

Psicológica (1998) 19, 387-400.

Redes neurales vs modelos estadísticos: Simulaciones sobre tareas de predicción y clasificación

Alfonso Pitarque*, Juan Francisco Roy y Juan Carlos Ruiz

Universitat de València

En este trabajo se comparan mediante simulación redes neuronales (del tipo perceptrón multicapa) con modelos estadísticos (regresión múltiple, análisis discriminante y regresión logística) en tareas de predicción y clasificación (binaria o no binaria), manipulando los patrones de correlación existentes entre los predictores (o variables de entrada) por un lado, y entre predictores con el criterio (variable de salida) por otro. Los resultados muestran que en tareas de predicción redes neuronales y modelos de regresión múltiple tienden a rendir por igual. Por contra en las tareas de clasificación en todo tipo de condiciones las redes neuronales rinden mejor que los modelos estadísticos de análisis discriminante y regresión logística. Se discuten los resultados en el marco de la polémica redes neuronales vs modelos estadísticos convencionales.

Palabras clave: redes neuronales, modelos estadísticos, predicción, clasificación.

En investigación aplicada es muy común encontrar situaciones en las que debemos estimar o predecir el comportamiento de una variable criterio en función de una o varias variables predictoras. Cuando el criterio es una variable cuantitativa se suele hablar de *problemas de predicción* o estimación, mientras que cuando es una variable cualitativa/categorial se habla entonces de *problemas de clasificación*.

Tradicionalmente la solución a estos problemas se ha llevado a cabo desde la óptica de modelos estadísticos de regresión: regresión simple o múltiple para

* Los autores quieren expresar su gratitud a dos revisores del trabajo por sus sugerencias. Correspondencia: Alfonso Pitarque; Area de Metodología; Facultad de Psicología; Avda. Blasco Ibáñez 21; 46010 Valencia; E-mail: pitarque@uv.es

problemas de predicción; análisis discriminante o modelos de regresión logística para problemas de clasificación. Casi todas estas técnicas descansan en el llamado *modelo de regresión general lineal*, pudiendo ser conceptualizadas a su vez como casos particulares del análisis de correlación canónica (ver p.e. Knapp, 1978; Allison, Gorman y Primavera, 1993).

Este tipo de modelos estadísticos han sido suficientemente analizados en la literatura (ver p.e. Cohen y Cohen, 1983; Tabachnick y Fidell, 1989; Tatsuoka, 1989). Cuando los datos satisfagan a priori las condiciones de aplicabilidad de estas técnicas (normalidad e igualdad de varianzas; independencia; linealidad; normalidad de los residuos; tamaño muestral suficiente,...), el rendimiento dependerá del patrón de correlaciones que mantengan los predictores entre sí, y cada predictor con el criterio: será óptimo cuando existan altas correlaciones de los predictores con el criterio, pero bajas correlaciones entre los predictores. Por contra, el rendimiento será peor a medida que aquellas correlaciones descendan, y/o que éstas últimas aumenten (colinealidad).

A partir de 1986 (Rumelhart y McClelland, 1986; McClelland y Rumelhart, 1986) las llamadas redes neurales o modelos conexionistas han ido progresivamente utilizándose como herramientas de predicción y clasificación. De forma breve una red neural es un sistema informático reticular (de inspiración neuronal) que aprende de la experiencia mediante la auto-modificación de sus conexiones (ver p.e. para revisiones Heccht-Nielsen, 1990; Hertz, Krogh y Palmer, 1991; Wasserman, 1989, entre otros muchos buenos manuales; y en castellano Hilera y Martínez, 1995; Martín y Sanz, 1997).

Como herramientas de predicción/clasificación, muchos autores están estudiando las relaciones entre las técnicas estadísticas convencionales y los modelos conexionistas (ver p.e. Cherkassky, Friedman y Wechler, 1994; Flexer, 1995; Michie, Spiegelhalter y Taylor, 1994; Ripley, 1996; Sarle, 1994, 1998). De hecho las redes neurales han sido conceptualizadas como técnicas estadísticas no paramétricas al estar libres del cumplimiento de los supuestos teóricos de la estadística paramétrica (ver p.e. Smith, 1993), o también como técnicas de regresión no lineal (ver p.e. Sarle, 1998). El problema surge cuando encontramos resultados contradictorios a la hora de determinar qué modelos son más eficientes en la solución de problemas concretos de predicción/clasificación. Así mientras algunos trabajos empíricos no encuentran diferencias entre los resultados hallados por unos y otros modelos (Croall y Mason, 1992; Michie et al, 1994; Ripley, 1993; Thrun, Mitchell y Cheng, 1991), otros resultados tienden a apoyar una ligera superioridad de las redes neurales sobre las técnicas estadísticas (ver p.e. Garson, 1991; Huang y Lippman, 1987; White, 1994).

Los estadísticos explican estas discrepancias empíricas aduciendo una incorrecta aplicación de las técnicas estadísticas: análisis inadecuado de los supuestos teóricos en los que se basan (homogeneidad de la matriz de covarianzas, normalidad,...), utilización de matrices de datos sesgadas, ausencia de pre-procesamiento de los datos (outliers, missing data,...) y transformación de variables, etc. Por su parte los expertos en redes aducen que pese a que las redes neurales a priori son capaces de asociar cualquier patrón de entrada con cualquier patrón de salida, su rendimiento depende del ajuste heurístico de numerosos parámetros (número de unidades de entrada, salida y ocultas, si procede; funciones de activación: lineal, sigmoideal, tangencial,...; regla de aprendizaje: Hebb, delta, retropropagación,...; coeficientes de aprendizaje y momentum, etc.), ajuste que no siempre garantiza la solución deseada, dada además la estructura de "caja negra" (Cherkassky, Friedman y Wechler, 1994) de este tipo de modelos.

Nuestro objetivo será contrastar sistemáticamente el rendimiento de redes neurales (en concreto redes multi-capas de propagación hacia atrás, también llamadas redes de perceptrón multi-capas; ver Rumelhart, Hinton y Williams, 1986; Rumelhart, Durbin, Golden y Chauvin, 1995) con modelos estadísticos convencionales (regresión múltiple, regresión logística y análisis discriminante) bien en problemas de predicción de una variable criterio cuantitativa, bien en problemas de clasificación de una variable cualitativa (de dos o tres categorías de clasificación), previa verificación de los supuestos teóricos en los que se basan las técnicas estadísticas.

Dado que los resultados pueden ser además modulados por el patrón de correlaciones que mantengan las variables predictoras (o variables de entrada en terminología neural) con el criterio (variable de salida), manipularemos sistemáticamente tales correlaciones desde los casos a priori más favorables a los modelos de regresión (altas correlaciones predictores-criterio, pero bajas correlaciones entre predictores) hasta los casos más comunes en investigación aplicada.

De este modo esperamos obtener información precisa sobre la controversia planteada en relación a qué modelos (estadísticos o neurales) son más eficientes en la solución de problemas de predicción/clasificación de patrones.

MÉTODO

Materiales y procedimiento Se generaron mediante simulación 300 ficheros de datos de 150 filas (por sujetos) y 8 columnas (por 8 variables: 7 variables predictoras más un criterio para los modelos de regresión; 7 variables de entrada más una, dos o tres de salida para las redes neurales, en función de la

condición; ver más abajo). Los datos generados eran puntuaciones típicas (z) correspondientes a muestras aleatorias de variables de distribución normal multivariante correlacionadas de 4 formas distintas en función de otras tantas condiciones experimentales:

- (a) en la condición baja-alta (desde ahora condición A), las 7 variables predictoras correlacionaban de forma baja entre sí (r_{XY} promedio=.05), pero cada una de ellas correlacionaba significativamente ($r = 0.05$) con el criterio (r_{XY} promedio=.35). Esta era sin duda la condición a priori más favorable a los modelos estadísticos convencionales.
- (b) en la condición alta-baja (desde ahora condición B), por el contrario los predictores correlacionaban significativamente entre sí ($r = 0.05$; r_{XY} promedio=.35), pero de forma baja con el criterio (r_{XY} promedio=.05).
- (c) en la condición baja-baja (desde ahora condición C), los predictores correlacionaban de forma baja entre sí (r_{XY} promedio=.05), así como con el criterio (r_{XY} promedio=.05).
- (d) en la condición alta-alta (desde ahora condición D), las 7 variables predictoras correlacionaban significativamente entre sí ($r = 0.05$; r_{XY} promedio=.35), así como con el criterio ($r = 0.05$; r_{XY} promedio=.35).

A su vez sobre estos 4 conjuntos de ficheros se hicieron distintas transformaciones en función del tipo de tarea a realizar (predicción o clasificación) y el tipo de herramienta utilizada (modelos de regresión o redes neurales) tal y como se explica a continuación:

- En tareas de predicción a través de modelos de regresión múltiple los anteriores ficheros no sufrieron modificación alguna.
- En tareas de predicción a través de redes neurales las variables predictoras (variables de entrada) fueron acotadas, mediante transformación lineal, en el rango $[-1,1]$, mientras que el criterio se acotó en el rango $[0,1]$, sin que las correlaciones mencionadas se vieran, lógicamente, afectadas.

En tareas de clasificación distinguimos a su vez entre variables categoriales dicotómicas y variables categoriales de tres categorías de agrupamiento (como representativas de las variables cualitativas no dicotómicas).

- En tareas de clasificación de variables dicotómicas a través de modelos estadísticos (análisis discriminante y regresión logística) la variable criterio de los mencionados ficheros se transformó en 0 (para valores originales

negativos) ó 1 (para valores positivos) para representar los dos niveles de clasificación de tal variable. Se obtuvo así una variable binaria procedente de una distribución binomial con $p=0.50$. Por su parte la homogeneidad de la matriz de varianzas-covarianzas de cada uno de los grupos de clasificación fue puesta a prueba por medio del test M de Box (ver SPSS, 1993a).

- En tareas de clasificación de variables dicotómicas a través de redes neurales la variable criterio se desdobló en dos variables de salida, codificando con (0,1) los valores originales negativos y con (1,0) los valores originales positivos. Esto es así porque para representar en redes neurales variables cualitativas de k categorías se necesitan como mínimo k-1 vectores de salida (ver Smith, 1993). Así mismo, las 7 variables de entrada se acotaron mediante transformación lineal en el rango ± 1 . De este modo los ficheros originales se transformaron para esta condición en ficheros de 150 filas por 9 columnas (7 variables de entrada y 2 unidades de salida).

- En tareas de clasificación de variables cualitativas de tres categorías a través de modelos de regresión (análisis discriminante) la variable criterio de los mencionados ficheros se transformó en 0 (para valores inferiores o iguales al centil 33 de dicha variable), 1 (para valores comprendidos entre los centiles 33 y 67) o 2 (para valores iguales o superiores al centil 67), consiguiendo así una variable procedente de una distribución multinomial con $p_1 = p_2 = p_3 = 1/3$. La homogeneidad de la matriz de varianzas-covarianzas de cada uno de los grupos de clasificación fue así mismo puesta a prueba por medio del test M de Box (ver SPSS, 1993a).

- Por su parte en tareas de clasificación de variables cualitativas de tres categorías a través de redes neurales la variable criterio de los mencionados ficheros se desdobló en tres variables de salida: codificando con (0,0,1) los valores inferiores o iguales al centil 33 de dicha variable, con (0,1,0) los valores comprendidos entre los centiles 33 y 67, y con (1,0,0) los valores iguales o superiores al centil 67. Así mismo, las 7 variables de entrada se acotaron en el rango ± 1 . De este modo los ficheros originales se transformaron para esta condición en ficheros de 150 filas por 10 columnas (7 variables de entrada y 3 unidades de salida).

En resumen, el objetivo de este trabajo fue confrontar modelos de regresión estadística contra redes neurales en problemas de predicción de una variable cuantitativa y en problemas de clasificación, bien de una variable cualitativa dicotómica, bien de una variable cualitativa de tres niveles de

clasificación. Los modelos estadísticos utilizados fueron: regresión múltiple para la predicción de una variable cuantitativa; análisis discriminante y modelos de regresión logística para problemas de clasificación binaria; y análisis discriminante para problemas de clasificación no binaria.

Por su parte se utilizaron redes neurales de estructura 7/25/1 (7 unidades de entrada; 25 unidades ocultas; 1 unidad de salida) para la predicción de una variable cuantitativa; redes de estructura 7/25/2 para la clasificación de una variable cualitativa binaria; y redes de estructura 7/25/3 para la clasificación de variables cualitativas de tres niveles, por las razones expuestas más arriba. El número de predictores así como el tipo de arquitectura neural elegidos coinciden básicamente con lo observado en otros trabajos (ver p.e. Wilson y Hardgrave, 1995).

Todas las redes neurales empleadas eran redes heteroasociativas que aprendían mediante el algoritmo de retro-propagación (también llamada regla delta; ver Rumelhart et al 1986; 1995) durante 25000 ciclos de aprendizaje. La función de activación utilizada fue la sigmoidea. Los coeficientes de aprendizaje fueron, respectivamente, de .30 y .15 para las conexiones de las unidades de entrada-unidades ocultas, y unidades ocultas-unidades de salida, mientras que el coeficiente *momentum* era de .40. La elección de estos parámetros fue llevada a cabo de forma heurística a través de ensayos pilotos en simulaciones previas. Los pesos de las conexiones de cada capa se aleatorizan en el rango ± 1 antes de cada entrenamiento.

Las variables dependientes que medimos fueron el coeficiente de determinación para problemas de predicción y la proporción de clasificaciones correctas para problemas de clasificación binaria y no binaria. Recordemos que el primero se define como

$$R^2 = \frac{MC_{reg}}{MC_{total}} = r_{yy'}^2$$

es decir, la proporción de varianza total explicada por el modelo (de regresión o red neural, dependiendo del caso), o también el coeficiente de correlación de Pearson elevado al cuadrado entre las puntuaciones del criterio y las predichas por los modelos (de regresión o red neural, dependiendo del caso).

En las redes neurales se consideraba que una clasificación era correcta cuando la respuesta de la unidad de salida adecuada era la mayor y además excedía la probabilidad de acertar por puro azar (.50 para clasificaciones binarias; .33 para clasificaciones de 3 categorías; ver p.e. Smith, 1993; Wilson y Hardgrave, 1995).

RESULTADOS

Analizamos por separado los datos correspondientes a la tarea de predicción cuantitativa (regresión múltiple vs redes neurales de 1 unidad de salida), clasificación cualitativa binaria (análisis discriminante y regresión logística vs redes neurales de 2 unidades de salida) y clasificación cualitativa no binaria (análisis discriminante vs redes neurales de 3 unidades de salida).

Los resultados correspondientes a la tarea de predicción cuantitativa aparecen en la tabla 1 y figura 1.

Globalmente considerados estos resultados apuntan a la idea de que redes neurales y modelos de regresión múltiple tienden a rendir por igual en todas las condiciones (A, B, C, D), lo que estaría de acuerdo con lo hallado por otros laboratorios (ver p.e. Ripley, 1993). Analizando el efecto de las distintas condiciones de relacionalidad se observan los mejores resultados en la condición A, seguida de la condición D, y por último en las condiciones B y C, indistintamente. Estos resultados muestran que las distintas condiciones de relacionalidad tienden a afectar de forma similar a modelos estadísticos y redes neurales, siendo el patrón de correlaciones predictores-criterio el que determina la calidad del rendimiento de ambos sistemas.

Los resultados correspondientes a la tarea de clasificación cualitativa binaria (proporción de clasificaciones correctas) aparecen en la tabla 2 y figura 2.

Tabla 1. Medias, los coeficientes de determinación, desviaciones típicas (D.T.) e intervalos de confianza (al 95%) de medias (I.C.) para las distintas condiciones (A, B, C, D) de la tarea de predicción cuantitativa del criterio.

	REDES NEURALES				REGRESION MULTIPLE			
	A	B	C	D	A	B	C	D
MEDIA	.83	.07	.07	.41	.84	.07	.07	.41
D.T.	.02	.04	.04	.06	.02	.04	.04	.06
I.C.	.82; .84	.05; .08	.05; .08	.39; .44	.83; .84	.05; .08	.05; .08	.39; .44

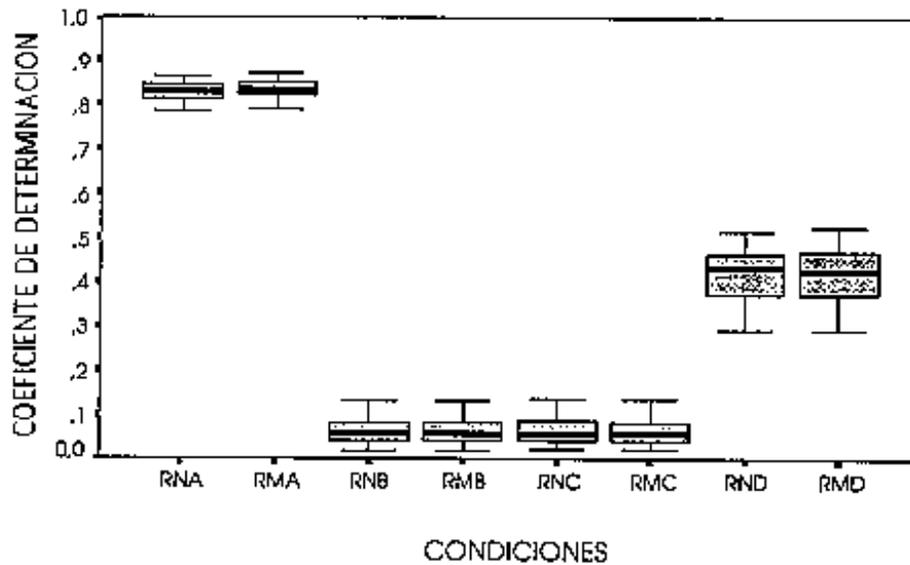


Figura 1. Diagrama de cajas (plot) de las distintas condiciones de predicción cuantitativa del criterio (donde las letras iniciales RN y RM denotan, respectivamente, las abreviaturas de 'red' y 'regresión múltiple', mientras que las letras finales A, B, C, D, representan las condiciones de relacionalidad referidas en el texto).

Tabla 2. Proporciones medias de clasificaciones correctas, desviaciones típicas (D.T.) e intervalos de confianza (al 95%) de medias (I.C.) para las distintas condiciones (A, B, C, D) de la clasificación binaria.

	REDES NEURALES				ANALISIS DISCRIMINANTE			
	A	B	C	D	A	B	C	D
MEDIA	.90	.65	.64	.75	.88	.59	.60	.73
D.T.	.02	.05	.07	.04	.02	.04	.04	.03
I.C.	.89; .91	.62; .67	.61; .66	.74; .77	.87; .88	.57; .61	.59; .62	.71; .74

REGRESION LOGISTICA				
	A	B	C	D
MEDIA	.88	.60	.61	.73
D.T.	.02	.03	.03	.04
I.C.	.87; .89	.59; .62	.59; .62	.71; .74

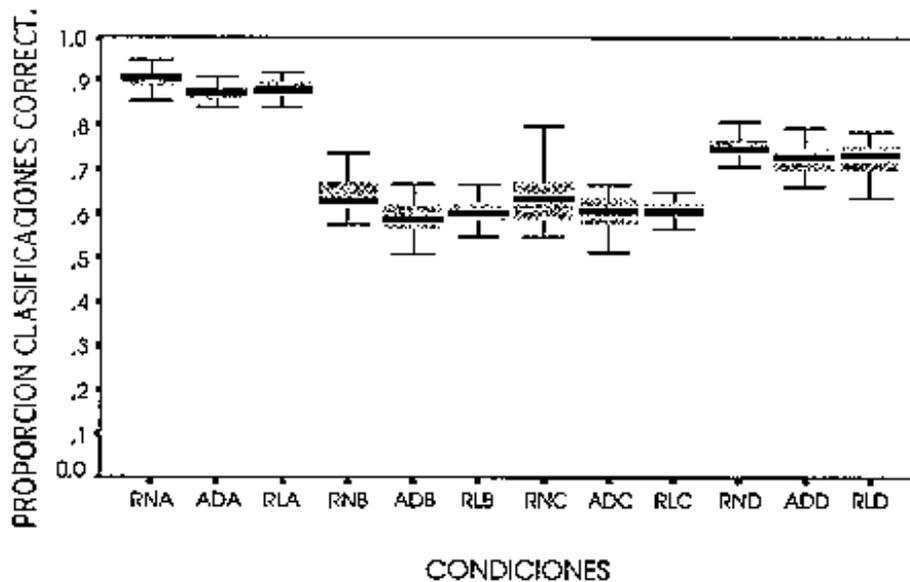


Figura 2. Diagrama de cajas de las distintas condiciones de clasificación binaria (donde las dos letras iniciales RN, ADA y RL

respectivamente, las abreviaturas de ‘red neuronal discriminante’ y ‘regresión logística’, mientras que las letras finales D, R, representan las condiciones de relacionalidad referidas en el texto).

Globalmente considerados estos resultados muestran que las redes neurales tienden a producir una proporción de clasificaciones correctas superior a la obtenida por los modelos estadísticos (sin que existan diferencias entre regresión logística y análisis discriminante), e independientemente del patrón de correlaciones que mantengan las variables. Esta afirmación se vio corroborada por análisis estadísticos convencionales ($p < 0.01$). Por otra parte, analizando el efecto de las distintas condiciones de relacionalidad se observa la misma tendencia de resultados que la observada en la tarea de predicción.

Por último los resultados correspondientes a la tarea de clasificación cualitativa no binaria (proporción de clasificaciones correctas) aparecen en la tabla 3 y figura 3.

Al igual que en el análisis anterior (modelos de clasificación binaria) estos resultados muestran que las redes neuronales consiguen una proporción de clasificaciones correctas significativamente superior a la obtenida por la técnica de análisis discriminante, indistintamente del patrón de correlaciones que mantengan las variables. Esta afirmación se vio así mismo corroborada por análisis estadísticos ($p < 0.01$).

DISCUSIÓN

De la consideración global de los resultados anteriores se desprenden una serie de conclusiones:

Únicamente en la tarea de predicción cuantitativa y bajo condiciones idóneas de aplicabilidad (condición A; condiciones que por otro lado pocas veces satisfacen las datos procedentes de investigaciones aplicadas), el procedimiento clásico de regresión obtuvo mejores resultados (del orden de un punto porcentual) que las redes neurales. En el resto de condiciones redes neurales y modelos de regresión múltiple rinden por igual en este tipo de tarea. Estos resultados coinciden básicamente con los hallados en otros trabajos como el de Wilson y Hardgrave (1995) quienes no encontraron diferencias entre redes neurales y modelos de regresión en una tarea de predicción del éxito académico de estudiantes. Esto sugeriría que un investigador debería realizar un estudio previo del patrón de correlaciones que sustienden las variables predictores y criterio antes de decidirse por utilizar un modelo de regresión o una red neural:

sólo en el caso reflejado en la condición A podría ser aconsejable utilizar modelos de regresión, pudiendo utilizar indistintamente redes neurales y técnicas de regresión en todos los demás casos.

Sin embargo en todo tipo de tareas de clasificación (binaria o no) las redes neurales obtuvieron resultados mejores en todas las condiciones que los modelos estadísticos convencionales (bien análisis discriminante, bien regresión logística). La superioridad de las redes queda especialmente de manifiesto en la condición A, condición que satisface de forma idónea las condiciones de aplicabilidad de las técnicas estadísticas convencionales. En

Tabla 3. Proporciones medias de clasificaciones correctas, desviaciones típicas (D.T.) e intervalos de confianza (al 95%) de medias (I.C.) para las distintas condiciones (A, B, C, D) de la tarea de clasificación no binaria.

	REDES NEURALES				ANALISIS DISCRIMINANTE			
	A	B	C	D	A	B	C	D
MEDIA	.79	.48	.47	.59	.78	.45	.45	.57
D.T.	.04	.04	.05	.04	.04	.04	.04	.04
I.C.	.77; .81	.46; .49	.45; .48	.57; .61	.76; .79	.44; .47	.43; .47	.56; .59

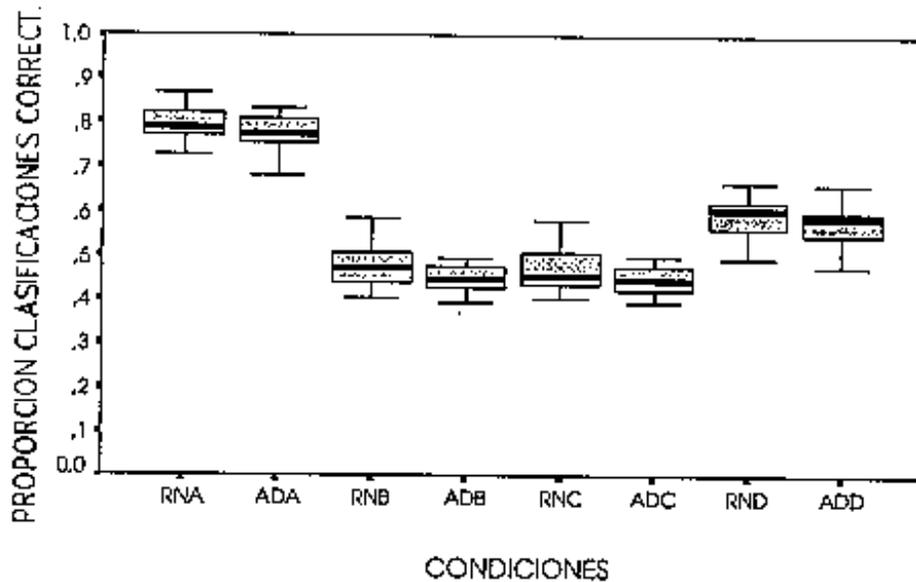


Figura 3. Diagrama de cajas de las distintas condiciones de clasificación no binaria (donde las letras iniciales RN y AD denotan, respectivamente, las abreviaturas de 'red y análisis discriminante', mientras que las letras finales A, B, C y D representan condiciones de relacionalidad referidas en el texto).

este sentido nuestros resultados coinciden con los hallados por otros autores como Fogelman (1994) quien obtuvo así mismo mejores resultados utilizando redes neurales que modelos ARIMA en la predicción de series temporales. De igual modo Navarro (1998) ha demostrado recientemente cómo la estimación de valores faltantes (*missing values*) a través de redes neurales suele ser mejor que la conseguida por procedimientos estadísticos convencionales.

Este patrón de resultados favorece claramente a las redes neurales sobre los modelos estadísticos clásicos como técnicas de clasificación, afirmación que se ve fortalecida por su mayor versatilidad de uso, al no depender su aplicabilidad del cumplimiento de los supuestos teóricos sobre los que se basan las técnicas estadísticas (normalidad, homoscedasticidad, independencia,...). Otra ventaja adicional (importante en investigación aplicada) de las redes neurales sobre los modelos estadísticos es que admiten como variables de entrada conjuntos mixtos de variables cuantitativas y cualitativas.

Un aspecto menor que llama la atención en relación a la tarea de clasificación binaria son los resultados similares obtenidos a través de las dos técnicas estadísticas empleadas (análisis discriminante y regresión logística), lo que contrasta con los resultados obtenidos por Press y Wilson (1978) para quienes la regresión logística es una mejor herramienta de categorización que el análisis discriminante, al no depender del cumplimiento de los supuestos paramétricos de la estimación mínimo-cuadrática.

De todo lo anterior no se desprende que propugnemos el destierro de los métodos estadísticos convencionales a la hora de realizar tareas de clasificación, pues las redes neurales, pese a su mejor rendimiento, presentan una serie de inconvenientes que el investigador debe sopesar antes de decidirse por su utilización. En primer lugar, el entrenamiento de una red neural es un proceso *demasiado creativo* (Martín y Sanz, 1997) que generalmente se soluciona por un método heurístico de ensayo y error. Además, y esto es un aspecto especialmente delicado, la calidad de las soluciones dadas por la red elegida no puede ser siempre garantizada (Cherkassky et al, 1994) debido a su naturaleza de "caja negra" y otras causas: interferencia catastrófica, sobreaprendizaje, mínimos locales, etc. No hay que olvidar tampoco que una red neural no da información explícita sobre la importancia relativa de los distintos predictores. Ni, por último, obviar el elevado costo computacional requerido en el entrenamiento de las redes neurales, muy superior al de los modelos estadísticos. En última instancia deberá ser pues el investigador quien, sopesando tales limitaciones, decida si compensa decidirse por la utilización de una arquitectura de uno u otro tipo.

La consideración de todo lo dicho nos lleva a sugerir, con Fogelman (1994), Cherkassky et al (1994), Sarle (1994, 1998) y otros muchos autores, que técnicas estadísticas y redes neurales deben comenzar a ser usadas conjuntamente, tal y como comienzan a utilizarse en determinadas aplicaciones técnicas (ver Martín y Sanz, 1997). De este modo, la estadística, centrada tradicionalmente en funciones lineales, y las redes neurales, más acostumbradas a tratar con problemas mal definidos o no lineales (Smith, 1993), se verán mutuamente enriquecidas.

ABSTRACT

Neural networks vs statistical models in prediction and classification tasks
Neural networks (multi-layer perceptrons) and statistical models (multiple regression, discriminant analysis and logistic regression) are compared in both prediction and classification tasks. The degree of correlation within predictors variables, and between predictors with the

criterion is also manipulated. Results show no differences between neural nets and regression models in prediction tasks. However in classification tasks neural networks yield better results than discriminant analysis and logistic regression along all the conditions. Results are discussed in terms of the neural networks vs statistical models controversy.

Key wordsneural networks, statistical models, prediction, classification.

REFERENCIAS

- Allison, D.B., Gorman, B. y Primavera, L.H. (1993). Some of the most common questions asked of statistical consultants: Favorite responses and recommended readings. *Genetic, Social, and General Psychology Monographs*, 119 , 153-185.
- Chater, N. (1995). Neural networks: The new statistical models of mind. En Levy, J.P. y Bairaktaris, D. (eds.), *Connectionist models of memory and language*. Ucl Press. London.
- Cherkassky, V., Friedman, J.H. y Wechsler, H. (eds.) (1994). *From statistics to neural networks*. Springer- Verlag. Berlin.
- Cohen, J. y Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.
- Croall, I.F. y Mason, J.P. (1992). *Industrial applications of neural networks*. Springer-Verlag. N.Y.
- Flexer, A. (1995). *Connectionist and statisticians, friends or foes?*. The Austrian Research Institute for Artificial Intelligence. Acceso FTP. Servidor: ai.univie.ac.at.
- Fogelman, F. (1994). Neural networks architectures for pattern recognition. En Cherkassky, V., Friedman, J.H. y Wechsler, H. (eds.), *From statistics to neural networks*. Springer-Verlag. Berlin.
- Garson, G.D. (1991). A camparison of neural network and expert systems algorithms with common multivariate procedures for analysis of social science data. *Social Science Computer Review*, 9, 399-434.
- Hastie y Tibshirani (1994). Nonparametric Regression and Classification. Part 1- Nonparametric Regresion. En Cherkassky, V., Friedman, J.H. y Wechsler, H. (eds.), *From statistics to neural networks*. Springer- Verlag. Berlin.
- Hectht-Nielsen, R. (1990). *Neurocomputing*. Addison-Wesley. Cal.
- Hertz, J. Krogh, A. y Palmer, R. (1991). *Introduction tom the theory of neural computation*. Addison-Wesley. Cal.
- Hilera, J.R. y Martínez, V.J. (1995). *Redes neuronales artificiales: Fundamentos, modelos y aplicaciones*. Ra-ma. Madrid.
- Huang, W.Y. y Lippmann, R.P. (1987). Comparisons between neural nets and conventional clasifiers. *Proceedings of the IEEE International Conference on Neural Networks*, I, 485-494.

- Knapp, T.R. (1978). Canonical correlation analysis: A general parametric significance testing system. *Psychological Bulletin*, 85, 410-416.
- Martín, B. y Sanz, A. (1997). *Redes neuronales y sistemas borrosos*. Ra-ma. Madrid.
- McClelland, J.L. y Rumelhart, D.E., (eds) (1986). *Parallel distributed processing*. Vol. 2. Cambridge, Mass.: MIT Press.
- Michie, D., Spiegelhalter, D.J. y Taylor, C.C. (eds) (1994). *Machine learning, neural and statistical classification*. Londres: Ellis Horwood.
- Navarro, J.B. (1998). *Aplicación de redes neuronales artificiales al tratamiento de datos incompletos*. Tesis doctoral no publicada. Universitat Autònoma de Barcelona.
- Press, S.J. y Wilson, S. (1978). Choosing between logistic regression and discriminant analysis. *Journal of the American Statistical Association*, 73, 699-705.
- Ripley, B.D. (1993). Statistical aspects of neural networks. En O.E. Barndorff-Nielsen, J.L. Jensen y W.S. Kendall (eds), *Networks and chaos: Statistical and probabilistic aspects*. Londres: Chapman and Hall.
- Ripley, B.D. (1996). *Pattern recognition and neural networks*. Cambridge Univ. Press. Cambridge, G.B.
- Rosen, B. E. (1996). Ensemble learning using decorrelated neural networks. *Connection Science: Journal of Neural Computing, Artificial Intelligence and Cognitive Research*, 8, 373-383.
- Rumelhart, D.E., y McClelland, J.L. (eds) (1986). *Parallel distributed processing*. Vol. 1. Cambridge, Mass.: MIT Press.
- Rumelhart, D.E., Hinton, G.E. y Williams, R.J. (1986). Learning internal representations by error propagation. En Rumelhart, D.E., McClelland, J.L. and the PDP Research Group (eds.) (1986). *Parallel distributed processing: Explorations in the microstructure of cognition*. Cambridge, MA: MIT Press.
- Rumelhart, D.E., Durbin, R., Golden, R. y Chauvin, Y. (1995). Backpropagation: The basic theory. En Y. Chauvin y D.E. Rumelhart (eds.), *Backpropagation: Theory, architectures, and applications*. LEA: Hillsdale, N.J.
- Sarle, W.S. (1994). *Neural networks and statistical models*. Proceedings of the 19th Annual SAS Group conference, Cary, NC. pps. 1538-1550.
- Sarle, W.S. (1998). *Neural network FAQ*. Acceso FTP, parts 1 to 7. Acceso FTP: <ftp://ftp.sas.com/pub/neural/FAQ.html>
- Singh, S. (1997). Quantitative classification of conversational language using artificial neural networks. *Aphasiology*. 11(9): 829-844.
- Smith, M. (1993). *Neural networks for statistical modelling*. Van Nostrand Reinhold. N.Y.
- SPSS Inc (1993a). *SPSS for Windows. Professional Statistics (6.0)*. Chicago. Ill.
- SPSS Inc (1993b). *SPSS for Windows. Advanced Statistics (6.0)*. Chicago. Ill.
- Tabachnick, B.G. y Fidell, L.S. (1989). *Using multivariate statistics*. N.Y.: Harper & Row.
- Tatsuoka, M.M. (1989). *Multivariate analysis (2a ed.)*. Macmillan: N.Y.

- Thrun, S., Mitchell, T. y Cheng, J. (1991). The MONK's comparison of learning algorithms. Introduction and survey. En S. Thrun, J. Bala, E. Bloedorn e I. Bratko (eds.), *The MONK's problem. A performance comparison of different learning algorithms*. Pittsburg: Carnegie-Mellon Univ.
- Wasserman, P.D. (1989). *Neural computing: Theory and practice*. Van Nostrand Reinhold. N.Y.
- White, H. (1994). Parametric statistical estimation with artificial neural networks. En Smolensky, Mozer y Rumelhart (eds.), *Mathematical perspectives on neural networks*. Hillsdale, NJ: Erlbaum.
- Wilson, R. L. y Hardgrave, B. C. (1995). Predicting graduate student success in a MBA program: Regression vs. Clasification. *Educational and Psychological Measurement*. 55(2), 186-195.

(Revisión aceptada: 14/12/98)