

## Redes neuronales artificiales: abriendo la caja negra

Juan José Montaña Moreno<sup>1</sup> Alfonso Palmer Pol

Universidad de las Islas Baleares

Carlos Fernández Provencio

InfoMallorca, S.L.

### Resumen

El mayor esfuerzo en la investigación sobre Redes Neuronales Artificiales (RNA) se ha centrado en el desarrollo de nuevos algoritmos de aprendizaje, la exploración de nuevas arquitecturas de redes neuronales y la expansión de nuevos campos de aplicación. Sin embargo, se ha dedicado poca atención a desarrollar procedimientos que permitan comprender la naturaleza de las representaciones internas generadas por la red para responder ante una determinada tarea. En el presente trabajo, se plantea un doble objetivo: a) revisar los diferentes métodos interpretativos propuestos hasta el momento para determinar la importancia o efecto de cada variable de entrada sobre la salida de una red perceptrón multicapa, b) validar un nuevo método, denominado análisis de sensibilidad numérico (NSA, numeric sensitivity analysis), que permite superar las limitaciones de los métodos anteriormente propuestos. Los resultados obtenidos mediante simulación ponen de manifiesto que el método NSA es el procedimiento que, en términos globales, mejor describe el efecto de las variables de entrada sobre la salida de la red neuronal.

PALABRAS CLAVE: *Redes neuronales artificiales, análisis de sensibilidad.*

### Abstract

ARTIFICIAL NEURAL NETWORKS: OPEN THE BLACK BOX. The biggest effort in the research about Artificial Neural Networks (ANN) has been centered in the development of new learning algorithms, the exploration of new architectures of neural networks and the expansion of new application fields. However, little attention has been dedicated on to develop procedures that allow to understand the nature of the internal representations generated by the neural network to respond to a certain task. The present work outlined two objectives: a) to revise the different interpretative methods proposed until the moment to determine the importance or effect of each input variable on the output of a multilayer perceptron, b) to validate a new method, called numeric sensitivity analysis (NSA), that allows to overcome the limitations of the previously proposed methods. The results obtained by simulation show that the NSA method is, in general terms, the best procedure for the analysis of the effect of the input variables on the neural network output.

KEY WORDS: *artificial neural networks, sensitivity analysis.*

En los últimos quince años, las redes neuronales artificiales (RNA) han emergido como una potente herramienta para el modelado estadístico orientada principalmente al reconocimiento de patrones -tanto en la vertiente de clasificación como de predicción. Las RNA poseen una serie de características admirables, tales como la habilidad para procesar datos con ruido o incompletos, la alta tolerancia a fallos que permite a la red operar satisfactoriamente con neuronas o conexiones dañadas y la capacidad de responder en tiempo real debido a su paralelismo inherente.

Actualmente, existen unos 40 paradigmas de RNA que son usados en diversos campos de aplicación (Sarle, 1998). Entre estos paradigmas, el más ampliamente utilizado es el perceptrón multicapa asociado al algoritmo de aprendizaje *backpropagation error* (propagación del error hacia atrás), también denominado *red backpropagation* (Rumelhart, Hinton y Williams, 1986). La popularidad del perceptrón

<sup>1</sup>Dirección postal: Juan José Montaña Moreno. Facultad de Psicología. Universidad de las Islas Baleares. Carretera de Valldemossa, Km. 7,5. 07071 Palma de Mallorca (Spain). e-mail: juanjo.montano@uib.es

multicapa se debe principalmente a que es capaz de actuar como un aproximador universal de funciones (Funahashi, 1989; Hornik, Stinchcombe y White, 1989). Más concretamente, una red conteniendo al menos una capa oculta con suficientes unidades no lineales puede aprender cualquier tipo de función o relación continua entre un grupo de variables de entrada y salida. Esta propiedad convierte a las redes perceptrón multicapa en herramientas de propósito general, flexibles y no lineales; mostrando un rendimiento superior respecto a los modelos estadísticos clásicos en numerosos campos de aplicación.

El mayor esfuerzo en la investigación sobre RNA se ha centrado en el desarrollo de nuevos algoritmos de aprendizaje, la exploración de nuevas arquitecturas de redes neuronales y la expansión de nuevos campos de aplicación. Sin embargo, se ha dedicado poca atención a desarrollar procedimientos que permitan comprender la naturaleza de las representaciones internas generadas por la red para responder ante un problema determinado. En lugar de eso, las RNA se han presentado al usuario como una especie de "caja negra" cuyo complejísimo trabajo, de alguna forma mágico, transforma las entradas en salidas predichas. En otras palabras, no se puede saber inmediatamente cómo los pesos de la red o los valores de activación de las neuronas ocultas están relacionados con el conjunto de datos manejados. Así, a diferencia de los modelos estadísticos clásicos, no parece tan evidente conocer en una red el efecto que tiene cada variable explicativa sobre la/s variable/s de respuesta.

Esta percepción acerca de las RNA como una "caja negra", sin embargo, no es del todo cierta. De hecho, desde finales de los años 80 han surgido diversos intentos por desarrollar una metodología que permitiera interpretar lo aprendido por la red, aunque son escasos los trabajos orientados a la validación de tales procedimientos. Estos esfuerzos no tienen únicamente por objeto determinar las variables de entrada con mayor peso o importancia sobre la salida de la red, sino también identificar y eliminar del modelo las variables redundantes o irrelevantes, ésto es, aquellas variables que pueden expresarse en términos de otras variables de entrada o aquellas que simplemente no contribuyen en la predicción. Este último aspecto no es trivial en el campo de las RNA, debido a que las RNA -al igual que otros tipos de modelado-, se ven afectadas por la denominada maldición de la dimensionalidad (*curse of dimensionality*) (Bishop, 1995; Sarle, 1998). Esto significa que el número de datos necesarios para especificar una función, en general, crece exponencialmente con la dimensión del espacio de entrada. Por tanto, la reducción de la dimensión del espacio de entrada mediante la eliminación de variables redundantes o irrelevantes permite trabajar con un menor número de datos, acelera el proceso de convergencia de los pesos de la red y, en base a las demostraciones de Baum y Haussler (1989), podemos tener la expectativa de obtener un error de generalización más bajo (Rzempoluck, 1998).

En el presente trabajo nos planteamos dos objetivos: a) revisar los diferentes métodos interpretativos propuestos hasta el momento para determinar la importancia o efecto de cada variable de entrada sobre la salida de una red perceptrón multicapa, b) presentar y validar un nuevo método, denominado análisis de sensibilidad numérico (NSA, *numeric sensitivity analysis*), que permite superar algunas limitaciones de los métodos anteriormente propuestos.

### El perceptrón multicapa

Consideremos un perceptrón multicapa compuesto por una capa de entrada, una capa oculta y una capa de salida como el mostrado en la figura 1. Un patrón de entrada  $p$  formado por un conjunto de valores en las variables de entrada  $x_i$  está representado por el vector  $X_p : x_{p1}, \dots, x_{pi}, \dots, x_{pN}$ . Por su parte,  $w_{ij}$  es el peso de conexión desde la neurona de entrada  $i$  a la neurona oculta  $j$  y  $v_{jk}$  es el peso de conexión desde la neurona oculta  $j$  a la neurona de salida  $k$ . Respecto a las señales de entrada y los valores de activación de las neuronas,  $net_{pj}$  y  $net_{pk}$  son las entradas netas que reciben las neuronas ocultas y de salida para un patrón  $p$  dado, respectivamente;  $b_{pj}$  e  $y_{pk}$  son los valores de activación o de salida de las neuronas ocultas y de salida, respectivamente, para un patrón  $p$  dado, como resultado de aplicar una función de activación  $f(\cdot)$  sobre la entrada neta de la neurona.

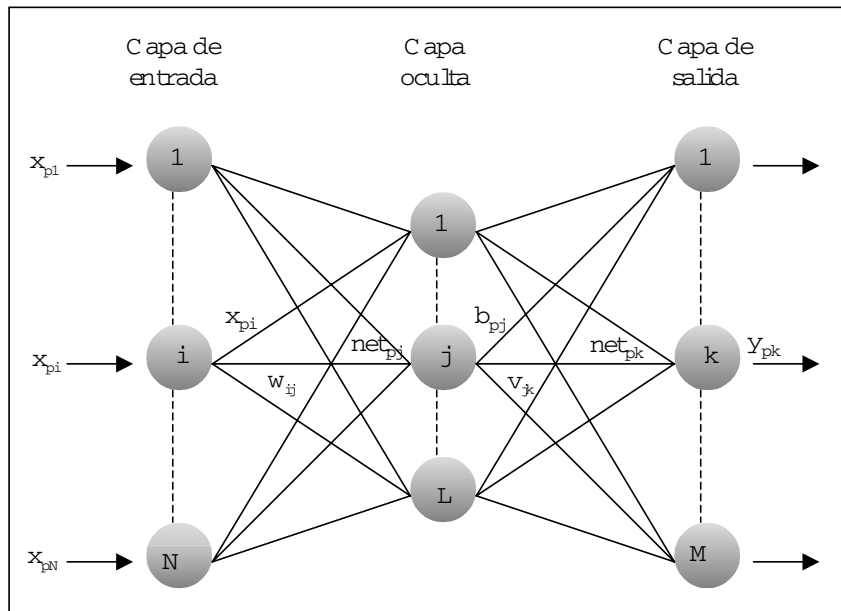


Figura 1. Perceptrón multicapa compuesto por una capa de entrada, una capa oculta y una capa de salida.

### Abriendo la caja negra

En el presente apartado se exponen dos tipos de metodologías generales que permiten conocer lo que ha aprendido un perceptrón multicapa con una capa oculta a partir del valor de los pesos y los valores de activación de las neuronas, esto es, lo que se pretende es conocer el efecto o importancia de cada variable de entrada sobre la salida de la red. Estas dos metodologías corresponden al análisis basado en la magnitud de los pesos y al análisis de sensibilidad.

### *Análisis basado en la magnitud de los pesos*

El análisis basado en la magnitud de los pesos agrupa aquellos procedimientos que se basan exclusivamente en los valores almacenados en la matriz estática de pesos con el propósito de determinar la influencia relativa de cada variable de entrada sobre cada una de las salidas de la red. Este tipo de análisis tiene su origen en el examen de los pesos de conexión entre las neuronas de entrada y ocultas. Así, se podría decir que las entradas con pesos de valor absoluto alto son importantes, mientras que aquellas cuyo valor de pesos es próximo a cero no son importantes. Sin embargo, este método es poco fiable para medir la importancia de las variables de entrada, debido a que la presencia de pesos altos en las conexiones entre la capa de entrada y oculta, no necesariamente significa que la entrada sea importante, y viceversa (Masters, 1993). Este método inicial ha dejado paso a expresiones matemáticas más elaboradas que tienen en cuenta no solo los pesos de conexión entre la capa de entrada y oculta, sino también los pesos de conexión entre la capa oculta y de salida.

En este sentido, se han propuesto diferentes ecuaciones basadas en la magnitud de los pesos (Yoon, Brobst, Bergstresser y Peterson, 1989; Baba, Enbutu y Yoda, 1990; Garson, 1991a; Garson, 1991b; Yoon, Swales y Margavio, 1993; Milne, 1995; Gedeon, 1997; Tsaih, 1999), aunque todas ellas se caracterizan por calcular el producto de los pesos  $w_{ij}$  y  $v_{jk}$  para cada una de las neuronas ocultas y obtener el sumatorio de los productos calculados. A continuación, se presenta una de las ecuaciones más utilizadas, la propuesta por Garson (Garson, 1991a, 1991b; Modai, Saban, Stoler, Valevski y Saban, 1995):

$$Q_{ik} = \frac{\sum_{j=1}^L \left( \frac{w_{ij}}{\sum_{r=1}^N w_{rj}} v_{jk} \right)}{\sum_{i=1}^N \left( \sum_{j=1}^L \left( \frac{w_{ij}}{\sum_{r=1}^N w_{rj}} v_{jk} \right) \right)} \quad (1)$$

donde  $\sum_{r=1}^N w_{rj}$  es la suma de los pesos de conexión entre las  $i$  neuronas de entrada y la neurona oculta  $j$ .

En esta ecuación debemos tener en cuenta, por una parte, que el valor de los pesos se usa en valor absoluto para que los pesos positivos y negativos no se cancelen y, por otra parte, que el valor del umbral de las neuronas ocultas y de salida no se tienen en cuenta, asumiendo que su inclusión no afecta al resultado final (Garson, 1991a). El índice  $Q_{ik}$  representa el porcentaje de influencia de la variable de entrada  $i$  sobre la salida  $k$ , en relación a las demás variables de entrada, de forma que la suma de este índice para todas las variables de entrada debe dar como valor el 100%.

### *Análisis de sensibilidad*

El análisis de sensibilidad está basado en la medición del efecto que se observa en una salida  $y_k$  o en el error cometido debido al cambio que se produce en una entrada  $x_i$ . Así, cuanto mayor efecto se observe sobre la salida, mayor sensibilidad podemos deducir que presenta respecto a la entrada. A continuación, se presentan diferentes formas de realizar el análisis de sensibilidad.

*Análisis de sensibilidad basado en el error*

Una forma de aplicar el análisis de sensibilidad consiste en analizar el efecto producido sobre el error debido a cambios en la entrada. Normalmente, la función de error que se utiliza es la raíz cuadrada de la media cuadrática del error (RMC error) que viene dada por la siguiente expresión:

$$RMC_{error} = \sqrt{\frac{\sum_{p=1}^P \sum_{k=1}^M (d_{pk} - y_{pk})^2}{P \cdot M}} \quad (2)$$

donde  $d_{pk}$  es la salida deseada para el patrón  $p$  en la neurona de salida  $k$ .

La aplicación de este tipo de análisis sobre un conjunto de datos –en general, el grupo de entrenamiento–, consiste en ir variando el valor de una de las variables de entrada a lo largo de todo su rango mediante la aplicación de pequeños incrementos, mientras se mantienen los valores originales de las demás variables de entrada (Frost y Karri, 1999). Una manera sencilla de determinar el tamaño de los incrementos en las variables de entrada se basa en dividir el rango de la variable por el número de patrones con el que se cuenta. Una vez aplicados los incrementos a una determinada variable de entrada, se procede a entrenar la red neuronal calculando el valor de RMC error. Siguiendo este procedimiento para todas las variables de entrada, se puede establecer una ordenación en cuanto a importancia sobre la salida. Así, la variable de entrada que proporcione el mayor RMC error será la variable con más influencia en la variable de respuesta, mientras que la variable de entrada con menor RMC error asociado será la que menos contribuya en la predicción de la red.

Existen otras variantes de este método como, por ejemplo, evaluar el error cometido por la red restringiendo la entrada de interés a un valor fijo (por ejemplo, el valor promedio) para todos los patrones o directamente eliminando esa entrada (Masters, 1993). Si el error aumenta sensiblemente ante el cambio provocado, se puede concluir que la entrada es importante.

*Análisis de sensibilidad basado en la salida*

Una forma más común de realizar el análisis de sensibilidad consiste en estudiar el efecto que se observa directamente en una variable de salida debido al cambio que se produce en una variable de entrada. Esta aproximación ha sido aplicada de forma intuitiva en la identificación de variables relevantes en un variado número de tareas como la predicción del comportamiento de la bolsa (Bilge, Refenes, Diamond y Shadbolt, 1993), la predicción de las auto-expectativas negativas en niños (Reid, Nair, Kashani y Rao, 1994; Kashani, Nair, Rao, Nair y Reid 1996), el análisis de supervivencia (De Laurentiis y Ravdin, 1994), la predicción del resultado de un tratamiento psiquiátrico (Modai, Saban, Stoler, Valevski y Saban, 1995) y la predicción de resultados farmacológicos (Opara, Primozic y Cvelbar, 1999). De esta forma, sobre la red entrenada se fija el valor de todas las variables de entrada a su valor medio y, a continuación, se puede optar por añadir ruido o ir variando el valor de una de las variables a lo largo de todo su rango mediante la aplicación de pequeños incrementos. Esto permite registrar los cambios producidos en la salida de la red y aplicar sobre estos cambios un índice resumen que dé cuenta de la magnitud del efecto de las variaciones producidas en la entrada  $x_i$  sobre la salida  $y_k$ .

Esta forma de proceder es sencilla de aplicar, sin embargo, desde esta aproximación se debe tomar una decisión bastante arbitraria acerca de la cantidad de ruido a añadir o de la magnitud del incremento y también del valor al que quedan fijadas las demás variables. A continuación, se presentan dos métodos basados en el análisis de sensibilidad sobre la salida que gozan de un fundamento matemático más sólido: la matriz de sensibilidad Jacobiana y el método de sensibilidad numérico.

#### Matriz de sensibilidad Jacobiana

El algoritmo de aprendizaje *backpropagation* (Rumelhart, Hinton y Williams, 1986) aplicado a un perceptrón multicapa se basa en el cálculo de la derivada parcial del error con respecto a los pesos para averiguar qué dirección tomar para modificar los pesos con el fin de reducir el error de forma iterativa. Mediante un procedimiento similar, los elementos que componen la matriz Jacobiana  $S$  proporcionan, de forma analítica, una medida de la sensibilidad de las salidas a cambios que se producen en cada una de las variables de entrada. En la matriz Jacobiana  $S$  -de orden  $I \times K$ -, cada fila representa una entrada de la red y cada columna representa una salida de la red, de forma que el elemento  $S_{ik}$  de la matriz representa la sensibilidad de la salida  $k$  con respecto a la entrada  $i$ . Cada uno de los elementos  $S_{ik}$  se obtiene calculando la derivada parcial de una salida  $y_k$  con respecto a una entrada  $x_i$ , esto es,  $\frac{\partial y_k}{\partial x_i}$ . En este caso, la derivada parcial representa la pendiente instantánea de la función subyacente entre  $x_i$  e  $y_k$  para unos valores dados en ambas variables. Aplicando la regla de la cadena sobre  $\frac{\partial y_k}{\partial x_i}$  tenemos que:

$$S_{ik} = \frac{\partial y_k}{\partial x_i} = f'(net_k) \sum_{j=1}^L v_{jk} f'(net_j) w_{ij} \quad (3)$$

Si las neuronas de la capa oculta y de salida utilizan la función de activación sigmoideal logística, la expresión final de cálculo de la sensibilidad de la salida  $k$  con respecto a la entrada  $i$  sería:

$$S_{ik} = y_k(1 - y_k) \sum_{j=1}^L v_{jk} b_j(1 - b_j) w_{ij} \quad (4)$$

Así, cuanto mayor sea el valor absoluto de  $S_{ik}$ , más importante es  $x_i$  en relación a  $y_k$ . El signo de  $S_{ik}$  indica si el cambio observado en  $y_k$  va en la misma dirección o no que el cambio provocado en  $x_i$ . Cuando la discrepancia entre la salida  $y_k$  calculada por la red y la salida deseada  $d_k$  para un patrón dado es mínima, el término derivativo  $f'(net_k)$  es próximo a cero con funciones sigmoideas y, como consecuencia, el valor de la derivada parcial queda anulado. Con el fin de solucionar este problema, algunos autores (Lisboa, Mehridehnavi y Martin, 1994; Rzempoluck, 1998) proponen suprimir el término derivativo  $f'(net_k)$  asumiendo que no afecta a la comparación entre las sensibilidades de las diferentes entradas respecto a la salida. Una solución alternativa que proponemos podría consistir en utilizar la función lineal en las neuronas de salida. De esta forma, evitamos la cancelación del valor de la derivada parcial ya que la derivada de la función lineal es igual a la unidad.

Como se puede observar, los valores de la matriz Jacobiana dependen no solo de la información aprendida por la red neuronal, que está almacenada de forma dis-

tribuida y estática en las conexiones  $w_{ij}$  y  $v_{jk}$ , sino también de la activación de las neuronas de la capa oculta y de salida que, a su vez, dependen de las entradas de la red. Como diferentes patrones de entrada pueden proporcionar diferentes valores de pendiente, la sensibilidad necesita ser evaluada a partir de todo el conjunto de entrenamiento. Considerando el valor de sensibilidad entre  $i$  y  $k$  para el patrón  $X_p$  como  $S_{ik}(p)$ , podemos definir la sensibilidad a partir de la esperanza matemática,  $E(S_{ik}(p))$ , y la desviación estándar,  $SD(S_{ik}(p))$ , que pueden ser calculadas mediante:

$$E(S_{ik}(p)) = \frac{\sum_{p=1}^P S_{ik}(p)}{P} \quad (5)$$

$$SD(S_{ik}(p)) = \sqrt{\frac{\sum_{p=1}^P (S_{ik}(p) - E(S_{ik}(p)))^2}{P - 1}} \quad (6)$$

El cálculo de la matriz Jacobiana ha sido expuesto en multitud de ocasiones (Hwang, Choi, Oh y Marks, 1991; Hashem, 1992; Fu y Chen, 1993; Zurada, Malinowski y Cloete, 1994; Bishop, 1995; Rzepoluck, 1998) y ha sido aplicado en campos tan variados como el reconocimiento de imágenes (Takenaga, Abe, Takatoo, Kayama, Kitamura y Okuyama, 1991), la ingeniería (Guo y Uhrig, 1992; Bahbah y Girgis, 1999), la meteorología (Castellanos, Pazos, Ríos y Zafra, 1994) y la medicina (Harrison, Marshall y Kennedy, 1991; Engelbrecht, Cloete y Zurada, 1995; Rambhia, Glenn y Hwang, 1999).

Respecto a los estudios de validación de este método, Gedeon (1997) realizó recientemente una comparación entre el análisis basado en la magnitud de los pesos y el análisis de sensibilidad basado en el cálculo de la matriz Jacobiana, mostrando este último tipo de análisis un mayor acuerdo con la técnica de ir eliminando cada vez una variable de entrada y observar el efecto que tiene en el rendimiento de la red neuronal. Estos resultados parecen indicar que las técnicas basadas en propiedades dinámicas como el análisis de sensibilidad son más fiables que las técnicas basadas en propiedades estáticas.

#### Método de sensibilidad numérico

En los apartados anteriores, se han descrito los diferentes métodos propuestos para analizar la importancia que tienen las variables explicativas sobre la/s variable/s de respuesta en una red perceptrón multicapa. Tales métodos han demostrado su utilidad en determinadas tareas de predicción, sin embargo, cuentan con una serie de limitaciones que pasamos a comentar.

El análisis basado en la magnitud de los pesos no ha demostrado ser sensible a la hora de ordenar las variables de entrada en función de su importancia sobre la salida (Garson, 1991a; Sarle, 2000) y, en los estudios comparativos (Gedeon, 1997), el análisis de sensibilidad basado en el cálculo de la matriz Jacobiana ha demostrado ser siempre superior. Por su parte, el análisis de sensibilidad consistente en añadir

incrementos o perturbaciones se basa en la utilización de variables de entrada cuya naturaleza es continua, ya que no sería del todo correcto añadir incrementos a variables nominales, esto es, variables que toman valores discretos (Hunter, Kennedy, Henry y Ferguson, 2000). Por su parte, la versión analítica del análisis de sensibilidad, el cálculo de la matriz Jacobiana, parte del supuesto de que todas las variables implicadas en el modelo son continuas (Sarle, 2000). Este supuesto limita el número de campos de aplicación de las RNA en las Ciencias del Comportamiento y de la Salud donde es muy común el manejo de variables discretas (por ejemplo, género: 0 = varón, 1 = mujer ó estatus: 0 = sano, 1 = enfermo).

En este trabajo, se presenta un nuevo método, denominado análisis de sensibilidad numérico (NSA, *numeric sensitivity analysis*), que permite superar las limitaciones comentadas de los métodos anteriores. Este nuevo método se basa en el cálculo de las pendientes que se forman entre entradas y salidas, sin realizar ningún supuesto acerca de la naturaleza de las variables y respetando la estructura original de los datos.

Para analizar el efecto de una variable de entrada  $x_i$  sobre una variable de salida  $y_k$  mediante el método NSA, en primer lugar, debemos ordenar ascendentemente los  $p$  patrones a partir de los valores de la variable de entrada  $x_i$ . En función de tal ordenación, se genera un número  $G$  determinado de grupos de igual o aproximado tamaño. El número idóneo de grupos dependerá del número de patrones disponible y de la complejidad de la función que se establece entre la entrada y la salida, aunque en la mayoría de casos será suficiente un valor de  $G = 30$  o similar. Para cada grupo formado se calcula la media aritmética de la variable  $x_i$  y la media aritmética de la variable  $y_k$ . A continuación, se obtiene el índice NSA basado en el cálculo numérico de la pendiente formada entre cada par de grupos consecutivos,  $g_r$  y  $g_{r+1}$ , de  $x_i$  sobre  $y_k$  mediante la siguiente expresión:

$$NSA_{ik}(g_r) \equiv \frac{\bar{y}_k(g_{r+1}) - \bar{y}_k(g_r)}{\bar{x}_i(g_{r+1}) - \bar{x}_i(g_r)} \quad (7)$$

donde

- $\bar{x}_i(g_r)$  y  $\bar{x}_i(g_{r+1})$  son las medias de la variable  $x_i$  correspondientes a los grupos  $g_r$  y  $g_{r+1}$ , respectivamente,
- $\bar{y}_k(g_r)$  e  $\bar{y}_k(g_{r+1})$  son las medias de la variable  $y_k$  correspondientes a los grupos  $g_r$  y  $g_{r+1}$ , respectivamente.

Una vez calculados los  $G-1$  valores NSA, se puede obtener el valor de la esperanza matemática del índice NSA o pendiente entre la variable de entrada  $i$  y la variable de salida  $k$  mediante:

$$E(NSA_{ik}(g_r)) = \sum_{r=1}^{G-1} NSA_{ik}(g_r) \cdot f(NSA_{ik}(g_r)) = \frac{\bar{y}_k(g_G) - \bar{y}_k(g_1)}{\bar{x}_i(g_G) - \bar{x}_i(g_1)} \quad (8)$$

donde

- $f(NSA_{ik}(g_r)) \equiv \frac{\bar{x}_i(g_{r+1}) - \bar{x}_i(g_r)}{\bar{x}_i(g_G) - \bar{x}_i(g_1)}$  representa la función de probabilidad del índice NSA,

- $\bar{x}_i(g_G)$  y  $\bar{x}_i(g_1)$  son los valores promedio de la variable  $x_i$  para el último grupo  $g_G$  y el primer grupo  $g_1$ , respectivamente,
- $\bar{y}_k(g_G)$  e  $\bar{y}_k(g_1)$  son los valores promedio de la variable  $y_k$  para el grupo  $g_G$  y el grupo  $g_1$ , respectivamente.

Cuando la variable de entrada  $x_i$  es binaria, la esperanza del índice  $NSA$  se obtiene calculando la media de la variable  $y_k$  cuando la variable  $x_i$  toma el valor mínimo y la media de la variable  $y_k$  cuando la variable  $x_i$  toma el valor máximo, y aplicando la siguiente expresión:

$$E(NSA_{ik}(g_r)) = \frac{\bar{y}_k(x_{imax}) - \bar{y}_k(x_{imin})}{x_{imax} - x_{imin}} \quad (9)$$

donde

- $x_{imax}$  y  $x_{imin}$  son los valores máximo y mínimo que toma la variable  $x_i$ , respectivamente,
- $\bar{y}_k(x_{imax})$  e  $\bar{y}_k(x_{imin})$  son la media de la variable  $y_k$  cuando la variable  $x_i$  toma el valor máximo y la media de la variable  $y_k$  cuando la variable  $x_i$  toma el valor mínimo, respectivamente.

El valor de la esperanza matemática del índice  $NSA$  representa el efecto promedio que tiene un incremento de  $x_i$  sobre  $y_i$ . Cuando la variable de entrada es binaria, la esperanza matemática representa el efecto promedio provocado por el cambio del valor mínimo al valor máximo en la variable  $x_i$ . Al igual que en el caso de la matriz Jacobiana, cuanto mayor sea el valor absoluto de  $E(NSA_{ik}(g_r))$ , más importante es  $x_i$  en relación a  $y_k$ . El signo de  $E(NSA_{ik}(g_r))$  indica si el cambio observado en  $y_k$  va en la misma dirección o no que el cambio provocado en  $x_i$ .

Los estudios piloto en los que hemos aplicado el método  $NSA$  muestran que la forma más adecuada de representar las variables de entrada de naturaleza discreta (binarias o politómicas) es mediante la utilización de codificación ficticia, tal como se hace habitualmente en el modelado estadístico. La introducción de los valores originales de una variable politómica (por ejemplo, nivel social: bajo = 1, medio = 2 y alto = 3) en la red neuronal, no permite reflejar de forma adecuada el efecto o pendiente que tiene el cambio de una categoría a otra categoría. Por otro lado, es conveniente que las variables de entrada y salida sean reescaladas al mismo rango de posibles valores (por ejemplo, entre 0 y 1). Esto último evita posibles sesgos en la obtención del índice  $NSA$  debido a la utilización de escalas de medida diferentes entre las variables de entrada y, además, permite obtener un valor de esperanza matemática estandarizado a diferencia de la matriz Jacobiana. Así, el rango de posibles valores que puede adoptar  $E(NSA_{ik}(g_r))$  oscila entre  $-1$  y  $+1$ . Estos dos límites indicarían un efecto máximo de la variable de entrada sobre la salida, con una relación negativa en el primer caso ( $-1$ ) y con una relación positiva en el segundo caso ( $+1$ ). Los valores iguales o próximos a cero indicarían ausencia de efecto de la variable de entrada.

El cálculo de la desviación estándar del índice  $NSA$ , cuando las variables implicadas son de naturaleza continua, se puede realizar mediante:

$$SD(NSA_{ik}(g_r)) = \sqrt{E(NSA_{ik}^2(g_r)) - (E(NSA_{ik}(g_r)))^2} \quad (10)$$

El valor de la desviación estándar se debe interpretar como el grado de oscilaciones que ha sufrido la pendiente que se establece entre  $x_i$  e  $y_k$ , de manera que a mayor valor de la desviación estándar, mayor comportamiento caótico o aleatorio tiene la función entre las dos variables implicadas.

Por último, es conveniente aportar junto a los índices esperanza matemática y desviación estándar, la representación gráfica de la pendiente entre cada par de variables de entrada y salida. Esto es especialmente útil en aquellos casos en los que el valor de la esperanza matemática puede enmascarar la función subyacente entre la variable de entrada y la variable de salida.

Una vez descritos los principales métodos interpretativos dirigidos al análisis del efecto de las variables de entrada en una red perceptrón multicapa, a continuación, nos proponemos realizar un estudio comparativo acerca de su rendimiento. Los métodos bajo estudio son: análisis basado en la magnitud de los pesos a través del método de Garson (1991a), análisis de sensibilidad basado en el cálculo del incremento observado en la función RMC error al eliminar una variable de entrada, análisis de sensibilidad basado en el cálculo de la matriz Jacobiana y, finalmente, el método NSA.

## Método

### *Materiales y procedimiento*

Se han generado mediante simulación cuatro matrices de datos compuesta cada una de ellas por 1000 registros y cuatro variables con rango entre 0 y 1. Las tres primeras variables (X1, X2 y X3) actúan como variables predictoras o variables de entrada a la red, mientras que la última variable (Y) es una función de las variables predictoras y actúa como variable de salida. El valor del coeficiente de correlación de Pearson entre las variables predictoras oscila entre 0 y 0.40. En todos los casos, la variable X1 no tiene ningún tipo de contribución o efecto en la salida Y de la red, seguida de la variable X2 con un efecto intermedio y la variable X3 que presenta el mayor efecto sobre la salida de la red. A fin de analizar el comportamiento de los diferentes métodos dependiendo del tipo de variable implicada, se ha manipulado en cada una de las matrices la naturaleza de las variables de entrada. A continuación, se presenta una descripción detallada de cada matriz utilizada:

*Matriz 1: Variables cuantitativas.* Las variables de entrada son de naturaleza cuantitativa con distribución normal. Para generar la variable Y como función de las variables de entrada se utilizó la siguiente expresión:

$$Y = \tanh(X2) - \exp(2.4 \cdot X3) + \text{Error } N(0, 0.1) \quad (11)$$

De esta forma, X1 no contribuye en la variable Y. La función que se establece entre X2 y la variable de salida es la tangente hiperbólica con rango entre  $-1$  y  $1$  para la variable de salida, mientras que la función entre X3 y la variable de salida es de tipo exponencial negativo con rango entre  $-1$  y  $-10$  para la variable de salida. Por último, se añadió un error aleatorio con distribución normal, media 0 y desviación estándar 0.1. La variable de salida resultante se reescaló a valores entre 0 y 1.

*Matriz 2: Variables discretas binarias.* Las variables de entrada y salida son de naturaleza discreta binaria con valores 0 y 1. La relación entre las entradas y la salida se determinó a partir del coeficiente Phi. Así, los valores del coeficiente Phi entre X1, X2, X3 y la salida fueron  $-0.001$ ,  $0.346$  y  $0.528$ , respectivamente.

*Matriz 3: Variables discretas politómicas.* Las variables de entrada son de naturaleza discreta politómica con valores 1, 2 y 3, mientras que la variable de salida es discreta binaria con valores 0 y 1. La relación entre las entradas y la salida se determinó a partir del coeficiente V de Cramer. Así, los valores del coeficiente V entre X1, X2, X3 y la salida fueron  $0.023$ ,  $0.380$  y  $0.593$ , respectivamente. Cada variable predictora se codificó mediante la utilización de dos variables ficticias binarias (valor 1 = 0 0, valor 2 = 1 0 y valor 3 = 0 1), de forma que cada variable estaba representada por dos neuronas de entrada.

*Matriz 4: Variables cuantitativas y discretas.* La variable X1 es de naturaleza cuantitativa con distribución normal, la variable X2 es discreta binaria con valores 0, 1 y la variable X3 es continua con distribución normal. Para generar la variable Y como función de las variables de entrada se utilizó la siguiente expresión:

$$\begin{cases} \text{Si } X2 = 0, \text{ entonces } Y = 1 + 1.1 \times \exp(X3) + \text{Error } N(0, 0.1) \\ \text{Si } X2 = 1, \text{ entonces } Y = 1.1 \times \exp(X3) + \text{Error } N(0, 0.1) \end{cases} \quad (12)$$

De esta forma, X1 no contribuye en la variable Y. Por su parte, cuando X2 toma el valor 0, la variable Y se incrementa en una unidad. La función que se establece entre X3 y la variable Y es de tipo exponencial con rango entre 1 y 3 para la variable de salida. Al igual que en el caso de la matriz 1, se añadió un error aleatorio con distribución normal, media 0 y desviación estándar 0.1. La variable de salida resultante se reescaló a valores entre 0 y 1.

Para llevar a cabo el estudio, se utilizó el programa informático *Sensitivity Neural Network 1.0*, creado por nosotros, que, a través de un interfaz de sencillo manejo, permite simular el comportamiento de un perceptrón multicapa entrenado con el algoritmo de aprendizaje *backpropagation* e incorpora como novedad los métodos interpretativos descritos.

Para el entrenamiento de las redes neuronales, cada matriz de datos fue dividida en tres grupos: 500 patrones actuaron como conjunto de entrenamiento, 250 patrones actuaron como conjunto de validación y 250 patrones actuaron como conjunto de test. Las redes neuronales simuladas estaban compuestas por tres o seis neuronas de entrada dependiendo de la matriz de datos, dos neuronas ocultas y una neurona de salida. Como funciones de activación se utilizó la función tangente hiperbólica en las neuronas ocultas y la función lineal en la neurona de salida. Como parámetros de aprendizaje se utilizó un valor de 0.25 para la tasa de aprendizaje y un valor de 0.8 para el factor momentum. Los pesos de conexión y umbral fueron inicializados con diferentes valores semilla. La red neuronal que obtuvo el mejor rendimiento ante el conjunto de validación de la matriz correspondiente, fue la seleccionada para pasar a la fase de test. Los cuatro modelos de red finalmente obtenidos, mostraron un buen ajuste ante los datos de test.

## Resultados

La Tabla 1 muestra los resultados obtenidos tras aplicar los métodos interpretativos sobre los modelos de red seleccionados en la fase de validación.

Tabla 1. Resultados obtenidos tras aplicar los métodos interpretativos sobre los modelos de red

Métodos interpretativos						
	Método de Garson	$\Delta$ error	Matriz de sensibilidad Jacobiana		Análisis de sensibilidad numérico	
			Media	S.D.	Media	S.D.
Matriz 1						
X1	1.102	0.001	0.001	0.003	0.004	1.934
X2	38.980	0.043	0.388	0.027	0.468	1.560
X3	59.920	0.113	-0.786	0.311	-0.865	0.923
Matriz 2						
X1	24.890	0.001	0.007	0.028	-0.001	-
X2	30.140	0.037	0.339	0.048	0.345	-
X3	44.960	0.081	0.501	0.069	0.527	-
Matriz 3						
X1						
X1V1	5.422	-	-0.022	0.032	-0.021	-
X1V2	6.650	-	0.023	0.054	0.002	-
Suma	12.072	0.001	0.045		0.023	
X2						
X2V1	13.630	-	0.306	0.229	0.080	-
X2V2	16.560	-	0.450	0.317	0.303	-
Suma	30.190	0.050	0.756		0.383	
X3						
X3V1	7.049	-	0.302	0.186	0.027	-
X3V2	50.690	-	0.464	0.534	0.531	-
Suma	57.739	0.109	0.766		0.558	
Matriz 4						
X1	17.020	0.000	0.001	0.014	0.094	2.810
X2	29.930	0.132	-0.312	0.006	-0.325	-
X3	53.050	0.076	0.599	0.086	0.585	1.770

Se puede observar que, en general, los cuatro métodos analizados establecen correctamente la jerarquía entre las variables de entrada en función de su importancia.

Un análisis en detalle permite observar que los métodos se comportan de forma correcta cuando las variables de entrada son cuantitativas (matriz 1), y cuando son cuantitativas y discretas (matriz 4), a excepción del método basado en el análisis del incremento del error el cual otorga más importancia a X2 frente a X3 en la matriz 4. Por tanto, este método no establece de forma correcta la jerarquía de importancia entre las variables de entrada. Por su parte, el método de Garson sobrevalora el efecto de la variable X1 en la matriz 4 proporcionando para esta variable un 17.02% de importancia relativa, cuando en realidad su efecto es nulo. Finalmente, el cálculo

de la matriz Jacobiana y el método NSA proporcionan valores promedio muy similares con las matrices 1 y 4. También se puede observar que estos dos métodos identifican aquellos casos en los que la relación entre entrada y salida es negativa.

Cuando las variables de entrada son discretas binarias (matriz 2), se puede observar que el método NSA es el que representa con más exactitud la realidad, debido a que este índice coincide con bastante precisión con los valores reales del índice Phi entre las variables de entrada y la salida de la red. Por otra parte, se puede observar que el cálculo de la matriz Jacobiana, proporciona valores promedio próximos al índice Phi, pero no con la exactitud dada por el método NSA. Este aspecto se puede apreciar mejor observando la tabla 2 donde se proporcionan los valores Phi y los promedios de la matriz Jacobiana y el método NSA para la matriz 2. Por su parte, tal como expresa la tabla 1, el método de Garson en este caso también sobrevalora la importancia de X1, proporcionando un valor de 24.89%.

Tabla 2. Valores Phi, V e índices obtenidos mediante la matriz de sensibilidad Jacobiana y el análisis de sensibilidad numérico para las matrices 2 y 3.

	X1	X2	X3
Matriz 2			
Phi	-0.001	0.346	0.528
Jacobiana	0.007	0.339	0.501
Numérico	-0.001	0.345	0.527
Matriz 3			
V	0.023	0.380	0.593
Jacobiana	0.045	0.756	0.766
Numérico	0.023	0.383	0.558

Cuando las variables de entrada son discretas politómicas (matriz 3), se puede comprobar que el método NSA también es el que mejor se ajusta a la realidad, debido a que es el método que más se aproxima a los valores del índice V. Para ello, como se puede observar en la Tabla 1, se calcula la suma de los valores absolutos obtenidos con las dos variables ficticias que representan cada variable predictora. La Tabla 2 permite comparar el valor real del índice V con los valores promedio de la matriz Jacobiana y el método NSA para la matriz 3. En este caso, la matriz Jacobiana otorga prácticamente la misma importancia a X2 y a X3 con valores 0.756 y 0.766, respectivamente, cuando en realidad X3 es más importante que X2. Al igual que en los casos anteriores, el método de Garson sobrevalora la importancia de X1, proporcionando un valor de 12.07%.

Respecto al grado de fiabilidad, se puede observar que el cálculo de la matriz Jacobiana proporciona en general una gran precisión en la estimación de la media en función del valor de la desviación estándar. Recordemos que en el caso del método NSA, el valor de la desviación estándar se debe interpretar como el grado de oscilación que ha sufrido la pendiente que se establece entre una entrada cuantitativa y la salida. Así, se puede observar en la matriz 1 (ver Tabla 1) que cuanto mayor aleatoriedad existe en la función que se establece entre las dos variables implicadas, mayor es el valor de la desviación estándar de la pendiente.

Por otra parte, con el método NSA la interpretación del efecto de una variable es mucho más sencilla ya que la media que proporciona está acotada a valores entre  $-1$  y  $1$ . Recordemos que esta condición se cumplirá, siempre que se acoten las variables

a un mismo rango de valores. En cambio, los valores promedio proporcionados por la matriz Jacobiana pueden oscilar, a nivel teórico, entre  $-\infty$  y  $+\infty$ .

Finalmente, a modo de ejemplo, se presenta en la Figura 2 la representación gráfica del método NSA proporcionada por el programa *Sensitivity Neural Network 1.0* para la variable X3 de la matriz 1. Recordemos que en la simulación se determinó que la función subyacente entre esta variable de entrada y la salida era una función exponencial negativa. Se puede observar que la red neuronal ha aprendido correctamente la función existente entre ambas variables.

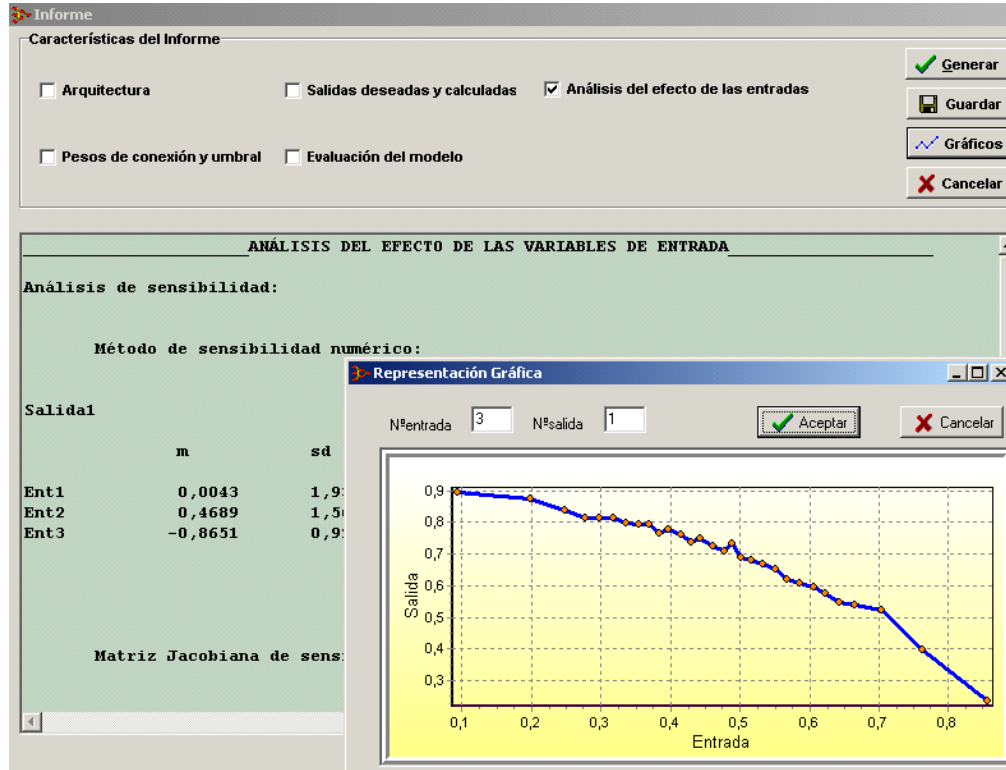


Figura 2. Representación gráfica del método NSA proporcionada por el programa Sensitivity Neural Network 1.0 para la variable X3 de la matriz 1.

## Conclusiones

Las redes neuronales del tipo perceptrón multicapa han obtenido, en los últimos años, excelentes resultados como predictores en tareas de clasificación de patrones y estimación de variables cuantitativas. Sin embargo, este tipo de arquitectura no permite analizar, al menos directamente, el papel desempeñado por cada variable predictora sobre la salida de la red.

Se ha realizado un estudio comparativo acerca de la utilidad de cuatro métodos dirigidos a evaluar la importancia relativa de las variables de entrada en un perceptrón multicapa. Para ello, se han obtenido mediante simulación cuatro matrices de datos en donde se ha manipulado el grado de relación o importancia entre las variables y la naturaleza de las mismas.

De los resultados obtenidos se desprende una serie de ventajas en la aplicación del método NSA presentado en este trabajo, con respecto a los métodos anteriormente propuestos.

En primer lugar, el método NSA es el que, en cuanto a grado de generalidad, permite describir mejor el efecto o importancia de las variables de entrada sobre la salida. Se ha podido observar que el método NSA y la matriz Jacobiana proporcionan valores promedio similares cuando las variables implicadas son cuantitativas. La matriz Jacobiana es perfectamente válida en estos casos, teniendo en cuenta además la baja variabilidad que proporciona. En este sentido, trabajos anteriores corroboran esta conclusión (Harrison, Marshall y Kennedy, 1991; Takenaga, Abe, Takatoo, Kayama, Kitamura y Okuyama, 1991; Guo y Uhrig, 1992; Castellanos, Pazos, Ríos y Zafra, 1994; Engelbrecht, Cloete y Zurada, 1995; Bahbah y Girgis, 1999; Rambhia, Glenly y Hwang, 1999).

Sin embargo, cuando las variables implicadas son discretas (binarias y politómicas), el método NSA es el más adecuado debido a que los valores que proporciona se aproximan considerablemente al índice de asociación Phi en el caso de variables binarias y al índice de asociación V en el caso de variables politómicas.

El método de Garson sobrevalora en la mayoría de casos la importancia de variables que son irrelevantes para la salida de la red. La baja validez del método de Garson observada en nuestro estudio es coincidente con estudios anteriores (Gedeon, 1997; Sarle, 2000). Por su parte, el método basado en el cálculo del incremento observado en la función RMC error al eliminar una variable de entrada no ha sido capaz de establecer correctamente la jerarquía de importancia cuando las variables implicadas son cuantitativas y discretas (matriz 4). Además, este método no permite una interpretación sencilla respecto al grado de relación entre una variable de entrada y la salida de la red.

En segundo lugar, con el método NSA la interpretación del efecto de una variable es mucho más sencilla ya que el índice NSA que proporciona está acotado en el intervalo  $[-1, 1]$  a diferencia de los valores que teóricamente puede proporcionar la matriz Jacobiana.

Por último, el método NSA incorpora un procedimiento que permite representar gráficamente la función aprendida por la red entre una variable de entrada y la salida. Esta representación gráfica aporta información relevante que complementa la información proporcionada por los índices numéricos, debido a que en muchos casos un índice de resumen no es suficiente para reflejar la función subyacente entre variables.

A modo de resumen, se puede decir que el método NSA es el que, de forma global, permite evaluar con mayor exactitud la importancia o efecto de las variables de entrada con independencia de su naturaleza (cuantitativa o discreta), superando las limitaciones de los métodos propuestos hasta el momento. Con este método se amplía considerablemente el número de campos de aplicación potenciales donde la utilización de variables discretas es muy común: medicina, sociología, psicología, biología, economía, etc. Ahora no solo podremos construir potentes instrumentos de predicción mediante redes perceptrón multicapa sino que a partir de las redes obtenidas también podremos analizar el impacto de las variables predictoras en cualquier conjunto de datos.

El método NSA junto a los métodos anteriormente propuestos se encuentran im-

plementados, como se ha comentado, en el programa *Sensitivity Neural Network 1.0*, creado por nuestro equipo. De esta forma, el usuario puede analizar las cualidades de cada uno de los métodos interpretativos con respecto a los demás para diferentes configuraciones de datos. Para este fin, *Sensitivity Neural Network 1.0* se encuentra disponible poniéndose en contacto con los autores via correo electrónico (juanjo.montano@uib.es).

## Referencias

- Baba, K., Enbutu, I. y Yoda, M. (1990). Explicit representation of knowledge acquired from plant historical data using neural network. En IEEE (Ed.), *Proceedings of the International Joint Conference on Neural Networks* (pp. 155-160). New York: IEEE.
- Bahbah, A.G. y Girgis, A.A. (1999). Input feature selection for real-time transient stability assessment for artificial neural network (ANN) using ANN sensitivity analysis. En IEEE (Ed.), *Proceedings of the 21st International Conference on Power Industry Computer Applications* (pp. 295-300). Piscataway, NJ: IEEE.
- Baum, E.B. y Haussler, D. (1989). What size net gives valid generalization? *Neural Computation*, 1, 151-160.
- Bilge, U., Refenes, A.N., Diamond, C. y Shadbolt, J. (1993). Application of sensitivity analysis techniques to neural network bond forecasting. En A.N. Refenes (Ed.), *Proceedings of 1st International Workshop on Neural Networks in the Capital Markets* (p. 12). London: London Business School.
- Bishop, C.M. (1995). *Neural networks for pattern recognition*. Oxford: Oxford University Press.
- Castellanos, J., Pazos, A., Ríos, J. y Zafra, J.L. (1994). Sensitivity analysis on neural networks for meteorological variable forecasting. En J. Vlontzos, J.N. Hwang y E. Wilson (Eds.), *Proceedings of IEEE Workshop on Neural Networks for Signal Processing* (pp. 587-595). New York: IEEE.
- De Laurentiis, M. y Ravdin, P.M. (1994). A technique for using neural network analysis to perform survival analysis of censored data. *Cancer Letters*, 77, 127-138.
- Engelbrecht, A.P., Cloete, I. y Zurada, J.M. (1995). Determining the significance of input parameters using sensitivity analysis. En J. Mira y F. Sandoval (Eds.), *Proceedings of International Workshop on Artificial Neural Networks* (pp. 382-388). New York: Springer.
- Frost, F. y Karri, V. (1999). Determining the influence of input parameters on BP neural network output error using sensitivity analysis. En B. Verma, H. Selvaraj, A. Carvalho y X. Yao (Eds.), *Proceedings of the Third International Conference on Computational Intelligence and Multimedia Applications* (pp.45-49). Los Alamitos, CA: IEEE Computer Society Press.
- Fu, L. y Chen, T. (1993). Sensitivity analysis for input vector in multilayer feedforward neural networks. En IEEE (Ed.), *Proceedings of IEEE International Conference on Neural Networks* (pp. 215-218). New York: IEEE.
- Funahashi, K. (1989). On the approximate realization of continuous mapping by neural networks. *Neural Networks*, 2, 183-192.
- Garson, G.D. (1991a). Interpreting neural-network connection weights. *AI Expert*, April, 47-51.
- Garson, G.D. (1991b). A comparison of neural network and expert systems algorithms with common multivariate procedures for analysis of social science data. *Social Science Computer Review*, 9(3), 399-434.
- Gedeon, T.D. (1997). Data mining of inputs: analysing magnitude and functional measures. *International Journal of Neural Systems*, 8(2), 209-218.
- Guo, Z. y Uhrig, R.E. (1992). Sensitivity analysis and applications to nuclear power plant. En IEEE (Ed.), *International Joint Conference on Neural Networks* (pp. 453-458). Piscataway, NJ: IEEE.
- Harrison, R.F., Marshall, J.M. y Kennedy, R.L. (1991). The early diagnosis of heart attacks: a neuro-computational approach. En IEEE (Ed.), *Proceedings of IEEE International conference on Neural Networks* (pp. 231-239). New York: IEEE.
- Hashem, S. (1992). Sensitivity analysis for feedforward artificial neural networks with differentiable activation functions. En IEEE (Ed.), *International Joint Conference on Neural Networks* (pp. 419-424). New York: IEEE.

- Hornik, K., Stinchcombe, M. y White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5), 359-366.
- Hunter, A., Kennedy, L., Henry, J. y Ferguson, I. (2000). Application of neural networks and sensitivity analysis to improved prediction of trauma survival. *Computer Methods and Programs in Biomedicine*, 62, 11-19.
- Hwang, J.N., Choi, J.J., Oh, S. y Marks, R.J. (1991). Query based learning applied to partially trained multilayer perceptron. *IEEE Transactions on Neural Networks*, 2(1), 131-136.
- Kashani, J.H., Nair, S.S., Rao, V.G., Nair, J. y Reid, J.C. (1996). Relationship of personality, environmental, and DICA variables to adolescent hopelessness: a neural network sensitivity approach. *Journal of American Children and Adolescent Psychiatry*, 35(5), 640-645.
- Lisboa, P.J.G., Mehridehnavi, A.R. y Martin, P.A. (1994). The interpretation of supervised neural networks. En P.J.G. Lisboa y M.J. Taylor (Eds.), *Proceedings of the Workshop on Neural Network Applications and Tools* (pp. 11-17). Los Alamitos, CA: IEEE Computer Society Press.
- Masters, T. (1993). *Practical neural networks recipes in C++*. London: Academic Press.
- Milne, K. (1995). Feature selection using neural networks with contribution measures. En IEEE (Ed.), *Proceedings of Australian Conference of Artificial Intelligence* (pp. 124-136). Sydney: IEEE West Australian Section.
- Modai, I., Saban, N.I., Stoler, M., Valevski, A. y Saban, N. (1995). Sensitivity profile of 41 psychiatric parameters determined by neural network in relation to 8-week outcome. *Computers in Human Behavior*, 11(2), 181-190.
- Opara, J., Primozic, S. y Cvelbar, P. (1999). Prediction of pharmacokinetic parameters and the assessment of their variability in bioequivalence studies by artificial neural networks. *Pharmaceutical Research*, 16(6), 944-948.
- Rambhia, A.H., Glenney, R. y Hwang, J. (1999). Critical input data channels selection for progressive work exercise test by neural network sensitivity analysis. En IEEE (Ed.), *IEEE International Conference on Acoustics, Speech, and Signal Processing* (pp. 1097-1100). Piscataway, NJ: IEEE.
- Reid, J.C., Nair, S.S., Kashani, J.H. y Rao, V.G. (1994). Detecting dysfunctional behavior in adolescents: the examination of relationships using neural networks. En P.W. Lefley (Ed.), *Proceedings of Annual Symposium of Computational Applications on Medical Care* (pp. 743-746). New York: Springer.
- Rumelhart, D.E., Hinton, G.E. y Williams, R.J. (1986). Learning internal representations by error propagation. En D.E. Rumelhart y J.L. McClelland (Eds.), *Parallel distributed processing* (pp. 318-362). Cambridge, MA: MIT Press.
- Rzempeluck, E.J. (1998). *Neural network data analysis using Simulnet*. New York: Springer-Verlag.
- Sarle, W.S. (Ed.) (1998). *Neural network FAQ*. Recuperado 20/10/01, desde la dirección Internet <ftp://ftp.sas.com/pub/neural/FAQ.html>.
- Sarle, W.S. (2000). *How to measure importance of inputs?* Recuperado 2/11/01, desde la dirección Internet <ftp://ftp.sas.com/pub/neural/importance.html>.
- Takenaga, H., Abe, S., Takatoo, M., Kayama, M., Kitamura, T. y Okuyama, Y. (1991). Input layer optimization of neural networks by sensitivity analysis and its application to recognition of numerals. *Transactions of the Institute of Electrical Engineers Japan*, 111(1), 36-44.
- Tsaih, R. (1999). Sensitivity analysis, neural networks, and the finance. En IEEE (Ed.), *International Joint Conference on Neural Networks* (pp. 3830-3835). Piscataway, NJ: IEEE.
- Yoon, Y.O., Brobst, R.W., Bergstresser, P.R. y Peterson, L.L. (1989). A desktop neural network for dermatology diagnosis. *Journal of Neural Network Computing*, 1, 43-52.
- Yoon, Y., Swales, G. y Margavio, T.M. (1993). A comparison of discriminant analysis versus artificial neural networks. *Journal of the Operational Research Society*, 44(1), 51-60.
- Zurada, J.M., Malinowski, A. y Cloete, I. (1994). Sensitivity analysis for minimization of input data dimension for feedforward neural network. En IEEE (Ed.), *Proceedings of IEEE International Symposium on Circuits and Systems* (pp. 447-450). New York: IEEE.