

Sobre núcleos, distancias y similitudes entre intervalos

Luis González, Francisco Velasco
Dpto. Economía Aplicada I
Universidad de Sevilla
Avda Ramón y Cajal s/n SEVILLA
{luisgon, velasco}@us.es

Juan A. Ortega
Dpto. Lenguajes y sistemas Informáticos
Universidad de Sevilla
Avda Reina Mercedes s/n SEVILLA
ortega@us.es

Cecilio Angulo
Grup Recerca Enginyeria Coneixement
UPC
Vilanova i la Geltrú, BARCELONA
cangulo@esaii.upc.es

Francisco Ruiz
Grup Recerca Enginyeria Coneixement
Ramon Llull University (ESADE-URL)
ESADE Building Av. Pedralbes, BARCELONA
fruib1@pie.xtec.es

Resumen

Existe un buen número de aplicaciones en las que la información a codificar viene expresada en forma de un intervalo de valores. Esto sucede, en especial, cuando se intentan procesar los datos por predicción a un cierto tiempo, como en el estudio del transitorio de un sistema de control, o en el análisis de la evolución financiera de una empresa. En este trabajo se presenta y desarrolla la definición de una medida entre intervalos a partir de una distancia euclídea. Esta distancia tendrá en consideración como características esenciales el tamaño y la posición relativa entre los intervalos dentro de la recta real.

Gracias a la estrecha relación entre distancia y núcleo en un espacio de Hilbert, se hace posible definir un núcleo de tipo intervalar, sobre el espacio de los intervalos abiertos de dimensión finita en la recta real, un espacio que, originariamente, no posee ningún tipo de estructura algebraica de trabajo. El análisis sobre diferentes ejemplos prácticos pone de manifiesto las prestaciones de los instrumentos aquí presentados.

Palabras clave: Distancia, Similitud, Análisis Intervalar, Razonamiento Cualitativo, Máquina Núcleo.

1. Introducción

La indeterminación en la medida de los datos relativos a un sistema puede venir impuesta por diferentes fuentes: ruido en el sensor, incertidumbre en la variable asociada, o necesidad de incorporar todo un rango de valores de trabajo posibles. Esta imprecisión motiva la necesidad de tratar con espacios de trabajo definidos sobre elementos cualitativos: intervalos, órdenes de magnitud, intervalos modales, funciones de pertenencia. Más aún, incluso cuando la información del sistema bajo estudio es exacta, puede resultar interesante considerar sólo el valor cualitativo con objeto de generalizar resultados. Otra causa de indeterminación es la predicción del comportamiento de un sistema, incertidumbre que se manifiesta en la ventana de predicción a través de los parámetros asociados al modelo.

Son numerosos los trabajos y áreas de investigación que tratan con elementos básicos que recogen esta incertidumbre desde diferentes perspectivas. Tomando como punto de referencia la definición en [4] y [5] de una medida de similitud entre sucesos, aparece la posibilidad de realizar la construcción de una distancia entre intervalos, considerando éstos como sucesos de una determinada σ -álgebra. Sin embargo, para ello es necesario disponer de una variable aleatoria perfectamente definida, con objeto de calcular probabilidades que posteriormente permita determinar las distancias. Así, el problema fundamental de aplicar de forma directa este enfoque es la necesidad de conocer la distribución de la variable aleatoria, lo que reduce sustancialmente su campo de aplicación.

Por otro lado, también se debería vencer otro inconveniente: esta distancia tiene en cuenta el tamaño re-

lativo de los intervalos, en términos probabilísticos, pero no su posición relativa sobre la recta real. La imposibilidad de conseguir dentro del espacio de sucesos una distancia que tenga en cuenta esta característica de los intervalos conduce a definir la distancia entre éstos directamente sobre \mathcal{R} . En la Sección 2 se presenta una primera forma de distancia que da respuesta a parte de la problemática planteada, pero no por completo. La distancia adecuada se obtendrá en la siguiente Sección a partir de un enfoque distinto, utilizando conceptos de la teoría de aprendizaje estadístico y las máquinas núcleo. Su generalización en forma ℓ^p permite englobar de forma unificada la distancia definida en la Sección 2, y aquella considerada como adecuada distancia intervalar en la Sección 3. Además se prueba que la formulación en forma cuadrática permite diseñar distancias que conllevan cambios de ordenación sobre el espacio original. Al final se presentan las conclusiones y el trabajo en desarrollo.

2. Distancia Intervalar del Máximo

Una forma natural de definir una distancia aplicable sobre intervalos consiste en diseñar la ‘distancia’ entre conjuntos como “la mínima distancia entre dos puntos cualesquiera de ambos conjuntos”, una generalización de la distancia en la métrica de origen. Si A y B son dos conjuntos de un determinado espacio métrico, se define

$$d_c(A, B) = \min \{d(x_i, x_j) : x_i \in A, x_j \in B\} \quad (1)$$

donde d es la distancia en la correspondiente métrica. Como caso particular, especialmente interesante en este discurso, dados los intervalos abiertos de dimensión finita⁽¹⁾, $I_1 = (a_1, b_1) = B(c_1, r_1)$ e $I_2 = (a_2, b_2) = B(c_2, r_2)$, la aplicación d_c se expresa en la forma

$$\begin{aligned} d_c(I_1, I_2) &= \max \{0, a_2 - b_1, a_1 - b_2\} \\ &= \max \{0, \Delta c - \Sigma r, -\Delta c - \Sigma r\} \end{aligned} \quad (2)$$

donde $\Delta c = c_2 - c_1$ y $\Sigma r = r_1 + r_2$ ⁽²⁾.

⁽¹⁾Los intervalos se describirán indistintamente a partir de sus extremos, o en forma de bola abierta, según notación boreliana. Así, en función de su centro y su radio, se cumple que $a_i = c_i - r_i$, $b_i = c_i + r_i$ con $r_i \geq 0$.

⁽²⁾Por extensión, se define $\Delta r = r_2 - r_1$.

Sin embargo esta aplicación no es una distancia, ya que para cualquier A y B con $A \cap B \neq \emptyset$, se tiene $d_c(A, B) = 0$. Además, si se considera el caso $a_1 \leq a_2$, entonces, comenzando desde $b_2 = a_2$ y $a_1 = b_1$, se puede aumentar b_2 y disminuir a_1 tanto como se desee sin que la distancia entre intervalos cambie: d_c “no tiene en consideración el tamaño de los intervalos, sólo su proximidad respecto de los extremos más cercanos”. Por ejemplo, si se piensa que los conjuntos de trabajo son dos países, A y B , que comparten frontera, entonces $d_c(A, B)$ es cero, por lo que para dos personas situadas en distintos puntos del país A que deseen conocer B , esta medida no les proporciona apenas información sobre la longitud del viaje.

Ante esta utilidad, casi necesidad, de introducir el tamaño de los intervalos en la definición adecuada de distancia, en vez de considerarse la Ec. 1, puede pensarse en definir la distancia entre conjuntos como “la mínima distancia para alcanzar un conjunto desde cualquier punto del otro conjunto, d_m ”. Esta posibilidad aporta más información: siguiendo el ejemplo anterior, si $B \subset C$, con A compartiendo frontera con C a través de B , entonces, $d_c(A, B) = d_c(A, C) = 0$, pero no necesariamente $d_m(A, B) = d_m(A, C)$. Así, lo más interesante de esta definición sería que “la distancia tiene en cuenta no sólo la posición de los conjuntos sino también sus tamaños relativos”.

A partir de este momento el discurso se centra en el trabajo con intervalos abiertos de dimensión finita definidos sobre la recta real.

Definición 2.1 *Dados dos intervalos $I_1 = (a_1, b_1)$ e $I_2 = (a_2, b_2)$ se define la distancia del máximo entre ambos intervalos, y se denota por $d_m(I_1, I_2)$, a*

$$\begin{aligned} d_m(I_1, I_2) &= \max \{|a_2 - a_1|, |b_2 - b_1|\} \\ &= \max \{|\Delta c - \Delta r|, |\Delta c + \Delta r|\} \end{aligned} \quad (3)$$

La notación boreliana, y su extensión a incrementos de radios y de centros, se aprecia que es más adecuada que aquella que utiliza los extremos de los intervalos. Como ejemplo sirvan estos dos casos particulares:

- Sean dos intervalos $I_1 = (a_1, b_1)$ e $I_2 = (a_2, b_2)$, con igual extremo superior, es decir, $b_1 = b_2$ o equivalentemente $\Delta c + \Delta r = 0$, entonces de la Definición 2.1 se sigue que,

$$d_m(I_1, I_2) = |a_2 - a_1| = |\Delta c - \Delta r| \quad (4)$$

- Si los intervalos son de la misma amplitud, $a_2 - a_1 = b_2 - b_1$, $\Delta r = 0$, entonces la distancia es,

$$d(I_1, I_2) = |a_2 - a_1| = |\Delta c| \quad (5)$$

En las dos ocasiones el valor numérico de la distancia es el mismo, pero mientras la notación sobre los extremos no permite distinguir los casos, la notación en forma de incrementos es plenamente ilustrativa.

Proposición 2.2 *La aplicación d_m es una distancia.* ■

Proposición 2.3 *La aplicación d_m cumple que $d_m(I_1, I_2) = |\Delta c| + |\Delta r|$.* ■

La demostración de estas dos proposiciones se desarrolla en el anexo.

Aunque ahora d_m es una distancia que considera la posición relativa de los centros de las bolas y la diferencia de radios, no es aún todo lo útil que se desearía ya que no tiene en cuenta la distancia entre extremos cercanos. Para aclararlo, considérense el intervalo $I_1 = (0, 1)$ y los intervalos de la forma $I_h = (2 - h, 3)$ con $h \geq 0$. El cálculo de la distancia del máximo indica que,

$$d_m(I_1, I_h) = \max\{|2 - h|, |3 - 1|\} = 2 \quad \forall h \in [0, 4]$$

lo que significa que no se tiene en cuenta el tamaño relativo de los intervalos de la forma I_h para cualquier valor de h entre 0 y 4, entre los cuales se hallan intervalos tan dispares como $(2, 3)$, de amplitud 1, ó $(-2, 3)$, de amplitud 5.

3. Distancia Intervalar a partir de un Núcleo

Para mejorar la distancia del máximo, d_m , definida en la Sección 2, se recurre a una forma de construcción, habitual dentro de la teoría del aprendizaje estadístico, a partir de la definición de funciones núcleo (*kernels*). Para una mejor comprensión de este tipo de funciones se puede acudir a [6] y [2], o a los textos en castellano [4], [3] y [1].

En esencia, el objetivo de la construcción de funciones núcleo es asegurar la existencia de una aplicación

ϕ definida desde el conjunto de trabajo (el cual no necesariamente está provisto de ninguna estructura matemática a priori) a un espacio vectorial dotado de un producto escalar, como por ejemplo puede ser \mathbb{R}^n , denominado espacio de características, \mathcal{F} .

A partir de esta función ϕ , en general no lineal, se define la función núcleo, que se denota $k(\cdot, \cdot)$, sobre pares de elementos del conjunto de trabajo como el producto escalar de sus transformados dentro del espacio de características,

$$k(\cdot, \cdot) = \langle \phi(\cdot), \phi(\cdot) \rangle_{\mathcal{F}}$$

La función núcleo $k(\cdot, \cdot)$ permite establecer similitudes entre los elementos originales a partir de sus transformados, lo que conlleva en ocasiones poder definir una distancia entre los puntos origen. La aplicación ϕ , por tanto, debe ser capaz de resaltar las características fundamentales de los elementos del conjunto inicial a ser consideradas a la hora de elaborar una medida de similitud y distancia; es por ello que al espacio imagen de la aplicación ϕ se le denomina espacio característico o espacio de características.

Siguiendo el enfoque de investigación intervalar, se denotará por \mathcal{I} la familia formada por todos los intervalos abiertos (a, b) contenidos en la recta real⁽³⁾ de dimensión finita,

$$\mathcal{I} = \{\text{intervalos } (a, b) \subset \mathbb{R} : a < b, a \neq -\infty, b \neq \infty\}$$

Desde la perspectiva de las máquinas núcleos, la elección primera de la función ϕ es aquella que asocia a cada intervalo, un vector en \mathbb{R}^2 de la forma más natural⁽⁴⁾,

$$\phi : \mathcal{I} \rightarrow \mathbb{R}^2$$

$$\phi((a, b)) = \phi(I) = \frac{1}{\sqrt{2}}(a, b) \in \mathbb{R}^2$$

o en notación de Borel,

$$\begin{aligned} \phi((c - r, c + r)) &= \phi(I) = \frac{1}{\sqrt{2}}(c - r, c + r) \\ &= \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} c \\ r \end{pmatrix} \\ &= P \begin{pmatrix} c \\ r \end{pmatrix} \in \mathbb{R}^2 \end{aligned}$$

⁽³⁾Por defecto, se trabajará con intervalos abiertos, pero el estudio se puede trasladar de forma natural a intervalos cerrados.

⁽⁴⁾El término $\frac{1}{\sqrt{2}}$ es introducido para conseguir una transformación ortogonal, como se verá posteriormente.

donde $P = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$ es una matriz ortogonal⁽⁵⁾ ($P^t P = I$). Nótese que cada intervalo define un único punto en el plano. El recorrido de la aplicación ϕ es la región $\{(a, b) \in \mathbb{R}^2 : a < b\}$.

La función núcleo asociada a esta aplicación ϕ , que permite establecer una función núcleo de similitud⁽⁶⁾ entre intervalos, es

$$k : \mathcal{I} \times \mathcal{I} \rightarrow \mathbb{R}$$

definida como

$$k(I_1, I_2) = \langle \phi(I_1), \phi(I_2) \rangle = \frac{1}{2} (a_1 a_2 + b_1 b_2) = c_1 c_2 + r_1 r_2$$

Lo más interesante de esta función núcleo de similitud es que da lugar a definir una distancia entre intervalos a partir de la distancia entre sus transformados en \mathbb{R}^2 .

Definición 3.1 *Dados dos intervalos, $I_1 = (a_1, b_1) = B(c_1, r_1)$ e $I_2 = (a_2, b_2) = B(c_2, r_2)$ se define la distancia entre ambos intervalos, y se denotará por $d(I_1, I_2)$, como*

$$\begin{aligned} d(I_1, I_2) &= \frac{1}{\sqrt{2}} \sqrt{(a_2 - a_1)^2 + (b_2 - b_1)^2} \\ &= \sqrt{\Delta^2 c + \Delta^2 r} \end{aligned} \quad (6)$$

pues

$$\begin{aligned} d^2(I_1, I_2) &= \langle \phi(I_2) - \phi(I_1), \phi(I_2) - \phi(I_1) \rangle \\ &= \frac{1}{2} (\Delta c - \Delta r)^2 + \frac{1}{2} (\Delta c + \Delta r)^2 \\ &= \Delta^2 c + \Delta^2 r \end{aligned}$$

Como corolario más interesante indicar cómo partiendo de entornos de trabajo bien distantes, se ha llegado a mostrar que la distancia definida se engloba dentro de la familia de las distancias ℓ^p sobre las componentes Δc y Δr . Obsérvese que la distancia ahora definida coincide con ℓ^2 , mientras que aquella de la Definición 2.1 se identifica con ℓ^1 , en virtud de la proposición 2.3.

La notación boreliana, en forma de bolas abiertas, ha permitido hacer evidente que en la definición de esta

⁽⁵⁾De hecho, para asegurar que la función es inyectiva, basta que P sea no singular.

⁽⁶⁾Si la función núcleo se utiliza para comparar se suele denominar función núcleo de similitud.

distancia a partir de un núcleo se tiene en consideración la distancia entre los centros de los intervalos, $\Delta^2 c$, y el tamaño relativo de ellos, $\Delta^2 r$, ponderándose ambos valores de forma equilibrada. Su expresión en forma cuadrática sería la siguiente

$$d^2(I_1, I_2) = (\Delta c, \Delta r) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \Delta c \\ \Delta r \end{pmatrix}$$

Volviendo al ejemplo de la Sección 2 del intervalo $I_1 = (0, 1)$ y los intervalos de la forma $I_h = (2 - h, 3)$ con $h \geq 0$, el cálculo de la distancia ahora definida establece,

$$d(I_1, I_h) = \sqrt{4 + (2 - h)^2} \quad \forall h \in [0, 4]$$

que es mínima en el caso $h = 2$, $d((0, 1), (0, 3))$, y máxima cuando $h = 0$ ó 4 , $d((0, 1), (2, 3))$ – intervalos pequeños pero distantes –, ó $d((0, 1), (-2, 3))$ – intervalos con intersección total pero de tamaño desigual –. Por tanto, se ha solventado la problemática a la que no podía responder la distancia del máximo, pero manteniendo las características deseables en la definición de distancia.

4. Generalizaciones de la distancia intervalar

Una vez concretada una primera definición de distancia intervalar dentro del entorno de núcleos, diversas generalizaciones son posibles de forma más o menos directa.

4.1. Distancia intervalar ponderada

Una generalización de la distancia intervalar puede considerarse para los casos en que la ponderación entre el tamaño de los intervalos y la distancia entre centros quisiera realizarse de forma no equilibrada. Para ello, se define la función ϕ proyectando el espacio original en el de características, de la forma

$$\phi(I) = A \begin{pmatrix} c \\ r \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} c \\ r \end{pmatrix}$$

de donde

$$\begin{aligned} k(I_1, I_2) &= \left\langle \left(A \cdot \begin{pmatrix} c_1 \\ r_1 \end{pmatrix} \right)^t, \left(A \cdot \begin{pmatrix} c_2 \\ r_2 \end{pmatrix} \right) \right\rangle \\ &= \begin{pmatrix} c_1 & r_1 \end{pmatrix} S \begin{pmatrix} c_2 \\ r_2 \end{pmatrix} \end{aligned}$$

por lo que,

$$\begin{aligned} d^2(I_1, I_2) &= \left(A \cdot \begin{pmatrix} \Delta c \\ \Delta r \end{pmatrix} \right)^t \cdot \left(A \cdot \begin{pmatrix} \Delta c \\ \Delta r \end{pmatrix} \right) \\ &= \begin{pmatrix} \Delta c & \Delta r \end{pmatrix} S \begin{pmatrix} \Delta c \\ \Delta r \end{pmatrix} \end{aligned}$$

siendo A una matriz no singular con objeto de que ϕ sea una aplicación inyectiva, y $S = A^t A$ es una matriz simétrica y definida positiva.

De esta forma se puede controlar el peso que corresponde dar a la posición de los intervalos, c , y a su tamaño, r . Así, por ejemplo, si

$$S = A^t A = \begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix}$$

y se consideran los intervalos $(0, 1)$ y $(1, 3)$, entonces se observa que

$$\begin{aligned} d^2((0, 1), (1, 3)) &= (0,5 - 2)^2 + 3(0,5 - 1)^2 = 3 \\ d^2((0, 1), (0, 3)) &= (0,5 - 1,5)^2 + 3(0,5 - 1,5)^2 = 4 \end{aligned}$$

mientras que sobre la distancia definida de origen

$$\begin{aligned} d_1^2((0, 1), (1, 3)) &= (0,5 - 2)^2 + (0,5 - 1)^2 = 2,5 \\ d_1^2((0, 1), (0, 3)) &= (0,5 - 1,5)^2 + (0,5 - 1,5)^2 = 2 \end{aligned}$$

es decir, que la elección de una ponderación u otra permite realizar cambios de ordenación sobre el conjunto de los intervalos.

4.2. Distancia intervalar definida sobre soporte compacto

La función núcleo de similitud que ha permitido definir la distancia intervalar en la Definición 3.1 no se encuentra acotada, por lo que aún podría mejorarse para que resultara más práctica a la hora de interpretar los resultados. Además, se trabaje con notación boreliana o con los extremos de intervalos, a la distancia no le es posible tratar con intervalos semiabiertos si el soporte de trabajo es toda la recta real, \mathbb{R} .

Con objeto de tener acotadas la similitud y la distancia, se estudia la posibilidad de definir el núcleo de similitud entre intervalos contenidos en un compacto de \mathbb{R} , en general $[\alpha, \beta]$, con $\alpha, \beta \in \mathbb{R}$ y $\alpha < \beta$.

Sea el compacto (intervalo cerrado en \mathbb{R}) $[\alpha, \beta]$ y consideramos el conjunto

$$\mathcal{I}_1 = \{\text{intervalos } (a, b) : \alpha \leq a < b \leq \beta\}$$

En este conjunto se define una nueva medida de similitud a partir de la definición de una función ϕ_1 semejante a la original

$$\begin{aligned} \phi_1 : \mathcal{I}_1 &\rightarrow \mathbb{R}^2 \\ \phi_1((a, b)) &= \frac{1}{\sqrt{2}}(a - \alpha, \beta - b) \end{aligned} \quad (7)$$

Nótese que esta aplicación ϕ_1 se puede escribir en la forma:

$$\begin{aligned} \phi_1(a, b) &= \frac{1}{\sqrt{2}}(c - r - \alpha, \beta - c - r) \\ &= \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ -1 & -1 \end{pmatrix} \begin{pmatrix} c \\ r \end{pmatrix} + \frac{1}{\sqrt{2}} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \\ &= P_1 \begin{pmatrix} c \\ r \end{pmatrix} + \Lambda \end{aligned}$$

donde P_1 es ortogonal. Aunque en este caso las componentes características son el espacio complementario al intervalo (a, b) respecto a la base compacta, se sigue que la distancia que se obtiene no varía respecto a aquella de la Definición 3.1 ya que

$$\phi_1(I_1) - \phi_1(I_2) = P_1 \begin{pmatrix} \Delta c \\ \Delta r \end{pmatrix}$$

y $P_1^t P_1 = P^t P = I$. Además, tal como se pretendía, la distancia ahora se encuentra acotada ya que si identificamos cada intervalo como un punto en \mathbb{R}^2 , se sigue que ahora ϕ_1 es una función continua en un compacto y por el Teorema de Weierstrass la función esta acotada, demostrándose que $d(I_1, I_2) \leq \beta - \alpha$ para cualquier par de intervalos.

Nótese que, aunque la distancia definida es la misma que la inicial, y presenta una formulación simple, la función núcleo de similitud asociada es bastante más compleja

$$k((a_1, b_1), (a_2, b_2)) = c_1 c_2 + r_1 r_2 + C^2 + R^2 - 2C(c_1 + c_2) - 2R(r_1 + r_2)$$

con $C = \frac{\alpha + \beta}{2}$ el centro del compacto base, y $R = \frac{\beta - \alpha}{2}$ su radio.

5. Conclusiones y trabajo futuro

Una forma eficiente de tratar con incertidumbre en la información es a través del uso de intervalos. Por

lo general, su tratamiento algebraico suele referirse a su aplicación punto a punto sobre todos los valores posibles del intervalo para posteriormente acotarse de forma conjunta. En el presente trabajo se plantea el tratamiento puramente intervalar de estos elementos básicos, para lo cual se hace necesario definir una distancia apropiada. La solución propuesta enfoca el problema mediante la utilización de núcleos ('kernels'), lo que lleva a definir, por vez primera en la literatura, núcleos intervalares, los cuales son la base de definición de la distancia. Se ha comprobado que la distancia definida cumple con una serie de características que la hacen particularmente útil en el tratamiento intervalar.

Se han presentado algunas generalizaciones y ejemplos ilustrativos de aplicación, pero sin duda este trabajo es sólo el inicio de una línea de investigación prometedora que auna el área emergente de investigación en máquinas núcleo junto con otros más tradicionales, como el Análisis Intervalar, el Razonamiento Cualitativo o la Lógica Difusa.

Desde el punto de vista de las máquinas núcleo, ha de referirse que los núcleos de similitud asociados a las distancias no son independientes de su posición sobre el intervalo base, hecho que debe mejorarse en el futuro. Sin embargo, la distancia es plenamente válida como útil de trabajo en entornos cualitativos, y su aplicación sobre diversos casos de estudio es una de las líneas futuras de investigación. En particular, se está realizando su desarrollo para el análisis y medición del riesgo crediticio.

Agradecimientos

Este trabajo ha sido soportado en parte por las ayudas ACC-944-SEJ-2002 concedida por la Junta de Andalucía y TIC2002-04371-C02-02, proyecto MERITO, del Ministerio de Ciencia y Tecnología.

Referencias

- [1] C. Angulo. *Learning with Kernel Machines into a Multi-Class Environment*. Doctoral thesis, Technical University of Catalonia, April 2001. In Spanish.
- [2] N. Cristianini and J. Shawe-Taylor. *An introduction to Support Vector Machines and other*

kernel-based learning methods. Cambridge University press 2000, 2000.

- [3] L. González. *Teoría del aprendizaje estadístico de la regresión. Máquinas de regresión de vector base*. Trabajo interno del departamento, Facultad de Ciencias Económicas y Empresariales, Universidad de Sevilla, Diciembre 2000.
- [4] L. González. *Análisis discriminante utilizando máquinas núcleos de vectores soporte. Función núcleo similitud*. Tesis doctoral, Universidad de Sevilla, Marzo 2002.
- [5] B. Schölkopf. *Statistical learning and kernel methods*. Technical Report MSR-TR-2000-23, Microsoft Research Limited, Febrero 2000.
- [6] V.N. Vapnik. *Statistical Learning Theory*. John Wiley & Sons, Inc, 1998.

A. Anexo

Demostración Proposición 2.2:

Se debe comprobar que se verifican las propiedades de distancia,

1. $0 \leq d_m(I_1, I_2)$ para I_1 e I_2 intervalos. Trivial.
2. $d_m(I_1, I_2) = 0 \Leftrightarrow \left\{ \begin{array}{l} a_2 = a_1 \\ b_2 = b_1 \end{array} \right\} \Leftrightarrow I_1 = I_2$.
3. $d_m(I_1, I_2) = d_m(I_2, I_1)$. Evidente, por el uso del valor absoluto en la definición.
4. Para cualesquiera tres intervalos I_1, I_2 e I_3 se cumple la desigualdad triangular ya que es cierto que:

$$|a + b| \leq |a + c| + |c + b|, \quad \forall a, b, c \in \mathbb{R}$$

y se tiene que:

$$\left. \begin{array}{l} |a_2 - a_1| \leq |a_2 - a_3| + |a_3 - a_1| \\ |b_2 - b_1| \leq |b_2 - b_3| + |b_3 - b_1| \end{array} \right\} \Rightarrow$$

$$\max\{|a_2 - a_1|, |b_2 - b_1|\} \leq \max\{|a_2 - a_3|, |b_2 - b_3|\} +$$

y de aquí se sigue que:

$$d_m(I_1, I_2) \leq d_m(I_1, I_3) + d_m(I_2, I_3)$$

■

Demostración Lema 2.3:

En primer lugar se tiene que si $a, b \in \mathbb{R}$ entonces:

$$\max\{|a|, |b|\} = \frac{1}{2}(|a| + |b| + ||b| - |a||)$$

ya que si $a \geq 0$ y $b \geq 0$ entonces si:

$$a \geq b \Rightarrow \max\{a, b\} = a = \frac{1}{2}(a + b + (a - b)) = \frac{1}{2}(a + b + |b - a|)$$

Análogamente se tiene si $a \leq b$.

Por otro lado, como $|a| \geq 0$ y $|b| \geq 0$ se cumple la afirmación anterior.

De esta forma se sigue:

$$\begin{aligned} 0 \leq d_m(I_1, I_2) &= \max\{|\Delta c - \Delta r|, |\Delta c + \Delta r|\} \stackrel{?}{=} |\Delta c| + |\Delta r| \Leftrightarrow \\ \max\{|\Delta c - \Delta r|, |\Delta c + \Delta r|\}^2 &= (|\Delta c| + |\Delta r|)^2 \Leftrightarrow \\ \max\{|\Delta c - \Delta r|^2, |\Delta c + \Delta r|^2\} &= \\ \frac{1}{2} \left(|\Delta c - \Delta r|^2 + |\Delta c + \Delta r|^2 + \left| |\Delta c - \Delta r|^2 - |\Delta c + \Delta r|^2 \right| \right) &= \\ \frac{1}{2} \left((\Delta c - \Delta r)^2 + (\Delta c + \Delta r)^2 + \left| (\Delta c - \Delta r)^2 - (\Delta c + \Delta r)^2 \right| \right) &= \\ \Delta^2 c + \Delta^2 r + 2|\Delta c \Delta r| &= (|\Delta c| + |\Delta r|)^2 \end{aligned}$$

■