

Las redes neuronales como herramientas estadísticas no paramétricas de clasificación

Alfonso Pitarque, Juan Carlos Ruiz y Juan Francisco Roy
Universidad de Valencia

Recientemente diversos trabajos (ver p.e. Cherkassky, Friedman y Wechsler, 1994; Ripley, 1996) han analizado las relaciones entre redes neuronales y técnicas estadísticas convencionales. Aunque los resultados no son concluyentes, en general las redes neuronales han mostrado una capacidad clasificatoria igual o superior que las técnicas estadísticas, con la ventaja de poderse utilizar independientemente del cumplimiento de los supuestos teóricos relativos a estas técnicas (de ahí que se haya hablado de ellas como de «técnicas no paramétricas»). Se presentan una serie de simulaciones y aplicaciones sobre datos reales que apoyan esta idea. En concreto contrastamos redes neuronales del tipo perceptrón multi-capas contra modelos de regresión logística y análisis discriminante en tareas de clasificación, 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 las redes neuronales clasifican mejor que las técnicas estadísticas, incluso en la condición a priori más favorable a éstas, es decir, cuando existen altas correlaciones entre los predictores con el criterio pero bajas correlaciones entre los predictores. Un patrón de resultados similares aparece en tareas de clasificación con datos reales. Se discuten los resultados en el marco de la polémica redes neuronales vs modelos estadísticos convencionales (ver p.e. Sarle, 1999), valorando si vale la pena ganar en capacidad clasificatoria a costa de incrementar el costo computacional y de recursos involucrados en el entrenamiento de las redes neuronales.

Neural networks as non-parametric classification statistical tools. The relationships between neural networks and statistical methods have been recently analyzed (see e.g. Cherkassky, Friedman and Wechsler, 1994; Ripley, 1996). In general terms neural networks have shown an equal or greater capacity to classify than statistical tools. Moreover they do not need to satisfy the parametric assumptions of the statistical techniques. Simulations about simulated and real data are shown: multi-layer perceptrons versus logistic regression and discriminant analysis statistical models are compared in classification tasks, manipulating the correlation patterns within input variables (predictors) and between the input variables with the output variable (criterion). Results show that neural networks classify better than statistical tools both in simulated data as in empirical data (González-Romá et al, 1999).

Recientemente se ha constatado un gran interés en la literatura en analizar las relaciones entre las redes neuronales (NN) y los modelos estadísticos (ME) tradicionales (ver p.e. Cherkassky, Friedman y Wechsler, 1994; Ripley, 1996) y, como se ha comentado, aunque la terminología es distinta a veces los conceptos a los que se apela son los mismos (Sarle, 1998). De hecho las actualmente tan popularizadas redes de función base radial son modelos híbridos estadístico/neuronales.

Entre las propiedades de las NN que han llamado la atención de los estadísticos destacan las relativas a su buen rendimiento ante problemas no lineales o datos con mucho «ruido», y el poderse utilizar independientemente del cumplimiento de los supuestos teóri-

cos relativos a las técnicas estadísticas (y de ahí que se haya hablado de ellas como de «técnicas de distribución libre o no paramétricas»). Por ello las RN han sido aplicadas a problemas de tradición estadística como predicción y clasificación (a través de las llamadas redes hetero-asociativas: perceptrón multi-capas y redes de función base radial), reducción de la dimensionalidad (a través de las llamadas redes auto-asociativas: Hopfield, Kohonen, etc), series temporales, etc.

El hecho es que multitud de trabajos han comparado RN y ME en estos y otros tipos de tareas. Nuestra línea de investigación surgió cuando al revisar la literatura 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 neuronales sobre las técnicas estadísticas (ver p.e. Garson, 1991; Huang y Lippman, 1987; Navarro, 1998; Whi-

te, 1994). Por su parte este mismo año Lim, Loh y Shih (1999) comparando varios algoritmos estadísticos y 2 basados en redes neurales (LVQ —learning vector quantization- y RBF— función de base radial—) encontraron que éstas no quedaron precisamente muy bien situadas en una ordenación en su poder clasificatorio.

Ante esta disparidad de resultados decidimos en un trabajo reciente (Pitarque, Roy y Ruiz, 1998) comparar mediante simulación redes neuronales (del tipo perceptrón multicapa; MLP) con modelos estadísticos, en tres tareas:

- predicción de una variable criterio cuantitativa por medio bien de una red MLP o bien de un modelo de regresión múltiple;
- clasificación de una variable criterio categorial binaria por medio de una red MLP, por análisis discriminante o por regresión logística
- y clasificación de una variable criterio categorial de tres niveles por medio de una red MLP o por análisis discriminante.

Trabajamos siempre con 7 variables predictoras cuantitativas tipificadas extraídas aleatoriamente de una población con normalidad multivariante. El número de registros fue siempre de 150.

Dado que 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, puede ser determinante en los resultados hallados, manipulamos éstos en cuatro niveles de la variable:

- (a) en la condición baja-alta, o *condición A*, las 7 variables predictoras correlacionaban de forma baja entre sí (r_{xy} promedio=.05), pero cada una de ellas correlacionaba significativamente ($\alpha \leq 0.05$) 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 (*condición B*), por el contrario los predictores correlacionaban significativamente entre sí ($\alpha \leq 0.05$; r_{xy} promedio=.35), pero de forma baja con el criterio (r_{xy} promedio=.05).
- (c) en la condición baja-baja (*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 (*condición D*), las 7 variables predictoras correlacionaban significativamente entre sí ($\alpha \leq 0.05$;

r_{xy} promedio=.35), así como con el criterio ($\alpha \leq 0.05$; r_{xy} promedio=.35).

Los principales resultados que hallamos en nuestro trabajo, fueron:

- a) En la tarea de predicción de la variable criterio cuantitativa (Figura 1).
- b) En la tarea de clasificación de una variable criterio categorial binaria (Figura 2).
- c) En la tarea de clasificación de una variable criterio categorial no binaria (Figura 3).

De estos resultados se desprenden dos conclusiones básicas:

- En tareas de predicción las NN no superan nunca la efectividad de los ME, y de hecho bajo condiciones idóneas de aplicabilidad, el procedimiento clásico de regresión múltiple mostró ser significativamente mejor que las redes neurales.
- Por el contrario en las tareas de clasificación y en todas las condiciones las NN superan a los ME, incluida la condición a priori ideal para los ME.

Ello sugiere que las NN son mejores herramientas de clasificación que los ME, pero no mejores herramientas de predicción.

Esto por lo que respecta a datos simulados donde, recordemos, las variables cumplían, entre otros, los requisitos de normalidad multivariante e igualdad de las matrices de varianzas-covarianzas entre los distintos grupos de clasificación. ¿Se mantendría este patrón de resultados ante datos empíricos reales?

Para responder a esta pregunta utilizamos la matriz de datos de un trabajo de Psicología Social (González-Romá, Peiró, Lloret y Zornoza, 1999) donde 153 empleados de una agencia de administración respondieron a 66 ítems (escalas tipo Likert) que evaluaban 4 grandes rasgos: clima organizacional, satisfacción en el trabajo, involucración en el trabajo y compromiso con la organización. Estos ítems nos sirvieron como variables predictoras. Como variable dependiente tomamos las respuestas a un ítem categorial de 4 niveles que medía el bienestar psicológico general de dichos empleados.

Un análisis descriptivo de tales ítems nos mostró cómo algunos se distribuían normalmente, mientras que otros muchos mostraban una clara asimetría. Por su parte las correlaciones entre dichos ítems

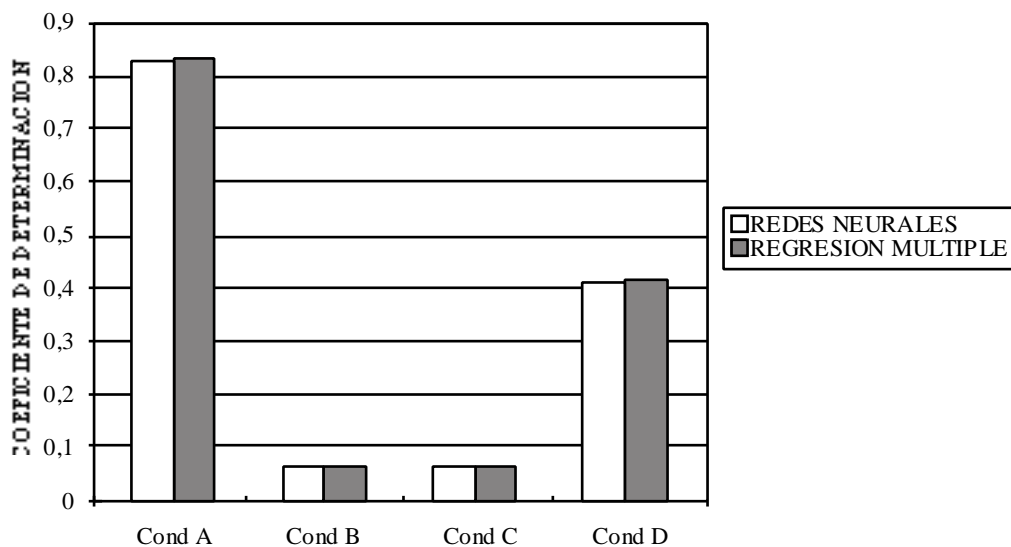


Figura 1. Coeficientes de determinación promedio de los modelos de regresión múltiple y redes neurales en las distintas condiciones en tareas de predicción cuantitativa del criterio

oscilaron entre 0 y .80 en valores absolutos. Estos resultados muestran, como preveíamos, que en investigación real difícilmente satisfacen los datos los supuestos teóricos de la estadística paramétrica.

Comenzamos confrontando la capacidad clasificatoria de los ME y las RN entrenándolos sobre el 100% de los datos de dicha matriz. En concreto confrontamos un ME de análisis discriminante (programa SSPS 7.0), con una red neural MLP (66/30/4) y una red RBF —función de base radial— (módulo Neural Connection 2.0 de SPSS) cuyos resultados aparecen abajo (figura 4).

Como se aprecia en la figura 4 una red MLP obtuvo los mejores resultados clasificando correctamente un 99% de los sujetos. Un ME de análisis discriminante obtuvo un 97.5% de clasificaciones correctas.

Contrastada la capacidad clasificatoria de dichos modelos probamos su capacidad de generalización ante nuevos registros.

Para ello entrenamos dichos modelos con el 90% de los registros anteriores (registros de entrenamiento) y dejamos que los propios modelos ya calculados clasificaran al 10% de los datos restantes (registros de validación). Los registros a formar parte de una u otra muestra fueron seleccionados aleatoriamente, replicando 10 veces tal proceso. Los resultados aparecen abajo (fig. 5).

Aunque, lógicamente la capacidad clasificatoria de estos modelos ante datos nuevos ha descendido, dicho descenso es proporcional a los resultados obtenidos en la figura 4.

De la consideración global de todos estos resultados caben entresacar una serie de conclusiones:

- En primer lugar las redes neurales no son sustitutos válidos de los ME en tareas de predicción, pues nunca consiguen superar la capacidad predictiva de éstos.

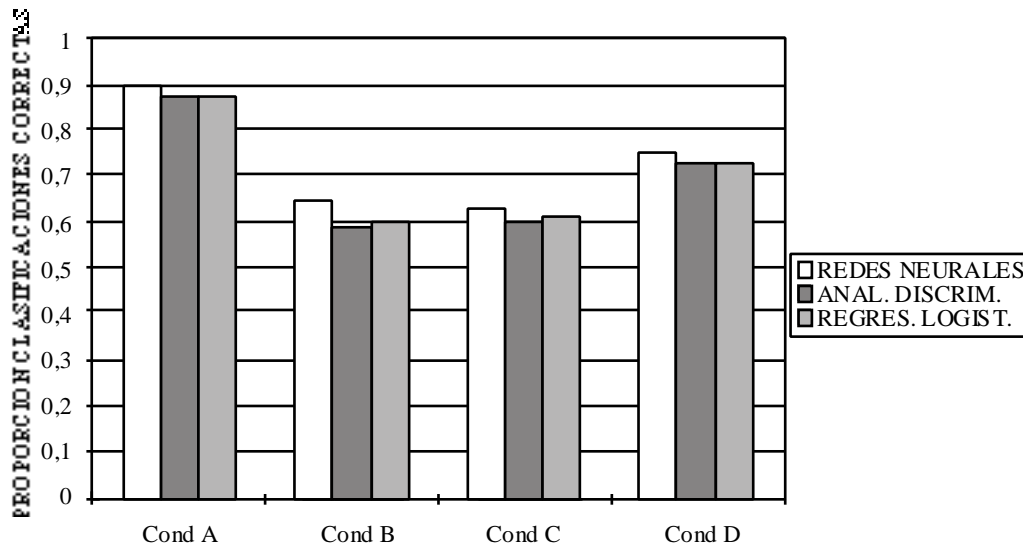


Figura 2. Proporciones medias de clasificaciones correctas llevadas a cabo por los modelos de clasificación binaria (redes neurales, análisis discriminante y regresión logística) en las distintas condiciones

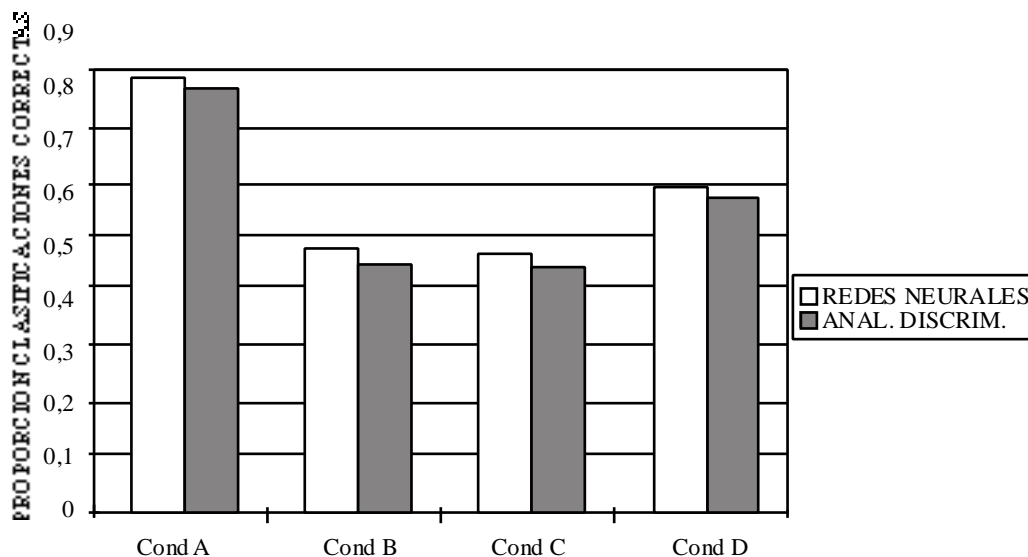


Figura 3. Proporciones medias de clasificaciones correctas llevadas a cabo por los modelos de clasificación no binaria (redes neurales y análisis discriminante) en las distintas condiciones

– En segundo lugar las NN (y en concreto las MLP) sí que muestran una capacidad clasificatoria mejor que los ME convencionales, independientemente de la naturaleza de las variables predictoras o de las relaciones que sustiendan entre ellas.

¿Quiere esto decir que propugnemos el abandono de la utilización de los ME convencionales en los problemas de clasificación? Las respuesta es 'no' por varias razones:

a) En primer lugar, gracias al fácil acceso por parte de los usuarios al software estadístico (SPSS, SYSTAT, STATISTICA, etc) los ME convencionales resultan hoy día más económicos de aplicar que las redes neurales en cuanto a los recursos temporales y computacionales involucrados. El usuario deberá pues plantearse si ganar unos pocos puntos porcentuales en capacidad clasificatoria compensa el mayor trabajo que deberá invertir.

b) Otro aspecto importante que deberá tener en cuenta es que así como un ME da una solución única ante unos mismos datos (siendo además la mejor solución posible, definida como el mejor ajuste a un modelo lineal), una RN no garantiza que la solución dada sea la óptima. De hecho deberá ir probando mediante un pro-

ceso «artesanal» distintas tipologías, número de unidades ocultas, funciones de activación, reglas de aprendizaje, etc., hasta dar con la solución adecuada. Como hemos podido comprobar nosotros mismos trabajando con el módulo Neural Connection aplicar una topología de NN con las opciones que por defecto sugiere el programa da resultados hasta 10 puntos porcentuales inferiores a los que se obtiene elaborando una red «artesanal». Ello por contrapartida, ya lo hemos señalado, requiere de grandes costos temporales.

c) No hay olvidar tampoco que una red neural no da información explícita sobre la importancia relativa de los distintos predictores, ni sobre el grado de ajuste del modelo a los datos.

La consideración de todo lo dicho nos lleva a sugerir con Cherkassky et al (1994), Sarle (1994, 1998) y otros muchos autores, que técnicas estadísticas y redes neurales pueden complementarse adecuadamente en la solución de tareas de predicción y clasificación, tal y como comienzan a utilizarse en determinadas aplicaciones técnicas (ver Martín y Sanz, 1997). De este modo, la estadística, centrada tradicionalmente en problemas lineales, y las redes neurales, más acostumbradas a tratar con problemas de categorías

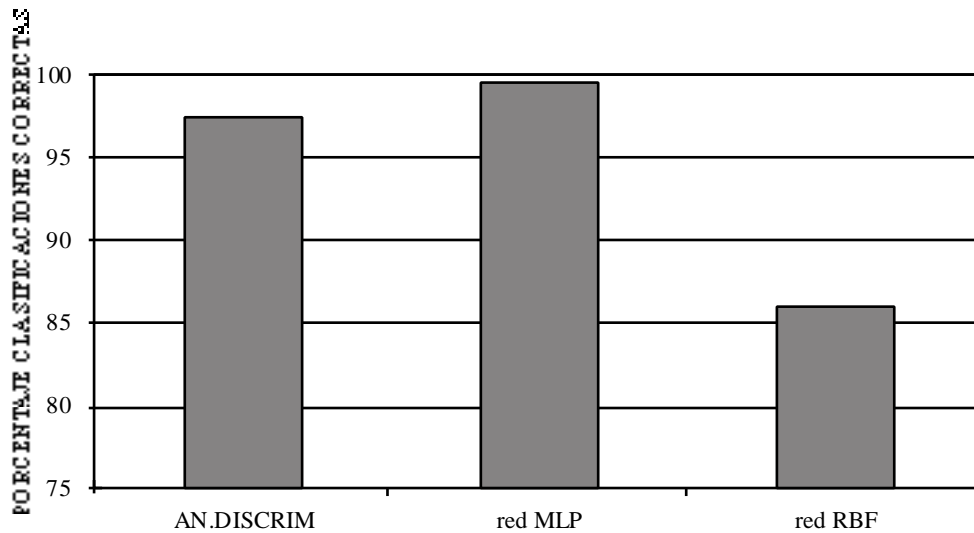


Figura 4. Tarea de clasificación de una variable categorial ante el 100% de los datos de entrenamiento

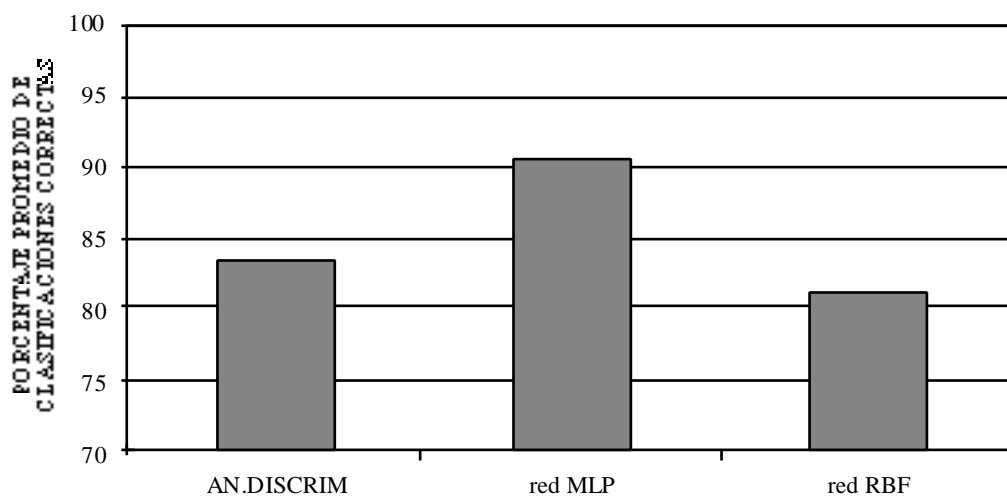


Figura 5. Tarea de clasificación de una variable categorial ante un 10% de datos nuevos

mal definidas, relaciones no lineales o datos con mucho «ruido» se verán mutuamente enriquecidas. No es de extrañar por ello que pa-

quetes estadísticos (SPSS, STATISTICA, MATLAB) ya incorporen en sus últimas versiones módulos de redes neurales.

Referencias

- Cherkassky, V., Friedman, J.H. y Wechsler, H. (eds.) (1994). *From statistics to neural networks*. Springer-Verlag, Berlín.
- Croall, I.F. y Mason, J.P. (1992). *Industrial applications of neural networks*. Springer-Verlag, N.Y.
- Garson, G.D. (1991). A comparison of neural network and expert systems algorithms with common multivariate procedures for analysis of social science data. *Social Science Computer Review*, 9, 399-434.
- González-Romá, V., Peiró, J.M., Lloret, S. y Zornoza, A. (1999). The validity of collective climates. *Journal of Occupational and Organizational Psychology*, 72, 25-40.
- Huang, W.Y. y Lippmann, R.P. (1987). Comparisons between neural nets and conventional classifiers. *Proceedings of the IEEE International Conference on Neural Networks*, 1, 485-494.
- Lim, T., Loh, W. Y Shih, Y. (1999). A comparison of prediction accuracy, complexity and training time of thirty-three old and new algorithms. *Machine Learning*, 1-27.
- Martín, B. y Sanz, A. (1997). *Redes neuronales y sistemas borrosos*. Rama. Madrid.
- 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.
- Pitarque, A., Roy, J.F. y Ruiz, J.C. (1998). Redes neuronales vs modelos estadísticos: Simulaciones sobre tareas de predicción y clasificación. *Psicología*, 19, 387-400.
- 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.
- Sarle, W.S. (1994). *Neural networks and statistical models*. Proceedings of the 19th Annual SAS Group conference, Cary, NC. pp. 1.538-1.550.
- Sarle, W.S. (1998). *Neural network FAQ. Acceso FTP, parts 1 to 7*. Acceso FTP: <ftp://ftp.sas.com/pub/neural/FAQ.html>
- Smith, M. (1993). *Neural networks for statistical modelling*. Van Nostrand Reinhold. 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.
- White, H. (1994). Parametric statistical estimation with artificial neural networks. En Smolensky, Mozer y Rumelhart (eds.), *Mathematical perspectives on neural networks*. Hillsdale, NJ: Erlbaum.