rango más amplio de valores. Entonces, no tendríamos la dificultad de que todos los datos tienen valores muy similares a las variables dependientes o que la regresión predice probabilidades que son menores que 0 o mayores que 1. Esta función es el logaritmo de la probabilidad proporcional. Este método de

regresión logística fue empleado por Wiginton (1980) [13] en uno de los primeros resultados publicados de credit scoring. En la regresión logística, se toma el logaritmo de la probabilidad ampliada como una combinación lineal de las características de los clientes.

$$\text{Ln}(p_i/1-p_i) = w_0 + x_1 w_1 + x_2 w_2 + \dots + x_p w_p = \mathbf{w}'\mathbf{x} \quad (2.30)$$

Si

$$p_i/1-p_i$$

toma valores entre 0 e ∞ , entonces,

$$\text{Ln}(p_i/1-p_i)$$

toma valores entre $-\infty$ y $+\infty$. Tomando exponenciales en ambos lados de (2.30) obtenemos la ecuación

$$p_i = e^{w'x} / 1 + e^{w'x} \quad (2.31)$$

Este es el supuesto de la regresión logística. Es interesante notar que si asumimos que la distribución de los valores de las características de los clientes tanto buenos como malos es una normal multivariante, como veíamos en la sección 2.2, entonces este ejemplo satisface el supuesto de regresión logística. De nuevo asumimos que las medias son μ_G para los buenos clientes y μ_B para los malos con una matriz de covarianzas común Σ .

Capítulo 2. Una revisión de los métodos estadísticos empleados en ...

La función de densidad condicional correspondiente a este caso viene dada por la expresión (2.12)

$$f(\mathbf{x}|G) = (2\pi)^{-p/2} (\det \Sigma)^{-1/2} \exp(-(\mathbf{x} - \boldsymbol{\mu}_G)' \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}_G) / 2) \quad (2.32)$$

Si $p(B)$ es la proporción de la población que son malos y $p(G)$ es la proporción de la población que son buenos, entonces el logaritmo de la probabilidad proporcional para el cliente i cuyas características son \mathbf{x} es

$$\begin{aligned} \ln(p_i / 1 - p_i) &= \ln(p(G)f(\mathbf{x}|G) / p(B)f(\mathbf{x}|B)) \quad (2.33) \\ &= \mathbf{x}' \Sigma^{-1} 2(\boldsymbol{\mu}_G - \boldsymbol{\mu}_B) + (\boldsymbol{\mu}_B' \Sigma^{-1} \boldsymbol{\mu}_B + \boldsymbol{\mu}_G' \Sigma^{-1} \boldsymbol{\mu}_G) + \ln(p(G) / p(B)) \end{aligned}$$

Si esta es una combinación lineal de los x_i , entonces satisface el supuesto de regresión logística. De todos modos, otras clases de distribución también satisfacen el supuesto logístico, incluidas algunas que no cumplen las funciones lineales discriminantes si se les aplica el método de la pérdida de Bayes visto en la sección 2.2. Consideremos por ejemplo, el caso donde las características son todas binarias e independientes entre sí. Entonces

$$\begin{aligned} P(X_i = 1|G) &= p(G(i)); \quad P(X_i = 0|G) = 1 - p(G(i)); \\ P(X_i = 1|B) &= p(B(i)); \quad P(X_i = 0|B) = 1 - p(B(i)); \end{aligned}$$

Así si p_G, p_B son las probabilidades anteriores de los buenos y malos clientes en la población

$$P(G|\mathbf{x})=P(\mathbf{x}|G)p(G)/P(\mathbf{x})=\prod_i p(G(i))^{x_i} (1-p(G(i))^{1-x_i} p(G))/P(\mathbf{x}) \quad (2.34)$$

entonces

$$\begin{aligned} \ln(P(G|\mathbf{x})/P(B|\mathbf{x})) &= \sum_i x_i (\ln(p(G(i)))-\ln(p(B(i)))) \\ &+ \sum_i (1-x_i) (\ln(1-p(G(i)))-\ln(1-p(B(i)))) + \ln(p(G)/p(B)) = \\ &= \sum_i x_i (\ln(p(G(i)(1-p(B(i)))/(p(B(i))(1-p(G(i)))))) + \\ &\sum_i \ln(1-p(G(i)))/(1-p(B(i))) + \ln(p(G)/p(B)) \end{aligned}$$

Esta es otra forma de representar la expresión (2.30) y satisface el supuesto de regresión logística.

La única dificultad, comparada con la regresión ordinaria es que no es posible utilizar el método de los mínimos cuadrados ordinarios para calcular los coeficientes \mathbf{w} . Se utiliza el método de la máxima verosimilitud para estimar estos coeficientes y se emplea Newton-Raphson para resolver las ecuaciones que aparecen. Con el poder de los ordenadores modernos esto no es un problema, sobre todo para las grandes muestras que normalmente se utilizan cuando se construyen credit scoring.

Uno de los resultados sorprendentes que se obtienen es que, teóricamente, la regresión logística es mejor para aplicarla a las muchas clases de distribuciones que las resoluciones de clasificación de la regresión lineal. Y, sin embargo, cuando se realizan comparaciones de los desarrollos de scoring utilizando los dos métodos con el mismo conjunto de datos, aparecen unas diferencias muy pequeñas en sus

Capítulo 2. Una revisión de los métodos estadísticos empleados en...

clasificaciones. En términos generales, la diferencia entre los métodos es, que la regresión lineal trata de ajustar la probabilidad p de impago como una combinación lineal de los atributos mientras la regresión logística trata de ajustar el $\ln(p/(1-p))$ como una combinación lineal de los atributos.

2.6 Otros métodos de regresión no lineal

Otras dos funciones no lineales han sido utilizadas para construir un credit scoring. La primera es la función probit, que fue usada para un credit scoring por Grablowsky y Talley [8] (1981). En el análisis probit, si $N(x)$, representa la función de una distribución normal, esto es,

$$N(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-y^2/2} dy$$

Entonces la clave consistirá en estimar $N^{-1}(p_i)$ como una función lineal de las características de la aplicación, así consideraremos

$$N^{-1}(p_i) = \mathbf{w}'\mathbf{x}_i = w_0 + w_1 x_{1i} + w_2 x_{2i} + \dots + w_p x_{pi} \quad (2.35)$$

De nuevo p_i toma valores entre 0 y 1, y $N^{-1}(p_i)$ toma valores entre $-\infty$ y $+\infty$ y esto permite a la función lineal variar sobre todos los valores. Se podría pensar sobre el método probit como que la variable

Otros métodos de regresión no lineal

$$W = \mathbf{w}'\mathbf{x}_i = w_1 X_1 + w_2 X_2 + \dots + w_p X_p,$$

es una medida de la bondad de una aplicación, si tratamos de predecir si una nueva aplicación es buena o mala esto va a depender de si el valor de W es mayor o menor que una cierta barrera con un valor C . Si suponemos que C no está fijada pero es una variable con una distribución normal estándar, entonces la probabilidad de que la nueva aplicación sea buena es la misma que la de la ecuación (2.35). Como en el caso de la regresión logística, se utiliza la estimación de la máxima verosimilitud para obtener los valores de \mathbf{w} y de nuevo habría que utilizar técnicas iterativas para resolver estos valores. Hay situaciones en las que un proceso de iteración puede no converger, y entonces deberían probarse procesos iterativos alternativos.

El otro método usado en credit scoring y que es muy común en modelos económicos es el análisis tobit. La transformación tobit supone que se puede estimar p_i como

$$p_i = \max\{\mathbf{w}'\mathbf{x}_i, 0\} = \max\{w_0 + w_1 x_{1i} + w_2 x_{2i} + \dots + w_p x_{pi}, 0\}. \quad (2.36)$$

En este caso, se trata de limitar los valores entre los dos lados de la ecuación (2.21) del método de regresión del análisis discriminante, limitando además, que el lado derecho sea positivo. Hay muchas situaciones económicas donde la variable que estamos interesados en mostrar su valor solo tiene significado si es positiva, los análisis estadísticos de las regresiones tobit sirven para ser aplicados en estos casos.

En el contexto de credit scoring encontramos algunas cosas insatisfactorias sobre la asimetría de la transformación tobit. Se termina

Capítulo 2. Una revisión de los métodos estadísticos empleados en ...

con la dificultad de estimar probabilidades negativas pero no con probabilidades mayores de uno. Un modelo más simétrico sería

$$p_i = \min\{1, \max\{\mathbf{w}'\mathbf{x}_i, 0\}\}, \quad (2.37)$$

pero desafortunadamente el análisis y la estimación de los parámetros para esta transformación no es posible con los métodos disponibles.

Los métodos probit y tobit no encuentran muchos adeptos entre los estudiosos de credit scoring ya que estos están más preocupados por encontrar la decisión correcta para casos en que los valores de decisión no están muy claros que en el hecho de que la probabilidad de que un cliente sea malo sea claramente 0.05 ó -0.05. De todas formas, esto tiene que ser recordado por las técnicas estadísticas que no conocen esto, y su objetivo es minimizar la suma total de errores. Así, un caso extraño debe afectar los parámetros considerablemente y hacer cambiar las clasificaciones en la región de dificultad. Esto sugiere que debe ser un método en dos etapas. En la primera etapa, se trata de estimar la región de corte posible, mientras en la segunda etapa, se trata de proceder a la clasificación en las regiones de la forma más exacta posible.

2.7 Clasificación en árbol (método de partición recursiva)

Un método estadístico completamente diferente para la clasificación y discriminación es el denominado clasificación en árbol, algunas veces también llamado algoritmo de partición recursiva. La idea es dividir el conjunto de respuestas de las aplicaciones en diferentes conjuntos e identificar cada uno de ellos como bueno o malo dependiendo de lo que

Clasificación del árbol (método de partición recursiva)

mayoritariamente ocurra en ese conjunto. La idea fue desarrollada para problemas de clasificación general por Breiman y Friedman en 1973. Estos autores describieron un número de aplicaciones estadísticas (pero no de credit scoring) en su libro (Breiman 1984) [1]. Estas técnicas fueron empleadas para credit scoring rápidamente por Makowski en 1985 [11] y Coffman en 1986 [2]. La idea aparece recogida también por la literatura de la inteligencia artificial. Ideas similares fueron utilizadas en problemas de clasificación, y actualmente se dispone de abundante software para su implementación.

El conjunto de asignaciones crediticias A es dividido primero en dos subconjuntos basándonos en la muestra de aplicaciones anteriores, estos dos nuevos subconjuntos de atributos de la aplicación son mucho más homogéneos con respecto al riesgo de crédito que los del conjunto original. Cada uno de estos conjuntos es de nuevo dividido en dos para obtener cada vez subconjuntos más homogéneos, y así sucesivamente. Esta es la razón por la cual el método se llama partición recursiva. El proceso se detiene cuando todos los subconjuntos cumplen los requerimientos para ser nodos terminales del árbol. Cada nodo terminal es clasificado como componente de los conjuntos A_G o A_B .

Hay tres decisiones importantes en el proceso de realización de la clasificación en árbol:

- Cuál es la regla a utilizar para dividir los conjuntos en dos, la denominada regla de división
- Cómo decidir cuando un conjunto es un nodo terminal: la regla de detención
- Cómo asignar los nodos terminales a las categorías de buenos y malos

Capítulo 2. Una revisión de los métodos estadísticos empleados en ...

La decisión de asignación en bueno o malo es la más fácil de realizar. Normalmente, se asignaría el nodo como bueno si la mayoría de los casos de la muestra en este nodo son buenos. Una alternativa es minimizar el coste de equívoco en la clasificación. Si D es el débito en el que se incurre por clasificar un malo como bueno y L es el coste por clasificar un bueno como malo, entonces seguiremos el criterio propuesto por Edelman, Crook y Thomas [5] donde se minimiza el coste si se clasifica el nodo como bueno cuando el ratio de buenos sobre malos en este nodo en la muestra excede (D/L) .

Lo que se busca es la mejor división para cada característica que nos permita obtener alguna medida que nos indique la bondad de la división. Entonces se decide que división de las características es mejor según esta medida. Para cualquier característica continua X_i , se busca en las divisiones $\{x_i < s\}$ donde quedarán clasificados los clientes insolventes, y en $\{x_i \geq s\}$ donde encontraremos a los buenos clientes para todos los valores de s y buscamos el valor de s cuya medida es la mejor. Si X_i es una variable categórica, entonces buscamos todas las posibles divisiones de las categorías en dos y probamos la medida según estas divisiones diferentes. Normalmente, se pueden ordenar las categorías en ratios crecientes y conocer cuál será la mejor división que los ordene en dos grupos. Pero no sabemos qué medida utilizar. La más común es la basada en el estadístico de Kolmogorov-Smirnov, pero hay al menos otras cuatro, las basadas en el índice básico de impureza, el índice de Gini, el índice de entropía, y la suma media de cuadrados.

2.7.1 Medidas de división del método del árbol

El estadístico de Kolmogorov-Smirnov

Para una característica continua X_i , denotaremos por $F(s|G)$ a la función de distribución condicional de X_i entre los buenos clientes y $F(s|B)$ es la respectiva entre los malos. Suponemos que los malos clientes tienen una mayor propensión a tomar valores pequeños de X_i que los buenos clientes y que las definiciones de los costes D y L son como explicamos en la sección anterior. La regla que dividiría en el valor es la que minimiza

$$L F(s|G)p(G)+D (1-F(s|B))p(B) \quad (2.38)$$

Si $Lp(G)=Dp(B)$, esto es lo mismo que elegir la distancia de Kolmogorov-Smirnov entre las dos distribuciones: se quiere minimizar $F(s|G)-F(s|B)$ o más obviamente se quiere maximizar $F(s|B)-F(s|G)$.

Si se piensa en los dos subgrupos como el de la derecha (r) y la izquierda (l), entonces esto es como maximizar la diferencia entre $p(l|B)$, la probabilidad de que un malo aparezca en el grupo de la izquierda ($F(s|B)$ en el caso continuo), y $p(l|G)$, la probabilidad de que un buen cliente aparezca en la izquierda ($F(s|G)$ en el caso continuo). Se puede reescribir $p(l|B)$ como

$$p(B|l)p(l)/p(B)$$

utilizando el teorema de Bayes, y así es más fácil el cálculo. Para variables continuas y categóricas, el criterio de Kolmogorov-Smirnov (KS) es el siguiente. Buscamos la división entre izquierda y derecha que maximiza

$$KS=|p(l|B)-p(l|G)|=|(p(B|l)/p(B))-(p(G|l)/p(G))|p(l) \quad (2.39)$$

El índice básico de impureza $i(v)$

Este índice nos proporciona una medida de impureza para una clase completa, que trata de determinar cuánta de esta hay en cada nodo v del árbol. Si se divide el nodo principal en un nodo a la izquierda l y un nodo a la derecha r con las proporciones $p(l)$ y $p(r)$ respectivamente, se puede medir el cambio de impureza que la división ha ocasionado de la siguiente manera

$$I=i(v)-p(l)i(l)-p(r)i(r) \quad (2.40)$$

El máximo de esta diferencia es el máximo del cambio de impureza, lo cual significa que los nuevos nodos son mucho más puros. Por lo tanto, lo que buscamos es elegir la división que maximice esta expresión y esto es equivalente a minimizar $p(l)i(l)+p(r)i(r)$. Evidentemente, si la división no permitiera encontrar una diferencia positiva, entonces quedaría descartada.

Otro aspecto importante en el cálculo de este índice es el valor de $i(v)$ que queda determinado por

Clasificación en árbol (método de partición recursiva)

$$i(v) = \begin{cases} p(G/v) & \text{si } p(G/v) \leq 0.5 \\ p(B/v) & \text{si } p(B/v) < 0.5 \end{cases} \quad (2.41)$$

donde v es el conjunto completo antes de ninguna división.

El índice de Gini

En vez de suponer linealidad de la probabilidad de impureza, el índice de Gini es cuadrático y así pone relativamente más peso en la pureza de los nodos. Este se define como

$$i(v) = p(G/v)p(B/v) \quad (2.42)$$

El índice de entropía

Otro índice no lineal es el de la entropía, donde

$$i(v) = -p(G/v)\ln(p(G/v)) - p(B/v)\ln(p(B/v)) \quad (2.43)$$

Como su nombre indica, el valor del índice nos da la entropía, o la cantidad de información en la división entre los buenos y malos clientes en el nodo. Este es una medida de cuantas formas diferentes se podría explicar bien con la actual división de buenos y malos en el nodo, y esto está relacionado con la información estadística utilizada en inmensas medidas de clasificación.

El método de la maximización de la suma media de cuadrados

La última medición que consideramos no es un índice pero lo utilizamos como tal, con la prueba de la χ^2 , con la que probamos si la proporción de buenos clientes es el mismo en los dos nuevos nodos resultantes de la división. Si el estadístico χ^2 es grande, normalmente decimos que la hipótesis es nula; las dos proporciones no son la misma. El valor más grande, el más improbable de que las proporciones sean las mismas, sin embargo se podría reinterpretar para decir que es la mayor diferencia entre ellas. Esto es lo que nosotros vamos buscando en una división, así que procederíamos con la siguiente prueba.

Si $n(l)$ y $n(r)$ son los números totales que tenemos a la derecha y a la izquierda, entonces se maximiza

$$\chi^2 = n(l)n(r) \left(\frac{(p(G|l) - p(G|r))^2}{n(l) + n(r)} \right)$$

Hay otros métodos de división. Breiman (1984) [1] sugiere que un criterio mejor para buscar la siguiente división sería considerar cual es la situación que hay después de las r generaciones de divisiones. Esto no nos conduce a un perfeccionamiento inmediato pero si es una estrategia importante a largo plazo de esta división. Este es el caso que normalmente se plantea utilizando diferentes características en diferentes niveles del árbol, y también otras características diferentes van por delante en diferentes subconjuntos en un mismo nivel del árbol. Esto nos dice que hay una relación no lineal entre las características de la aplicación.

Clasificación en árbol (método de partición recursiva)

Sin embargo, sobre esto tenemos que decir que cuando paramos el árbol y reconocemos un nodo como un nodo terminal, es quizás más honesto hablar de parada o regla de corte. Esto enfatiza que lo que se espera es que el árbol inicial sea demasiado grande y necesitamos cortarlo para que el árbol sea robusto. Si se obtiene un árbol donde cada nodo terminal tiene solo un caso perteneciente a la muestra, entonces el árbol sería un perfecto discriminador de la muestra pero sería un muy pobre clasificador para cualquier otro conjunto. Entonces decimos que un nodo es un nodo terminal por dos razones. La primera razón es el número de casos del nodo, si este es muy pequeño no tiene sentido volver a dividirlo. Esto se considera así cuando el nodo tiene menos de 10 casos. La segunda razón es el valor de la medida de la división, si la mejor división entre dos nodos hijos es fuertemente diferente a cualquier otro valor de medida entonces lo dejaremos aquí. Que sea fuertemente diferente en este contexto significa que hay que dar un valor barrera de la diferencia de medida que denominaríamos β .

Habiendo obtenido el árbol más grande, se pueden cortar y retocar algunas divisiones. La mejor manera de hacer esto es utilizar una extensión de la muestra, que no fue utilizada en la construcción del árbol. Esta muestra es utilizada para estimar empíricamente la pérdida esperada para diferentes cortes del árbol. Utilizando esta extensión de la muestra y una clasificación de árbol T , definimos $T_G(T_B)$ como el conjunto de nodos que son clasificados como buenos (malos). Entonces $r(t,B)$ es la proporción de la extensión de la muestra que está en el nodo t y que es clasificada como malo y $r(t,G)$ es la proporción de la muestra que está en t y es buena. Entonces un estimador de la pérdida esperada será

$$r(T) = \sum_{t \in T_G} D r(t,B) + \sum_{t \in T_B} L r(t,G) \quad (2.44)$$

Si $n(T)$ es el número de nodos en el árbol T , definimos $c(T)=r(T)+dn(T)$ y cortamos un árbol T^* para buscar todos los subárboles de T^* y elegir el árbol T que minimice $c(T)$. Si $d=0$, nuestro árbol final coincidirá con el árbol original sin ningún corte, mientras que si d es muy grande, el árbol consistirá en un único nodo. Así, el valor de d nos indicará cuánto de grande debe ser el árbol que buscamos.

2.8 El método de vecindad

El método de vecindad es un método no paramétrico estándar para los problemas de clasificación, sugerido por primera vez por Fix y Hodges (1952) [7]. Este se aplicó a credit scoring primero por Chatterjee y Barcun (1970) [3] y más tarde por Henley y Hand (1996) [9]. La idea consiste en que tenemos dos clases de elementos representados en espacios separados (espacio 1, y espacio 2) y deseamos decidir en cual de los dos clasificaremos un nuevo elemento.

La técnica de la vecindad más cercana, en esencia, es tomar una decisión basada en la distancia más corta de ese elemento a cada una de las muestras vecinales ya representadas, como podemos ver en la figura 1.

El método de vecindad

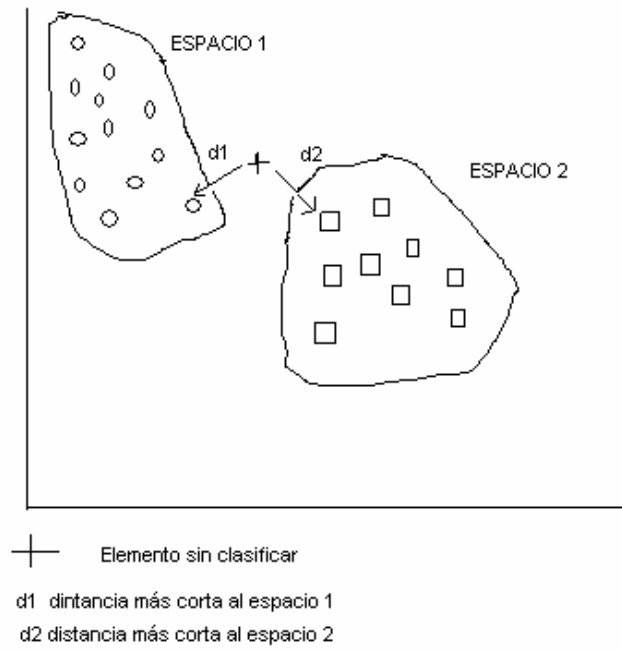


Figura 1

Así, se mide la distancia al punto más cercano del espacio 1 y del espacio 2 y el nuevo elemento quedará clasificado en el espacio a cuya distancia sea menor.

Con esta técnica el nuevo elemento quedará asignado a la clase 1 si la diferencia es negativa y a la clase 2 si es positiva.

Pero este método puede presentar problemas si un elemento de alguna de las dos muestras está mal clasificado, como se ve en la figura 2.

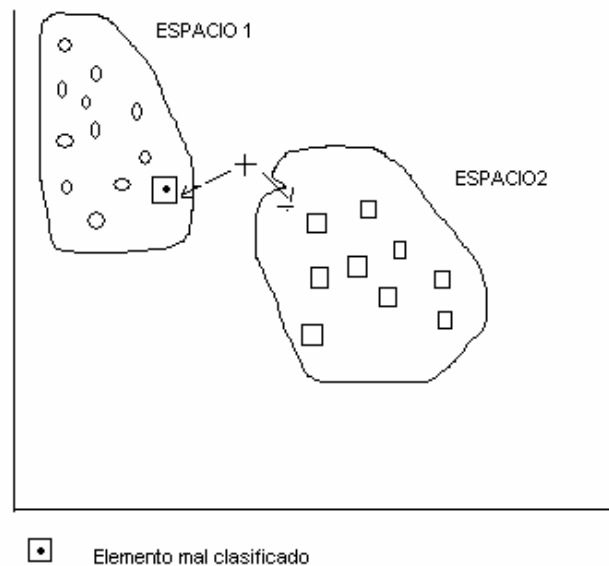


Figura 2

Si el elemento por clasificar mide su distancia con el erróneamente clasificado el resultado que obtenemos es una mala clasificación. La solución a este problema es tomar varias distancias, de manera que el efecto de un elemento mal clasificado pierda su influencia, esto es lo que se denomina clasificación por k-vecindades, donde k es el número de medidas que se toman.

Los tres parámetros necesarios para utilizar este método son la métrica, cuántas aplicaciones k constituyen el conjunto de los vecinos próximos, y cuántos de estos deberían ser buenos para que la aplicación se clasificase como buena. Normalmente, la respuesta a esta última pregunta es que si la mayoría de los vecinos son buenos, la aplicación es buena; en otro caso, la aplicación se clasifica como mala. Sin embargo, si como en la

El método de vecindad

sección (2.2), el promedio de la pérdida por impago es D y el promedio del beneficio perdido por rechazar a un bueno es L , entonces se puede clasificar una nueva aplicación como buena si al menos (D/L) de los vecinos próximos son buenos. Este criterio minimizaría la pérdida esperada si la probabilidad de que una nueva aplicación sea buena es la proporción de vecinos que son buenos, como quedó visto en la expresión (2.5).

El método de la vecindad incluye, además, el problema de encontrar el mejor método de medición y la elección de la métrica es claramente crucial en este tipo de problemas. En la práctica se utilizan varias medidas.

La medida más básica es la distancia de Hamming, esta consiste en lo siguiente:

Tenemos dos vectores $\mathbf{x}_1=(x_1,\dots,x_p)$ e $\mathbf{y}_2=(y_1,\dots,y_p)$, la distancia de Hamming se calcula mediante la suma de la diferencia pareada entre las componentes de los vectores, esto es,

$$H = \sum \|x_i - y_i\|$$

Otra medida muy común es la distancia Euclídea, si consideramos los dos vectores anteriores, y queremos encontrar la distancia entre ambos $d(\mathbf{x},\mathbf{y})$, se utilizará la siguiente expresión:

$$d(\mathbf{x},\mathbf{y}) = \sqrt{\left(\sum_{i=1}^n (x_i - y_i)^2\right)}$$

Capítulo 2. Una revisión de los métodos estadísticos empleados en ...

Una versión simplificada de esta medida es la llamada medición de Manhattan que es una simplificación de la distancia Euclídea y se calcula de la siguiente manera:

$$D_M = \sum |x_i - y_i|$$

El efecto de esta distancia, aparte de que obviamente es más fácil de calcular, es que supone que tomando un vector con origen en el centro de un cuadrado, las distancias desde ese punto a la frontera del cuadrado son iguales. Esto lo ilustramos en la siguiente figura:

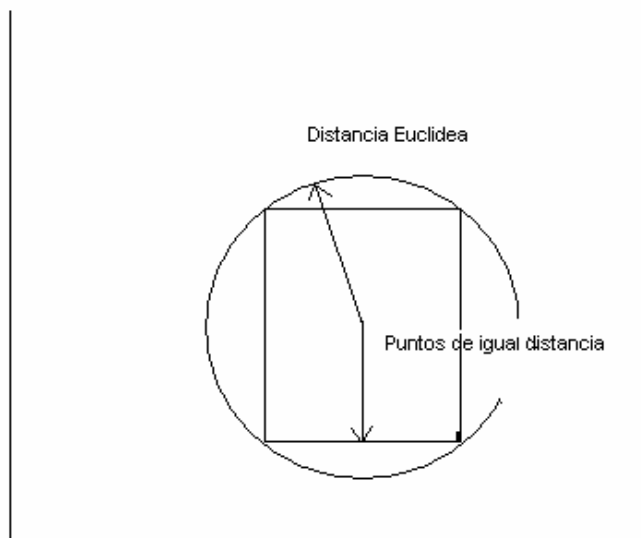


Figura 3

El método de vecindad

El círculo mostrado nos indica la distancia Euclídea desde el punto central a otros puntos equidistantes mostrada por la dimensión del vector. La distancia de Manhattan introduce un error que es aceptado como compensación a la rapidez en el cálculo.

Por último, otra medida utilizada es la distancia a la medida geométrica que también se deriva de la distancia Euclídea, pero cometiendo aún un error mayor, esta consiste en que la distancia entre dos vectores se define como el máximo de las diferencias entre cada elemento de los mismos, esto es:

$$D_c = \text{Max } |x_i - y_i|$$

Fukanaga y Flick [6] (1984) introdujeron una métrica general definida a partir de la expresión

$$d(\mathbf{x}_i, \mathbf{y}_i) = ((\mathbf{x}_i - \mathbf{y}_i)' \mathbf{A} (\mathbf{x}_i - \mathbf{y}_i))^{1/2} \quad (2.45)$$

donde \mathbf{A} es una $p \times p$ matriz simétrica definida positiva. A dicha matriz, se le denomina métrica local si depende del vector y se denomina métrica global si es independiente del mismo. La dificultad de la métrica local es que la tónica de las características del conjunto de entrenamiento no son apropiadas en general, y por esto muchos autores se concentran en la métrica global.

Los exámenes más detallados del método de vecindad en el contexto de credit scoring los realizaron Henley y Hand [9] (1996), y se

Capítulo 2. Una revisión de los métodos estadísticos empleados en ...

concentraron en métricas que mezclaban la distancia Euclídea y la distancia en la dirección que mejor separa los buenos y los malos clientes. Esta dirección, se toma a partir de la función lineal discriminante de la sección (2.3). Así, si \mathbf{w} es el vector p -dimensional que define la dirección, dada en (2.20), Henley y Hand sugirieron una métrica de la forma

$$d(\mathbf{x}_i, \mathbf{y}_i) = \{(\mathbf{x}_i - \mathbf{y}_i)'(\mathbf{I} + \lambda(\mathbf{w}\mathbf{w}'))(\mathbf{x}_i - \mathbf{y}_i)\}^{1/2} \quad (2.46)$$

donde \mathbf{I} es la matriz identidad. En este trabajo anteriormente citado, los autores realizan un gran número de experimentos para identificar cual es la elección más conveniente de λ . De forma similar, eligen k , el número ideal de vecindades, y experimentan con varios valores de k . Sin embargo no obtuvieron grandes variaciones en el resultado acerca de la mejor elección de λ . Esta se movía en un rango entre 1,4 y 1,8. La elección de k dependía claramente de la forma de la muestra de entrenamiento, y en algunos casos la k elegida daba notables diferencias.

Los métodos de vecindad, no han sido utilizados habitualmente en credit scoring, en comparación con la amplia utilización de los métodos de regresión logística y lineal. Estos métodos tienen algunas características potencialmente atractivas para la implementación actual. Sería fácil adaptar dinámicamente la muestra añadiendo casos nuevos a esta en cuando conociéramos si estos son buenos o malos e ir añadiendo uno a uno estos casos hasta conseguir una muestra más grande. Con esto se superaría parcialmente la necesidad de adaptar el sistema de scoring regularmente debido a los cambios de la población, sin embargo una métrica d como la obtenida en (2.46) necesitaría ser adaptada teniendo en cuenta los cambios de la población y esto no sería dinámico. En cuanto a

El método de vecindad

Lo concerniente a la gran cantidad de cálculos que hay que realizar para clasificar los nuevos casos en buenos y malos clientes, según los valores de k se puede solventar con los ordenadores modernos que realizan los cálculos en pocos segundos. Sin embargo, en muchos casos, encontrar una buena métrica viene a ser equivalente a la construcción de un scoring por el método de la regresión.

2.9 El método de la discriminación múltiple

Hay muchas circunstancias en credit scoring en las cuales las aplicaciones se tienen que dividir en más de dos grupos. Hay que ser cuidadoso con esto porque la formación de grupos puede deberse al modo en que la entidad toma sus decisiones. Algunas entidades pueden considerar, según aplicaciones anteriores, buenos a aquellos que quiere como clientes, a otros no los quiere porque incurren en impago, y otros que tampoco quiere como clientes porque el beneficio que obtienen no les compensa. Otra forma de dividir los clientes puede ser en buenos, con impago, y en bancarota, sabiendo que las características de los dos últimos grupos son diferentes.

Si se quiere dividir la población en g grupos, se pueden utilizar los métodos utilizados anteriormente aplicando las modificaciones convenientes. Por ejemplo las modificaciones para poder aplicar el análisis discriminante serían las siguientes:

La base del análisis discriminante es una base de datos que consiste en los valores de n observaciones/individuos/casos en los que se han medido p variables cuantitativas X_1, \dots, X_p . Los n casos están agrupados en g grupos establecidos por los valores de una variable dependiente $U: G_1, \dots, G_g$. El grupo G_i consiste en n_i casos ($i=1, \dots, g$). Por tanto,

Capítulo 2. Una revisión de los métodos estadísticos empleados en ...

$n = \sum_{i=1}^g n_i$. La tabla de datos establecida por las observaciones tendrá n filas y p columnas. Cada fila puede ser considerada como un punto en un espacio de p dimensiones donde las coordenadas de cada punto se obtendrán a partir de los valores en las p variables para el individuo correspondiente: $x_{ij} = (x_{ij,1}, x_{ij,2}, \dots, x_{ij,p}) \in \mathbb{R}^p$, $i=1, \dots, g$, $j=1, \dots, n_i$. A partir de la representación de las n filas se trata de extraer un nuevo espacio de pequeña dimensión tal que, al proyectar la nube de puntos sobre dicho espacio, por un lado, los puntos correspondientes a individuos del mismo grupo estén próximos y, por otro, los correspondientes a individuos de distintos grupos estén alejados. Los ejes de este nuevo espacio se llamarán funciones discriminantes. La expresión de una función discriminante Y_k será

$$Y_k = a_{k1} X_1 + a_{k2} X_2 + \dots + a_{kp} X_p \quad \{p\} \quad k=1, \dots, s \quad (2.47)$$

suponiendo que se extraen s funciones discriminantes. Las puntuaciones discriminantes para el individuo j del grupo G_i son

$$y_{ij,k} = a_{k1} x_{ij,1} + a_{k2} x_{ij,2} + \dots + a_{kp} x_{ij,p} \quad k=1, \dots, s \quad (2.48)$$

A partir de las puntuaciones discriminantes, un individuo para el que se conoce a cuál de los grupos pertenece, será clasificado en uno de ellos. El porcentaje de casos correctamente clasificados será un índice de la efectividad de las funciones discriminantes. Si dichas funciones son efectivas sobre la muestra observada, es de esperar que también lo sean

El método de la discriminación múltiple

cuando se trate de clasificar a un individuo para el que se desconoce a cuál de los grupos pertenece.

Si trabajamos con los coeficientes estandarizados el vector

$$\mathbf{x}_{ij} = [x_{ij,1}, x_{ij,2}, \dots, x_{ij,p}]' \in \mathbb{R}^p, \quad i=1, \dots, g, \quad j=1, \dots, n_i$$

es sustituido por

$$\mathbf{z}_{ij} = [z_{ij,1}, z_{ij,2}, \dots, z_{ij,p}]' \in \mathbb{R}^s$$

Y cada grupo puede ser representado por sus valores promedio en las funciones discriminantes. Es decir, el grupo $G_i (i=1, \dots, g)$ está representado por el vector $\mathbf{z}_i = [z_{i,1}, z_{i,2}, \dots, z_{i,p}]' \in \mathbb{R}^s$, el llamado centroide de dicho grupo.

Según Krzanowski [10]

$$n_i \sum_{k=1}^s (v_{i,k} - \bar{z}_{i,k})^2 \text{ sigue una distribución } \chi_s^2$$

donde $\mathbf{v}_i = (v_{i,1}, v_{i,2}, \dots, v_{i,s})$ es la verdadera media poblacional del grupo G_i (con respecto a las funciones discriminantes), y χ_s^2 es la distribución chi-cuadrado con s grados de libertad. Por tanto si $\chi_s^2(\alpha)$ es el valor excedido por exactamente el $(100\alpha)\%$ de una distribución χ_s^2 (obtenible de una tabla χ_s^2), entonces

Capítulo 2. Una revisión de los métodos estadísticos empleados en ...

$$P\left(\sum_{k=1}^s (v_{i,k} - \overline{z_{i,k}})^2 < (\chi_s^2(\alpha)/n_i)\right) = 1 - \alpha$$

Observamos que $P\left(\sum_{k=1}^s (v_{i,k} - \overline{z_{i,k}})^2 < (\chi_s^2(\alpha)/n_i)\right)$ define un entorno alrededor del punto z_i , y de radio $(\chi_s^2(\alpha)/n_i)^{1/2}$. En otras palabras, el entorno alrededor del punto z_i , y de radio $(\chi_s^2(\alpha)/n_i)^{1/2}$ define una región de confianza del $100(1-\alpha)\%$ para la verdadera media poblacional del grupo G_i . Un razonamiento parecido nos muestra que es de esperar que el $100(1-\alpha)\%$ de la población G_i caiga en el entorno alrededor del punto z_i , y de radio $(\chi_s^2(\alpha)/n_i)^{1/2}$:

$P\left(\sum_{k=1}^s (q_{i,k} - \overline{z_{i,k}})^2 < (\chi_s^2(\alpha)/n_i)\right)$ para un individuo q_i de la población $G_i = 1 - \alpha$

En otras palabras, la distancia $\left(\sum_{k=1}^s (q_{i,k} - \overline{z_{i,k}})^2\right)^{1/2}$ de un individuo cualquiera q_i de la población G_i hasta su centroide sigue una distribución χ_s^2 .

Entonces, para poder clasificar al individuo ij ($i=1,\dots,g, j=1,\dots,n_i$) nos basamos en lo precedente y calculamos para cada población $G_l, l=1,\dots,g$, la probabilidad $\alpha \in (0,1)$ tal que el punto $z_{ij} \in R^s$ está justo en el límite

El método de la discriminación múltiple

de la región en la que podemos esperar el 100(1- α)% de dicha población. Es decir,

$$P\left(\chi^2_s \leq \sum_{k=1}^s (z_{ij,k} - \bar{z}_{l,k})^2\right) = 1 - \alpha \quad (2.49)$$

Escribimos la probabilidad en (2.49) como probabilidad condicionada: $P(D \leq d_l | G_l) = 1 - \alpha$. La letra D denota la palabra "Distancia", y d_l representa el número $\sum_{k=1}^s (z_{ij,k} - \bar{z}_{l,k})^2$, y la expresión G_l significa que consideramos cualquier individuo del grupo G_l . Si $P(D \leq d_l | G_l)$ es pequeña, entonces es bastante probable que el caso ij pertenezca al grupo G_l .

Está claro que $P(D > d_l | G_l) = 1 - P(D \leq d_l | G_l) = \alpha$, la cual es la probabilidad que aparece en los análisis discriminantes en el programa SPSS. Si $P(D > d_l | G_l)$ es grande, entonces es bastante probable que el caso ij pertenezca al grupo G_l . Luego, usamos el teorema de Bayes para calcular la probabilidad

$$P(G_l | D > d_l) = \frac{(P(D > d_l | G_l)P(G_l))}{(P(D > d_1 | G_1)P(G_1) + \dots + P(D > d_g | G_g)P(G_g))} \quad (2.50)$$

donde normalmente se supone que $P(G_1) = \dots = P(G_g) = 1/g$ (los grupos son a priori equiprobables). (En el SPSS se escribe la probabilidad de (2.30) como $P(G_l | D = d_l)$). La probabilidad $P(G_l | D > d_l)$ es la probabilidad de

Capítulo 2. Una revisión de los métodos estadísticos empleados en...

que un individuo arbitrario pertenezca al grupo G_l . El Análisis Discriminante clasifica el caso ij en el grupo G_l para el que la probabilidad $P(G_l|D>d_l)$ sea máxima¹².

. Este es el llamado grupo mayor para el caso ij . El grupo G_k para el que la probabilidad correspondiente $P(G_k|D>d_k)$ es la segunda mayor, se llamará el segundo mayor, etc.

Suponemos que $c(i,j)$ es el coste de asignar una aplicación que responde como los del grupo j al grupo i , y sabemos que p_j es la proporción de población incluida en el grupo j . Entonces $p(\mathbf{x}|j)$ es la probabilidad de que las aplicaciones en el grupo j tengan los atributos \mathbf{x} (similar a $p(\mathbf{x}|G)$ en el punto 2.2). Entonces la probabilidad de que una aplicación con los atributos \mathbf{x} esté en el grupo j , que expresamos como $P(j|\mathbf{x})$, satisface

$$P(j|\mathbf{x}) = \frac{p_j p(\mathbf{x}|j)}{\sum_i p_i p(\mathbf{x}|i)} \quad (2.51)$$

Si se quiere minimizar la pérdida esperada, se asignaría la aplicación con atributos \mathbf{x} al grupo i si

$$\sum_j c(i,j) p_j p(\mathbf{x}|j) < \sum_j c(k,j) p_j p(\mathbf{x}|j) \quad \text{para todo } k, k \neq i \quad (2.52)$$

¹² Nótese que esta clasificación coincide con la que está basada (solamente y directamente) en la distancia $\sum_{k=1}^s (z_{ij,k} - \bar{z}_{l,k})^2$

El método de la discriminación múltiple

De forma similar, se puede extender al método de regresión logística para clasificar tres o más grupos siguiendo la ecuación (2.31). A pesar de que los métodos estadísticos para la clasificación en múltiples grupos están bien desarrollados, hay muy pocas aplicaciones de estos en credit scoring puesto que en la mayoría de las entidades bancarias su interés se centra en una división de los clientes en solventes o insolventes lo cual nos lleva casi siempre a una división de los mismos en solo dos grupos.

Bibliografía

[1] : Breiman, L., Friedman, J., Olshen, R., Stone, C. Classification and Regression trees, Wadsworth, belmont, CA. (1984).

[2] : Coffman, J. The Proper Role of Tree Analysis in Forecasting the Risk Behavior of Borrowers. MDS Reports, Management Decision System, Atlanta, GA, 3, 4, 7, 9.(1986).

[3] : Chatterjee, S., Barcun, S. A nonparametric approach to credit screening. J.Amer.Statist. Assoc., 65, 150-154. (1970).

[4] : Fisher, R. The use of multiple measurements in taxonomic problems, Ann. Eugenics. 7, 179-188. (1936).

[5] : Thomas, L., Edelman, D., Crook, J. Credit Scoring and its Applications SIAM monographs on mathematical modeling and computation. (2002).

Capítulo 2. Una revisión de los métodos estadísticos empleados en ...

[6] : Fukanaga, K., Flick, T. An optimal global nearest neighbour metric. IEEE Trans. Pattern Anal. Mach.Intell. PAMI-1. 25-37. (1984).

[7] : Fix, E., Hodges, J. Discriminatory Analysis, Nonparametric Discrimination, Consistency Properties. Report 4, Project 21-49-004, School of Aviation Medicine, Randolph Field; TX. (1952).

[8] : Grablowsky, B., Talley, W. Probit and discriminant functions for classifying credit applicants: A comparison. J. Econom.Business, 33, 254-261. (1981).

[9] : Henley, W., Hand, D. A k-NN classifier for assessing consumer credit risk. Statistician, 65, 77-95. (1996).

[10] : Krzanowski, W.J. Principles of Multivariate Analysis. Clarendon Press, Oxford (1998).

[11] : Makowski, P. Credit scoring branches out. Credit World, 75, 30-37. (1985).

[12] : Reichert, A., Cho, C., Wagner,G. An examination of the conceptual issues involved in developing credit scoring models, J. Business Econom. Statist., 1, 101-114. (1983).

[13] : Wiginton, J. A note on the comparison of logit and discriminant models of consumer credit behaviour. J. Financial Quantitative Anal., 15, 757-770. (1980).

Capítulo III

UNA REVISIÓN SOBRE ALGUNOS MÉTODOS NO ESTADÍSTICOS EMPLEADOS EN LA CONSTRUCCIÓN DE CLASIFICACIONES DE CRÉDITO

3.1 Introducción

La idea original para desarrollar un credit scoring fue utilizar el análisis estadístico de una muestra de clientes anteriores para ayudarnos a decidir como se puede clasificar a un nuevo cliente satisfactoriamente. Se pueden utilizar también métodos no estadísticos para resolver el mismo problema. Estos se han utilizado para cualquier problema de clasificación en los últimos 25 años. Hasta la década de los 80, los únicos métodos utilizados fueron los estadísticos, pero en 1981 Freed y Glover [7] realizaron un estudio donde encontraron que la función lineal de las características que mejor discrimina entre grupos se podía modelar como

Capítulo 3. Una revisión sobre algunos métodos no estadísticos ...

un problema de programación lineal. Si se quiere tomar el número de casos donde la discriminación es incorrecta como una medida de la bondad del ajuste, entonces se tienen que introducir variables enteras en la programación lineal, esto lo veremos en la Sección (3.3).

En la década de los años 70 se produjeron grandes avances en el área de la inteligencia artificial, estos avances trataban de programar las computadoras para que reprodujeran las destrezas naturales de los humanos. Uno de los éxitos obtenidos fue la construcción de los llamados sistemas expertos, en los cuales se introducían en el ordenador una base de datos con información de una cierta materia obtenida de expertos con cierta credibilidad y un mecanismo para generar reglas a partir de esta información. Los programas de los ordenadores entonces utilizaban esta combinación para analizar nuevas situaciones y llegar a una conclusión sobre esta nueva situación, la cual sería tan buena como la que los expertos hubieran obtenido. Los sistemas pilotos se desarrollaron en el ámbito del diagnóstico médico, que es, en esencia, un problema de clasificación, de modo que es obviamente trasladable a un sistema de credit scoring.

En la década de los 80 aparece otra variante de inteligencia artificial basada en los métodos de problemas de clasificación: las redes neuronales. Estas han experimentado un desarrollo rápido y continuado ya que es un área de avances muy activa. Las redes neuronales son algoritmos que modelan los procesos de decisión al igual que las células del cerebro utilizan neuronas para comunicarse unas con otras y así llegar a un mecanismo de aprendizaje. Es un sistema de unidades de proceso que están conectadas y, donde cada una de ellas emite una señal de salida cuando recibe señales de entrada.

El sistema consiste en dar un conjunto de datos donde cada uno tiene asociado un conjunto de señales de entrada y una señal de salida

específica. Se trata de aprender, a partir de estos datos, cómo reproducir la relación entre las señales de entrada y salida, ajustando la manera en que cada unidad de proceso relaciona su señal de salida a la correspondiente señal de entrada. Si las señales de entrada son las características de un cliente y la señal de salida responde si el desarrollo de su crédito ha sido bueno o malo, tenemos la aplicación del método al credit scoring. En la sección (3.4) exploraremos estas posibilidades.

Se puede pensar también en el desarrollo de scoring como un tipo de problema de optimización combinatoria. Se tiene un número de parámetros, las posibles calificaciones dadas a los distintos atributos, y una manera de medir la bondad de cada conjunto de parámetros. Hay un número de métodos genéricos para resolver esta clase de problemas que son los algoritmos genéticos. De estos hay algunos proyectos pilotos en credit scoring. Estas ideas las veremos en la sección (3.5).

3.2 Programación lineal

Mangasarian [20] (1965) fue el primero en ver que la programación lineal podía utilizarse en problemas de clasificación donde hay dos grupos y un hiperplano de separación (una función lineal discriminante), que puede separar exactamente a los dos grupos. Freed, Glover, [8] y Hand [13] en 1981 estudiaron que la programación lineal podía ser utilizada también para discriminar cuando los dos grupos no son necesariamente linealmente separables, utilizando objetivos tales como la minimización de la suma de los errores absolutos o minimizando el máximo error.

Recordamos que la decisión que tiene que tomar la entidad prestamista es dividir A , el conjunto de todas las combinaciones de los valores que pueden tomar las variables de las asignaciones de crédito

Capítulo 3. Una revisión sobre algunos métodos no estadísticos ...

$\mathbf{X}=(X_1, X_2, \dots, X_p)$, en dos conjuntos: A_G , correspondiente a las respuestas dadas por los buenos clientes, y A_B , el conjunto de las respuestas dadas por los malos clientes. Suponemos que tenemos una muestra de aplicaciones previa n . Si n_G de la muestra son buenos, para facilitar la notación suponemos que son los n_G primeros de la muestra, los restantes n_B de la muestra n_G+1, n_G+2, \dots, n son malos. Asumimos que la asignación crediticia i tiene las características

$$\mathbf{x}_i=(x_{i1}, x_{i2}, \dots, x_{ip})'$$

como respuestas a las variables

$$\mathbf{X}=(X_1, X_2, \dots, X_p)'$$

Queremos obtener los valores

$$(w_1, w_2, \dots, w_p)'$$

de manera que el valor de la suma resultante

$$w_1 X_1 + w_2 X_2 + \dots + w_p X_p$$

esté por encima de un valor de corte c para las buenas aplicaciones y por debajo de este valor de corte para las malas aplicaciones. Si las variables de la aplicación se transformaran en variables binarias, de forma que

tomaran valor 0 cuando las variables inciden negativamente en la solvencia del crédito y 1 cuando las variables son buenas para la devolución del mismo, entonces podemos pensar que la suma de los w_i nos está dando una calificación del cliente o aplicación que estará por encima o por debajo del valor de c .

Normalmente, no se puede esperar obtener una división perfecta de los buenos y malos clientes, por esto se introducen las variables a_i , las cuales son positivas o cero, con lo que se tienen en cuenta posibles errores. Así, si la aplicación i de la muestra es buena, se tendrá que cumplir que

$$w_1 x_{i1} + w_2 x_{i2} + \dots + w_p x_{ip} \geq c - a_i,$$

mientras que si la aplicación j es mala, se tendrá que cumplir que

$$w_1 x_{j1} + w_2 x_{j2} + \dots + w_p x_{jp} \leq c + a_j.$$

Para encontrar los valores $(w_1, w_2, \dots, w_p)'$ que minimizan la suma de los valores absolutos de esta desviación, se tiene que resolver el siguiente problema de programación lineal

$$\min a_1 + a_2 + \dots + a_{n_G + n_B} \quad (3.1)$$

$$\text{s.a. } w_1 x_{i1} + w_2 x_{i2} + \dots + w_p x_{ip} \geq c - a_i, \quad 1 \leq i \leq n_G$$

$$w_1 x_{i1} + w_2 x_{i2} + \dots + w_p x_{ip} \leq c + a_i, \quad n_G + 1 \leq i \leq n_G + n_B$$

$$a_i \geq 0, \quad 1 \leq i \leq n_G + n_B$$

Capítulo 3. Una revisión sobre algunos métodos no estadísticos ...

Si se quisiera minimizar la máxima desviación, se simplifica al mismo término de error en cada restricción, de la siguiente forma

$$\begin{aligned}
 & \min a && (3.2) \\
 & \text{s.a. } w_1 x_{i1} + w_2 x_{i2} + \dots + w_p x_{ip} \geq c - a, && 1 \leq i \leq n_G. \\
 & w_1 x_{i1} + w_2 x_{i2} + \dots + w_p x_{ip} \leq c + a, && n_G + 1 \leq i \leq n_G + n_B \\
 & a \geq 0
 \end{aligned}$$

En credit scoring, una ventaja de la programación lineal sobre los métodos estadísticos es que si se quiere construir un scoring con un sesgo particular, esta metodología puede incluir el sesgo en el desarrollo del mismo. Suponemos por ejemplo, que X_1 es la variable binaria ser menor de 25 años o no, y X_2 es la variable binaria ser mayor de 65 años, y queremos dar una calificación mayor a ser menor de 25 que a las personas jubiladas. Para esto es necesario añadir la restricción $w_1 \geq w_2$ a las restricciones ya utilizadas en (3.1) y (3.2) y así obtener nuestro scoring. De la misma forma se podría asegurar que el peso de algunas variables en la aplicación debe ser mayor que el resto, donde las primeras serían las variables desde X_1 hasta X_s y las segundas son las variables desde X_{s+1} hasta X_p , por lo tanto tendríamos que añadir a cada restricción (3.1) o (3.2) las restricciones para cada aplicación de forma que

$$w_1 x_{i1} + w_2 x_{i2} + \dots + w_s x_{is} \geq w_{s+1} x_{i,s+1} + w_{s+2} x_{i,s+2} + \dots + w_p x_{ip} \quad (3.3)$$

para todo i .

Sin embargo, se debe ser estrictamente cuidadoso porque en la formulación de la programación lineal se requiere que los clientes buenos

tengan una valoración mayor o igual que un valor de corte y que los clientes malos tengan una valoración menor o igual que el valor de corte, esto es, la programación lineal no obtiene soluciones con desigualdades estrictas. Esto significa que si elegimos el valor de corte, así como las w_i , hay siempre una solución trivial donde el valor de corte es cero y además las w_i son todas cero. Así, todos obtienen una calificación total cero y están situados exactamente en el valor de corte. Una forma de solucionar esto podría ser forzar que el valor de corte fuese un valor distinto de cero. Sin embargo, Freed y Glover (1986) [9] demostraron que esto no es correcto y que se necesita resolver el modelo dos veces, una con valor de corte positivo y otra con un valor de corte negativo.

Para solucionar este problema han sido sugeridas varias modificaciones en la formulación que nos llevarían a reconvertir la restricción (3.1) en

$$w_1 x_{i1} + w_2 x_{i2} + \dots + w_p x_{ip} \geq c - e + a_i, \quad n_G + 1 \leq i \leq n_G + n_B \quad (3.4)$$

$$e \geq 0$$

quedando así un vacío entre los dos grupos generados. Sin embargo, entonces se tienen que desarrollar técnicas para decidir en que grupo se incluyen los puntos pertenecientes al vacío creado. Estas dificultades han sido superadas en el caso en el que los valores principales de las características de los buenos y los malos clientes son diferentes utilizando la normalización propuesta por Glover (1990) [12], añadiendo la restricción:

Capítulo 3. Una revisión sobre algunos métodos no estadísticos ...

$$\sum_{j=1}^p (n_B \sum_{i=1}^{n_G} x_{ij} - n_G \sum_{i=n_G+1}^{n_G+n_B} x_{ij}) w_j = 1. \quad (3.5)$$

Una de la formulaciones más generales de la programación lineal fue propuesta por Freed y Glover [10] en 1986. En el momento que los errores o desviaciones externas a_i , no satisfacen alguna restricción, entonces utilizaremos las desviaciones internas e_i , que nos indican la bondad de las calificaciones de las aplicaciones que están correctamente clasificadas. Freed y Glover propusieron como objetivo una combinación ponderada del máximo interno y las desviaciones externas y la suma de los valores absolutos de las desviaciones internas y externas. Esto se resuelve con el siguiente programa lineal:

$$\min k_0 a_0 - l_0 e_0 + \sum_{i=1}^{n_G+n_B} k_i a_i + \sum_{i=1}^{n_G+n_B} l_i e_i \quad (3.6)$$

$$\text{s.a. } w_1 x_{i1} + w_2 x_{i2} + \dots + w_p x_{ip} \geq c - a_0 - a_i + e_0 + e_i, \quad 1 \leq i \leq n_G$$

$$w_1 x_{i1} + w_2 x_{i2} + \dots + w_p x_{ip} \leq c + a_0 + a_i - e_0 - e_i, \quad n_G + 1 \leq i \leq n_G + n_B$$

$$\sum_{j=1}^p (n_B \sum_{i=1}^{n_G} x_{ij} - n_G \sum_{i=n_G+1}^{n_G+n_B} x_{ij}) w_j = 1$$

$$a_i, e_i \geq 0, \quad 0 \leq i \leq n_G + n_B.$$

Esta formulación del problema siempre da una solución no trivial y es invariante bajo la transformación lineal de los datos.

La conclusión práctica sobre los métodos de programación lineal es que aunque no tienen una base estadística, se puede evaluar si los

parámetros son estadísticamente significativos. Ziari [40], en 1997 sugiere un método para estimar los parámetros utilizando las técnicas de remuestreo entre las que se encuentra el método jackknifing. Estas técnicas suponen que la muestra utilizada en la programación lineal representa la población vacía de clientes. Para remuestrear un subconjunto desde la muestra original, cada nueva submuestra generará diferentes parámetros estimados. La distribución de estos estimadores refleja alguna información sobre la distribución de los parámetros obtenidos desde la muestra completa. En el proceso de la jackknifing, un subconjunto de t clientes es apartado de la muestra (normalmente $t=1$) y se resuelve la programación lineal con los restantes. Esto se repite utilizando diferentes subconjuntos de t clientes para obtener estimadores de los parámetros. En otros métodos, se utiliza una muestra de $n=n_G+n_B$ que es tomada con reemplazo desde el conjunto de datos original de la misma forma y son calculados los parámetros de la programación lineal. Esto se repite un número de veces, y la media y la desviación típica de estos estimadores de los parámetros dan aproximaciones de la desviación típica de los estimadores de los parámetros originales.

Una ventaja clave de la regresión sobre la programación lineal es que en el primer método se pueden introducir de una sola vez las m características más importantes y la regresión encuentra las m que son más discriminantes. Nath y Jones [23] en 1988 mostraron como el método de la jackknifing aplicado en programación lineal podía utilizarse para seleccionar las m características más poderosas. Sin embargo esto lleva consigo la resolución de programaciones lineales un gran número de veces.

Erenguc y Koehler [6] en 1990 y Nath, Jackson y Jones [24] en 1992 compararon la programación lineal y métodos de regresión para

Capítulo 3. Una revisión sobre algunos métodos no estadísticos ...

clasificar varios conjuntos de datos (ninguno contenía datos de créditos). Sus resultados sugieren que la programación lineal no clasifica tan bien como los métodos estadísticos.

3.3 Programación entera

Los modelos de programación lineal minimizan la suma de la desviación en el credit scoring de aquellos clientes mal clasificados. Sin embargo, un criterio más práctico debería ser minimizar el número de clientes mal clasificados, o el coste total de clasificar mal, si tenemos en cuenta que D , el coste de clasificar un malo como bueno, es muy diferente de L , el coste de clasificar un buen cliente como malo. La construcción de credit scoring bajo este criterio podía llevarnos de nuevo a la programación lineal pero, como algunas variables tienen que ser enteras esta técnica se denomina programación lineal entera. Este modelo fue expuesto por Koehler y Erenguc [18] en 1990 como sigue:

$$\begin{aligned} \min & L(d_1+d_2+\dots+d_{n_G})+D(d_{n_G+1}+\dots+d_{n_G+n_B}) & (3.7) \\ \text{s.a} & w_1x_{i1}+w_2x_{i2}+\dots+w_px_{ip} \geq c-Md_i, & 1 \leq i \leq n_G \\ & w_1x_{i1}+w_2x_{i2}+\dots+w_px_{ip} \leq c+Md_i, & n_G+1 \leq i \leq n_G+n_B \\ & d_i \in \{0,1\} \end{aligned}$$

donde d_i es una variable que vale 1 si el cliente i de la muestra está mal clasificado y 0 en caso contrario, y $M > 0$. De nuevo, obtenemos que la

expresión (3.7) se minimiza para $c=0$, $w_i=0$, $i=1,\dots,p$, esto implica que se tienen que añadir condiciones de normalización, como por ejemplo

$$\sum_{j=1}^p w_j = 1 \text{ y } c \in \{-1,+1\} \quad (3.8)$$

Joachimsthaler y Stam [17] en 1990 y Koehler y Erenguc [18] en el mismo año, demostraron que el modelo (3.7) es un modelo mejor de clasificación que los modelos de programación lineal, sin embargo, tiene dos grandes desventajas. La primera es que su resolución es mucho más larga que la programación lineal y la segunda es que frecuentemente se obtiene un número de soluciones óptimas con el mismo número de casos mal clasificados en el conjunto de entrenamiento pero con bastantes diferencias en el resultado aún manteniendo las muestras.

La complejidad computacional para resolver los algoritmos de programación entera es realmente grande, lo cual hace que quede descartado para comercializar un modelo de credit scoring.

Varios autores añaden condiciones extra y objetivos secundarios a (3.7) para asegurar una clasificación que sea robusta en la muestra ofrecida, así como minimizar el número de mal clasificados en el entrenamiento de la muestra. Pavar, Wanarat y Loucopoulus [26] en 1997 sugieren modelos que además maximizan la diferencia entre la media de las calificaciones discriminantes; Bajgier y Hill [4] en 1982 sugirieron minimizar la suma de las desviaciones (la diferencia entre el valor de los mal clasificados y el valor de corte) así como el número de los mal clasificados. Rubin [33] (1990) encontró un modo secundario de

Capítulo 3. Una revisión sobre algunos métodos no estadísticos ...

maximizar la mínima desviación interior, la diferencia entre el valor de los clasificados correctamente y el valor de corte.

La programación entera ha sido utilizada para resolver los problemas de clasificación con el criterio del mínimo número de mal clasificados, pero también puede utilizarse para superar dos de las dificultades que aparecen en la programación lineal para minimizar la suma de los errores absolutos. La primera es el problema de la eliminación de las soluciones triviales y asegurar que sus valores son invariantes si se cambia el orden de los datos. Esto se superó en la formulación del mínimo coste de los mal clasificados (3.7) utilizando la normalización (3.8). Una normalización similar se puede aplicar al problema del mínimo de la desviación típica o al mínimo de la máxima desviación que fue desarrollada por Glen en [11] (1999).

Otra crítica al método de programación lineal era la dificultad para elegir el mejor scoring utilizando solo m características. El método jakknifing sugiere como resolver este problema, pero también se puede utilizar la formulación de Glenn e imponiendo las condiciones acerca de la suma de los coeficientes w_i .

3.4 Redes neuronales

Las redes neuronales son modelos que intentan reproducir el comportamiento del cerebro. Como tal modelo, realiza una simplificación, averiguando cuáles son los elementos relevantes del sistema, bien porque la cantidad de información de que se dispone es excesiva o bien porque es redundante. Una elección adecuada de sus características, más una estructura conveniente, es el procedimiento convencional utilizado para construir redes capaces de realizar una

determinada tarea.

En la búsqueda de literatura sobre estudios o elaboraciones de credit scoring utilizando redes neuronales, nos encontramos que la mayoría de los trabajos realizados versan sobre la solvencia de las empresas y no de clientes particulares que es lo que nos ocupa. La única mención obtenida es el trabajo de Soldevilla y Guillén (1997) [35], que modelizan el impago de préstamos concedidos a particulares por una entidad bancaria utilizando una red perceptrón entrenada a través de retropropagación y se comparan los resultados con los del análisis logit, resultando ligeramente superior el modelo neuronal, queremos sin embargo, destacar el artículo de Bell, Ribar y Verchio [3] ya que es la primera referencia de aplicación de redes neuronales en el estudio de solvencias.

3.4.1 Elementos y arquitectura de una red neuronal artificial

Cualquier modelo de red neuronal consta de dispositivos elementales de proceso: las neuronas. La neurona artificial pretende mimetizar las características más importantes de las neuronas biológicas. Cada neurona i -ésima está caracterizada en cualquier momento por un valor numérico denominado valor o estado de activación $a_i(t)$; asociado a cada unidad, existe una función de salida f_i , que transforma el estado actual de activación en una señal de salida, y_i , tal que $y_i(t)=f_i(a_i(t))$. Dicha señal es enviada a través de los canales de comunicación unidireccionales a otras unidades de la red; en estos canales la señal se modifica de acuerdo con la sinapsis (el peso, w_{ij}) asociada a cada uno de ellos según una determinada regla. Las señales moduladas que han llegado a la unidad j -ésima se combinan entre ellas, generando así la entrada total, Net_j .

Capítulo 3. Una revisión sobre algunos métodos no estadísticos ...

$$\text{Net}_j = \sum_{i=1}^n y_i w_{ij}$$

Una función de activación, F , determina el nuevo estado de activación $a_j(t+1)$ de la neurona, teniendo en cuenta la entrada total calculada y el anterior estado de activación $a_j(t)$.

Si se tienen N unidades (neuronas), podemos ordenarlas arbitrariamente y designar la j -ésima unidad como U_j . Su trabajo es simple y único, y consiste en recibir las entradas de las células vecinas y calcular el valor de salida, el cual es enviado a todas las células restantes.

Con N unidades el estado de activación se expresa como:

$$A(t) = (a_1(t), a_2(t), \dots, a_i(t), \dots, a_N(t))$$

y el vector que contiene las salidas de todas las neuronas en un instante t es:

$$Y(t) = (f_1(a_1(t)), f_2(a_2(t)), \dots, f_i(a_i(t)), \dots, f_N(a_N(t)))$$

En algunos modelos, esta salida es igual al nivel de activación de la unidad, en cuyo caso la función f_i es la función identidad, $f_i(a_i(t)) = a_i(t)$.

En cualquier sistema que se esté modelando, es útil caracterizar tres tipos de unidades: entradas, salidas y ocultas. Las unidades de entrada reciben señales desde el entorno, las de salida las envían fuera del sistema y las unidades ocultas son aquellas cuyas entradas y salidas se encuentran dentro del sistema.

Se conoce como capa o nivel a un conjunto de neuronas cuyas entradas provienen de la misma fuente y cuyas salidas se dirigen al mismo destino.

Existen cuatro funciones de transferencia típicas que determinan distintos tipos de neuronas:

- Función escalón
- Función lineal
- Sigmoidal
- Función gaussiana

Dado el estado de activación $a_i(t)$ de la unidad U_i y la entrada total que llega a ella, Net_i , el estado de activación siguiente, $a_i(t+1)$, se obtiene aplicando una función F , llamada función de activación.

$$a_i(t+1)=F(a_i(t),Net_i)$$

En la mayoría de los casos, F es la función identidad, por lo que el estado de activación de la neurona en $t+1$ coincidiría con el Net de la misma en t . En este caso, el parámetro que se le pasa a la función salida, f , de la neurona será directamente el Net . El estado de activación anterior

Capítulo 3. Una revisión sobre algunos métodos no estadísticos ...

no se tiene en cuenta. Según esto, la salida de una neurona i , (y_i) quedará según la expresión:

$$y_i(t+1) = f(\text{Net}_i) = f\left(\sum_{j=1}^N w_{ij} y_j(t)\right)$$

Por tanto, y en lo sucesivo, consideraremos únicamente la función f , que denominaremos indistintamente de transferencia o de activación. Además, normalmente la función de activación no está centrada en el origen del eje que representa el valor de la entrada neta, sino que existe cierto desplazamiento debido a las características internas de la propia neurona y que no es igual en todas ellas. Este valor se denota como θ_i y representa el umbral de activación de la neurona i , obteniendo entonces:

$$y_i(t+1) = f(\text{Net}_i - \theta_i) = f\left(\sum_{j=1}^N w_{ij} y_j(t) - \theta_i\right)$$

La salida que se obtiene en una neurona para las diferentes formas de la función serán:

a) Función de activación escalón

Si el conjunto de los estados de activación es $E = \{0,1\}$, tenemos que:

$$y_i(t+1) = \begin{cases} 1 & \text{si } (Net > \theta_i) \\ y_i(t) & \text{si } (Net = \theta_i) \\ 0 & \text{si } (Net < \theta_i) \end{cases}$$

Si el conjunto de los estados de activación es $E = \{-1, 1\}$, tenemos que:

$$y_i(t+1) = \begin{cases} 1 & \text{si } (Net > \theta_i) \\ y_i(t) & \text{si } (Net = \theta_i) \\ -1 & \text{si } (Net < \theta_i) \end{cases}$$

b) Función de activación lineal o identidad

El conjunto de estados E puede contener cualquier número real; el estado de activación coincide con la entrada total que ha llegado a la unidad

$$y_i(t+1) = Net_i - \theta_i$$

c) Función de activación lineal-mixta

$$y_i(t+1) = \begin{cases} b & \text{si } (Net_i \leq b + \theta_i) \\ Net_i - \theta_i & \text{si } (b + \theta_i < Net_i < B + \theta_i) \\ B & \text{si } (Net_i \geq B) \end{cases}$$

Capítulo 3. Una revisión sobre algunos métodos no estadísticos ...

Con esta función, el estado de activación de la unidad está obligado a permanecer dentro de un intervalo de valores reales prefijados, en este caso [b,B]

d) Función de activación sigmodial

Es una función continua, por tanto el espacio de los estados de activación es un intervalo del eje real

$$y_i = 1 / (1 + e^{-(Net_i - \theta_i)})$$

Para simplificar la expresión de la salida de una neurona i , es habitual considerar la existencia de una neurona ficticia, con valor de salida la unidad, asociada a la entrada de cada neurona i mediante una conexión con peso de valor $-\theta_i$. De esta forma la expresión de salida quedará:

$$y_i(t+1) = f\left(\sum_{j=1}^N w_{ij} y_j(t) - \theta_i\right) = f\left(\sum_{j=1}^N w_{ij} y_j(t)\right) = f(Net_i)$$

3.4.2 Mecanismo de aprendizaje

El aprendizaje es el mecanismo por el cual una red neuronal modifica sus pesos en respuesta a una información de entrada. Los cambios que se producen durante el proceso de aprendizaje se reducen a la destrucción,

modificación y creación de conexiones entre neuronas. En los modelos de redes neuronales artificiales, la creación de una nueva conexión implica que el peso de la misma pasa a tener un valor distinto de cero. De la misma forma, una conexión se destruye cuando su peso pasa a ser cero.

Durante el proceso de aprendizaje, los pesos de las conexiones de la red sufren modificaciones, por tanto se puede afirmar que este proceso ha terminado (la red ha aprendido) cuando los valores de los pesos permanecen estables $dw_{ij}/dt=0$.

Un aspecto importante respecto al aprendizaje en las redes neuronales es el conocer cómo se modifican los valores de los pesos; es decir, cuales son los criterios que se siguen para cambiar el valor asignado a las conexiones cuando se pretende que la red aprenda una nueva información.

Estos criterios determinan lo que se conoce como la regla de aprendizaje de la red. De forma general, se suelen considerar dos tipos de reglas: las que responden a lo que habitualmente se conoce como aprendizaje supervisado, y las correspondientes a un aprendizaje no supervisado.

Es por ello, por lo que una de las clasificaciones que se realizan de las redes neuronales obedece al tipo de aprendizaje utilizado por dichas redes. Así, se pueden distinguir:

- Redes neuronales con aprendizaje supervisado
- Redes neuronales con aprendizaje no supervisado

La diferencia fundamental entre ambos tipos estriba en la existencia o no de un agente externo (supervisor) que controle el proceso de aprendizaje de la red.

Capítulo 3. Una revisión sobre algunos métodos no estadísticos ...

Otro criterio que se puede utilizar para diferenciar las reglas de aprendizaje se basa en considerar si la red puede aprender durante su funcionamiento habitual o si el aprendizaje supone la desconexión de la red: es decir, su inhabilitación hasta que el proceso termine. En el primer caso, se trataría de un aprendizaje ON LINE, mientras que el segundo es lo que se conoce como aprendizaje OFF LINE.

Cuando el aprendizaje es OFF LINE, se distingue entre una fase de aprendizaje o entrenamiento y una fase de operación o funcionamiento, existiendo un conjunto de datos de entrenamiento y un conjunto de datos de test o prueba que serán utilizados en la correspondiente fase. En estas conexiones los pesos permanecen fijos después que termina la etapa de entrenamiento de la red. Debido precisamente a su carácter estático, estos sistemas no presentan problemas de estabilidad en su funcionamiento.

En las redes con aprendizaje ON LINE no se distingue entre fase de entrenamiento y de operación, de tal forma que los pesos varían dinámicamente siempre que se presente una nueva información al sistema. En este tipo de redes, debido al carácter dinámico de las mismas, el estudio de la estabilidad suele ser un aspecto fundamental de estudio.

3.4.3 Redes con aprendizaje supervisado

El aprendizaje supervisado se caracteriza porque el proceso de aprendizaje se realiza mediante un entrenamiento controlado por un agente externo (supervisor) que determina la respuesta que debería generar la red a partir de una entrada determinada. El supervisor comprueba la salida de la red y en el caso de que esta no coincida con la deseada, se procederá a modificar los pesos de las conexiones, con el fin de conseguir que la salida obtenida se aproxime a la deseada.

En este tipo de aprendizaje se suelen considerar, a su vez, tres formas de llevarlo a cabo que dan lugar a los siguientes aprendizajes supervisados:

- Aprendizaje por corrección de error
- Aprendizaje por refuerzo
- Aprendizaje estocástico

Aprendizaje por corrección de error

Consiste en ajustar los pesos de las conexiones de la red en función de la diferencia entre los valores deseados y los obtenidos en la salida de la red; es decir, en función del error cometido en la salida.

Una regla o algoritmo simple de aprendizaje por corrección de error podría ser el siguiente:

$$\Delta w_{ji} = \alpha y_i (d_j - y_j)$$

Siendo:

Δw_{ji} : Variación en el peso de la conexión entre las neuronas i y j

$$(\Delta w_{ji} = w_{ji}^{actual} - w_{ji}^{anterior})$$

y_i : Valor de salida de la neurona i

d_j : Valor de salida deseado para la neurona j

α : Factor de aprendizaje ($0 < \alpha < 1$) que regula la velocidad del aprendizaje.

Un ejemplo de este tipo de algoritmo lo constituye la regla de aprendizaje del Perceptrón, utilizada en el entrenamiento de la red del mismo nombre desarrollada por Rosenblatt en 1958 [34].

Capítulo 3. Una revisión sobre algunos métodos no estadísticos ...

Un algoritmo muy conocido que mejora el del Perceptrón y permite un aprendizaje más rápido y un campo de aplicación más amplio es el propuesto por Widrow y Hoff en 1960 [38], denominado regla delta o regla del mínimo error cuadrado, también conocida como la regla de Widrow -Hoff, que se aplicó en las redes desarrolladas por los mismos autores, conocidas como ADALINE, con una única neurona de salida, y MADALINE, con varias neuronas de salida.

Widrow y Hoff definieron una función que permitía cuantificar el error global cometido en cualquier momento durante el proceso de entrenamiento de la red, lo cual es importante, ya que cuanta más información se tenga sobre el error cometido, más rápido se puede aprender.

Este error medio se expresa de la siguiente forma.

$$\text{Error}_{global} = (1/2P) \sum_{k=1}^P \sum_{j=1}^N (y_j^{(k)} - d_j^{(k)})^2$$

Siendo:

N: Número de neuronas de salida

P: Número de informaciones que aprende la red

$1/2 \sum_{j=1}^N (y_j^{(k)} - d_j^{(k)})^2$: Error cometido en el aprendizaje de la información

k-ésima

Por tanto, de lo que se trata es de encontrar unos pesos para las conexiones de la red que minimicen esta función de error. Para ello, el ajuste de los pesos de la red se puede hacer de forma proporcional a la

variación relativa del error que se obtiene al variar el peso correspondiente:

$$\Delta w_{ji} = k(\partial \text{Error}_{global} / \partial w_{ji})$$

mediante este procedimiento, se llegan a obtener un conjunto de pesos con los que se consigue minimizar el error medio.

Otro algoritmo de aprendizaje por corrección de error lo constituye el denominado regla delta generalizada o algoritmo de retropropagación del error, también conocido como regla LMS (Least-Mean-Square Error) multicapa. Se trata de una generalización de la regla delta para poder aplicarla a redes con conexiones hacia delante (feedforward) con capas o niveles internos u ocultos de neuronas que no tienen relación con el exterior. Son redes con capa de entrada, capas ocultas y capa de salida.

Estas redes multicapa pueden utilizarse en muchas más aplicaciones que las ya mencionadas Perceptrón, ADALINE y MADALINE, pero su proceso de aprendizaje es mucho más lento, debido a que durante el mismo se debe explorar el espacio de posibles formas de utilización de las neuronas de las capas ocultas; es decir, se debe establecer cuál va a ser su papel en el funcionamiento de la red.

Las bases de este nuevo aprendizaje (al que Rumelhart, Hinton, y Williams llamaron backpropagation) fueron sentadas por diferentes investigadores ([36], [25], [19], [29]) que propusieron soluciones al problema del entrenamiento de redes multicapa, de forma independiente y sin conocimiento de la existencia de otros trabajos paralelos.

Existe también una versión recurrente del algoritmo de retropropagación que se puede utilizar en redes multicapa que presentan conexiones recurrentes con el fin de que estas redes aprendan la naturaleza temporal de algunos datos ([27], [2], [31]).

Capítulo 3. Una revisión sobre algunos métodos no estadísticos ...

Para concluir con los algoritmos por corrección de error, hay que mencionar que también se utilizan en algunas redes monocapa con conexiones laterales y autorrecurrentes.

Aprendizaje por refuerzo

Se trata de un aprendizaje supervisado, más lento que el anterior, que se basa en la idea de no disponer de un ejemplo completo del comportamiento deseado; es decir, de no indicar durante el entrenamiento exactamente la salida que se desea que proporcione la red ante una determinada entrada.

En el aprendizaje por refuerzo la función del supervisor se reduce a indicar mediante una señal de refuerzo si la salida obtenida en la red se ajusta a la deseada (éxito=+1 o fracaso=-1), y en función de ello se ajustan los pesos basándose en un mecanismo de probabilidades. Se podría decir que en este tipo de aprendizaje la función del supervisor se asemeja más a la de un crítico (que opina sobre la respuesta de la red) que a la de un maestro (que indica a la red la respuesta concreta que debe generar), como ocurría en el caso de la supervisión por corrección de error.

Aprendizaje estocástico

Este tipo de aprendizaje consiste básicamente en realizar cambios aleatorios en los valores de los pesos de las conexiones de la red y evaluar su efecto a partir del objetivo deseado y de distribuciones de probabilidad.

En el aprendizaje estocástico se suele hacer una analogía en términos termodinámicos, asociando la red neuronal con un sólido físico que tiene cierto estado energético. En el caso de la red, la energía de la misma representaría el grado de estabilidad de la red, de tal forma que el estado

de mínima energía correspondería a una situación en la que los pesos de las conexiones consiguen que su funcionamiento sea el que más se ajusta al objetivo deseado.

Según lo anterior, el aprendizaje consistiría en realizar un cambio aleatorio de los valores de los pesos y determinar la energía de la red. Si la energía es menor después del cambio; es decir, si el comportamiento de la red se acerca al deseado, se acepta el cambio. Si, por el contrario, la energía no es menor, se aceptaría el cambio en función de una determinada y preestablecida distribución de probabilidad.

3.4.4 Redes con aprendizaje no supervisado

Las redes con aprendizaje no supervisado (también conocido como autosupervisado) no requieren influencia externa para ajustar los pesos de las conexiones entre sus neuronas. La red no recibe ninguna información por parte del entorno que le indique si la salida generada en respuesta a una determinada entrada es o no correcta; por ello, suele decirse que estas redes son capaces de autoorganizarse. Estas redes deben encontrar las características, regularidades, correlaciones o categorías que se puedan establecer entre los datos que se presenten en su entrada. Puesto que no hay supervisor que indique a la red la respuesta que debe generar ante una entrada concreta, cabría preguntarse precisamente por lo que la red genera en estos casos. Existen varias posibilidades en cuanto a la interpretación de la salida de estas redes, que dependen de su estructura y del algoritmo de aprendizaje empleado.

En algunos casos, la salida representa el grado de familiaridad o similitud entre la información que se le está presentando en la entrada y las informaciones que se le han mostrado hasta entonces (en el pasado). En otro caso, podría realizar una clusterización (clustering) o

Capítulo 3. Una revisión sobre algunos métodos no estadísticos ...

establecimiento de categorías, indicando la red a la salida a qué categoría pertenece la información presentada a la entrada, siendo la propia red quien debe encontrar las categorías apropiadas a partir de correlaciones entre las informaciones presentadas. Una variación de esta categorización es el prototipado. En este caso, la red obtiene ejemplares o prototipos representantes de las clases a las que pertenecen las informaciones de entrada.

Finalmente, algunas redes con aprendizaje no supervisado lo que realizan es un mapeo de características (feature mapping), obteniéndose en las neuronas de salida una disposición geométrica que representa un mapa topográfico de las características de los datos de entrada, de tal forma que si se presentan a la red informaciones similares, siempre sean afectadas neuronas de salida próximas entre sí, en la misma zona del mapa.

En cuanto a los algoritmos de aprendizaje no supervisado, en general se suelen considerar dos tipos que dan lugar a los siguientes aprendizajes:

- Aprendizaje hebbiano
- Aprendizaje competitivo y cooperativo

En el primer caso, normalmente se pretende medir la familiaridad o extraer características de los datos de entrada, mientras que en el segundo suele orientarse hacia la clasificación de dichos datos.

Aprendizaje hebbiano

Este tipo de aprendizaje se basa en el siguiente postulado formulado por Hebb en 1949: Cuando un axón de una celda A está suficientemente cerca como para conseguir excitar una celda B y repetida o persistentemente toma parte en su activación, algún proceso de

crecimiento o cambio metabólico tiene lugar en una o ambas celdas, de tal forma que la eficiencia de A, cuando la celda a activar es B, aumenta. Por celda, Hebb entiende un conjunto de neuronas fuertemente conexas a través de una estructura compleja. La eficiencia podría identificarse con la intensidad o magnitud de la conexión; es decir, con el peso.

Se puede decir, por tanto, que el aprendizaje hebbiano consiste básicamente en el ajuste de los pesos de las conexiones de acuerdo con la correlación (multiplicación en el caso de valores binarios +1 y -1) de los valores de activación (salidas) de las dos neuronas conectadas:

$$\Delta w_{ij} = y_i y_j$$

Esta expresión responde a la idea de Hebb, puesto que si las dos unidades son activas (positivas), se produce un reforzamiento de la conexión. Por el contrario, cuando una es activa y la otra pasiva (negativa), se produce un debilitamiento de la conexión. Se trata de una regla de aprendizaje no supervisado, pues la modificación de los pesos se realiza en función de los estados (salidas) de las neuronas obtenidos tras la presentación de cierto estímulo (información de entrada a la red), sin tener en cuenta si se deseaba obtener o no esos estados de activación.

Aprendizaje competitivo y cooperativo

En las redes con aprendizaje competitivo y cooperativo, suele decirse que las neuronas compiten y cooperan unas con otras con el fin de llevar a cabo una tarea dada. Con este tipo de aprendizaje, se pretende que cuando se presente a la red cierta información de entrada, sólo una de las neuronas de salida de la red, o una por cierto grupo de neuronas, se active (alcance su valor de respuesta máximo). Por tanto, las neuronas compiten

Capítulo 3. Una revisión sobre algunos métodos no estadísticos ...

por activarse, quedando finalmente una, o una por grupo, como neurona vencedora, quedando anuladas el resto, que son forzadas a sus valores de respuesta mínimos.

La competición entre neuronas se realiza en todas las capas de la red, existiendo en estas neuronas conexiones recurrentes de autoexcitación y conexiones de inhibición (signo negativo) por parte de neuronas vecinas. Si el aprendizaje es cooperativo, estas conexiones con las vecinas serán de excitación (signo positivo).

El objetivo de este aprendizaje es categorizar (clusterizar) los datos que se introducen en la red. De esta forma, las informaciones similares son clasificadas formando parte de la misma categoría, y por tanto deben activar la misma neurona de salida. Las clases o categorías deben ser creadas por la propia red, puesto que se trata de un aprendizaje no supervisado, a través de las correlaciones entre los datos de entrada.

Pensamos que las características de estas redes con aprendizaje competitivo y cooperativo son las adecuadas para construir un credit scoring que permita clasificar a los clientes en categorías según su probabilidad de impago, y concretamente de entre ellas el Algoritmo de Kohonen que desarrollaremos en el capítulo 4 de esta tesis.

3.5 Algoritmos genéticos

3.5.1 Introducción

En las últimas décadas, la comunidad científica internacional ha mostrado un creciente interés en una nueva técnica de búsqueda de óptimo basada en la teoría de la evolución y que se conoce como el algoritmo genético. Esta técnica se basa en los mecanismos de selección que utiliza la naturaleza, de acuerdo a los cuales los individuos más aptos

de una población son los que sobreviven, al adaptarse más fácilmente a los cambios que se producen en su entorno. Hoy en día se sabe que estos cambios se efectúan en los genes de un individuo, y que los atributos que les permite adaptarse mejor a su entorno se transmiten a sus descendientes cuando éste se reproduce.

Un investigador de la Universidad de Michigan llamado John Holland era consciente de la selección natural, y a fines de los años sesenta desarrolló una técnica que permitió incorporarla a un programa. Su objetivo era lograr que las computadoras aprendieran por sí mismas. A la técnica que inventó Holland se le llamó originalmente "planes reproductivos", pero se hizo popular bajo el nombre "algoritmo genético" tras la publicación de su libro en 1975.

Los algoritmos genéticos son métodos adaptativos que pueden usarse para resolver problemas de búsqueda y optimización. Están basados en el proceso genético de los organismos vivos. A lo largo de las generaciones, las poblaciones evolucionan en la naturaleza de acorde con los principios de la selección natural y la supervivencia de los más fuertes, postulados por Darwin. Por imitación de este proceso, los algoritmos genéticos son capaces de ir creando soluciones para problemas del mundo real. La evolución de dichas soluciones hacia valores óptimos del problema depende en buena medida de una adecuada codificación de las mismas.

Un algoritmo genético consiste en una función matemática o una rutina de software que toma como entradas a los ejemplares y retorna como salidas cuales de ellos deben generar descendencia para la nueva generación.

En la naturaleza los individuos de una población compiten entre sí en la búsqueda de recursos tales como comida, agua y refugio. Incluso los miembros de una misma especie compiten a menudo en la búsqueda de un compañero. Aquellos individuos que tienen más éxito en sobrevivir y

Capítulo 3. Una revisión sobre algunos métodos no estadísticos ...

en atraer compañeros tienen mayor probabilidad de generar un gran número de descendientes. Esto significa que los genes de los individuos mejor adaptados se propagarán en sucesivas generaciones hacia un número de individuos creciente. La combinación de buenas características provenientes de diferentes ancestros, puede a veces producir descendientes "superindividuos", cuya adaptación es mucho mayor que la de cualquiera de sus ancestros. De esta manera, las especies evolucionan logrando unas características cada vez mejor adaptadas al entorno en el que viven.

Los algoritmos genéticos usan una analogía directa con el comportamiento natural. Trabajan con una población de individuos, cada uno de los cuales representa una solución factible a un problema dado. A cada individuo se le asigna un valor o puntuación, relacionado con la bondad de dicha solución. En la naturaleza esto equivaldría al grado de efectividad de un organismo para competir por unos determinados recursos. Cuanto mayor sea la adaptación de un individuo al problema, mayor será la probabilidad de que el mismo sea seleccionado para reproducirse, cruzando su material genético con otro individuo seleccionado de igual forma. Este cruce producirá nuevos individuos, descendientes de los anteriores, los cuales comparten algunas de las características de sus padres. Cuanto menor sea la adaptación de un individuo, menor será la probabilidad de que dicho individuo sea seleccionado para la reproducción, y por tanto, de que su material genético se propague en sucesivas generaciones.

De esta manera se produce una nueva población de posibles soluciones, la cual reemplaza a la anterior y verifica la interesante propiedad de que contiene una mayor proporción de buenas características en comparación con la población anterior. Así, a lo largo de las generaciones, las buenas características se propagan a través de la

población. Favoreciendo el cruce de los individuos mejor adaptados, van siendo exploradas las áreas más prometedoras del espacio de búsqueda. Si el algoritmo genético ha sido bien diseñado, la población convergerá hacia una solución óptima del problema.

3.5.2 El algoritmo genético simple

Un algoritmo genético (AG) es un proceso de búsqueda sistemática a partir de una población de soluciones potenciales de un problema, así las soluciones candidatas que estén más próximas a la solución del problema tienen una mayor oportunidad de llegar a ser la solución candidata que otras. Los algoritmos genéticos fueron propuestos por Holland (1975) [16] y tiene analogía con el principio de la selección natural de la evolución de Darwin [5] (1859).

Como se verá a continuación, se necesita una codificación o representación del problema, que resulte adecuada al mismo. Además se requiere una función de ajuste o adaptación al problema, la cual asigna un número real a cada posible solución codificada.

La ecuación (3.9) es un ejemplo de una función de scoring que ha sido utilizada en la literatura. Un artículo de Yobas, Crook, y Ross [39] (2000) donde se obtienen algunos detalles tiene una función diferente.

$$f(x_i) = a_1 x_{i,1}^{b_1} + a_2 x_{i,2}^{b_2} + \dots + a_p x_{i,p}^{b_p} + c, \quad (3.9)$$

Suponemos que queremos calcular los parámetros $a_1, a_2, \dots, a_p, b_1, b_2, \dots, b_p$, y c en la siguiente ecuación de credit scoring para clasificar la

Capítulo 3. Una revisión sobre algunos métodos no estadísticos ...

asignación de un préstamo, donde x_{i1}, \dots, x_{ip} son los valores de las características de la aplicación i .

Cuando los parámetros son estimados, entonces se puede clasificar una aplicación como buena o mala si el valor de $f(x_i)$ es mayor o menor que cero.

El procedimiento seguido en un Algoritmo Genético se muestra en la Figura 4:

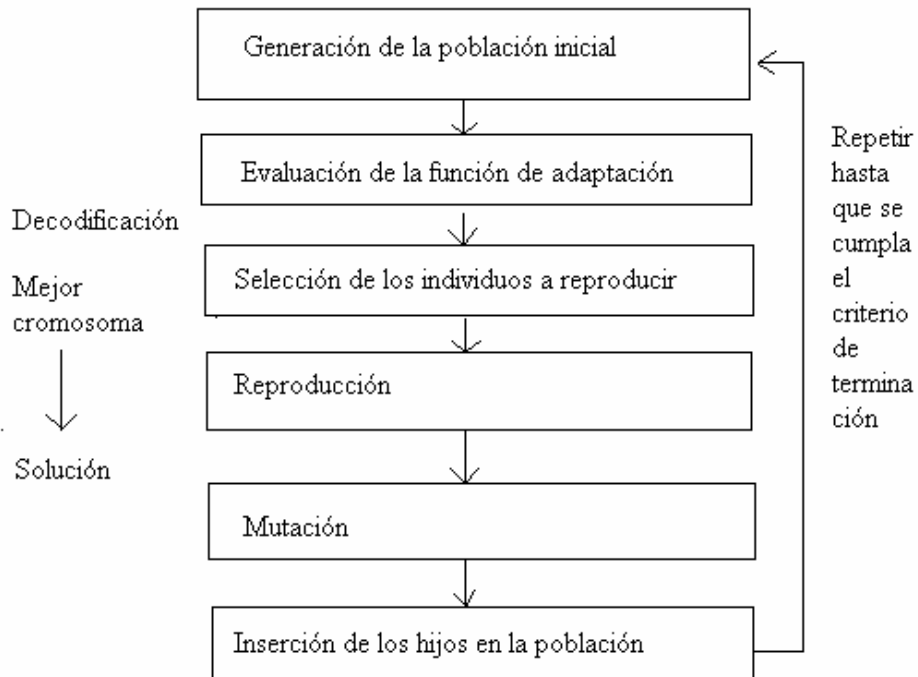


Figura 4: Esquema del funcionamiento del algoritmo genético

Durante la ejecución del algoritmo, los padres deben ser seleccionados para la reproducción, a continuación dichos padres seleccionados se cruzarán generando dos hijos, sobre cada uno de los cuales actuará un

operador de mutación. El resultado de la combinación de las anteriores funciones será un conjunto de individuos (posibles soluciones del problema), los cuales en la evolución del algoritmo genético formarán parte de la siguiente población.

En los dos apartados siguientes desarrollamos con más detalle el esquema de funcionamiento del algoritmo genético.

3.5.3 Codificación

Se supone que los individuos (posibles soluciones del problema), pueden representarse como un conjunto de parámetros, los cuales agrupados forman una cadena de valores (a menudo referida como cromosoma). Si bien el alfabeto utilizado para representar los individuos no debe necesariamente estar constituido por el $(0,1)$, buena parte de la teoría en la que se fundamentan los algoritmos genéticos utiliza dicho alfabeto. En términos biológicos, el conjunto de parámetros representando un cromosoma particular se denomina fenotipo. El fenotipo contiene la información requerida para construir un organismo, el cual se refiere como genotipo. Los mismos términos se utilizan en el campo de los algoritmos genéticos. La adaptación al problema de un individuo depende de la evaluación del genotipo. Esta última puede inferirse a partir del fenotipo, es decir, puede ser computada a partir del cromosoma, usando la función de evaluación. La función de adaptación debe ser diseñada para cada problema de manera específica. Dado un cromosoma particular, la función de adaptación le asigna un número real, que se supone refleja el nivel de adaptación al problema del individuo representado por el cromosoma.

Durante la fase reproductiva se seleccionan los individuos de la población para cruzarse y producir descendientes, que constituirán, una

Capítulo 3. Una revisión sobre algunos métodos no estadísticos ...

vez mutados, la siguiente generación de individuos. La selección de padres más utilizada, es la denominada función de selección proporcional a la función objetivo, en la cual cada individuo tiene una probabilidad de ser seleccionado como padre que es proporcional al valor de su función objetivo.

Denotando p_j la probabilidad de que un individuo sea seleccionado como padre, se tiene que

$$p_j = (f_j / \sum_{j=1}^{n_{pop}} f_j)$$

Una de las maneras de superar el problema relacionado con la rápida convergencia proveniente de los superindividuos que surge al aplicar la anterior función de selección, es el efectuar la selección proporcional al rango del individuo, con lo cual se produce una repartición más uniforme de la probabilidad de selección. Si denotamos por $\text{rango}(f_j)$ el rango de la función objetivo del individuo (j) cuando los individuos de la población han sido ordenados de menor a mayor (es decir, el peor individuo tiene rango 1, mientras que el individuo con mejor función objetivo tiene rango λ), y sea p_j^{rango} la probabilidad de que el individuo (j) sea seleccionado como padre cuando la selección se efectúa proporcionalmente al rango del individuo, se tiene que

$$p_{j,t}^{\text{rango}} = (\text{rango}(f_j) / (\lambda(\lambda+1)/2))$$

Una vez seleccionados los padres, sus cromosomas se combinan utilizando habitualmente los operadores cruce y mutación. Las formas básicas de dichos operadores se describen a continuación.

El operador de cruce, coge dos padres seleccionados y corta sus cadenas de cromosomas en una posición escogida al azar, para producir dos subcadenas iniciales y dos subcadenas finales. Después se intercambian las subcadenas finales, produciéndose dos nuevos cromosomas completos (ver Figura 5).

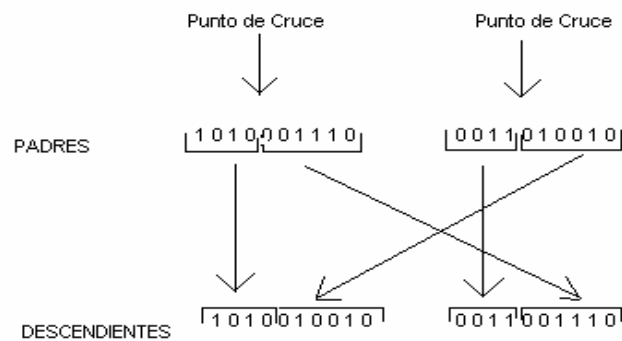


Figura 5: Operador de cruce basado en un punto

Ambos descendientes heredan genes de cada uno de los padres. Este operador se conoce como operador de cruce basado en un punto. Habitualmente el operador de cruce no se aplica a todos los pares de

Capítulo 3. Una revisión sobre algunos métodos no estadísticos...

individuos que han sido seleccionados para emparejarse, sino que se aplica de manera aleatoria, normalmente con una probabilidad comprendida entre 0.5 y 1. En el caso en que el operador de cruce no se aplique, la descendencia se obtiene simplemente duplicando los padres.

El operador de mutación se aplica a cada hijo de manera individual, y consiste en la alteración aleatoria (normalmente con probabilidad pequeña) de cada gen componente del cromosoma.

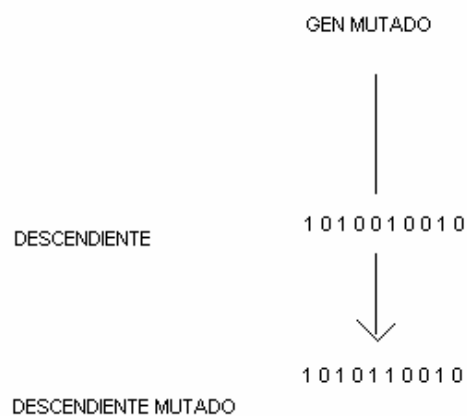


Figura 6: Operador de mutación

La Figura 6 muestra la mutación del quinto gen del cromosoma. Si bien, puede pensarse en principio que el operador de cruce es más importante que el operador de mutación, ya que proporciona una exploración rápida del espacio de búsqueda, este último asegura que ningún punto del espacio de búsqueda tenga probabilidad cero de ser

examinado, y es de capital importancia para asegurar la convergencia de los algoritmos genéticos.

Los parámetros que el analista cambia son el número de soluciones candidatas en la población, las probabilidades de cruce y de mutación, y el número de generaciones. Michalewicz [21] (1996) sugirió algunas reglas heurísticas y el tamaño de la población entre 50-100, con p_c entre 0.65 y 1.00 y p_m de 0.001 a 0.01. Albright [1] (1994) en su aplicación de Algoritmos genéticos en la construcción de un credit scoring de tarjetas utilizó una población de 100 soluciones y 450 generaciones con p_c alrededor de 0.75 y p_m de 0.05.

3.5.4 Patrones

En este apartado se tratará de explicar como y por qué funciona un Algoritmo Genético. Para esto, nos ayudaremos con un problema particular. Supongamos la siguiente información:

Nº de cadena	Cadena
1	1010
2	0111
3	1100
4	0011

Capítulo 3. Una revisión sobre algunos métodos no estadísticos ...

El término patrón fue utilizado por primera vez por Holland [15]. Si utilizamos el símbolo * para significar que no importa el valor de lo que sustituye, podemos decir que el patrón 0**1 lo poseen las cadenas 2 y 4 pero no lo contienen las cadenas 1 y 3. La similitud de las cadenas es utilizada para mostrar por qué un algoritmo genético es probable que progrese hacia una solución óptima.

Hay que advertir que cada cadena de longitud l contiene 2^l patrones. Si hay n cadenas en una población, se pueden encontrar entre 2^l y $n2^l$ patrones diferentes, dependiendo de cuantas cadenas tengan el mismo patrón.

Ahora consideramos los efectos de la creación de una nueva población con un número esperado de patrones. Se muestra como se crean nuevas generaciones, el número esperado de cadenas en una población dada que contenga el patrón más adecuado crecerá. Esto es, el patrón más adecuado llegará a dominar la población de soluciones.

Suponemos que la población es el conjunto de n_{pop} cadenas. El número de cadenas que contengan el patrón s que existirán en la nueva generación en el momento $(t+1)$ será igual a:

$$m(s,t+1) = m(s,t) \left(\frac{f(s,t)}{\sum_{j=1}^{n_{pop}} f_j} \right) n_{pop} \quad (3.10)$$

Donde $m(s,t)$ denota el número de cadenas que contienen a s en el momento t y $\frac{f(s,t)}{\sum_{j=1}^{n_{pop}} f_j}$ es la probabilidad de que una cadena que contiene el patrón s sea seleccionada. Si \bar{f} es la media de la frecuencia

absoluta de todas las cadenas de la población, podemos reescribir (3.10) como

$$m(s,t+1)=m(s,t)(f(s,t)/\bar{f}) \quad (3.11)$$

La ecuación (3.11) muestra como se crea cada generación, si la frecuencia absoluta media del patrón s es mayor que el de la población, entonces el número esperado de cadenas en la población con este patrón crecerá. De la misma manera, el número esperado de cadenas de la población con un patrón con una frecuencia absoluta relativamente baja decrecerá. De hecho, se puede ver que si la frecuencia relativa de un patrón excede a la media, el número de cadenas con este patrón crece exponencialmente conforme crece el número de generaciones. Para mostrar esto suponemos que $f(s,t) = (1+k) \bar{f}$, esto es, la frecuencia absoluta media de un patrón es una proporción constante de \bar{f} mayor que \bar{f} . Entonces la expresión (3.11), $m(s,1)=m(s,0)(1+k)$, $m(s,2)=m(s,1)(1+k)=m(s,0)(1+k)^2$, y en general $m(s,t)=m(s,0)(1+k)^t$.

Ahora rectificamos la expresión (3.11) para incluir el cruce y la mutación. Para comprender esto, necesitamos definir el orden y la longitud definida de un patrón. El orden de un patrón s , denominado $O(s)$ es el número de ceros y unos en posición fija del patrón. La longitud definida de un patrón denominada δ , es la diferencia entre las posiciones del primer número fijado en el patrón y el último.

El punto de cruce es elegido aleatoriamente y cada punto tiene la misma probabilidad de ser elegido. Así, cuanto mayor sea la longitud definida, mayor es la probabilidad de que el patrón sea destruido. En particular, si la localización del cruce está elegido en el lugar $(L-1)$, la

Capítulo 3. Una revisión sobre algunos métodos no estadísticos ...

probabilidad de que el patrón sea destruido es $\delta/(L-1)$ y la probabilidad de que este sobreviva es

$$1-(\delta/(L-1)) \quad (3.12)$$

Dado que la probabilidad de que una cadena sea seleccionada para el cruce es p_c , la probabilidad de sobrevivir puede ser rectificada por

$$1-p_c(\delta/(L-1))$$

Es posible que a causa de las posiciones de los números particulares (ceros y unos) en dos cadenas que son cruzadas, después del cruce el hijo mantenga las dos cadenas.

Se puede rectificar la expresión (3.11) para considerar el cruce. Siendo $m(s,t+1)$ el número esperado de cadenas que contienen s después de la creación de una nueva generación. La ecuación (3.11) considera únicamente la selección. De ese modo, cuando consideramos el cruce, el número esperado de cadenas que contengan s en la nueva generación es igual al número en la población reproducida multiplicado por el cambio de las cadenas supervivientes. Así la expresión (3.11) se puede reemplazar por

$$m(s,t+1) \geq m(s,t)(f(s,t)/\bar{f})(1-(p_c \delta/(L-1))) \quad (3.13)$$

Así, podemos decir que el número esperado de cadenas que contienen el patrón s crecerá entre generaciones si la probabilidad del cruce decrece y/o si la longitud definida de las cadenas en la solución decrece.

La ecuación (3.13) además puede modificarse para considerar la mutación. Intuitivamente se podría apreciar que como el orden (número de bits en posición fija) aumenta, dada la probabilidad de que un bit mute, el número esperado de cadenas que contengan un patrón en la generación siguiente será menor. Hay una mayor posibilidad de que uno de los bits fijados mute y que el hijo falle en contener el patrón s .

Podemos concluir fijándonos en la expresión (3.13), que el número de cadenas que contengan un patrón con una aptitud superior a la media y que tengan una pequeña longitud definida y un bajo orden, crecerá exponencialmente en las nuevas generaciones. Esto es conocido por el teorema del patrón.

Bibliografía

[1] : Albright, H. Construction of a Polynomial Classifier for Consumer Loan Applications Using Genetic Algorithms. Working Paper, Department of System Engineering. University of Virginia, Charlottesville. VA. (1994).

[2] : Almeida, L. A learning rule for asynchronous perceptrons with feedback in a combinatorial environment. Proceedings of the IEEE First International Conference on Neural Networks. Vol II, pags. 609-618, (1987).

[3] : Bell, T.; Ribar, G.; Verchio, J. Neural nets versus logistic regression: a comparison of each model's ability to predict commercial bank failures.

Bibliografia

Deloitte & Touche/University of Kansas Symposium of Auditing Problems, Kansas City, pp.29-53 (1990).

[4] : Bajgier, S., Hill, A. An experimental comparison of statistical and linear programming approaches to the discriminant problem. *Decision Sci.*, 13, 604-611. (1982).

[5] : Darwin, C. *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. John Murray, London. (1859).

[6] : Erenguc, S., Koehler, G. Survey of mathematical programming models and experimental results for linear discriminant analysis. *Managerial Decision Econom.*, 11, 215-225. (1990).

[7] : Freed, N., Glover, F. A linear programming approach to the discriminant problem. *Decision Sci.*, 12, 68-74. (1981).

[8] : Freed, N., Glover, F. Simple but powerful goal programming formulations for the discriminant problem. *European J. Oper. Res.*, 7, 44-60. (1981 B).

[9] : Freed, N., Glover, F. Evaluating alternative linear programming models to solve the two-group discriminant problem. *Decision Sci.*, 17, 151-162. (1986 A).

[10] : Freed, N., Glover, F. Resolving certain difficulties and improving classification power of LP discriminant analysis formulations. *Decision Sci.*, 17, 589-595. (1986 B).

Bibliografia

- [11] : Glen, J. Integer programming models for normalisation and variable selection in mathematical programming models for discriminant analysis. *J. Oper.Res. SOC.*, 50, 1043-1053. (1999).
- [12] : Glover, F. Improved linear programming models for discriminant analysis. *Decision Sci.*, 21, 771-785. (1990).
- [13] : Hand, D. *Discrimination and Classification*. John Wiley, Chichester, U.K. (1981). H2 : Hand, D. *Construction and Assessment of Classification Rules*. John Wilwy, Chichester, U.K. (1997).
- [14] : Haykin, S. *Neural Networks: A Comprehensive Foundation*. Prentice-Hall International. London. (1999).
- [15] : Holland, J. *Hierarchical Description of Universal and Adaptive System*. Department of Computer and Communication Sciences, University of Michigan, Ann. Arbor. (1968).
- [16] : Holland, J. *Adaptation in Natural and Artificial Systems*. University of Michigan Press. Ann Arbor. (1975).
- [17] : Joachimsthaler, E., Stam, A. "Mathematical programming approaches for the classication problem in two-group discriminant analysis". *Multivariate Behavioural Res.*, 25,427-454. (1990).
- [18] : Koehler, G., Erenguc, S. Minimising misclassification in linear discriminant analysis *Decision Sci.*, 21, 63-85. (1990).

Bibliografia

[19] : Le Cun, Y. Une procedure d'apprentissage pour resau a seuil assymetrique. Proceedings of Cognitiva 85: À la Frontiere de l'Intelligence Artificielle, des Sciences de la Connaissance et des Neurosciences. CESTA, Paris, 500-604. (1985).

[20] : Mangasarian, O. Linear and nonlinear separation of patterns by linear programming. Oper. Res., 13, 444-452. (1965).

[21] : Michalewicz, Z. Genetic Algorithms +Data Structures = Evolution Programs. Springer-Verlag, Berlin. (1996).

[22] : Minsky, M., Papert, S. Perceptrons. MIT Press, Cambridge, MA. (1969).

[23] : Nath, R., Jones, T. A variable selection criterion in the linear programming approaches to discriminant analysis. Decision Sci., 19, 554-563. (1988).

[24] : Nath, R., Jones, T., Jackson, W. A compararison of the classical and the linear programming approaches to the classification problem in discriminant analysis. J. Statist. Comput. Simul., 41, 73-93. (1992).

[25] : Parker, D. Learning Logic. Invention report S81-64, File 1, Office of Technology Licencing, Stanford University, Stanford, CA. (1982).

[26] : Pavar, R., Wanart, P., Loucopoulos, C. Examination of the classification performance of MIP models with secondary goals for the two group discriminant problem. Ann. Oper. Res., 74, 173-189. (1997).

Bibliografía

[27] : Pineda, F. Generalization of Back-propagation to recurrent neural networks. *Physical Review Letters*, 59, pags. 2229-2232, (1987).

[28] : Rosenblatt, F. On the Convergence of Reinforcement Procedures in Simple Perceptrons. Report VG-1196-G-4, Cornell Aeronautical Laboratory, Buffalo, NY. (1960).

[29] : Rumelhart, D., Hinton, G., Williams, R. Learning representation by back-propagating errors. *Nature*, 323, 533-536. (1986).

[30] : Richard, M., Lipman, R. Neural network classifiers estimate Bayesian a posteriori probabilities. *Neural Comput.*, 3, 461-483. (1991).

[31] : Rohwer, R., Forrest, B. Training time-dependence in neural networks. *Proceedings of the IEEE First International Conference on Neural Networks. Vol II*, pags. 701-708, (1987).

[32] : Rumelhart, D., McClelland, J. *Parallel Distributed processing: Explorations in the Microstructure of Cognition. Vol. 1*, MIT Press, Cambridge, MA. (1986).

[33] : Rubin, P. Heuristic solution procedures for a mixed-integer programming discriminant model, *Managerial Decision Econom.*, 11, 255-266. (1990).

[34] : Rosenblatt, F. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65, pags. 386-408, (1958). Reimpreso en el texto *Neurocomputing*. (J. Anderson y E. Rosenfeld ed), pags. 92-114, MIT Press, (1988).

Bibliografía

[35] : Soldevilla, C., Guillén, M. Consumer credit scoring using artificial neural networks. *Intelligent Technologies in Accounting and Business*, Huelva pp.117-128 (1997).

[36] : Werbos, P. Beyond Regression: New tools for prediction and analysis in the behavioral sciences. Ph. D. Thesis. Harvard University, (1974).

[37] : White, H. Learning in artificial neural networks: A statistical perspective. *Neural Comput.*,1, 425-464. (1989).

[38] : Widrow, B., Hoff, M. Adaptive Switching Circuits. *IREWESCON Convention Record, Part 4*, pags. 96-104, (1960). Reimpreso en el texto *Neurocomputing*. (J. Anderson y E. Rosenfeld ed), pags. 126-134, MIT Press, (1988).

[39] : Yobas, M., Crook, J., Ross, P. Credit Scoring Using Neural and Evolutionary Techniques. Working Paper 97/2, Credit Research Centre, University of Edinburgh, Edinburgh, Scotland. (1997).

[40] : Ziari, H., Leatham, D., Ellinger, P. Development of statistical discriminant mathematical programming model via resampling estimation techniques. *Amer. J. Agricultural Econom.* 79, 1352-1362. (1997).

Capítulo IV

CONSTRUCCIÓN DE UN MODELO MIXTO DE CREDIT SCORING

4.1 Introducción

La propuesta de este trabajo, que se desarrolla en este capítulo, consiste en la creación de un modelo de Credit Scoring utilizando dos técnicas, vistas en los capítulos 2 y 3, de forma conjunta, redes neuronales y métodos estadísticos, concretamente, el Algoritmo de Kohonen y el Análisis Discriminante.

El objetivo de este capítulo es desarrollar un modelo de Credit Scoring que, además de cumplir los objetivos tradicionales de calificación y clasificación de los clientes para poder tomar una decisión de conceder o no un préstamo solicitado, incorpore todos aquellos requisitos necesarios para ser considerado un modelo interno para la valoración de riesgo, como es requerido por Basilea II.

Así, el modelo de Credit Scoring que planteamos construir debe permitir obtener una clasificación de los clientes en segmentos. Cada uno

Capítulo 4. Construcción de un Modelo Mixto de Credit Scoring

de estos segmentos debe reunir unos mismos patrones de riesgo, de manera que el modelo construido permita calcular ó bien la pérdida esperada (EL), ó bien, los valores de la probabilidad de impago (PD) y de la pérdida dado impago (LGD), para cada segmento de clientes obtenido.

En la literatura consultada encontramos algunos modelos propuestos para la construcción de un Credit Scoring, tales como modelos estadísticos que emplean análisis discriminante, regresiones logísticas, clasificación en árbol, etc. Encontramos un ejemplo en el artículo de Roszbach [20], donde la autora hace un estudio para encontrar el método estadístico que mejor prediga el pago o impago de un crédito. Como modelos no estadísticos encontramos redes neuronales, programación lineal, etc, un ejemplo es el artículo de Galindo y Tamayo [10] donde emplean diversos métodos estadísticos y no estadísticos. Pero en ambos artículos y otros estudiados y citados a lo largo de este trabajo, siempre se emplean los métodos por separado. Entre la literatura consultada, un artículo clave para el desarrollo de este trabajo es el de Loreta Mester [18], Vicepresidenta de la Reserva Federal en Filadelfia donde se alude a un artículo de Altman [1]. En este último artículo el autor propone la idea de unir métodos estadísticos con redes neuronales ya que conjuntamente deberían conseguir mejorar las predicciones de un Credit Scoring.

Situados en este punto ya solo queda decidir qué método estadístico y qué modelo de red neuronal hay que conjuntar, y la decisión fue crear un modelo de Credit Scoring con Análisis Discriminante y el Algoritmo de Kohonen.

Una vez tomada esta decisión encontramos en la literatura un artículo de Shin et al.(1998) [21] donde describen un sistema de calificación del riesgo de crédito que está funcionando en la práctica en un banco coreano, el Boram Bank, consistente en un modelo híbrido de perceptrón multicapa y regresión logística. Este sistema, que incorpora modelos

separados para diferentes sectores de la economía, calcula para cada empresa una puntuación basada en los resultados del análisis logit y de un perceptrón con tres capas. Posteriormente, cada empresa es asignada con arreglo a sus puntuación a una categoría determinada de entre un total de siete, cada una de las cuales define un determinado nivel de riesgo de crédito.

En este estudio también se comparan la eficacia del perceptrón y el logit por separado, no resultando diferencias significativas entre la capacidad clasificadora de cada técnica. El conocimiento de la existencia de este trabajo nos ratificó en la idea que desarrollamos en esta tesis.

En cuanto a la aplicación del Algoritmo de Kohonen en este tema encontramos que ya ha sido empleado en alguna ocasión, lo que nos reafirma en la decisión de su empleo en este trabajo, dos ejemplos a mencionar son: el trabajo de Serrano Cinca y Martín del Brio (1993) [24] que analizan la crisis bancaria española durante los años 80 y el de López González y Flórez López (1999) [23] que caracterizan la solvencia empresarial.

En este capítulo trataremos de explicar y razonar nuestra propuesta de modelo de Credit Scoring y cuáles son las aportaciones al mismo de cada uno de los mecanismos empleados, por esto, el desarrollo del capítulo se ha configurado de la siguiente manera. En el apartado 2 se explica cómo se utilizan los métodos de Análisis Discriminante y el Algoritmo de Kohonen en la elaboración de nuestro modelo de Credit Scoring, detallando los conceptos fundamentales que se emplean. El método estadístico nos proporcionará las variables significativas del modelo, es decir, las características de los clientes que servirán para determinar la solvencia o insolvencia de los mismos, y para ello utilizaremos el concepto de los coeficientes estandarizados.

Capítulo 4. Construcción de un Modelo Mixto de Credit Scoring

A continuación, a los datos obtenidos les aplicaremos una formulación de elaboración propia, que nos proporcionará una calificación para cada cliente en función de la proporción de características significativas que cada uno posee.

Una vez obtenida la sucesión de calificaciones de todos los clientes aplicaremos el Algoritmo de Kohonen con el que obtendremos una clasificación de los clientes por segmentos.

En el apartado 3 incorporamos lo obtenido en el apartado anterior a los conceptos definidos por Basilea II. Lo más relevante a destacar es que el Algoritmo de Kohonen nos permitirá obtener una clasificación de los clientes en segmentos, y se realizará una correspondencia entre las calificaciones de los clientes y el valor de las componentes de riesgo (PD, LGD, y EL) en cada uno de los segmentos de clasificación, tal y como propone el Nuevo Acuerdo de Capital de Basilea, y para terminar se enumeran las principales conclusiones obtenidas.

4.2 Modelo mixto de Credit Scoring

4.2.1 Modelo de Análisis Discriminante

El desarrollo teórico, así como las características del Análisis Discriminante lo hemos visto en el capítulo 3 de esta tesis. En este epígrafe utilizamos este método estadístico con la pretensión de clarificar la relevancia de una serie de variables en el pago de las concesiones de créditos.

Aplicando este método a un Credit Scoring lo primero que tenemos que hacer es definir las variables, de manera que: X_1, X_2, \dots, X_p son las características que consideramos que a priori van a explicar que un

Modelo Mixto de Credit Scoring

cliente es solvente o insolvente. Por otro lado, Y , es la variable dependiente que nos indica si un cliente ha pagado o no, es decir, la podemos definir como un ratio de impago, siendo entonces $Y \in [0,1]$.

Si no tenemos información suficiente del desarrollo del crédito diremos que si este ha resultado impagado $Y=1$, y en caso contrario $Y=0$. Si tuviéramos más información entonces definiríamos Y como:

$$Y=1-(\text{tiempo hasta el impago}/\text{tiempo de duración del préstamo})$$

Si llamamos $T=$ tiempo de duración del préstamo y $\tau=$ momento de impago, podemos reescribir la variable Y como

$$Y=(T-\tau)/T$$

Si se hubiesen producido aplazamientos de pago pactados con la entidad bancaria se considerarían un mayor tiempo de duración del préstamo y por lo tanto quedaría modificado T al alza.

En nuestro problema de credit scoring sólo necesitamos dividir a la población de clientes en dos grupos, solventes e insolventes, esto es, $g=2$, por lo tanto, el número de funciones discriminantes que vamos a tener será $k=g=2$ y quedan expresadas de la siguiente forma:

$$Y_k = a_{k1} X_1 + a_{k2} X_2 + \dots + a_{kp} X_p \quad k=1,2 \quad (4.1)$$

obteniendo que las puntuaciones discriminantes para el individuo j del grupo g_i son

Capítulo 4. Construcción de un Modelo Mixto de Credit Scoring

$$y_{ij,k} = a_{k1} x_{ij,1} + a_{k2} x_{ij,2} + \dots + a_{kp} x_{ij,p} \quad k=1,2 \quad (4.2)$$

al mismo tiempo se ha calculado la matriz de covarianzas entre grupos.

$$B = (1/(g-1)) \sum_{i=1}^g n_i (\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})'$$

y la matriz de covarianzas dentro de grupos

$$W = (1/(n-g)) \sum_{i=1}^g \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)(x_{ij} - \bar{x}_i)'$$

Suponiendo que W es no singular, podemos obtener el vector propio de $W^{-1}B$, \mathbf{a}_1 , correspondiente al mayor valor propio $m = \lambda_1$. Sea $\mathbf{a}'_1 = [a_{11}, a_{12}, \dots, a_{1p}]$, obtenemos la primera función discriminante a partir de (4.1).

$$Y_1 = a_{11} X_1 + a_{12} X_2 + \dots + a_{1p} X_p$$

Análogamente, la segunda función discriminante viene del vector de $W^{-1}B$ que corresponde al segundo valor propio, de manera que los

Modelo Mixto de Credit Scoring

vectores $\mathbf{a}_1, \mathbf{a}_2$ que corresponden a los valores propios $\lambda_1 \geq \lambda_2 \geq 0$ de $W^{-1}B$ definen las funciones discriminantes de (4.1).

Pero lo que a nosotros nos interesa estudiar es la relación entre las funciones discriminantes, y a partir de aquí supondremos que los valores propios positivos son distintos, esto es, $\lambda_1 > \lambda_2$. Entonces podemos decir que la matriz AWA' es una matriz diagonal 2×2 . Siendo A

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \end{pmatrix}$$

Si definimos

$$\mathbf{a}_k^* = [\mathbf{a}_{k1}^*, \mathbf{a}_{k2}^*, \dots, \mathbf{a}_{kp}^*] / (\mathbf{a}_k' W \mathbf{a}_k)^{1/2} \quad k=1,2 \quad (4.3)$$

obtenemos la siguiente igualdad:

$$A^* W (A^*)' = I_k \quad (4.4)$$

que nos permite obtener una función discriminante

$$Z_k = \mathbf{a}_{k1}^* X_1 + \mathbf{a}_{k2}^* X_2 + \dots + \mathbf{a}_{kp}^* X_p \quad k=1,2$$

Capítulo 4. Construcción de un Modelo Mixto de Credit Scoring

(generada por el vector \mathbf{a}_k^* en (4.3) que da al individuo j del grupo g_k la puntuación discriminante,

$$Z_{ij,k} = \mathbf{a}_{k1}^* x_{ij,1} + \mathbf{a}_{k2}^* x_{ij,2} + \dots + \mathbf{a}_{kp}^* x_{ij,p}$$

siendo

$$\mathbf{A}^* = \begin{pmatrix} 1/(a_1 W a_1')^{1/2} & 0 \\ 0 & 1/(a_2 W a_2')^{1/2} \end{pmatrix} \mathbf{A}$$

Para cada Z_k , la varianza dentro de grupos es uno, y eso hace que en \mathbf{A}^* los coeficientes de una sola variable original en las funciones discriminantes sean comparables. Sin embargo, para poder comparar los coeficientes de diferentes variables originales en una misma función discriminante, conviene multiplicar la matriz \mathbf{A}^* por la matriz \mathbf{G} , que viene dada por:

$$\mathbf{G} = \begin{pmatrix} \sqrt{\text{Var}(X_1)} & 0 & 0 & 0 \\ 0 & \sqrt{\text{Var}(X_2)} & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sqrt{\text{Var}(X_p)} \end{pmatrix}$$

Modelo Mixto de Credit Scoring

La matriz A^*G hace posible comparar los coeficientes de diferentes variables originales en una misma función discriminante ya que hemos quitado la influencia de las diferentes variabilidades originales al multiplicar cada coeficiente por la desviación típica dentro de grupos de la variable original correspondiente. Los valores de la matriz A^*G son los coeficientes estandarizados de las funciones discriminantes. Esto es, si nosotros tomamos la función Z_1 como discriminante de los clientes solventes y multiplicamos los valores $a_{11}^*, a_{12}^*, \dots, a_{1p}^*$ por G , estamos encontrando unos coeficientes cuyos valores nos permiten decidir la influencia mayor o menor (según el valor que obtengamos) de las características en la solvencia del préstamo, y por lo tanto podemos establecer un orden de incidencia de las características sobre el pago del crédito.

Los coeficientes con los que definitivamente vamos a trabajar serán los calculados con A^*G , esto es, los coeficientes estandarizados de la función discriminante de los clientes solventes, que para no incrementar la nomenclatura los seguiremos denotando por $a_{11}^*, a_{12}^*, \dots, a_{1p}^*$.

Hay que advertir que la aplicación del análisis discriminante también nos puede proporcionar la eliminación de algunas de las variables X_1, X_2, \dots, X_p , ó bien por una alta correlación entre ellas ó sencillamente por falta de significatividad para la discriminación entre créditos solventes e insolventes. De manera que es posible que el número de características a tener en cuenta sea $d \leq p$.

Las principales aportaciones a nuestro credit scoring del análisis discriminante son:

Capítulo 4. Construcción de un Modelo Mixto de Credit Scoring

- La aportación fundamental es poder obtener los valores de los coeficientes estandarizados a_{1j}^* , porque nos permiten establecer un orden de incidencia de las características sobre el pago del crédito, tanto es así que a partir de ahora vamos a utilizar el valor de estos coeficientes para denotar el peso de las variables X_i en el pago o impago de los préstamos. La utilización de estos coeficientes solventa al mismo tiempo, el problema que presentan las redes neuronales, puesto que estas no distinguen las características influyentes en la solvencia del préstamo de las que no lo son.

- La utilización del análisis discriminante no solo nos cuantifica la importancia de las variables, sino que nos cualifica a las mismas, esto es, nos dice si las variables relevantes son p ó $d \leq p$

- Además, el análisis discriminante como técnica de clasificación de datos nos proporciona una clasificación inicial de clientes, y con la elaboración de nuestro modelo queremos aportar que al añadir el método de redes neuronales esta primera clasificación se puede mejorar, por lo tanto nos aporta la posibilidad de comparar los resultados obtenidos solo con análisis discriminante con los conseguidos por nuestro modelo mixto.

Sin embargo, el análisis discriminante no puede aportar una segmentación fundamentada del riesgo, y esta será la aportación principal del modelo de redes neuronales que vemos en la siguiente sección

4.2.2 Elaboración del Modelo

En la elaboración del modelo propuesto se planteaba la posibilidad de recurrir a valoraciones del cliente ya realizadas, o bien, construir una nueva forma de calificación. Nosotros optamos por la segunda por dos

razones, la primera es que esto supone una aportación, creemos que importante, para este trabajo y la segunda razón es que así, partíamos de una valoración que nos servía para la siguiente parte del modelo que consiste en la clasificación del cliente y en el cálculo de la probabilidad de impago.

La obtención de la nota del cliente no ha sido sencilla, en un primer momento nos planteamos aplicar el algoritmo de Kohonen en cada una de las características y hacer clasificaciones de cada una de ellas, de manera que cada clase conseguida en una característica le asignáramos tanto el valor de los coeficientes estandarizados obtenidos en el análisis discriminante como en el algoritmo, pero este método de calificación, además de ser relativamente complicado presentaba problemas con las características que inciden negativamente en el préstamo. Con la misma perspectiva de seguir utilizando los dos métodos para la calificación se elaboró una formulación parecida a la definitiva con el hándicap de que no permitía obtener una función continua de las mismas, lo que nos llevaba a perder muchas posibles propiedades con las que poder trabajar con posterioridad con esta función, y tras múltiples pruebas llegamos a la conclusión de que la más satisfactoria es la que expresamos a continuación

$$U=U(X_1, X_2, \dots, X_d) = (1 / \sum_{j=1}^d |a_{1j}^*|) \sum_{j=1}^d a_{1j}^* (I_{(0, \infty)}(a_{1j}^*) - X_j)$$

donde

$$I_A(x) = \begin{cases} 1 & \text{si } x \in A \\ 0 & \text{si } x \notin A \end{cases}$$

Capítulo 4. Construcción de un Modelo Mixto de Credit Scoring

Para obtener el valor de U para todos y cada uno de los clientes cuyas características ya conocemos y sabemos como han respondido ante el préstamo concedido, lo primero que tenemos que hacer es normalizar los valores de las variables X_j de manera que $0 \leq X_j \leq 1$ con $j=1, \dots, d$. Además, en el modelo propuesto entenderemos que el cliente tiene mejor calificación U cuando más se aproxime a 0 y tendrá peor calificación cuando U se acerque más a 1.

Nótese que si $a_{1j}^* < 0$, entonces el sumando correspondiente es

$$-a_{1j}^* X_j$$

lo que hace que cuanto más posea el cliente de esta característica ($X_j \cong 1$) que influye negativamente en la concesión de un crédito, más penalizaremos la nota. Recíprocamente, cuanto menos posea de esta característica ($X_j \cong 0$) menor será la nota obtenida con respecto a esta misma.

Si $a_{1j}^* > 0$, entonces el sumando correspondiente será

$$a_{1j}^* (1 - X_j).$$

En esta situación, cuanto más posea el cliente de la característica ($X_j \cong 1$), la calificación que esta aporta a U es cercana a cero.

Por lo tanto, siguiendo este criterio de valoración, podemos deducir que $0 \leq U \leq 1$.

Modelo Mixto de Credit Scoring

Uno de los objetivos planteados por Basilea II es la implantación de segmentación de créditos según el riesgo de concesión de los mismos. En nuestro modelo asociamos la medición del riesgo con la calificación que obtenga el cliente, en este caso, entendemos que cuanto menor sea la nota obtenida menor es el riesgo del capital prestado y viceversa. Esto nos conduce pues, a hacer una segmentación de las calificaciones obtenidas por los clientes mediante la aplicación del Algoritmo de Kohonen.

El proceso de aplicación del algoritmo está más detallado en el apartado siguiente y como se puede ver, al realizar la clasificación de las notas U podemos obtener un vector:

$$\mathbf{q}^* = [q_1^*, q_2^*, \dots, q_m^*]'$$

donde cada q_i^* representa la calificación hacia la que convergen las notas de los clientes próximos, y por lo tanto, decimos que esos están representados por la calificación q_i^* . Cada q_i^* representará a un segmento de riesgo quedando los extremos de los segmentos \tilde{q}_i configurados de la siguiente manera:

$$\tilde{q}_1 = 0, \quad \tilde{q}_k = (q_k^* + q_{k+1}^*)/2 \quad \text{para} \quad 2 \leq k \leq m, \quad \tilde{q}_{m+1} = 1$$

Al mismo tiempo cada q_i^* será el valor del riesgo de concesión de riesgo en el segmento al que representa y como $0 \leq U \leq 1$ diremos que cada valor q_i^* es la probabilidad de impago de los clientes que pertenecen a ese segmento de riesgo.

Capítulo 4. Construcción de un Modelo Mixto de Credit Scoring

Podemos asegurar que el vector \mathbf{q}^* es único siempre que la función de densidad de las calificaciones obtenidas f_Z cumpla la propiedad de log-concavidad y esto se cumplirá cuando:

$$f_Z = f_{-a_{11}X_1} * \dots * f_{-a_{1s}X_s} * f_{a_{1s+1}(1-X_{s+1})} \dots * f_{a_{1d}(1-X_d)}$$

Esto es, cuando la función de densidad de las calificaciones coincida con la convolución de las funciones de densidad de las calificaciones parciales.

4.2.3 Algoritmo de Kohonen

La utilización de este algoritmo como método de clasificación para un credit scoring es otra de las aportaciones de esta tesis, ya que aunque se han utilizado otros modelos de redes neuronales, no hemos encontrado en la literatura ninguna clasificación de créditos usando el Algoritmo de Kohonen unidimensional. Entre las propiedades de este algoritmo destacan que la solución que se obtiene es única y que nos permite obtener una clasificación por segmentos de los valores de la variable, que en este caso es la calificación de los clientes, requisito de Basilea II.

Como este es un método poco extendido en la literatura procedemos en primer lugar a explicar el algoritmo de forma general para después especificar su aplicación en credit scoring.

Generalidades del Algoritmo de Kohonen

En 1982 Kohonen [15] introdujo lo que él llamó mapas autoorganizativos (SOM, Self-Organizing Maps) que consisten en modelos muy simples basados en el comportamiento neuronal. El interés que despertaron en su momento viene motivado, fundamentalmente, por su extremada simplicidad. Las investigaciones matemáticas acerca de estos mecanismos comienzan con el trabajo anteriormente citado del mismo autor. Este proporcionó un esbozo de demostración acerca de la propiedad autoorganizativa del caso unidimensional y cuando el estímulo del modelo sigue una distribución de probabilidad uniforme. En realidad, es conocido que, incluso en este caso, el tratamiento matemático no es sencillo. Cotrell y Fort [5] dieron para este caso particular una demostración rigurosa de la autorganización y de la convergencia casi segura cuando el ratio de aprendizaje es decreciente. Estos resultados fueron extendidos para distribuciones de estímulos mucho más generales por Bouton y Pagès [4]. En 1996, Fort y Pagès [9] dieron un estudio detallado de la Ecuación Diferencial Ordinaria asociada al algoritmo. En particular, demostraron que todos los puntos de equilibrio eran asintóticamente estables. Estos resultados se obtuvieron bajo la hipótesis de que las distribuciones de los estímulos eran logarítmico-cóncavas.

La arquitectura de la versión original del modelo de Kohonen es una red de dos capas con N neuronas de entrada y M de salida. Cada una de las N neuronas de entrada se conecta a las M de salida a través de conexiones hacia delante (feedforward).

Entre las neuronas de la capa de salida, puede decirse que existen conexiones laterales de inhibición (peso negativo) implícitas, pues aunque no estén conectadas, cada una de estas neuronas va a tener cierta influencia sobre sus vecinas. El valor que se asigne a los pesos de las conexiones feedforward entre las capas de entrada y salida (q_{ji}) durante

Capítulo 4. Construcción de un Modelo Mixto de Credit Scoring

el proceso de aprendizaje de la red va a depender precisamente de esta interacción lateral.

El funcionamiento de esta red es relativamente simple. Cuando se presenta a la entrada una información $E_k = (e_1^{(k)}, \dots, e_N^{(k)})$, cada una de las M neuronas de la capa de salida la recibe a través de las conexiones feedforward con pesos q_{ji} . También estas neuronas reciben las correspondientes entradas debidas a las conexiones laterales con el resto de las neuronas de salida y cuya influencia dependerá de la distancia a la que se encuentren.

Así, la salida generada por una neurona de salida j ante un vector de entrada E_k sería:

$$s_j(t+1) = f\left(\sum_{i=1}^N q_{ji} e_i^{(k)} + \sum_{p=1}^M \text{Int}_{pj} s_p(t)\right)$$

donde Int_{pj} es una función del tipo sombrero mejicano que representa la influencia lateral de la neurona p sobre la neurona j . La función de activación de las neuronas de salida (f) será del tipo continuo, lineal o sigmoideal, ya que esta red trabaja con valores reales.

Es evidente que se trata de una red de tipo competitivo, ya que al presentar una entrada E_k la red evoluciona hasta una situación estable en la que se activa una neurona de salida, la vecendora. Por ello, la formulación matemática de su funcionamiento puede simplificarse mediante la siguiente expresión, que representa cuál de las M neuronas se activaría al introducir dicha información E_k :

$$s_j = \begin{cases} 1 & \text{MIN} \|E_k - Q_j\| = \text{MIN} \left(\sum_{i=1}^N (e_i^{(k)} - q_{ji})^2 \right)^{1/2} \\ 0 & \text{resto} \end{cases}$$

donde $\|E_k - Q_j\|$ es una medida (por ejemplo, distancia euclidea) de la diferencia entre el vector de entrada $E_k = (e_1^{(k)}, \dots, e_N^{(k)})$, y el vector de los pesos $Q_j = (q_{j1}, \dots, q_{jN})$ de las conexiones entre cada una de las neuronas de entrada y la neurona de salida j . En estos pesos se registran los datos almacenados en la red durante el proceso de aprendizaje. En la fase de funcionamiento, lo que se pretende es encontrar el dato aprendido más parecido a la entrada para, en consecuencia, averiguar qué neurona se activará y, sobre todo, en qué zona del espacio bidimensional de salida se encuentra.

Lo que hace la red de Kohonen, en definitiva, es realizar una tarea de clasificación, ya que la neurona de salida activada ante una entrada representa la clase a la que pertenece dicha información de entrada. Además, como ante otra entrada parecida se activa la misma neurona de salida, u otra cercana a la anterior, debido a la semejanza entre las clases, se garantiza que las neuronas topológicamente próximas sean sensibles a entradas físicamente similares. Por esta causa, la red es especialmente útil para establecer relaciones, desconocidas previamente, entre conjuntos de datos.

El aprendizaje en el modelo de Kohonen es OFF LINE, por lo que se distingue una etapa de aprendizaje y otra de funcionamiento. En la etapa de aprendizaje se fijan los valores de los pesos de las conexiones (feedforward) entre la capa de entrada y la de salida.

Capítulo 4. Construcción de un Modelo Mixto de Credit Scoring

Esta red utiliza un aprendizaje no supervisado de tipo competitivo. Las neuronas de la capa de salida compiten por activarse y sólo una de ellas permanece activa ante una determinada información de entrada a la red. Los pesos de las conexiones se ajustan en función de la neurona que haya resultado vencedora.

Durante la etapa de entrenamiento, se presenta a la red un conjunto de informaciones de entrada (vectores de entrenamiento) para que ésta establezca, en función de la semejanza entre los datos, las diferentes categorías (una por neurona de salida) que servirán durante la fase de funcionamiento para realizar clasificaciones de nuevos datos que se presenten a la red. Los valores finales de los pesos de las conexiones entre cada neurona de la capa de salida con las de entrada se corresponderán con los valores de los componentes del vector de aprendizaje que consigue activar la neurona correspondiente. En el caso de existir más patrones de entrenamiento que neuronas de salida, más de uno deberá asociarse con la misma neurona; es decir, pertenecerán a la misma clase. En tal caso, los pesos se obtienen como un promedio de dichos patrones.

En este modelo, el aprendizaje no concluye después de presentarle una vez todos los patrones de entrada, sino que habrá que repetir el proceso varias veces para refinar el mapa topológico de salida, de tal forma que cuantas más veces se presenten los datos, tanto más se reducirán las zonas de neuronas que se deben activar ante entradas parecidas, consiguiendo que la red pueda realizar una clasificación más selectiva.

El algoritmo de aprendizaje utilizado para establecer los valores de los pesos de las conexiones entre las N neuronas de entrada y las M de salida es el siguiente:

Modelo Mixto de Credit Scoring

1. En primer lugar, se inicializan los pesos (q_{ji}) con valores aleatorios pequeños y se fija la zona inicial de vecindad entre las neuronas de salida.

2. A continuación se presenta a la red una información de entrada (la que debe aprender) en forma de vector $E_k = (e_1^{(k)}, \dots, e_N^{(k)})$, cuyas componentes $e_i^{(k)}$ serán valores continuos.

3. Puesto que se trata de un aprendizaje competitivo, se determina la neurona vencedora de la capa de salida. Esta será aquella j cuyo vector de pesos Q_j (vector cuyas componentes son los valores de los pesos de las conexiones entre esa neurona y cada una de las neuronas de la capa de entrada) sea el más parecido a la información de entrada E_k (patrón o vector de entrada).

Para ello, se calculan las distancias o diferencias entre ambos vectores, considerando una por una todas las neuronas de salida. Suele utilizarse la distancia euclídea o la siguiente expresión, que es similar a aquella, pero eliminando la raíz cuadrada:

$$d_j = \sum_{i=1}^N (e_i^{(k)} - q_{ji})^2 \quad 1 \leq j \leq M$$

Siendo:

$e_i^{(k)}$: Componente i -ésimo del vector k -ésimo de entrada.

q_{ji} : Peso de la conexión entre la neurona i de la capa de entrada y la neurona j de la capa de salida.

Capítulo 4. Construcción de un Modelo Mixto de Credit Scoring

4. Una vez localizada la neurona vencedora (j^*), se actualizan los pesos de las conexiones entre las neuronas de entrada y dicha neurona, así como los de las conexiones entre las de entrada y las neuronas vecinas de la vencedora. En realidad, lo que se consigue con esto es asociar la información de entrada con una cierta zona de la capa de salida

$$q_{ji}(t+1) = q_{ji}(t) + \alpha(t)[e_i^{(k)} - q_{j^*i}(t)] \quad \text{para } j \in \text{Zona}_{j^*}(t)$$

Donde $\text{Zona}_{j^*}(t)$ es la zona de vecindad alrededor de la neurona vencedora j^* en la que se encuentran las neuronas cuyos pesos son actualizados. El tamaño de esta zona se puede reducir en cada iteración del proceso de ajuste de los pesos, con lo que el conjunto de neuronas que pueden considerarse vecinas cada vez es menor. Sin embargo, en la práctica es habitual considerar una zona fija en todo el proceso de entrenamiento de la red.

El término $\alpha(t)$ es un parámetro de ganancia o coeficiente de aprendizaje, con un valor entre 0 y 1, decrece con el número de iteraciones (t) del proceso de entrenamiento. De tal forma que cuando se ha presentado un gran número de veces todo el juego de patrones de aprendizaje ($500 \leq t \leq 10000$) su valor es prácticamente nulo, con lo que la modificación de los pesos es insignificante.

Suele utilizarse alguna de las siguientes expresiones:

$$\alpha(t) = (1/t) \quad \alpha(t) = \alpha_1 (1 - (t/\alpha_2))$$

Modelo Mixto de Credit Scoring

Siendo α_1 un valor de 0.1 ó 0.2 y α_2 un valor próximo al número total de iteraciones del aprendizaje. Suele tomarse un valor $\alpha_2 = 10000$

El proceso se debe repetir, volviendo a presentar todo el juego de patrones de aprendizaje E_1, E_2, \dots , un mínimo de 500 veces ($t \geq 500$)

Aplicación del Algoritmo en el modelo de Credit Scoring

Recordamos que estamos considerando un conjunto de aplicaciones de préstamos o prestatarios ya existentes en una institución bancaria a partir de los cuales se quiere crear una clasificación que le permita a la entidad disponer de un credit scoring y por lo tanto poder tomar decisiones de cobertura de capital. Así, ante la disyuntiva de conceder o no un préstamo a un nuevo cliente, el objetivo de la elaboración de nuestro credit scoring es obtener una calificación de este, de manera que podamos asociar esta calificación con el posible pago o impago del crédito solicitado y decidir si se le concede o no y en caso afirmativo medir el grado de riesgo que asume la entidad financiera en esta operación.

Si partimos de datos históricos sobre préstamos anteriores tendremos una sucesión de n calificaciones asociadas a los n clientes. Aplicaremos el Algoritmo de Kohonen a esta sucesión de calificaciones para poder realizar una partición de ellas en segmentos que nos reflejarán una tendencia de comportamiento similar de los clientes que pertenezcan a un mismo segmento.

Consideraremos que nuestra muestra aleatoria se extrae de una sucesión

$$\{U^{(n)}\}_{n \in N}$$

Capítulo 4. Construcción de un Modelo Mixto de Credit Scoring

donde U es una variable aleatoria independiente con distribución continua μ . Nuestro objetivo es buscar una segmentación en función de su distribución de probabilidad. Para obtener esta segmentación utilizaremos el Algoritmo de Kohonen unidimensional. Este algoritmo está basado en un proceso de aprendizaje adaptativo y no supervisado.

El objetivo de este algoritmo es la obtención de segmentos de calificaciones para la sucesión $\{U^{(n)}\}_{n \in N}$ de manera, que cada uno de ellos represente un valor de probabilidad de impago, y así, cuando la calificación de un cliente pertenezca a un segmento determinado, podremos decir que su probabilidad de impago es la que está asociada a este segmento. Así, el objetivo es encontrar un vector de valores de calificación

$$\mathbf{q}^* = [q_1^*, q_2^*, \dots, q_m^*]'$$

de manera que la sucesión $\{U^{(n)}\}_{n \in N}$ queda dividida en m clases o segmentos de calificación y cada q_i^* es la calificación representativa de cada segmento y que debe estar asociada a la probabilidad de impago que el mismo representa.

Ahora, introducimos el Algoritmo de Kohonen, para explicar como obtenemos los vectores \mathbf{q}^* . Consideremos

$$\mathbf{q}^0 = [q_1^0, q_2^0, \dots, q_m^0]'$$

un valor inicial que representa una segmentación de las calificaciones, ahora tomaríamos la calificación de un cliente U_1 y mediante la aplicación de la fase competitiva del algoritmo,

$$\arg \min_{k=1\dots m} |U^{(n+1)} - q_k^n|$$

mediríamos la distancia de esta con cada componente de \mathbf{q}^0 , y asociaríamos al cliente 1 con el q_i^0 cuya distancia haya resultado menor, pero ahora la componente q_i^0 se vería afectada en su valor mediante la aplicación de la fase cooperativa del Algoritmo de Kohonen.

$$q_k^{n+1} = q_k^n - \varepsilon_{n+1} (q_k^n - U^{(n+1)})$$

donde $(\varepsilon_n)_{n \geq 1}$ es una secuencia de números reales comprendidos en el intervalo $[0,1]$. Usualmente se conoce a esta sucesión como ratio de aprendizaje.

Esta operación se realizaría con todos los valores de la sucesión $\{U^{(n)}\}_{n \in N}$ y obtendríamos \mathbf{q}^1 , así, diríamos que se ha realizado un epoch.

Si este proceso lo realizamos n veces (n epochs), obtenemos:

$$\mathbf{q}^n = [q_1^n, q_2^n, \dots, q_m^n]'$$

que representará la segmentación en la iteración $n \in N$.

Habremos obtenido el vector \mathbf{q}^* cuando $q^n = q^{n+1}$ y por lo tanto se cumple que $q^n = q^{n+1} = \mathbf{q}^*$

Los valores q_i^* son valores de calificación hacia los que convergen las calificaciones cercanas a ellas, por lo tanto, serán los representantes del

Capítulo 4. Construcción de un Modelo Mixto de Credit Scoring

valor de la calificación de cada segmento, quedando los extremos de los segmentos \tilde{q}_i configurados de la siguiente manera:

$$\tilde{q}_1=0, \quad \tilde{q}_k=(q_k^*+q_{k+1}^*)/2 \quad \text{para} \quad 2 \leq k \leq m, \quad \tilde{q}_{m+1}=1,$$

Cada q_k^* con $k=1, \dots, m$ sería el valor representativo de todas las calificaciones englobadas en el intervalo $\tilde{q}_k < q_k^* \leq \tilde{q}_{k+1}$.

4.3 Cálculo de las componentes de riesgo

El Nuevo Acuerdo de Basilea II obliga a las entidades bancarias que quieran utilizar un método interno de medición de riesgo a que este permita calcular unas determinadas componentes de riesgo, que son: la Probabilidad de Impago y la Pérdida dada de Impago, o bien la Pérdida Esperada.

Si se trata de un método pensado para exposiciones al detalle (pequeños préstamos), el valor de estas componentes de riesgo se puede calcular por segmentos de riesgo.

Una vez definido nuestro modelo pasamos a explicar a continuación como podemos hallar los valores de dichas componentes.

Probabilidad de impago

La probabilidad de impago o tasa de morosidad se define como la probabilidad de que el cliente entre en mora en un determinado periodo de tiempo.

Cálculo de las componentes de riesgo

Como q_i^* es el valor de convergencia de las calificaciones U comprendidas en el intervalo $[\tilde{q}_{i-1}, \tilde{q}_i]$, podemos decir entonces, que el conjunto de clientes agrupados en cada segmento $[\tilde{q}_{i-1}, \tilde{q}_i]$ tiende a comportarse como si hubieran alcanzado cada uno de ellos la calificación q_i^* .

Supongamos que $q_1^* \cong 0$, podríamos decir que los clientes del segmento $[\tilde{q}_0, \tilde{q}_1]$ son solventes con una probabilidad de impago igual a cero. En el caso contrario, con $q_m^* \cong 1$, los clientes comprendidos en el segmento $[\tilde{q}_{m-1}, \tilde{q}_m]$ tendrían una probabilidad de impago igual a 1.

Si esto es válido para los extremos, nos hace pensar que también lo es para los segmentos intermedios, puesto que el Algoritmo de Kohonen nos ratifica que el valor $q_i^* \in [\tilde{q}_{i-1}, \tilde{q}_i]$ es el valor de convergencia del segmento. En nuestro caso, q_i^* representa calificaciones semejantes obtenidas por clientes de características parecidas y por lo tanto con una probabilidad de comportamiento semejante, comportamiento que el Algoritmo nos permite calcular y concretar en los valores q_i^* y por lo tanto definimos la probabilidad de impago para cada segmento como

$$P(D|U \in [\tilde{q}_{i-1}, \tilde{q}_i]) = q_i^*$$

Capítulo 4. Construcción de un Modelo Mixto de Credit Scoring

Pérdida dada de impago (LGD)

También denominada Severidad, representa el ratio entre las pérdidas efectivas incurridas, como consecuencia de la entrada en mora y la Exposición.

Recordamos lo ya expuesto en el capítulo primero de este trabajo que para su cálculo es necesario pasar todos los flujos asociados a la recuperación, tanto positivos (pagos/venta de activos) como negativos (costes internos/judiciales), al valor presente en el día de entrada en mora, de manera que

% Recuperación=((Valor en el momento de la mora de Ingresos-
Costes de recuperación)/(Importe de la deuda al momento de entrar en
mora))

% LGD=1-% Recuperación

Suponemos que la entidad bancaria ha concedido pequeños préstamos por un valor C , término que Basilea II denomina EAD (Exposure at default). En cada uno de los segmentos definidos por q_i^* podemos contabilizar la cantidad de préstamos realizados en el mismo, la cantidad de capital expuesto a impago en cada segmento será C_i , de manera que

$\sum_{i=1}^m C_i = C$. De cada cantidad C_i podemos conocer la cantidad que ha resultado impagada y la denominaremos \overline{C}_i suponiendo que este valor ya recoge la recuperación, de forma que $0 \leq \overline{C}_i \leq C_i$.

Cálculo de las componentes de riesgo

Podemos decir entonces que en cada segmento la pérdida dada de impago (LGD) será $(\overline{C}_i / C_i) \times 100$

Pérdida Esperada (EL)

La pérdida esperada se define como la media matemática de la pérdida crediticia y también se denomina pérdida anticipada. En el Nuevo Acuerdo de Basilea II encontramos la expresión.

$$EL = PD \times EAD \times LGD$$

para el cálculo de este concepto definido para exposiciones empresariales, pero no para pequeños préstamos, por lo tanto, no realizamos ningún estudio sobre esta componente de riesgo, ya que este trabajo está enfocado en la elaboración de un credit scoring para exposiciones al detalle. Aunque, a la espera de que Basilea II proponga una formulación alternativa para el cálculo de EL para pequeños préstamos, nosotros vamos a utilizarla para el caso práctico que desarrollamos en el capítulo 5 de esta tesis.

4.4 Conclusiones

La necesidad de las instituciones financieras de obtener métodos eficientes para la asignación de una clasificación del riesgo para un conjunto elevado de consumidores y que cumplan con las condiciones de idoneidad de Basilea II ha sido la razón del desarrollo del método de Credit Scoring de este capítulo. En este trabajo presentamos un modelo

Capítulo 4. Construcción de un Modelo Mixto de Credit Scoring

de Credit Scoring basado en un modelo de Análisis Discriminante en conjunción con un modelo de Redes Neuronales, concretamente, el Algoritmo de Kohonen. Las conclusiones más relevantes de este método presentado son:

- I. Los modelos elegidos son complementarios, uno permite corregir las carencias del otro y viceversa, esto es, el análisis discriminante permite discernir las variables relevantes de las que no lo son y cuantificar en que medida son significativas, propiedades estas, de las que carecen los modelos de redes neuronales que consideran todas las variables introducidas en el modelo con la misma importancia. Por otro lado el Algoritmo de Kohonen permite una clasificación más fidedigna de las variables de lo que a priori se consigue con el análisis discriminante.
- II. Los resultados obtenidos a partir del Algoritmo solo dependen de los datos observados y la solución es única y estable (en el sentido probabilístico), lo que permite una segmentación del riesgo única y estable. El modelo trabaja con una gran variedad de distribuciones estadísticas truncadas. La solución obtenida solo se da en el caso unidimensional.
- III. Según el modelo desarrollado decidimos o no conceder un préstamo en función de la calificación q_i^* . Si estamos ante un prestatario cuyas características indican que es un buen cliente, q_i^* tenderá a cero y viceversa, por lo tanto podemos definir la probabilidad de impago como: $PD=q_i^*$. Conclusión que nos parece puede ser aplicada como modelo para las exigencias del Nuevo Acuerdo de Basilea en exposiciones al detalle, ya que los valores de riesgo de impago se desglosan en intervalos o

Conclusiones

segmentos, como exige el Nuevo Acuerdo y hay que advertir que cuanto mayor sea el número de intervalos, es decir, cuanto mayor sea la precisión en el cálculo del ratio de impago es mayor.

IV. También concluimos, que podemos obtener otra de las variables o conceptos descritos por Basilea II como una componente de riesgo a calcular para cada segmento obtenido y esta es la pérdida dada de impago cuya forma general ya indicada en el capítulo sería: $LGD = (\overline{C}_i / C_i) \times 100$

V. Por último, y teniendo en cuenta el supuesto de que se puede aplicar la obtención de la pérdida esperada para exposiciones al detalle, obtendremos una estimación de esta componente de riesgo a la espera de que el Comité determine una forma de obtención de EL, para pequeños préstamos.

Bibliografía

[1] : Altman, Edward I., Giancarlo Marco y Franco Varetto. Corporate Distress Diagnosis: Comparisons Using Linear Discriminant Analysis and Neural Networks (The Italian Experience). Journal of Banking and Finance 18 , pp.505-29. (1994).

Bibliografia

[2] : Bagnoli, M. Bergstrom, T., Log-concave probability and its applications. University of Michigan. (1989).

[3] : Benaïm, M. Fort, J.C., y Pagès G., Convergence of the one-dimensional Kohonen Algorithm. Adv. Appl. Prob. 30, 850--869. (1998).

[4] : Bouton C., Pagès G., Self-organization and convergence of the one-dimensional Kohonen Algorithm with non uniformly distributed stimuli. Stochastic Process and Apl. 47, 80-103. (1993).

[5] : Cottrell, M. y Fort, J. C. Étude d'un algorithme d'auto-organization. Ann. Inst. H. Poincaré B. 23, 1--20. (1987).

[6] : Duflo, M., Random Iterative Models", Applications of Mathematics: Stochastic Modelling and Applied Probability, 34, Springer--Verlag. (1979).

[7] : Flanagan A., Self-organizing neural networks Thèse 1306, Ecole Polytechnique de Laussane. (1994).

[8] : Fort, J.C., y Pagès G. On the a.s. convergence of the Kohonen Algorithm with a general neighborhood function. Ann. Appl. Prob., 5, 1177--1216. (1995).

[9] : Fort, J.C., y Pagès G. Convergence on stochastics algorithms: from the Kushner & Clark theorem to the Lyapunov functional method Adv. Appl. Prob., 28, 1072--1094. (1996).

Bibliografía

[10] : Galindo J., Tamayo P., Credit Risk Assessment using Statistical and Machine Learning Methods as an Ingredient for Financial Intermediaries Risk Modeling. Departamento de economía de la Universidad de Harvard y Thinking Machines Corp. (1997).

[11] : Guillén M. y Arís M., Count Data Models for a credit scoring system. Universidad de Barcelona. (1992).

[12] : Jost, A. Data Mining, Credit Risk Modeling: Design and Application. Elizabeth Mays Editor. (1998).

[13] : Kushner H.J., Clark D.S., Stochastic Approximation Methods for Constrained and Unconstrained System. Springer-Verlag. (1978).

[14] : Krzanowski, W.J., Principles of Multivariate Analysis. Clarendon Press, Oxford. (1998).

[15] : Kohonen T., Analysis of a simple self-organizing process, Biol. Cybernet. 44, 135--140. (1982).

[16] : Kohonen T., Self-Organization and Associative Memory. 3rd. Springer-Verlag. (1989).

[17] : M. Benaïm, J-C. Fort, G. Pagès. Convergence of the one-dimensional Kohonen Algorithm, Adv. Appl. Prob. 30, 850-869. (1998).

[18] : Mester, L.J., What's the point of Credit Scoring?, Business Review, Septiembre/Octubre 1997, Federal Reserve Bank of Philadelphia. (1997).

Bibliografía

[19] : Malhotra, D.K., Rashmi Malhotra, y Robert W. McLeod. Artificial Neural Systems in Commercial Lending The Bankers Magazine, pp.40-44. (noviembre/ diciembre 1994).

[20] : Roszbach, K., Bank lending policy, credit scoring and the survival of loans, Escuela de economía de Estocolmo, (1998).

[21] : Shin,K., Shin,T.,Han,I. Intelligent corporate credit rating system using bankruptcy probability matrix. Proceedings of the IV International Conference on Artificial Intelligence and Emerging Technologies in Accounting, Finance and tax, Huelva. (1998).

[22] : Yuying An, M., Log-concave probability distributios: Theory and statistical testing. University of Duke. (1995).

[23] : López González, E., Flórez López, R. El análisis de solvencia empresarial utilizando redes neuronales autoasociativas Proceedings of the IV International Meeting on Advances in Computational Management, Reus (1999).

[24] : Serrano Cinca, C., Martín del Brio, B. Predicción de la quiebra bancaria mediante el empleo de redes neuronales artificiales Revista española de Financiación y Contabilidad Vol.22, nº 74, pp.153-176.(1993).

Capítulo V

APLICACIÓN PRÁCTICA DE CREDIT SCORING MIXTO

5.1 Introducción

Para poder realizar una aplicación del modelo propuesto necesitamos datos históricos de clientes, y el punto de mira para conseguirlos estuvo en las entidades bancarias, pero el resultado de nuestro esfuerzo fue nulo puesto que se acogen al secreto de datos para no ceder información sobre sus clientes. Ante tal negativa, y el empeño por nuestra parte de mostrar una evidencia empírica en este trabajo, recurrimos a los datos que encontramos en el CD del libro *Credit Scoring and its Applications* [2] que nos permite, que en este capítulo podamos aplicar el modelo teórico propuesto y analizar los resultados que de él se desprendan. Para ello ha sido necesario emplear los programas Excel, que ha sido utilizado en el proceso de manipulación de datos, con SPSS se han podido realizar los cálculos del análisis discriminante, y Matlab nos ha permitido aplicar

Capítulo 5. Aplicación práctica de Credit Scoring Mixto

el Algoritmo de Kohonen haciendo uso de la toolbox Neural Networks en el apartado One-dimensional self-organizing map.

Siguiendo los preceptos de Basilea II la entidad bancaria deberá dividir sus exposiciones de préstamo por productos crediticios. Suponemos que los datos que utilizamos están referidos a exposiciones al detalle, esto es, préstamos de bajo valor, que pueden ser créditos al consumo, tarjetas de crédito, etc., así como, préstamos de bajo riesgo a empresas.

Una vez seleccionado el producto crediticio tomamos los datos históricos que tenemos de los clientes a los que les fue concedido un préstamo y seguimos los siguientes pasos:

- En primer lugar, con la aplicación del análisis discriminante obtendremos aquellas características de los clientes que explican la solvencia o insolvencia de los mismos al afrontar el pago del crédito, esto es, determinamos las variables explicativas X_j con $j=1, \dots, d$ del modelo. Al mismo tiempo obtendremos el valor de los coeficientes estandarizados de las funciones canónicas de Fisher a_{1j}^* que nos permiten valorar la relevancia de una característica en el pago del crédito.
- A continuación definimos la variable U como la calificación obtenida por el cliente con la siguiente expresión

$$U=U(X_1, X_2, \dots, X_d) = (1 / \sum_{j=1}^d |a_{1j}^*|) \sum_{j=1}^d a_{1j}^* (I_{(0, \infty)}(a_{1j}^* - X_j))$$

donde

$$I_A(x) = \begin{cases} 1 & \text{si } x \in A \\ 0 & \text{si } x \notin A \end{cases}$$

- Después se aplica el Algoritmo de Kohonen a las calificaciones finales U obteniendo una segmentación de las mismas, requisito del Nuevo Acuerdo de Basilea para las exposiciones al detalle.
- Finalmente veremos como obtenemos los valores de la Probabilidad de Impago (PD), de la Pérdida dada de Impago (LGD), y de la Pérdida Esperada (EL) para cada uno de los segmentos de riesgo de las exposiciones, cumpliendo así, de una forma más completa con los requerimientos de Basilea II.

5.2 Análisis de los resultados empíricos

En el capítulo anterior, al presentar el modelo teórico, planteamos dos opciones para resolver el análisis discriminante ó bien utilizar la expresión:

$$Y_k = a_{k1} X_1 + a_{k2} X_2 + \dots + a_{kp} X_p \quad k=1,2$$

ó bien

$$Z_1 = a_{11}^* X_1 + a_{12}^* X_2 + \dots + a_{1p}^* X_p$$

$$Z_2 = a_{21}^* X_1 + a_{22}^* X_2 + \dots + a_{2p}^* X_p$$

Capítulo 5. Aplicación práctica de Credit Scoring Mixto

Se ha decidido trabajar con la segunda opción y las razones son las siguientes:

1. Los datos de la variable dependiente Y_k se limitan a los valores 1 cuando se ha producido impago y 0 cuando el préstamo ha sido amortizado, también en el modelo teórico planteamos como variable dependiente un ratio de tiempo de pago del crédito que nos permitiría tener un mejor ajuste de un modelo de regresión, no es lo mismo un impago al final del préstamo que al principio sobre todo desde el punto de vista de la pérdida ocasionada a la entidad. Sea como fuere, se planteaba el problema de que al calcular $a_{k1}, a_{k2}, \dots, a_{kp}$ obtuviéramos valores de $Y_k \in [0,1]$, y este se soluciona con la utilización de las funciones canónicas discriminantes de Fisher Z_k .

2. Además, en nuestro caso, solo utilizaremos una de las funciones del análisis discriminante que será la que explica los préstamos pagados, suponiendo que esta es

$$Z_1 = a_{11}^* X_1 + a_{12}^* X_2 + \dots + a_{1p}^* X_p$$

los términos $a_{11}^*, a_{12}^*, \dots, a_{1p}^*$ nos indican la significatividad de las variables para explicar la solvencia de un cliente ante un préstamo, independientemente del valor de Z_1 .

Hay que recordar que el análisis discriminante nos permite obtener una clasificación previa de los clientes que han pagado su préstamo y los que no lo han hecho, por lo que podemos establecer una comparación entre

Análisis de los resultados empíricos

los resultados obtenidos con la aplicación del modelo que presentamos y lo conseguido solo con el análisis discriminante, en definitiva, podemos comprobar si existe una mejora y por lo tanto, una aportación significativa.

Disponemos en principio de doce variables para explicar el pago o impago de los créditos concedidos por una entidad bancaria a 897 clientes (n=897) de los que conocemos que 632 préstamos han sido amortizados, mientras que los 265 restantes han resultado insolventes. Para trabajar más cómodamente, asociaremos a cada variable con una abreviatura que utilizaremos a partir de ahora para referirnos a ellas, así tendremos:

Tabla 1: Abreviatura de las variables utilizadas originalmente

Variable	Abreviatura
Año de nacimiento	Año
Nº de niños	Niñ.
Personas dependientes	Pdep.
Tenencia de teléfono	Telf.
Renta del cónyuge	Rtacy.
Renta del cliente	Rtacl.
Valor del hogar	Vhg.
Hipoteca pendiente	HPpen.
Gastos en hipoteca o alquiler	Gtohaq
Gastos en préstamos	Gtopr.
Gastos en compras a plazos	Gtocpz
Gastos en tarjetas de crédito	Gtocr.

Fte: Elaboración propia

Capítulo 5. Aplicación práctica de Credit Scoring Mixto

Estos datos los sometemos al análisis discriminante en el programa SPSS y obtenemos los siguientes resultados:

Tabla 2: Anovas Univariantes

Variable	F	Sig=p
Año	5.077	0.024
Niñ.	0.578	0.447
Pdep.	0.613	0.434
Telf.	0.828	0.363
Rtacy.	1.327	0.250
Rtacl.	8.169	0.004
Vhg.	0.022	0.881
HPpen.	0.404	0.525
Gtohaq	0.931	0.335
Gtopr	0.054	0.816
Gtocpz	3.955	0.047
Gtotcr.	1.327	0.250

Fte: Elaboración propia

Análisis de los resultados empíricos

Tabla 3: Valores de los estadísticos de significación de las variables sometidas a estudio.

Variable	Lambda de Wilks	F	Sig=p
Año	0.987	16.101	0.000
Niñ.	0.998	1.875	0.171
Pdep.	1	0.224	0.636
Telf.	0.998	2.291	0.130
Rtacy.	0.996	4.606	0.032
Rtacl.	0.966	43.614	0.000
Vhg	0.998	2.571	0.109
HPpen.	0.998	2.18	0.140
Gtohaq	0.995	5.633	0.018
Gtopr	0.999	1.278	0.258
Gtocpz	0.995	6.623	0.010
Gtotcr	0.996	4.606	0.032

Contraste Glogal	0.977	Significatividad en el modelo $\chi^2=30.795$	0.02
------------------	-------	---	------

Fte: Elaboración propia

Antes de sacar conclusiones de los resultados obtenidos recordamos las definiciones de los estadísticos que hemos utilizado:

Lambda de Wilks: Contraste de significación multivariado, algunas veces denominado el estadístico de U. Lambda toma valores entre 0 y 1. Los valores próximos a 0 indican que las medias de los grupos son diferentes,

Capítulo 5. Aplicación práctica de Credit Scoring Mixto

mientras que los valores próximos a 1 indican que las medias de los grupos no son diferentes, que sea igual a 1 indica que todas las medias son iguales.

F: Cociente entre dos medias cuadráticas. El estadístico F solo puede tomar valores positivos o cero. Así, cuando la hipótesis nula es cierta, esperamos que F tome valores próximos a uno. A medida que las medias están más separadas entre sí, el valor F se hace mayor. Los valores grandes de F constituyen una prueba fehaciente en contra de la hipótesis nula, inclinándonos a pensar que la hipótesis correcta es la alternativa.

sig=p: El p-valor o nivel de significación empírico del contraste es el dato obtenido a partir del valor del estadístico de contraste, en las observaciones que corresponden a la realización de la muestra de tamaño n extraída de una población, y que nos informa sobre cuál sería el nivel de significación más pequeño que nos hubiera permitido rechazar la hipótesis nula. Se rechazará la hipótesis nula si el p-valor es menor o igual al nivel de significación adoptado por el experimentador. De manera habitual, se considera significativo un valor menor que 0.05.

Observando los estadísticos la primera conclusión que obtenemos por los valores de la Lambda de Wilks es que en nuestros datos ($n=897$), no hay mucha diferencia en las características de los clientes que han pagado de los que no han pagado, lo cual supone una dificultad mayor. Con la ayuda de los otros dos estadísticos tomamos como características significativas aquellas con valor de $\text{Sig}=p<0.05$ que además coinciden con un mayor valor del estadístico F, y dentro de un margen con un menor valor de la Lambda de Wilks.

Para ratificarnos en la toma de decisión de las variables significativas sometemos a las mismas al análisis denominado Anova de un factor que genera un análisis de varianza de un factor para una variable cuantitativa respecto a una única variable de factor (la variable independiente). El

Análisis de los resultados empíricos

análisis de varianza se utiliza para contrastar la hipótesis de que varias medias son iguales. Los resultados que obtenemos quedan reflejadas en la Tabla 2.

La hipótesis nula que se plantea al aplicar Anova es que las medias son iguales, el programa SPSS responde con un estadístico F relativamente grande y con p pequeño para rechazar la hipótesis nula. Fijándonos entonces en la tabla 2 solo podríamos trabajar con tres variables, Año, Rtacl, y Gtocpz. Estos resultados nos vienen a confirmar que los datos disponibles no son buenos, que hay mucha igualdad en los valores de las variables que caracterizan a los clientes solventes de los que no lo son, que dificultan la aplicación del análisis discriminante, o bien, que los datos no se ajustan a un modelo lineal. Sin embargo, la no bondad de estos datos nos puede ayudar a ratificar la justificación de aplicar un segundo método que nos permita una mejor clasificación de los clientes.

Trabajar con solo tres variables reduciría mucho la posibilidad de separación de clientes, por lo tanto a pesar de conocer que el resto de variables no aporta demasiado a la solución del problema, y aún sabiendo que no son significativas vamos a añadir las tres variables siguientes en significación discriminante y que en principio la Lambda de Wilks nos daba como aceptables¹³.

Trabajaremos definitivamente con las siguientes variables significativas obtenidas en la tabla 2 cuya matriz de correlación es:

¹³ La hipótesis nula que se propone es que la Lambda de Wilks es igual a 1

Análisis de los resultados empíricos

Tabla 4: Matriz de covarianzas de las variables significativas

	Rtacl.	Año	Rtacy.	Gtotcr.	Gtopz	Gtohaq
Rtacl.	1	-0.085	-0.002	-0.002	0.142	0.383
Año	-0.085	1	-0.118	-0.118	-0.014	0.046
Rtacy.	-0.002	-0.118	1	1	0.021	0.131
Gtotcr	-0.002	-0.118	1	1	0.021	0.131
Gtopz	0.142	-0.014	0.021	0.021	1	0.017
Gtohaq	0.383	0.046	0.131	0.131	0.017	1

Fte: Elaboración propia

Si nos fijamos en los resultados obtenidos lo primero que llama nuestra atención es la similitud en el valor de los estadísticos de las variables Renta del cónyuge y Gastos en tarjetas de crédito, comprobando además que existe una correlación 1 entre ellas, esto significa que ambas tienen la misma capacidad explicativa en el pago o impago de los préstamos y por lo tanto descartamos una de ellas, y hemos optado por no utilizar la variable Gastos en tarjetas de crédito, y nos quedamos definitivamente con cinco variables para que nos expliquen los pagos o impagos de estos créditos. Hay que hacer notar también que las correlaciones entre las demás variables no son muy altas lo que nos da idea de la fuerte componente ortogonal de las características o variables explicativas.

El siguiente paso consistirá en determinar cuál es el grado de significatividad de estas cinco variables, esto es, cuál es el grado de incidencia de cada una de ellas en la explicación de la variable dependiente. Si nos fijamos en las tablas anteriores y en los valores de los estadísticos, serán más significativas aquellas variables que tengan una Lambda de Wilks menor, un estadístico F mayor y un estadístico p menor. Pero con estos datos solo obtendríamos una clasificación ordinal

Análisis de los resultados empíricos

y lo que buscamos es una valoración cardinal de su faceta explicativa, para ello contamos con los valores de tres coeficientes diferentes:

1. Coeficientes estandarizados de las funciones canónicas discriminantes
2. Coeficientes de las funciones canónicas discriminantes
3. Coeficientes de las funciones lineales de Fisher.

Para ello elegimos los valores de los coeficientes enumerados en primer lugar como ya queda explicado en la Sección (5.1).

Al aplicar sobre los datos de nuestra muestra el programa SPSS, además de los datos estudiados anteriormente, también obtenemos el valor de los coeficientes estandarizados de las funciones canónicas discriminantes.

Así, los valores, ya ordenados de mayor a menor, que se obtienen son los siguientes:

Tabla 5: Valoración del grado de incidencia de las variables en el pago de los préstamos

Variable	a_{1j}^*	Valor
Rtacl.	a_{16}^*	0.83
Año	a_{11}^*	0.58
Rtacy	a_{15}^*	0.33
Gtospz.	$a_{1,11}^*$	0.182
Gtohaq	a_{19}^*	-0.116

Fte: Elaboración propia

Análisis de los resultados empíricos

Nos llama la atención que la variable Gastos en compras a plazos tenga una incidencia positiva en el pago de los préstamos, ya que cabría esperar que cuanto mayor fuera el endeudamiento en el que incurre el cliente, menor fuera su capacidad de respuesta al pago del préstamo.

Como no tenemos ninguna información adicional en la que sustentar una explicación lo tomamos tal y como nos viene dado, aunque nuestra interpretación ante los resultados sería que esta variable indica la respuesta efectiva de los clientes ante el pago a plazo.

Además de toda esta información, el análisis discriminante también aporta una clasificación previa de los clientes que recogemos en la siguiente tabla, siguiendo la proporción original:

Tabla 6: Pronóstico de préstamos pagados e impagados utilizando Análisis Discriminante con proporciones iniciales del 70.5% y 29.5% respectivamente.

	Pronosticados buenos	Pronosticados malos	Total
Porcentaje buenos	97%	3%	100%
Porcentaje malos	88.9%	11.1%	100%

Fte: Elaboración propia

Estos resultados explican que del total de la muestra están clasificados correctamente el 74,4% de los casos.

Si la probabilidad a priori de impago es del 50% obtendríamos la siguiente clasificación.

Análisis de los resultados empíricos

Tabla 7: Pronóstico de préstamos pagados e impagados utilizando Análisis Discriminante

	Pronosticados buenos	Pronosticados malos	Total
Porcentaje buenos	41.5%	58.5%	100%
Porcentaje malos	26%	74%	100%

Fte: Elaboración propia

En este caso del total de la muestra quedarían clasificados correctamente solo el 48.8% de los casos, aunque hay que advertir que el porcentaje de acierto en los clientes insolventes es mucho mayor ya que aumenta del 11.1% al 74%

Con los datos iniciales se testaron varios modelos estadísticos como logit y probit y el que mejores resultados aportaba fue el análisis discriminante¹⁴.

Aún así, vemos que con nuestra muestra es fácil detectar los clientes solventes, pero no así, los insolventes y esto se debe, como ya hemos indicado anteriormente, a la similitud de las características entre buenos y malos clientes.

Nuestro objetivo radica en intentar conseguir mejorar estos resultados con la aplicación del modelo que proponemos.

Con toda la información que tenemos razonamos de la siguiente manera para calificar a los clientes:

¹⁴ Si el lector desea aplicar este modelo en otra muestra, se recomienda testar los datos con diferentes métodos estadísticos y elegir el que aporte mejores resultados

Capítulo 5. Aplicación práctica de Credit Scoring Mixto

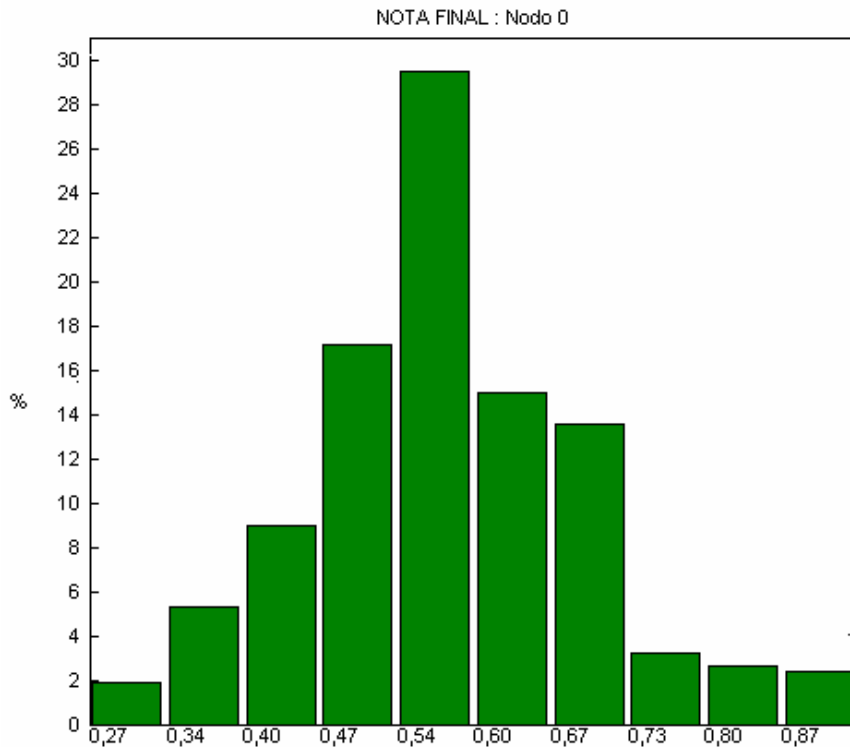
1. Cuanto mayor sea el valor de a_{1j}^* de una característica, más relevante será para el pago del préstamo y su efecto sobre el mismo estará relacionado directamente con el signo que le acompañe.
2. Cuanto más posea el cliente de aquella característica que afecta positivamente al pago, mejor puntuación debe obtener el cliente y viceversa.
3. Queremos calificar al cliente en un intervalo $[0,1]$ donde los mejores clientes tenderán a una nota de 0 y los no pagadores se acercarán al 1.

Con esto llegamos a la conclusión de que la calificación de cada cliente U vendrá dada por la expresión definida en la sección (5.1).

A continuación tomamos la sucesión de datos U que hemos obtenido como calificaciones de los clientes y le aplicamos el Algoritmo de Kohonen para obtener los segmentos de calificación que nos indiquen un mismo valor de riesgo, y con ello, calcular los componentes de riesgo definidos por Basilea II. Para poder hacer esto deben seguirse los siguientes pasos:

Análisis de los resultados empíricos

Gráfica de la distribución de las calificaciones de los clientes U



- 1) La función de distribución de U debe cumplir la propiedad de log-concavidad, esto es, si calculamos el logaritmo de la función de distribución esta debe ser cóncava. Como no conocemos la función de distribución de esta variable hacemos una exploración por SPSS y como se ve en la siguiente gráfica podríamos asegurar que la sucesión de calificaciones se distribuye con una función de distribución que cumple la log-concavidad.
- 2) Tenemos que decidir el número de segmentos o clases, que denominaremos m, en los que van a quedar divididas las notas U obtenidas por los clientes, y en nuestro caso decidimos, mirando la distribución, probar con $m=4$. La razón de esta elección y no

Capítulo 5. Aplicación práctica de Credit Scoring Mixto

otra es, que como los segmentos nos van a clasificar a los clientes por riesgo de impago, el orden de las clases nos indicarían clientes de riesgo nulo, clientes de bajo riesgo, clientes de riesgo medio y clientes de alto riesgo.

- 3) El siguiente paso entonces, es aplicar el Algoritmo de Kohonen, utilizando Matlab, a la distribución de calificaciones $U \in [0,1]$, indicándole que $m=4$, y el programa asigna de forma aleatoria unos valores iniciales de manera que el segmento $[0,1]$ queda dividido en cuatro clases. Así, el programa comienza a ejecutar el Algoritmo tal y como se explica en el capítulo 4 de este trabajo y obtenemos los siguientes resultados $\mathbf{q}^* = [0.4772, 0.5368, 0.6253, 0.7006]$. Por lo tanto los segmentos de agrupación de calificaciones serían: $[0, 0.507]$ $]0.507, 0.58105]$ $]0.58105, 0.66295]$ y $]0.66295, 1]$
- 4) Una vez obtenida la solución del problema, tenemos que comprobar si esta es correcta o no lo es. La solución no sería correcta si las clases o segmentos obtenidos no fueran representativos de los datos que contienen, es decir, podríamos encontrarnos con dos casos en los que la solución no fuera acertada.
 - a) Que el número de clases elegido sea menor que el que se debería haber cogido, y esto implicaría que habría datos que podrían dar más información, y en nuestro caso, representarían por ellos mismos otro segmento de riesgo y los estamos obligando a pertenecer a un nivel equivocado, y
 - b) Que el número de segmentos elegido es mayor que el que se debería haber tomado, y por lo tanto, estamos segmentando el riesgo en más niveles de los que verdaderamente se están dando en la entidad bancaria.

Análisis de los resultados empíricos

La solución a este problema lo encontramos en el mismo programa Matlab mediante la aplicación de la Función de Aprendizaje de pesos¹⁵.

Si los pesos han aprendido bien, es decir, si la solución es correcta, la Función de Aprendizaje nos proporciona una matriz ¹⁶ 4×4 en cuya diagonal principal tendremos unos y en el resto ceros, y ya adelantamos que este es el resultado obtenido, por lo tanto, queda certificada la bondad de la solución conseguida. Si el número de clases no hubiese sido el correcto aparecería un cero en la diagonal principal, y el uno que correspondería en ese lugar se trasladaría a aquella fila donde hubiera sido más significativo agrupar esa clase, esto nos estaría indicando que m debe ser menor que cuatro.

Por otra parte, repetimos el ejercicio suponiendo m=5 y los resultados obtenidos fueron:

$$\mathbf{q}^* = [0.3117 \ 0.482, \ 0.5418, \ 0.628, \ 0.7093]$$

pero al aplicar la Función de Aprendizaje la matriz 5×5 obtenida fue:

¹⁵ Los pesos son los valores obtenidos como solución

¹⁶ Hemos decidido que m=4

Capítulo 5. Aplicación práctica de Credit Scoring Mixto

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

lo que nos indicaba que los datos del primer segmento no eran significativos y que su tendencia era la misma que la del segundo segmento como así queda comprobado en la solución del problema.

A continuación presentamos una tabla donde representamos los pagos e impagos en cada uno de los segmentos

Tabla 8: Frecuencia absoluta de pagos e impagos en cada intervalo de calificación de los clientes

Segmento	Pagos	Impagos
[0, 0.507]	219	2
]0.507, 0.58105]	220	50
]0.58105, 0.66295]	127	78
]0.66295, 1]	66	135

Fte: Elaboración propia

Hemos representado las calificaciones obtenidas por los clientes en un gráfico utilizando el programa Excel que aparece en el anexo.

Análisis de los resultados empíricos

En este gráfico encontramos en el eje Definición solo dos valores, 0 y 1. El 0 representa aquellos clientes que han pagado su préstamo y el 1 los clientes que no pagaron su préstamo.

En el eje de Calificación representamos la puntuación obtenida por los clientes según el modelo desarrollado, que como hemos indicado está definida en el intervalo $[0,1]$. Si nos detenemos en el gráfico vemos que hay algunos valores relevantes en la calificación que vamos a ir detallando paulatinamente.

En el segmento $] 0.66295, 1]$ hay un valor importante que es 0.71443 puesto que no se produce ningún pago en clientes con $U > 0.71443$ que nos permite establecer una clasificación de los clientes que presentamos en la siguiente tabla.

Tabla 9: Pronóstico de préstamos pagados e impagados utilizando el Algoritmo de Kohonen tomando como referencia el último segmento

	Pronosticados buenos		Pronosticados malos		Total
Porcentaje buenos	100%	si	0%	si	100%
	$U < 0.71443$		$U < 0.71443$		
Porcentaje malos	64%	si	36%	si	100%
	$U < 0.7443$		$U > 0.71443$		

Fte: Elaboración propia

Si estudiamos los resultados obtenidos en el otro segmento extremo $[0, 0.507]$ vemos que todos los clientes que han obtenido su calificación en este segmento han resultado solventes con excepción de dos cuyas calificaciones son mayores que 0.4916, esto es, muy cercanas al valor de cierre del segmento 0.507.

Capítulo 5. Aplicación práctica de Credit Scoring Mixto

El valor aportado por el algoritmo a este segmento es 0.4772, calificación que se puede interpretar como valor indicativo a partir del cual los clientes pueden comenzar a ser insolventes y al mismo tiempo la entidad financiera siempre podrá esperar de todos aquellos clientes que obtengan una calificación inferior que sean solventes. Si tomamos como valor de referencia el del cierre del segmento 0.507, podemos elaborar otra tabla resumen de los porcentajes pronosticados que serían los siguientes:

Tabla 10: Pronóstico de préstamos pagados e impagados utilizando el Algoritmo de Kohonen tomando como referencia el primer segmento

	Pronosticados buenos		Pronosticados malos		Total
Porcentaje buenos	34.7% U<0.507	si	65.3% U>0.507	si	100%
Porcentaje malos	0.7% si U<0.507		99.3% U>0.507	si	100%

Fte: Elaboración propia

Si recordamos los datos obtenidos solo con la aplicación del análisis discriminante podemos establecer una comparación de los resultados de la Tabla 6 con los obtenidos en la Tabla 9 ya que las probabilidades a priori (70%) son muy cercanas a la nota de corte $U=0.71443$. Así, vemos que se ha obtenido una mejora en la predicción de malos clientes frente a la obtenida por el análisis discriminante que nos pronosticaba el 11,1%.

Análisis de los resultados empíricos

Además en este mismo tramo del gráfico, no encontramos ningún cliente que haya pagado su préstamo y que sea calificado como mal cliente, es decir nuestro error es del 0% mientras con el análisis discriminante se cometía un error del 3%. Lo que implica además un porcentaje de acierto global del 81% frente al 74,4% que obteníamos solo con el análisis discriminante.

Si seguimos analizando resultados en este último segmento apreciamos que este valor de corte entre pagos e impagos (0.71443) es muy próximo al valor aportado por el algoritmo de Kohonen 0.7006, que se puede considerar como un valor de alarma para la entidad financiera a la hora de conceder préstamos a clientes cuya calificación oscile entorno a ese valor.

De la misma manera establecemos una comparación de los datos obtenidos en la Tabla 7 con los de la Tabla 10 ya que la nota de corte $U=0.507$ es coincidente con la probabilidad a priori (50%) y vemos que se ha producido una gran mejora en el pronóstico de clientes insolventes que pasa de un error del 26% al 0.7%.

Podemos entonces resumir la información que se puede obtener del gráfico en la tabla siguiente:

Tabla 11: Frecuencias relativas de pagos e impagos por cada intervalo de calificación del cliente

Segmento	Pagos	Impagos
[0, 0.507]	34.7%	0.7%
]0.507, 0.58105]	34.8%	18.9%
]0.58105, 0.66295]	20%	29.4%
]0.66295, 1]	10.5%	51%

Fte: Elaboración propia

Capítulo 5. Aplicación práctica de Credit Scoring Mixto

Estos resultados nos sugieren que la población de esta entidad bancaria es bastante homogénea ya que el 53% está valorada en los intervalos intermedios,]0.507, 0.58105] y]0.58105, 0.66295] y tomando valores globales de los clientes situados en ellos, el 73% ha resultado solvente y el 27% incurre en impago. Por lo tanto sería en este tramo de calificación donde el banco tendría que afinar su toma de decisiones ya que hay que resaltar que las calificaciones intermedias de estos segmentos aportadas por el Algoritmo de Kohonen son 0.5368 y 0.6253, es decir, calificaciones cercanas a la media pero con tendencia a la insolvencia puesto que la superan.

Nuestro último objetivo es este trabajo es conseguir una relación entre las calificaciones obtenidas por los clientes y los componentes de valoración del riesgo que son la Probabilidad de impago, la Pérdida dada de impago y la Pérdida Esperada.

Para ello tomamos los valores obtenidos por el Algoritmo de Kohonen y que configuran el vector \mathbf{q}^*

$$\mathbf{q}^* = [0.4772, 0.5368, 0.6253, 0.7006]$$

Cada uno de estos valores reflejan las concentraciones de valor de las calificaciones en cada uno de los segmentos y es por esto que en nuestro modelo vamos a asociar la PD (probabilidad de impago), que se define como la probabilidad de que el cliente entre en mora en un determinado periodo de tiempo, con los valores q_i^* . (El razonamiento de esta asociación se explica en el capítulo anterior).

Por lo tanto

Análisis de los resultados empíricos

$$PD=[0.4772, 0.5368, 0.6253, 0.7006]$$

En cuanto al cálculo de la pérdida dada de impago que representa el ratio entre las pérdidas efectivas incurridas, como consecuencia de la entrada en mora, y la Exposición, tenemos que recordar que consideramos los préstamos concedidos en un mismo producto crediticio, y que estos son homogéneos en cantidad, esto es, a todos los clientes se les ha concedido un préstamo de la misma cantidad que suponemos igual a 1, por lo tanto $C = \sum_{l=1}^m C_l = 897$ es el capital total expuesto, por segmentos la cantidad prestada ha sido la siguiente:

Tabla 12: Capital concedido en préstamos en cada segmento

Segmento	C_l
[0, 0.507]	221
]0.507, 0.58105]	270
]0.58105, 0.66295]	205
]0.66295, 1]	201

Fte: Elaboración propia

Otro de los componentes de riesgo es la Pérdida dada de impago (LGD) también denominada Severidad que representa el ratio entre las pérdidas efectivas incurridas, como consecuencia de la entrada en mora, y la Exposición. Sabiendo esto, a partir de los datos de la tabla 9, podemos decir que la Pérdida dada de Impago (LGD) en cada uno de los segmentos de riesgo será la siguiente:

Capítulo 5. Aplicación práctica de Credit Scoring Mixto

Tabla 13: Valor de la Severidad en cada intervalo de riesgo

Segmento	LGD
[0, 0.507]	0.9%
]0.507, 0.58105]	18.5%
]0.58105, 0.66295]	38%
]0.66295, 1]	67%

Fte: Elaboración propia

En cuanto al último componente de riesgo, el cálculo de la Pérdida Esperada (EL) cuyo concepto coincide con la media matemática de la pérdida crediticia, viene determinada por Basilea II¹⁷, y se define como:

$$EL=PD \times EAD \times LGD$$

En nuestro caso el resultado que obtenemos es que por cada 100 unidades de préstamos concedidos en cada segmento la Pérdida Esperada es:

¹⁷ Basilea II propone esta fórmula de la Pérdida Esperada para exposiciones empresariales. Al no encontrar ninguna propuesta alternativa para exposiciones al detalle, proponemos la posibilidad de utilizar esta fórmula para el cálculo de la Pérdida Esperada en cada segmento de riesgo.

Tabla 14: Valor de la Pérdida Esperada por cada 100 unidades monetarias prestadas

Segmento	EL
[0, 0.507]	0.95 u.m.
]0.507, 0.58105]	26 u.m.
]0.58105, 0.66295]	48 u.m.
]0.66295, 1]	94 u.m.

Fte: Elaboración propia

5.3 Conclusiones

A pesar de no contar con los datos más idóneos para la resolución del problema planteado se ha podido llegar a las siguientes conclusiones:

1. Como apuntábamos en el apartado (4.1) uno de nuestros objetivos era comprobar si los resultados obtenidos en un Credit Scoring con la conjunción de modelos estadísticos tradicionales y redes neuronales [1] eran mejores que los obtenidos con la aplicación de estos modelos por separado y hemos visto en el desarrollo del ejercicio que así es. El análisis discriminante permite seleccionar las variables relevantes, propiedad que no poseen las redes neuronales, y la incorporación del Algoritmo de Kohonen permite conseguir una mejor diferenciación de los clientes solventes e insolventes que la obtenida solo con la aplicación del análisis discriminante.
2. Los resultados obtenidos a partir del Algoritmo, \mathbf{q}^* , son los valores de las calificaciones de los clientes que identifican cada segmento de riesgo, y estos valores, dependen al mismo tiempo, de los coeficientes estandarizados obtenidos en el Análisis Discriminante, por lo tanto los resultados que obtenemos se

Capítulo 5. Aplicación práctica de Credit Scoring Mixto

derivan de un modelo mixto compuesto por redes neuronales y análisis discriminante.

3. La solución \mathbf{q}^* obtenida es única y estable. La unicidad y estabilidad vienen confirmadas por la convergencia del Algoritmo de Kohonen en el caso unidimensional [3], que es el que hemos aplicado en este trabajo.
4. Según el modelo desarrollado decidimos o no conceder un préstamo en función de la calificación q_i^* . Si estamos ante un prestatario cuyas características indican que es un buen cliente, q_i^* tenderá a cero y viceversa, por lo tanto podemos definir la probabilidad de impago como: $PD=q_i^*$. En el ejemplo utilizado, hemos obtenido que los valores inferior y superior del vector \mathbf{q}^* nos permiten descartar con una alta fiabilidad los clientes solventes e insolventes respectivamente. Como podía esperarse, son los valores intermedios del vector \mathbf{q}^* los que más incertidumbre generan a la entidad bancaria a la hora de decidirse sobre la concesión o negación del préstamo, ya que estamos hablando de una probabilidad de impago entre el 53% y el 62%. El hecho de poder segmentar el riesgo de concesión de créditos es una conclusión que nos parece puede ser aplicada como modelo para las exigencias del Nuevo Acuerdo de Basilea en exposiciones al detalle, ya que los valores de riesgo de impago se desglosan en intervalos o segmentos, como exige el Nuevo Acuerdo y hay que advertir que cuanto mayor sea el número de intervalos, es decir, cuanto mayor sea m la precisión en el cálculo del ratio de impago es mayor, siempre que la función de aprendizaje del algoritmo nos certifique que las m clases son significativas.

Capítulo 5. Aplicación práctica de Credit Scoring Mixto

5. También concluimos, que podemos obtener otra de las variables o conceptos descritos por Basilea II como una componente de riesgo a calcular para cada segmento obtenido y esta es la pérdida dada de impago cuya forma general ya indicada en el capítulo sería: $LGD = (\overline{C_i}) / (C_i) \times 100$
6. Y por último, y teniendo en cuenta el supuesto de que podemos aplicar la obtención de la pérdida esperada, definida por Basilea II para exposiciones empresariales, a exposiciones al detalle en nuestro trabajo práctico los resultados han sido:

Segmento	EL
[0, 0.507]	0.95 u.m.
]0.507, 0.58105]	26 u.m.
]0.58105, 0.66295]	48 u.m.
]0.66295, 1]	94 u.m.

Donde observamos que la pérdida esperada va creciendo junto con la probabilidad de impago.

7. Por todo lo expuesto en las conclusiones anteriores, creemos haber elaborado un Credit Scoring que cumple las directrices de Basilea II para pequeños préstamos. Su incorporación en la banca lleva la necesidad de implementar nuevas técnicas, lo que supone incurrir en un coste técnico y de profesionales tanto en el ámbito de departamentos de riesgo de los bancos, como por parte del supervisor encargado de confirmar la idoneidad de estos métodos. Desde estas páginas, queremos hacer un llamamiento a la banca privada para una mayor colaboración Universidad-Banca. Este trabajo hubiera podido ser más concluyente en su parte práctica,

Capítulo 5. Aplicación práctica de Credit Scoring Mixto

si los bancos y cajas de ahorros a los que fueron solicitados datos anónimos, hubieran colaborado, y estas entidades podrían contar con un trabajo de aplicación a sus préstamos con las características exigidas por Basilea II.

8. Al realizar este trabajo, nos planteamos otras formas de aplicación del mismo, y vimos que para millones de personas los microcréditos son la única forma de salir de la pobreza. . Para impulsar esta forma de financiación, la Asamblea de la ONU ha querido que el año 2005 sea el Año Internacional del Microcrédito, y ha pedido a todos los involucrados en la erradicación de la pobreza que tomen medidas adicionales para hacer que estos recursos estén disponibles cada vez para más personas. El microcrédito mueve una cartera de unos 15000 \$. De ella se benefician entre 400 y 500 millones de familias. El préstamo promedio entre las instituciones microfinancieras es de 450\$. Y el 60% de las personas clientes de los microcréditos son mujeres, por todas estas razones la nueva vía de investigación y aplicación de credit scoring planteada a partir de este trabajo irá enfocada al estudio de los microcréditos.
9. Seguramente el trabajo presentado y concluido en este punto tenga imprecisiones o defectos que la doctoranda no alcanza ver, pero quedará agradecida a todas las personas a quienes llegue esta tesis doctoral y tengan correcciones, precisiones y aportaciones que realizar, y cuyo agradecimiento queda ya plasmado de antemano en este texto.

Capítulo 5. Aplicación práctica de Credit Scoring Mixto

Bibliografía

[1] : Altman, Edward I., Giancarlo Marco y Franco Varetto. Corporate Distress Diagnosis: Comparisons Using Linear Discriminant Analysis and Neural Networks (The Italian Experience). Journal of Banking and Finance 18, pp.505-29. (1994).

[2] : Thomas, L., Edelman, D., Crook, J. Credit Scoring and its Applications. SIAM monographs on mathematical modeling and computation. (2002).

[3] : Fort, J.C., y Pagès G., Convergence on stochastic algorithms: from the Kushner & Clark theorem to the Lyapunov functional method. Adv. Appl. Prob., 28, 1072--1094.(1996).