

Redes de Neuronas Artificiales versus Análisis de Conglomerados: un estudio comparativo en una muestra de niños con espina bífida

Ana María LÓPEZ JIMÉNEZ
José GARCÍA LUNA
Montserrat GÓMEZ DE TERREROS
José Enrique ROMERO GARCÍA
Universidad de Sevilla
Antonio R. GARCÍA TORRES
I.E.S. A Los Viveros@. Sevilla

Resumen

El uso de las redes de neuronas artificiales en análisis de datos en Psicología no es muy frecuente. Aunque la mayoría de las técnicas estadísticas de análisis multivariante pueden implementarse en redes de neuronas artificiales con distintas arquitecturas los investigadores prefieren las técnicas estadísticas. La ausencia de estudios que pongan de manifiesto las ventajas de unos procedimientos frente a otros puede explicar el escaso uso de las RNA. En la línea de los estudios comparativos, en este trabajo hemos utilizado una red competitiva y el algoritmo de clasificación no jerárquico de k-medias para realizar una partición en dos grupos de una muestra de niños con espina bífida en base a las puntuaciones obtenidas en el WISC-R.

Palabras clave: Análisis de conglomerados, redes de neuronas artificiales, aprendizaje no supervisado, k-medias.

Abstract

The use of ANN in the data analysis in Psychology is not very frequent. The prevalence of multivariate statistical techniques is evident although the majority of these techniques can be implemented in different architectures of ANN. In this work, we have compared the profiles of the clusters obtained using a non-hierarchical classification algorithm (k-means) and those obtained by means of a competitive net. These two analysis procedures have been applied to a sample of children with spinal bifida and who are undergoing treatment at the "Virgin del Rocío" Children's Hospital of seville.

Key words: cluster analysis, artificial neural network, unsupervised learning, k-means.

Dirección de la primera autora: Departamento de Psicología Experimental, Facultad de Psicología. Avda. Camilo José Cela s/n. 41005 Sevilla. ***Correo electrónico:*** analopez@cica.es

Aunque los conceptos básicos de las Redes de Neuronas Artificiales (en adelante RNA) fueron planteados en la década de los 50 su aplicación al ámbito de los problemas propios de la estadística es mucho más reciente y no puede decirse que, en este momento, sean una herramienta que forme parte del conjunto de herramientas de análisis de los investigadores en ciencias sociales y humanas. Se necesitan estudios que permitan establecer bajo qué condiciones las RNA son procedimientos superiores a las técnicas estadísticas convencionales de análisis de datos. La ausencia de este tipo de estudios no es, sin embargo, exclusiva del campo de las RNA. Una situación similar ocurre con muchas de las técnicas clásicas de análisis multivariante. En el análisis de conglomerados (*cluster analysis*), por ejemplo, se han propuesto múltiples algoritmos y reglas de clasificación pero pocos han sido los estudios realizados para determinar la efectividad de unas reglas frente a otras en aspectos tales como el número y características de los conglomerados (Balakrishnan, Cooper, Jacob, y Lewis, 1994). Concretamente, y en relación al algoritmo de clasificación no jerárquica de *k-medias* (*k-means*) se sabe que las soluciones dependen, en buena parte, del procedimiento utilizado para elegir las semillas, de las semillas elegidas y de la distribución de los datos. La selección aleatoria de las semillas, uno de los procedimientos más frecuente cuando se utiliza *k-medias*, a menudo proporciona soluciones subóptimas para los conglomerados (Hair *et al.*, 1999; Milligan y Sokol, 1980).

En este trabajo, comparamos los perfiles de los conglomerados obtenidos utilizando el algoritmo de clasificación no jerárquico *k-medias* -con selección aleatoria y no aleatoria de las semillas- con una red competitiva (Kohonen, 1984). Estos dos procedimientos de análisis los hemos aplicado a una muestra de niños con problemas de espina bífida que están siendo atendidos en el Hospital Infantil Virgen del Rocío de Sevilla.

En los apartados que siguen, y antes de comparar los resultados de la clasificación con los dos procedimientos, desarrollaremos aspectos conceptuales básicos de las RNA y del algoritmo de *k-medias*.

Redes de Neuronas Artificiales (RNA)

Las RNA son dispositivos de cálculo inspirados en las redes de neuronas biológicas (Bishop, 1995; Hiler y Martínez, 1995; Ripley, 1996). Como estas últimas, están constituidas por elementos simples denominados nodos o neuronas organizados en capas y altamente interconectados. A la forma particular de organizarse y conectarse las neuronas se le denomina arquitectura o topología de red. A cada conexión se le asigna un peso numérico que va a constituir el principal recurso de memoria a largo plazo de las RNA. El aprendizaje se realiza, usualmente, con la actualización de los pesos mediante una determinada regla de aprendizaje.

Una neurona es la unidad básica de procesamiento. Recibe y emite información de otras neuronas y del/hacia el mundo exterior (Sellés y Renom, 1998). En cada neurona se realiza un cálculo local y sencillo con las entradas que le proporcionan sus vecinas sin que sea necesario un control global en el conjunto de unidades. La neurona artificial procesa la información que le llega mediante la obtención de dos componentes (Russell y Norvig, 1996). El primero es un componente *lineal* denominado función de entrada (in_i) que calcula la suma ponderada de los valores de entrada a la neurona ($in_i = \sum w_{ij}$). El segundo es un componente, generalmente, *no lineal* conocido como función de activación o función de transferencia, f , que transforma la suma ponderada en el valor de salida de la neurona ($x_i = f(in_i)$). La figura 1 es un esquema del funcionamiento de una neurona artificial. En dicho esquema x_j representa las entradas a la neurona y w_{ij} los pesos de cada conexión.

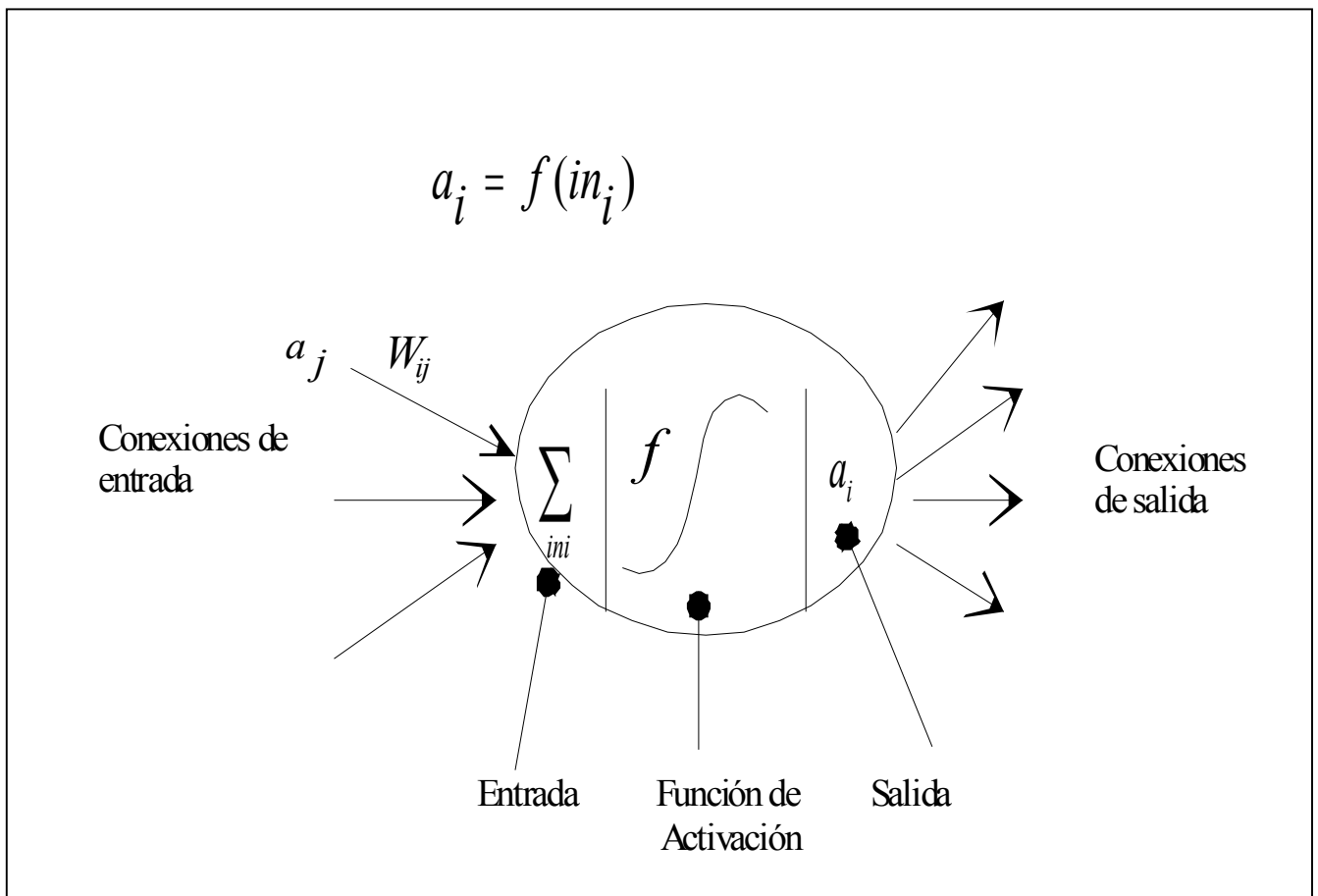


Figura 1. Representación esquemática de una neurona artificial.

La utilización de diferentes funciones matemáticas para f da lugar a distintos tipos de neuronas. Cuatro de las funciones de activación más comunes son: escalón, signo, lineal y sigmoide o logística.

La función escalón o umbral tiene un límite, u , de manera que produce un 1 cuando la entrada es mayor que u y, en caso contrario, produce 0. La función signo produce +1 si la entrada es mayor que u y -1 en caso contrario. La función lineal o identidad deja invariante la entrada. La función sigmoide o logística viene dada por

Las columnas de nodos como el representado en la Figura 1 se denominan capas. Se distinguen tres tipos de capas: de entrada, internas u ocultas y de salida. La capa de entrada (constituida por tantos nodos como variables medidas en la investigación) distribuye las señales entre los nodos de la siguiente capa. Cada neurona de la primera capa interna aplica la función suma y de activación a las entradas y las transmiten a la siguiente capa. La última capa interna emite señales que llegan a las neuronas de la capa de salida. En la figura 2 hemos representado una red con una capa oculta del tipo feed-forward o perceptrón multicapa caracterizada porque sólo tiene conexiones unidireccionales con todas las neuronas de la capa siguiente.

Este tipo de redes, de las más utilizadas en análisis de datos, es un *aproximador universal*. Con una sola capa oculta son capaces de aproximar cualquier función siempre y cuando las funciones de activación de los nodos de esta capa sean no lineales (Funahashi, 1989; Hartman, Keeler y Kowalski, 1990; Hornik, Stinchcombe y White, 1989).

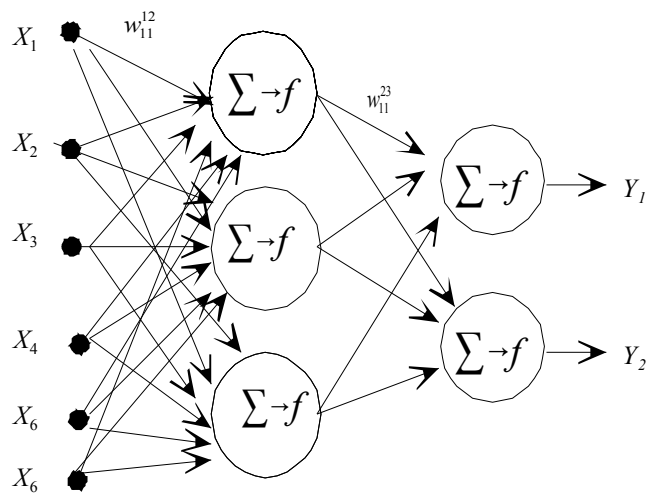


Figura 2. RNA feed-forward

Al procedimiento mediante el que se modifica el valor de los pesos de conexión para ajustarlo a la salida deseada se le denomina aprendizaje. Desde un punto de vista muy global se distinguen dos tipos de aprendizaje: supervisado y no supervisado. El aprendizaje supervisado (o aprendizaje con maestro) trata de conseguir que la red sea capaz de predecir, a partir de un conjunto de características suministradas como entradas, el valor que tomarán otras características, llamadas objetivo, habiendo sido observadas ambos tipos de características en cada uno de los patrones de entrenamiento (análisis discriminante). En el aprendizaje no supervisado existen características de entrada pero no características objetivo (aprendizaje sin maestro). No existe información que indique si la salida que produce la red es o no correcta. Las técnicas de aprendizaje no supervisado se utilizan sobre todo para obtener estructuras, relaciones o clasificaciones (*cluster analysis*). Existen muchos algoritmos de aprendizaje pero los más extendidos son la regla delta generalizada (*backpropagation*) para el aprendizaje supervisado y la regla estándar del aprendizaje competitivo en el caso no supervisado.

En el presente estudio hemos utilizado una red competitiva para obtener una clasificación en dos grupos de niños con espina bífida en base a doce características medidas. Las redes competitivas, a diferencia de la red *feed-forward*, se caracterizan por la presencia de relaciones bidireccionales entre los nodos de la misma capa (Kohonen, 1984).

La Figura 3 muestra la arquitectura de la red que hemos utilizado. La red tiene dos capas: a) entrada (representada por puntos negros en la figura 3) constituida por doce nodos correspondientes a las doce variables que hemos utilizado para clasificar a los niños, y b) una capa de salida con dos nodos (representados por círculos) correspondientes a los dos conglomerados en que queremos clasificar a los sujetos. Para cada vector de entrada, correspondiente al registro de un sujeto, solamente uno de los dos nodos de salida puede activarse.

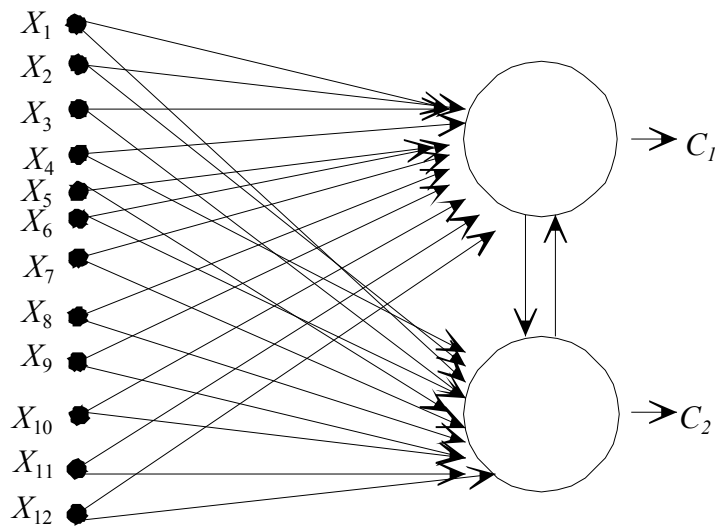


Figura 3. Red competitiva para la clasificación en dos conglomerados.

El algoritmo de aprendizaje comienza inicializando los pesos (w_{ij}) con valores aleatorios. A continuación se presenta a la red una información de entrada en forma de vector $(x_{1k}, x_{2k}, \dots, x_{pk})^1$ que activará a una sola de las dos neuronas de la capa de salida. A la neurona activada se le denomina *ganadora*. La neurona ganadora de la capa de salida es aquella cuyo vector de pesos sea más parecido al patrón de entrada $(x_{1k}, x_{2k}, \dots, x_{pk})$. Para evaluar el parecido se calcula la distancia entre el patrón de entrada y el vector de pesos de cada neurona. La distancia viene dada por la expresión (2), donde $i = 1, 2$ y $k = 1, \dots, N$.

Una vez localizada la neurona ganadora se actualizan los pesos según la regla estándar del aprendizaje competitivo que viene dada por la expresión (3). En la expresión (3), α es un valor entre (0, 1) que se conoce como tasa de aprendizaje. El proceso se repite hasta que la variación de los pesos sea insignificante.

Análisis de conglomerados

El análisis de conglomerados (*cluster analysis*) es un término genérico para una amplia variedad de algoritmos con un objetivo común: la formación de grupos de sujetos o variables similares (Aldenderfer y Blashfield, 1984; Anderberg, 1973; Everitt, 1993). Los algoritmos propuestos se clasifican en jerárquicos y no jerárquicos. Los algoritmos jerárquicos parten del conjunto inicial de unidades de análisis a agrupar y llegan a un sólo grupo que contiene a todos los elementos. Otra modalidad de los algoritmos de clasificación jerárquicos consiste en partir de un sólo grupo que contiene a todos los objetos o sujetos e ir subdividiendo, en etapas sucesivas, el grupo inicial (jerárquicos descendentes). Los algoritmos de clasificación no jerárquicos por el contrario, forman los grupos necesarios para optimizar alguna función de ajuste.

En este trabajo nos centraremos en el algoritmo de clasificación no jerárquico de *k-medias* por su similitud con la red competitiva representada en la figura 3. Brevemente, el algoritmo de *k-medias* procede como sigue (Lebart, Morineau y Fenelon, 1985):

1. Cada vector de entrada corresponde a las puntuaciones de un sujeto en las variables que se utilizan para la clasificación. Son las filas de la matriz de datos en forma estándar. Estos vectores denominados patrones de entrenamiento se le presentan hasta que la variación de los pesos sea insignificante.

1. Inicializa los valores de las semillas para un número (k) de conglomerados preespecificado. Esos valores que representan los centroides iniciales de los conglomerados inducen una primera partición. Generalmente, los valores iniciales se eligen aleatoriamente entre las unidades de análisis de la matriz de datos.
2. Se calcula la distancia de cada sujeto de la muestra a los centroides y se asigna al conglomerado de cuyo centroide esté más próximo.
3. Después de la primera clasificación, se vuelven a calcular los centroides.
4. Se repiten los pasos 2 y 3 hasta que no haya reasignación o la variación en los centroides sea menor que un nivel de tolerancia especificado previamente.

Comparación de las clasificaciones obtenidas con el procedimiento de k -medias y con una red competitiva

El objetivo de este estudio, como ya hemos comentado antes, es comparar los perfiles de los conglomerados obtenidos utilizando el algoritmo de k -medias y una red competitiva como la de la figura 3.

La muestra a clasificar estaba constituida por 68 niños con espina bífida atendidos en el hospital infantil Virgen del Rocío de Sevilla. Los niños (36 niños y 32 niñas), con edades comprendidas entre los 6 y los 16 años, habían realizado el test de inteligencia WISC-R. Este test consta de 12 subescalas: 6 miden habilidad verbal (información, comprensión, aritmética, semejanzas, vocabulario y dígitos) y las otras 6 miden habilidades manipulativas (figuras incompletas, historietas, cubos, rompecabezas, claves y laberintos). Las subescalas del WISC-R son las variables utilizadas para clasificar a los sujetos con el algoritmo k -medias y mediante la red de la figura 3. Además de las puntuaciones obtenidas en el WISC-R, disponíamos de información relativa a las siguientes variables:

- Variables de tipo médico: como la presencia o no de válvulas de derivación, presencia o no de mielinización y presencia o no de infecciones.
- Variables psicológicas: retraso o no en el desarrollo, presencia o ausencia de problemas escolares y presencia o no de ansiedad.
- Variables sociodemográficas: sexo, edad, hijo único o no, procedencia rural o urbana.

Estos tres tipos de variables las hemos utilizado para establecer el perfil de los conglomerados y analizar el posible efecto diferencial de los procedimientos de clasificación utilizados en dicho perfil.

Hemos utilizado el algoritmo de k -medias implementado en el programa estadístico SPSS (v. 10.0). Por análisis exploratorios previos, decidimos clasificar a los niños en dos conglomerados ($k = 2$) partiendo de dos conjuntos de semillas distintos: a) semillas aleatorias y b) semillas fijas. En la condición b), para un centroide utilizamos los valores correspondientes al percentil 25 de las doce subescalas del WISC-R y para el otro centroide los valores correspondientes al percentil 75.

Para entrenar la RNA de la figura 3, utilizamos el paquete de herramientas para la construcción de redes que incorpora el programa Matlab (v. 5.0) para una tasa de aprendizaje $\alpha = 0.1$ y 500 ciclos de aprendizaje.

Para las clasificaciones obtenidas representamos los centroides finales, calculamos las sumas de cuadrados intra-conglomerado y las distancias entre-conglomerados y cruzamos las partición obtenida con cada uno de los procedimientos con las variables de tipo médico, psicológicas y sociodemográficas descritas anteriormente.

Resultados

En la figura 4 hemos representado los centroides de los conglomerados obtenidos con el algoritmo de *k-medias* para los dos conjuntos de semillas y los obtenidos con la RNA. Los conglomerados se pueden etiquetar como Ainteligencia baja@ e Ainteligencia alta@. El perfil medio del conglomerado de Ainteligencia baja@ es similar al perfil de Ainteligencia alta@ independientemente del procedimiento de clasificación utilizado.

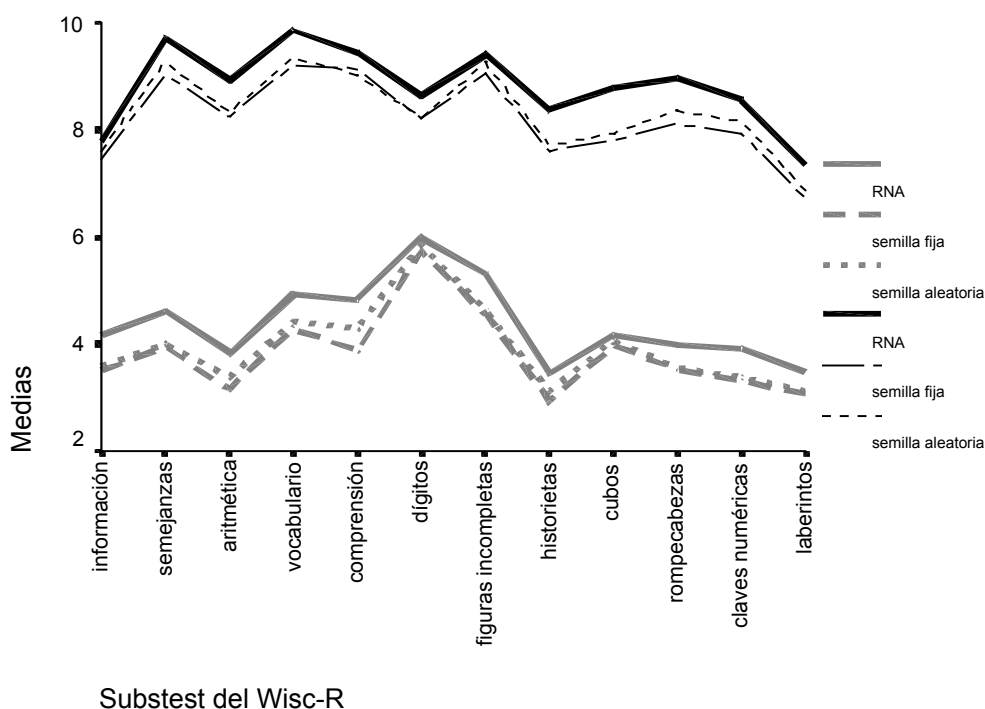


Figura 4. Perfiles medios de los conglomerados para los distintos procedimientos.

En cuanto al tamaño de los conglomerados si se observan diferencias. Sólo con el procedimiento de *k-medias* para la condición de semillas aleatorias se obtienen conglomerados de igual tamaño (34 sujetos). Cuando las semillas son fijas en el conglomerado de Ainteligencia baja@ se clasifican 32 sujetos y 36 en el conglomerado de Ainteligencia alta@. Las diferencias entre los tamaños de los conglomerados son mayores al clasificar con la RNA. En el conglomerado de Ainteligencia baja@ se clasifican 41 sujetos y 27 en el de Ainteligencia alta@. La tabla 1 muestra los acuerdos y desacuerdos en las clasificaciones realizadas por los dos procedimientos.

	K-medias(Semilla fija)		RNS	
	Baja	Alta	Baja	Alta
Baja	32	2	34	
K-medias(Semilla aleatoria)		34	7	27
Alta				
K-medias Baja			32	

(Semilla fija)	Alta	9	27
----------------	------	---	----

Tabla 1. Acuerdos y desacuerdos en la clasificación para las distintas condiciones.

De la tabla anterior cabe esperar que el conglomerado de Ainteligencia alta@ obtenido con la RNA sea más homogéneo que los de Ainteligencia alta@ obtenidos con el procedimiento de *k-medias* para los dos conjuntos de semillas. Para evaluar la homogeneidad de los conglomerados hemos calculado, utilizando la expresión propuesta por Ward en 1963 (cit. en Dillon y Goldstein, 1984), las sumas de cuadrados intra-conglomerado (ver tabla 2). De la tabla 2 no podría deducirse que valores mayores de homogeneidad estén vinculados a un procedimiento determinado. En este caso, el conglomerado de Ainteligencia alta@ obtenido con la RNA es más homogéneo que los obtenidos con el algoritmo de *k-medias* tanto para semillas aleatorias como fijas pero, por contra, el conglomerado de Ainteligencia baja@ obtenido con la RNA es más heterogéneo que los obtenidos con *k-medias*.

	SCintra
Inteligencia alta con semilla aleatoria	3346.54
Inteligencia baja con semilla aleatoria	2714.76
Inteligencia alta con semilla fija	3637.94
Inteligencia baja con semilla fija	2439.42
Inteligencia alta con RNA	2532.48
Inteligencia baja con RNA	3659.56

Tabla 2. Homogeneidad de los conglomerados a partir de la SC Intra-cluster.

Para evaluar la separación entre los conglomerados hemos calculado la distancia entre los centroides. La tabla 3 recoge estos los valores.

	Distancia Entre-clusters
k-medias con semilla aleatoria	15.44
k-medias con semilla fija	15.44
RNA	15.52

Tabla 3. Distancia entre-clusters.

El objetivo del análisis de conglomerados es conseguir grupos con la mínima variabilidad intra grupo como sea posible y la máxima variabilidad entre grupos. Una medida de variabilidad entre grupos o si se quiere de heterogeneidad es la distancia entre los centroides de los conglomerados. En este caso, aunque la separación entre los conglomerados obtenidos con la RNA es mayor que con el resto de los procedimientos estas diferencias son poco relevantes y pueden ser debidas a características concretas de la muestra.

En cuanto a los cruces con las variables de tipo médico, psicológicas y sociodemográficas obtuvimos los resultados representados en los diagramas de barras de la figuras 5, 6 y 7. Los perfiles de los conglomerados son similares para los distintos procedimientos salvo el conglomerado de Ainteligencia baja@ en el que la mayoría de los niños han tenido problemas de infecciones. En el resto la composición porcentual para los niveles de las distintas variables es similar. Desde un punto

de vista más sustantivo puede decirse que la presencia de válvulas con los mayores riesgos de infección que estas conllevan, los problemas de ansiedad asociados, probablemente, a la hospitalización prolongada, los retrasos en el desarrollo y los problemas escolares son los que hacen que la mayoría de los niños con estos problemas se clasifiquen en el conglomerados de Ainteligencia baja@.

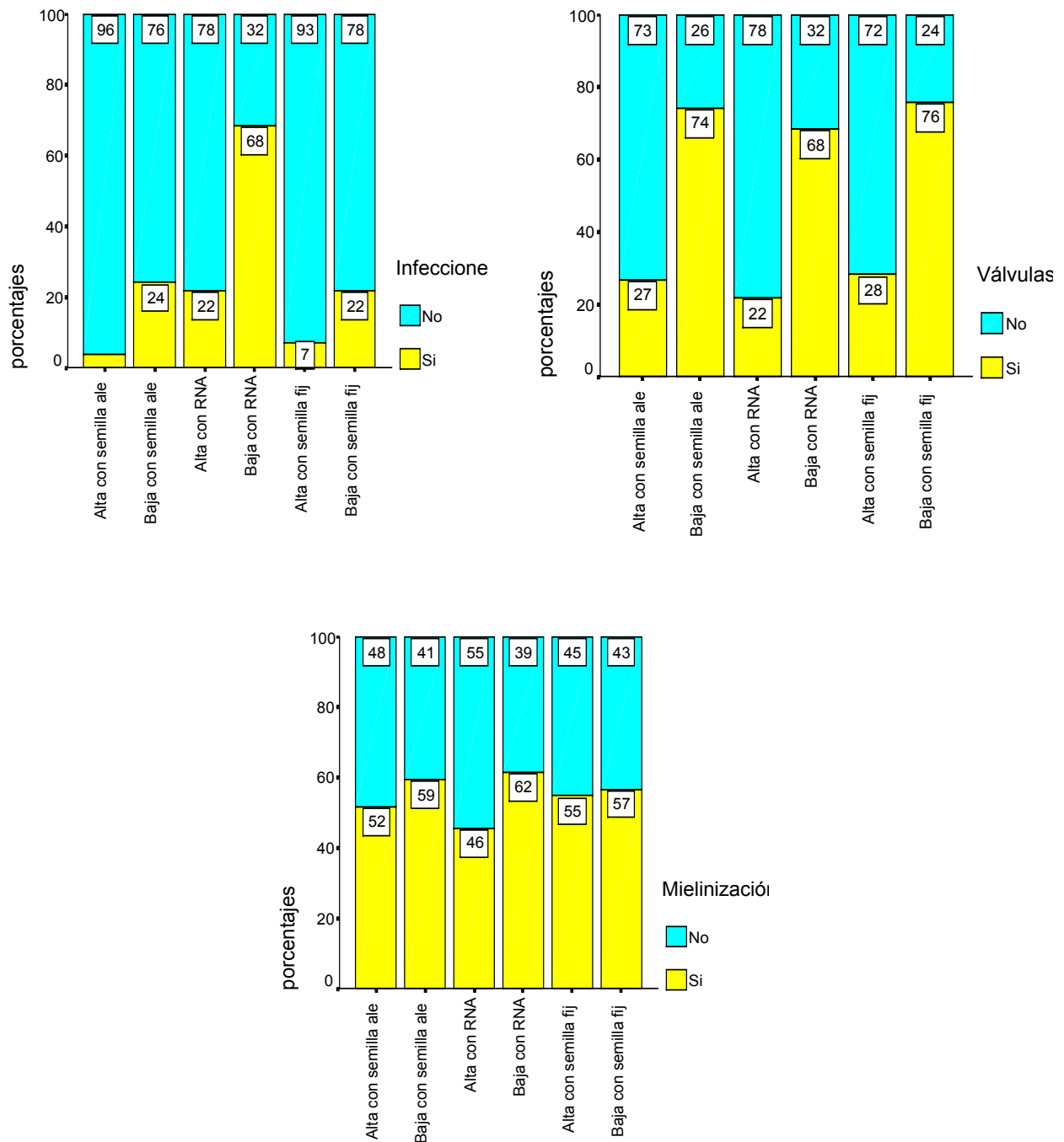


Figura 5. Perfiles de los conglomerados en función de los distintos procedimientos de clasificación para las variables de tipo médico.

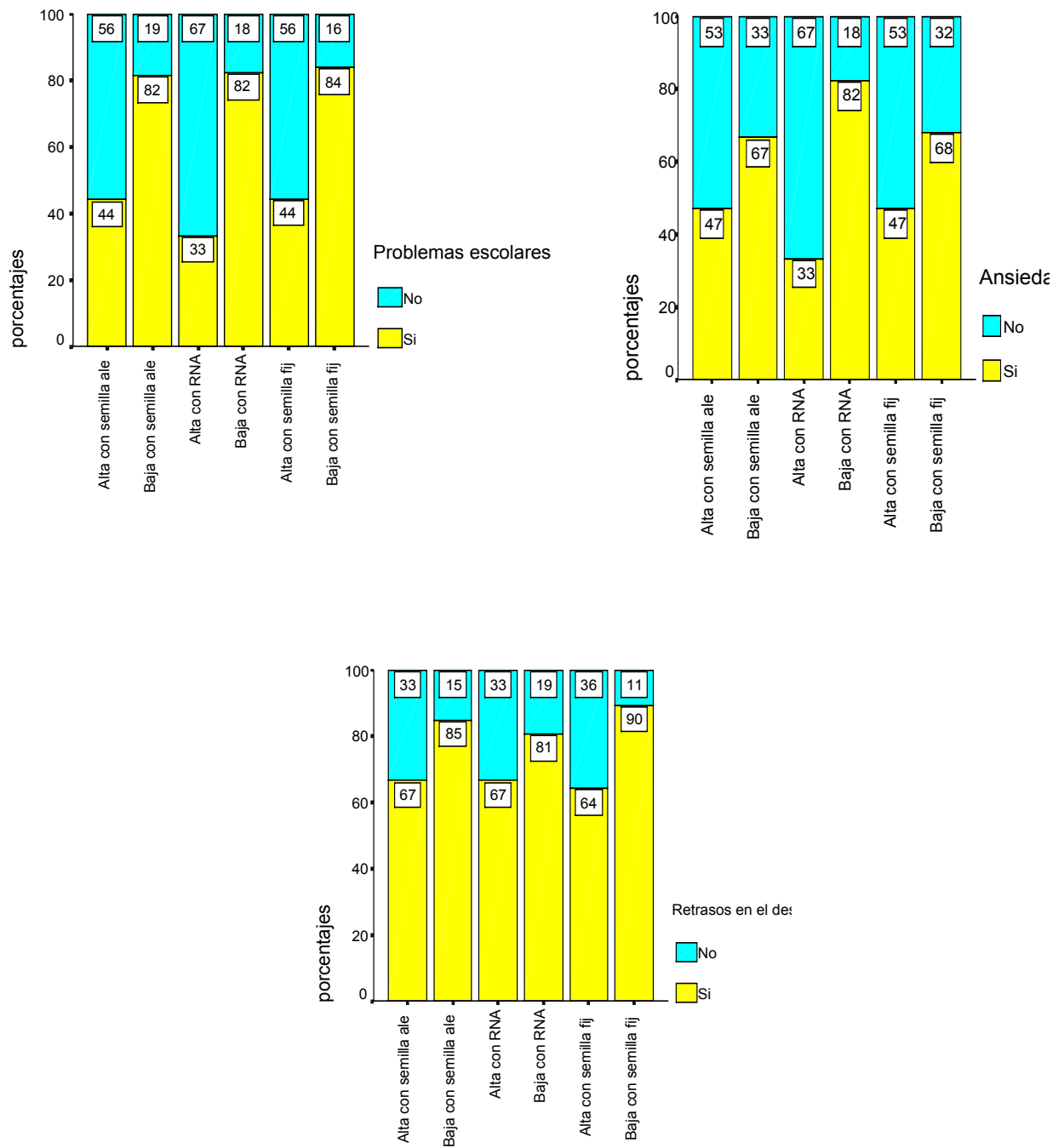


Figura 6. Perfiles de los conglomerados en función de los distintos procedimientos de clasificación para las variables psicológicas

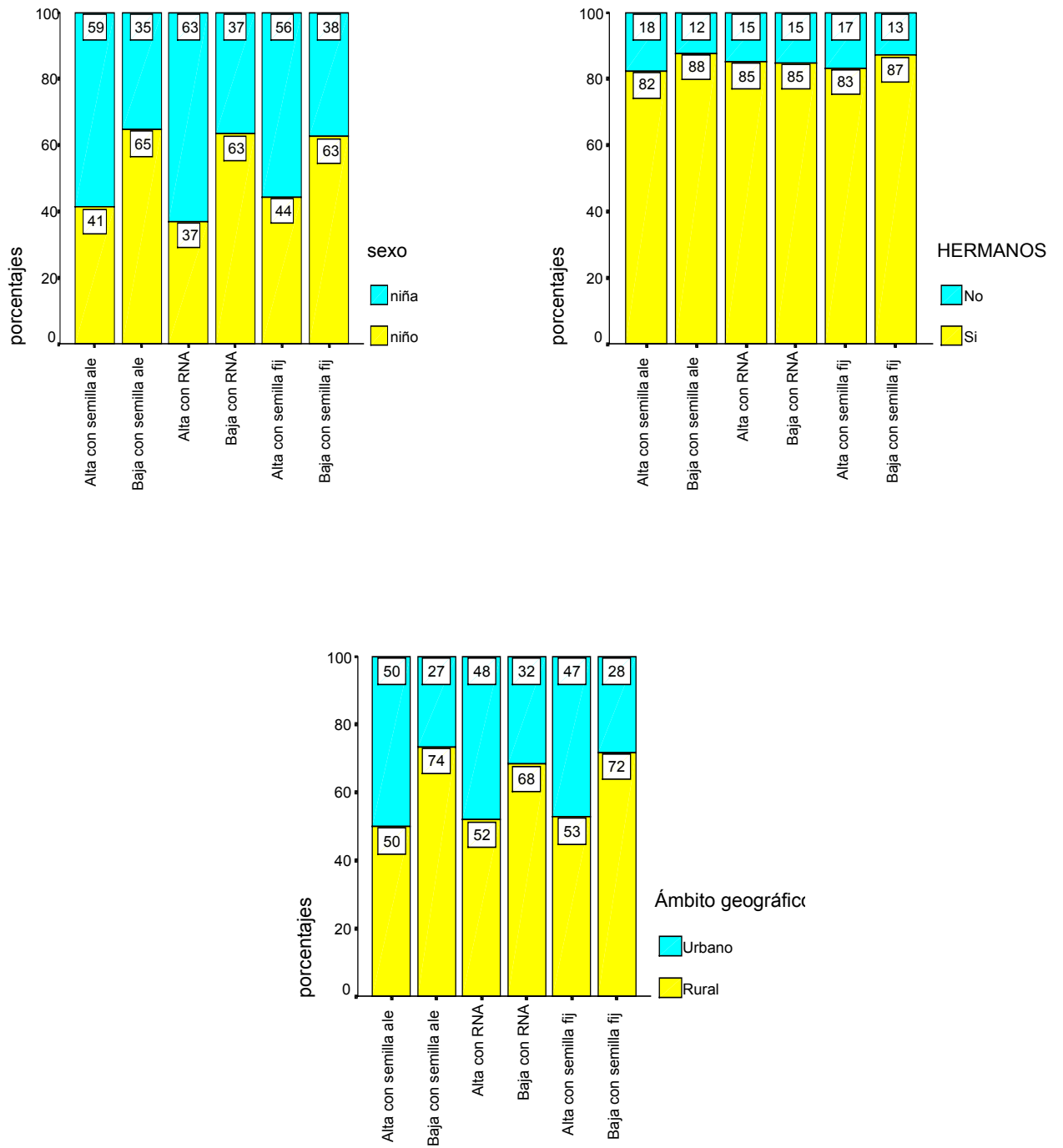


Figura 7. Perfiles de los conglomerados en función de los distintos procedimientos de clasificación para las variables sociodemográficas.

Conclusiones

El objetivo de este estudio fue comparar la ejecución del algoritmo de *k-medias* con selección aleatoria de las semillas y con selección fija frente a una RNA para la clasificación en dos grupos de una muestra de niños con espina bífida. Evaluamos las posibles diferencias en los perfiles medios, en la homogeneidad y en la separación entre los conglomerados obtenidos con los diferentes procedimientos. Comparamos, asimismo, los perfiles obtenidos bajo cada procedimiento utilizando una serie de variables externas a la clasificación. Los resultados obtenidos muestran que:

- 1) Los dos procedimientos proporcionan conglomerados similares en cuando a sus perfiles medios con pequeñas diferencias cuantitativas.
- 2) El algoritmo de *k-medias* proporciona conglomerados equilibrados sólo en la condición de selección aleatoria de las semillas. Para selección fija y con la RNA los tamaños de los conglomerados son distintos.
- 3) No puede decirse que alguno de los procedimientos favorezca la obtención de conglomerados más compactos. Si bien la suma de cuadrados del conglomerado que hemos etiquetado como de Ainteligencia alta@ obtenido con el algoritmo *k-medias* en la condición de semilla fija es superior al resto de los conglomerados de Ainteligencia alta@; la tendencia contraria ocurre para los conglomerados de inteligencia baja.
- 4) Los perfiles de los conglomerados, al cruzarlos con las variables externas, son similares. La única diferencia puede observarse con la variable presencia o ausencia de infecciones.

Las diferencias entre los índices utilizados en este trabajo, para evaluar la ejecución de los distintos procedimientos, no nos permiten afirmar que la ejecución de uno de ellos sea superior al resto. No obstante, los mayores requisitos computacionales de la RNA frente al algoritmo de *k-medias* no llevaría a recomendar este último. Somos conscientes de la necesidad de seguir en esta línea de comparación de procedimientos diseñando investigaciones que nos permitan evaluar los errores en la clasificación que puedan con los distintos procedimientos en función de los modelos de probabilidad que sigan los datos.

Referencias

- Aldenderfer, M.S., y Blashfield, R. (1984). *Cluster Analysis*. Beverly Hills: Sage.
- Anderberg, M. (1973). *Cluster analysis for applications*. Nueva York: Academic Press.
- Balakrishnan, P.V., Cooper, M. C., Jacob, V.S. y Lewis, P.A. (1994). A study of the classification capabilities of neural networks using unsupervised learning: A comparison with K-Means clustering. *Psychometrika*, 59, 509-525.
- Bishop, C. M. (1995). *Neural Networks for Pattern Recognition*. Oxford: Clarendon Press.
- Dillon, W.R., y Goldstein, M. (1984). *Multivariate analysis. Methods and applications*. Nueva York: Praeger.
- Everitt, B. (1993). *Cluster analysis (3rd ed.)*. Nueva York: Halsted Press.
- Funahashi, K. (1989). On the Aproximate Realization of Continuous Mappings by Neural Networks. *Neural Networks*, 2, 183-192.
- Hair, J.F., Anderson, R.E., Tatham, R.L. y Black, W.C. (1999). *Análisis Multivariante 50 ed*. Madrid: Prentice-Hall Iberia.
- Hartman, E.J., Keeler, J.D. y Kowalski, J.M. (1990). Layered Neural Networks with Gaussian Hidden Units as Universal Approximators. *Neural Computation*, 2, 210-215.
- Hilera, J.R, y Martínez, V.J. (1995). *Artificial Neural Networks*. Madrid: Ra-Ma.
- Hornik, K., Stinchcombe, M. y White, H. (1989). Multilayer Feedforward Networks are Universal Approximators. *Neural Networks*, 2, 359-366.

- Kohonen, T. (1984). *Self-organizing and Associative Memory. Vol 8*. Nueva York: Springer Verlag.
- Lebart, L, Morineau, A., y Fenelon, J.P. (1985). *Tratamiento estadístico de datos. Métodos y programas*. Barcelona: Marcombo.
- Matlab (v.5.0). The Math Works Inc. <http://www.mathworks.com>
- Milligan, G. W., y Sokol, L.M. (1980). A two-stage clustering algorithm with robust recovery characteristics. *Educational and Psychological Measurement*, 40, 755-759.
- Sellés, M. y Renom, J. (1998). Optimización de na batería de pruebas de lenguaje mediante una red neuronal artificial. En J. Renom (Ed.), *Tratamiento informatizado de datos*, 287-316. Barcelona: Masson.
- Ripley, B.D. (1996). *Pattern Recognition and Neural Networks*. Cambridge: University Press.
- Russell, S., y Norvig, P. (1996). *Inteligencia artificial. Un enfoque moderno*. México: Prentice-Hall Hispanoamericana.