

“Medición de eficiencia técnica con redes neuronales: una simulación”

Daniel Santín González (dsantin@ccee.ucm.es)

Departamento de Economía Aplicada VI, Universidad Complutense de Madrid

Francisco J. Delgado Rivero (fdelgado@correo.uniovi.es)

Departamento de Economía, Universidad de Oviedo

Aurelia Valiño Castro (ecap316@sis.ucm.es)

Departamento de Economía Aplicada VI, Universidad Complutense de Madrid

Resumen

En este trabajo se revisa la aplicación de las redes neuronales artificiales para la aproximación de funciones frontera y la medición de eficiencia técnica. En la segunda parte se presenta un ejercicio de simulación para comparar diferentes técnicas empleadas en la literatura de medición de eficiencia y productividad. Los resultados alcanzados muestran cómo las redes neuronales artificiales constituyen una ampliación del instrumental disponible complementando al resto de técnicas, como modelos de fronteras estocásticas y análisis envolvente de datos, especialmente en contextos no lineales y/o desconocimiento de formas funcionales.

Palabras clave: redes neuronales, eficiencia técnica, simulación

1. Introducción

Las redes neuronales artificiales (RNA) son un conjunto de técnicas que han sido aplicadas con cierto éxito en muchos ámbitos diferentes al económico como reconocimiento de voz y caracteres, ingeniería, robótica, meteorología, etc., fundamentalmente a partir del trabajo de Rumelhart *et al.* (1986) que supuso el redescubrimiento del algoritmo *backpropagation* (BP) extensamente empleado en la literatura de RNA.

Las RNA tienen su origen en el estudio de los (complejos) mecanismos de funcionamiento del cerebro humano. Así, destaca originariamente el trabajo de McCulloch y Pitts (1943) que introdujeron modelos sencillos con neuronas de carácter binario según un umbral de referencia. Posteriormente, Rosenblatt (1958) planteó la estructura multicapa característica de las redes más empleadas. A pesar de su origen modelizador del cerebro, las RNA han sido empleadas fundamentalmente en aquellos

casos en que la teoría no aporta demasiada luz en relación a la forma funcional que recoja la relación entre las variables objeto de análisis. Así, las RNA permiten llevar a cabo el estudio de tales relaciones a partir de las observaciones, tratándose de un aprendizaje orientado por los datos (*data-based* frente a *model-based*).

En este trabajo se pretende en primer lugar proporcionar una breve introducción a las RNA y suministrar una panorámica de las aplicaciones de estas técnicas en el ámbito económico y, especialmente, en el campo de la eficiencia, donde consideramos que las RNA constituyen un instrumento de análisis para añadir a los habituales tratamientos econométricos o de análisis de envolvente de datos (DEA). En la sección tercera se presenta la simulación llevada a cabo para comparar diferentes técnicas empleadas en la literatura de medición de eficiencia y productividad. Los resultados alcanzados muestran cómo las redes neuronales artificiales constituyen una ampliación del instrumental disponible complementando al resto de técnicas, especialmente en contextos no lineales y/o desconocimiento de formas funcionales. El trabajo se cierra con las principales conclusiones alcanzadas.

2. Redes neuronales artificiales: una panorámica

En el año 1989 varios autores demostraron que las RNA eran “aproximadores universales” de funciones (Carroll y Dickinson, 1989; Cybenko, 1989; Funahashi, 1989; Hecht-Nielsen, 1989; Hornik *et al.*, 1989; White, 1990). Posteriormente se comprobó la aproximación de sus derivadas (Hornik *et al.*, 1990). Estos resultados dieron por fin sustento teórico a los prometedores resultados alcanzados en las aplicaciones. Una revisión actual de la propiedad de aproximación universal de las redes se recoge en Scarselli y Chung (1998).

Las RNA (White, 1989a; Hertz *et al.*, 1991; Cheng y Titterington, 1994; Kuan y White, 1994; Bishop, 1995; Ripley, 1996; Zapranis y Refenes 1999) aparecen generalmente estructuradas en tres capas de unidades elementales o neuronas: entrada, para introducir las variables input, intermedia u oculta, y salida, que proporciona las estimaciones de la red.

Entre los diferentes tipos de redes existentes la más extendida es la red *feedforward*, *backpropagation network* o perceptrón multicapa (MLP, *Multi-Layer Perceptron*). A partir del planteamiento general:

$$y = f(\mathbf{x}, \mathbf{q}) + \mathbf{e} \quad (1)$$

donde \mathbf{x} es el vector de variables explicativas, \mathbf{e} es el componente aleatorio (se supone de “buen comportamiento”, es decir, i.i.d. con media nula y varianza constante), $f(\mathbf{x}, \mathbf{q}) = \hat{y}$ es la función desconocida a estimar a partir de la información disponible, la red consiste en:

$$\hat{y} = f(\mathbf{x}, \mathbf{q}) = F \left(\mathbf{b}_o + \sum_{j=1}^m G(\mathbf{g}_j + \sum_{i=1}^n x_i \mathbf{a}_{ij}) \mathbf{b}_j \right) \quad (2)$$

donde \hat{y} output de la red, F la función de activación de la capa de salidas, normalmente lineal, G la función de activación de la capa intermedia, con frecuencia la función sigmoide o logística¹, n el número de inputs, m el número de unidades en la capa intermedia, \mathbf{x} el vector de inputs ($i=1\dots n$), y \mathbf{q} el vector de pesos o conexiones (parámetros), que incluye el sesgo del output, \mathbf{b}_o , los sesgos de las unidades intermedias ($j=1\dots m$) \mathbf{g}_j , ponderaciones de la unidad de input i a la unidad intermedia j , \mathbf{a}_{ij} , y pesos de las unidades intermedias en el output \mathbf{b}_j .

En la figura 1 se recoge una representación intuitiva del funcionamiento de la red MLP con 3 capas de unidades y un solo output, MLP (n,m,1):

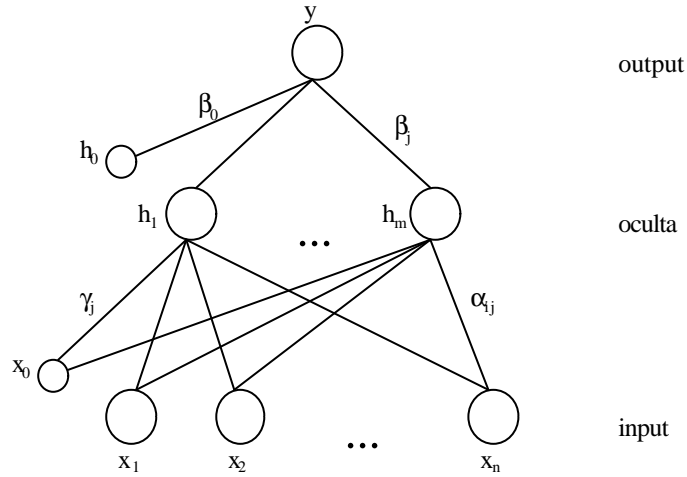


Figura 1. Perceptrón multicapa con tres capas y un output

¹ Las funciones de transferencia deben cumplir unas condiciones muy suaves para la aproximación universal. Normalmente se requieren funciones acotadas, diferenciables y monótonas crecientes.

También es posible establecer conexiones directas entre la capa de entradas y la de salidas, lo que implicaría la introducción de un término lineal, dando lugar a las llamadas “redes aumentadas” (*augmented single layer networks*):

$$\hat{y} = f(\mathbf{x}, \mathbf{q}) = F \left(\mathbf{b}_o + \sum_{i=1}^n x_i \mathbf{a}_i + \sum_{j=1}^m G(\mathbf{g}_j + \sum_{i=1}^n x_i \mathbf{a}_{ij}) \mathbf{b}_j \right) \quad (3)$$

La selección de la arquitectura de la red neuronal es una de las cuestiones más problemáticas y decisivas en las aplicaciones, y consiste básicamente en determinar posibles transformaciones de los datos², las variables input y su número, n , el número de unidades o nodos en la capa intermedia, m ³, la función de activación de la capa intermedia y de salida, y el subconjunto de ponderaciones distintas de cero (*pruning*)⁴.

Otro aspecto crucial de las RNA es el entrenamiento de la red o estimación del modelo, es decir, la búsqueda de los pesos o parámetros que minimicen alguna función de coste, habitualmente el error cuadrático. Sin duda, el protagonismo recae en el algoritmo *backpropagation* o algoritmo BP:

$$\mathbf{q}(k+1) = \mathbf{q}(k) + \mathbf{h} \nabla f(\mathbf{x}, \mathbf{q}) [y - f(\mathbf{x}, \mathbf{q})] \quad (4)$$

proceso iterativo (k indica la iteración), donde los parámetros se actualizan según el gradiente de la función de error a partir del coeficiente de aprendizaje \mathbf{h} hasta alcanzar el criterio de parada establecido, normalmente un cierto número (elevado) de iteraciones o un error admisible. Por tanto, las RNA son capaces de deducir los patrones de comportamiento de un fenómeno (complejo) a partir de un conjunto de observaciones o “aprendizaje de la experiencia”, y tanto sus ventajas como inconvenientes o limitaciones deben ser (re)conocidas por el investigador (figura 2).

Las RNA son modelos adaptativos en cuyo aprendizaje apenas se establecen supuestos sobre las relaciones entre las variables, por lo que a priori son un buen instrumento cuando se desconoce la naturaleza de tales relaciones (método no paramétrico), o tales relaciones son no lineales (método no lineal), siempre que se disponga de suficientes observaciones (formas flexibles).

² En general se llevan a cabo transformaciones de los datos iniciales: normalización a [0,1] ó [a,b], diferenciación y/o desestacionalización en series temporales, etc.

³ Las neuronas intermedias se determinan a través de procesos de “prueba y error” de forma que se procede al entrenamiento de la red considerando $m = 1, 2, 3, 4, \dots$

⁴ Por ejemplo, el análisis para la eliminación de inputs o unidades intermedias, o algunos enlaces entre ellos, “irrelevantes” desarrollado en White (1989b).

Sin embargo, la flexibilidad conlleva la facilidad para modelizar no sólo la señal sino también el ruido, inherente a la mayor parte de las series de datos. A menudo se requiere gran cantidad de datos, por lo que será necesario adecuar el tamaño de la red a las observaciones disponibles para mantener suficientes grados de libertad. Además, es frecuente citar las RNA como “cajas negras” ya que consisten en un complejo proceso de “prueba y error”.

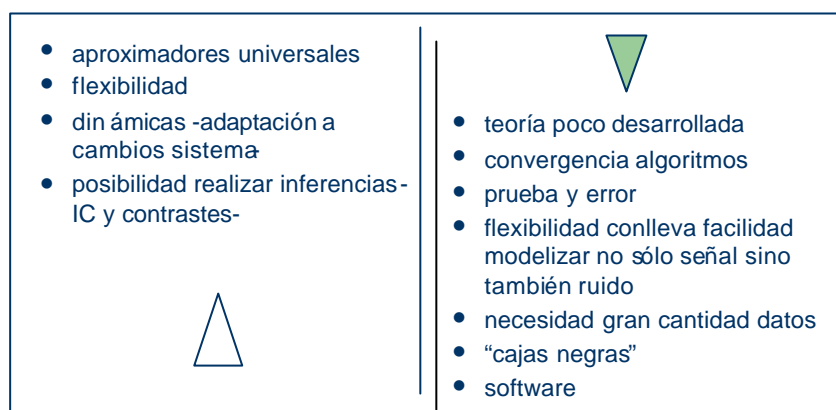


Figura 2. Ventajas e inconvenientes RNA

3. Redes neuronales en aproximación de funciones y medición de eficiencia

El análisis de la eficiencia en producción (Farrell, 1957) constituye un área de constante interés en Economía, al tratar el uso adecuado de los (escasos) recursos con la tecnología disponible. Frente a este concepto de eficiencia técnica⁵, surge la eficiencia asignativa cuando la tecnología varía y se toma la mejor combinación de inputs. Farè *et al.* (1987), Fried *et al.* (1993), Coelli *et al.* (1998) y Álvarez (2001) constituyen interesantes obras de referencia general en eficiencia.

En este análisis resulta crucial la estimación de la función frontera, que tradicionalmente se ha realizado por dos grandes vías, paramétricas y no paramétricas:

- Técnicas paramétricas, suponen adoptar una determinada forma funcional como Cobb-Douglas, translog, CES, generalizada de Leontief.... A su vez estas técnicas pueden ser determinísticas y estocásticas (Aigner *et al.* 1977):

* Determinísticas: toda desviación de la frontera se explica por ineficiencia.

⁵ La eficiencia técnica puede descomponerse en eficiencia de escala, eficiencia de congestión y eficiencia técnica pura (Farè *et al.*, 1985).

* Estocásticas: las desviaciones de la frontera se descomponen en componente aleatorio e ineficiencia, para lo que es necesario suponer una distribución, habitualmente semi-normal.

Las estimaciones pueden realizarse a través de COLS (*corrected ordinary least squares*), corrigiendo el término independiente por el máximo residuo positivo de la estimación MCO inicial, y máxima verosimilitud.

- Técnicas no paramétricas o de programación matemática, que no suponen una forma funcional concreta, entre las que destaca por su masiva utilización⁶:

* Análisis Envolvente de Datos (DEA), Charnes *et al.* (1978), emplea la programación lineal para configurar la frontera, normalmente determinística, que “envuelve” las observaciones disponibles adoptando supuestos sobre rendimientos a escala, constantes o variables, disponibilidad fuerte de inputs y outputs, y convexidad del conjunto de posibles combinaciones input/output.

A partir de la expresión habitual:

$$y_i = f(\mathbf{x}_i, \mathbf{q}) + \mathbf{e}_i - u_i \quad (5)$$

donde $u_i \geq 0$ recoge la ineficiencia técnica, se trata de asumir la expresión de la red neuronal (2) para la estimación de la frontera. Costa y Markellos (1997) plantean dos opciones para la determinación de la frontera con RNA: a) una vez estimada la red, proceder de igual modo que en COLS; b) mediante una red “sobreparametrizada” capaz de ajustarse de modo creciente a las observaciones, hasta alcanzar un cierto ratio señal/ruido. Posteriormente, la ineficiencia se determina por distancia observación-frontera.

En este contexto, las RNA suponen ampliar el conjunto de alternativas disponibles teniendo en cuenta que se trata de estimación no paramétrica (*free-model*), una técnica estocástica y la posibilidad de llevar a cabo inferencias como la construcción de intervalos de confianza⁷ para los índices de (in)eficiencia. En la figura 3 se representan genéricamente las posibilidades de estimación de la frontera. En ella se trata de recoger el carácter lineal de la aproximación econométrica y la gran flexibilidad de la RNA,

⁶ Otra técnica no paramétrica menos utilizada es *Free Disposal Hull* (FDH).

⁷ La construcción genérica de intervalos de confianza y de predicción en RNA puede verse en Hwang y Ding (1997), De Veaux *et al.* (1998) y Rivals y Personaz (2000), aunque no ha sido utilizada aún en el contexto de eficiencia.

capaz de detectar “áreas de congestión” caracterizadas por caídas del output ante incrementos del input. Por el contrario, las RNA carecen de estudios teóricos en este ámbito y presenta otros inconvenientes como la interpretación de los resultados y el requisito de recursos técnicos para su estimación. Una visión comparada de las alternativas se sintetiza en la tabla 2 (basada en Costa y Markellos, 1997), de la que se obtiene la esperada conclusión de que ninguna de las alternativas es globalmente superior al resto, de forma que las características de cada problema particular determinarán la elección.

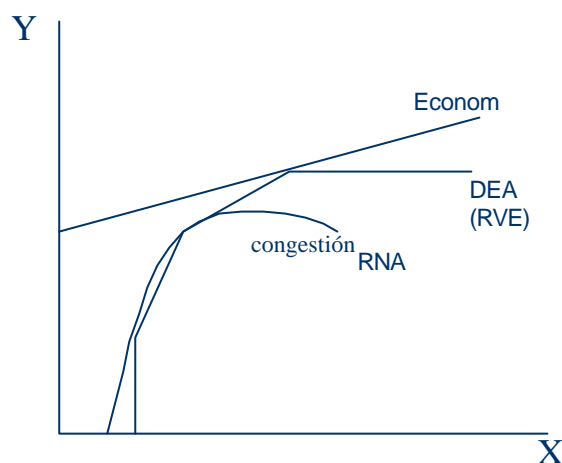


Figura 3. Estimación de funciones frontera

Tabla 2. Comparativa de las alternativas en medición de eficiencia

Criterio	Econométrica	DEA	RNA
Supuestos: forma funcional, datos...	Alto	Medio	Bajo
Flexibilidad	Bajo-Medio	Medio	Alto
Fundamento teórico	Alto	Alto	Bajo-Medio
Aplicaciones en Eficiencia	Alto	Alto	Bajo
Significación estadística	Sí	No	Si
Interpretación de resultados	Medio	Bajo	Medio
Estimaciones / proyecciones	Alto	No	Medio
Coste: software, tiempo estimación...	Bajo	Bajo	Alto

A continuación, se recoge una selección de trabajos publicados en esta materia, con una síntesis de los planteamientos y resultados alcanzados:

- Joerding *et al.* (1994), en el ámbito de funciones de producción. Estos autores estudian la imposición de propiedades sobre la tecnología productiva (positiva, monótona, cuasi-cóncava) a nivel teórico a través de restricciones en los pesos de la red, y concluyen que las RNA presentan unas características similares a las formas flexibles de Fourier. También analizan la estimación simultánea de función de producción y sistema de demanda de inputs. Sostienen que no es posible imponer rendimientos constantes a escala para todo el recorrido, ya que exigiría funciones de activación lineales, invalidando la aproximación universal.
- Costa y Markellos (1997) estudian la eficiencia en el sector del transporte, en concreto para el metro de Londres con datos anuales sobre 2 inputs, flota y trabajadores, y 1 output, kms. recorridos. Estos autores crean una muestra sintética para la estimación de la frontera añadiendo ruido $N(0, \sigma^2)$ a los datos iniciales. Concluyen que las RNA presentan resultados similares a COLS y DEA, pero presentan ventajas como facilitar la toma de decisiones, estudiar el impacto de rendimientos variables a escala versus rendimientos constantes, detección de áreas congestión, etc.
- Guermat y Hadri (1999) estudian las funciones frontera estocásticas mediante simulación Monte Carlo. Los datos (2 inputs y 1 output) son generados a partir de diversas funciones: Cobb-Douglas, CES y generalizada Leontief, y consideran varias funciones: RNA, Cobb-Douglas, translog, CES y generalizada de Leontief. Las RNA superan a la translog y Cobb-Douglas, si bien los resultados no son satisfactorios cuando los datos se generan a partir de funciones de Leontief o CES por problemas de especificación. Además la incorrecta especificación de la forma funcional (con RNA y translog) no afecta significativamente a la eficiencia media, máxima y mínima, pero sí a ineficiencias puntuales y ordenaciones.
- Santín y Valiño (2000) estudian la eficiencia en educación con un modelo en dos niveles, estimando en el primer nivel, de estudiante, la función de producción a través de una red neuronal. En su aplicación con datos de 7454 estudiantes y 12 inputs, las RNA superan los resultados de la aproximación econométrica en la estimación de la frontera.
- Fleissig *et al.* (2000) en el marco de funciones de coste realizan una comparación entre RNA, forma flexible Fourier, AIM –*asymptotically ideal model*–, translog y

Leontief generalizada. Los datos fueron simulados a partir de funciones CES y Box-Cox generalizadas. Sus conclusiones apuntan a que las RNA se comportan peor que Fourier al no imponer simetría y homogeneidad, características sí incorporadas en Fourier y AIM, y detectan dificultades de convergencia en la imposición de homogeneidad y simetría en las RNA.

En síntesis, consideramos que las redes neuronales constituyen un instrumento potencialmente útil para el análisis de la eficiencia. En especial serán relevantes cuando existan pocos indicios sobre la forma de la función de producción, o se detecten comportamientos no lineales. Así, en la construcción de funciones frontera, los resultados obtenidos se asemejan o superan a los correspondientes a la habitualmente empleada función translog, aunque no se han analizado aún cuestiones como por ejemplo los efectos de observaciones anómalas o outliers sobre la frontera estimada con la red neuronal. La disponibilidad de datos suficientes para poder estimar redes de cierta complejidad será clave para la aplicación de las RNA. En otro caso, será necesario aplicar mecanismos apropiados para la creación de muestras de mayor tamaño a partir de la original, como los métodos *bootstrap*.

Finalmente, resaltar la necesidad de avanzar tanto en el aspecto teórico como empírico para comprobar si finalmente es “eficiente” el uso de estas complicadas técnicas al campo de la eficiencia. La respuesta se obtendrá al evaluar los costes, como la complejidad de los modelos y algoritmos, recursos informáticos, interpretación económica, etc., y los posibles beneficios, como mejores resultados, mejoras en interpretación de los mismos, flexibilidad, etc.

4. Simulación

Sea $F(x)$ la siguiente función de producción continua con un input y un output:

$$F(x) = \begin{cases} \left(\frac{x}{e}\right)^2 & \text{si } x \in [0, e] \\ \ln(x) & \text{si } x \in [e, e^2] \\ A * \cos(x - e^2) + 2 - A & \text{si } x \in [e^2, e^2 + \pi], \text{ donde } A = 0.25 \\ \ln(x - 2\pi) & \text{si } x \in [e^2 + \pi, 26] \end{cases} \quad (6)$$

En esta función (figura 4) incorporamos todas las posibilidades de rendimientos a escala: en el primer tramo rendimientos crecientes (IRS), en el segundo y cuarto rendimientos decrecientes (DRS) y en el tercero una situación en la que aumentos de input significan reducciones de output, interpretado como área congestionada⁸.

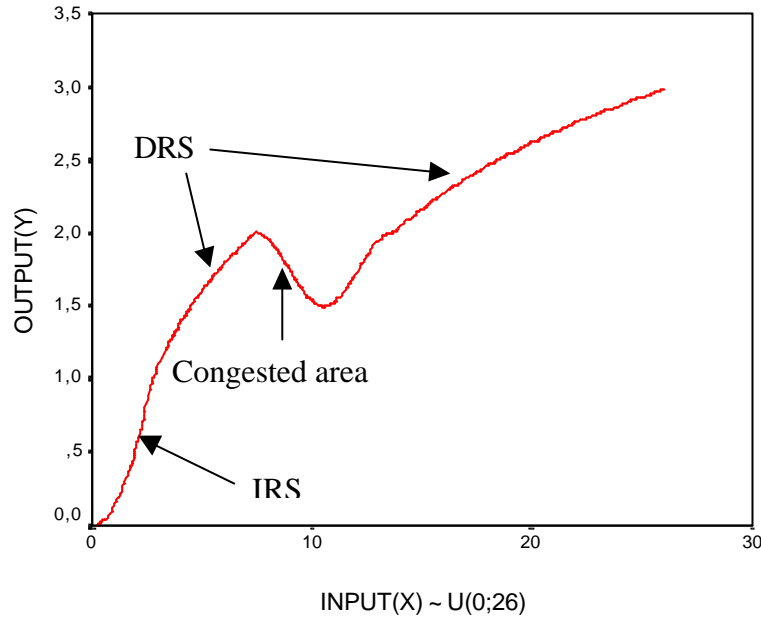


Figura 4. Función de producción no lineal

En el experimento se consideran 4 escenarios con 50, 100, 200 y 300 unidades (DMUs). Para cada escenario se generan números pseudo-aleatorios uniformemente distribuidos en el espacio de inputs:

$$X \sim U(0,26) \quad (7)$$

Después se calcula el verdadero output, que es la verdadera frontera de producción de la figura 4, y se generan ineficiencias inyectando distintas cantidades de ruido al output:

$$y^* \sim U(y + ay, y - by) \quad (8)$$

donde y^* es el output observado, $a= 0.05$ si $b= 0.1, 0.2, 0.3$; y $a=0.15$ si $b= 0.35, 0.6$, y se mide la verdadera eficiencia técnica (te) como:

$$te = (y^* / y) \quad (\text{se permite } te > 1 \text{ para recoger outliers}) \quad (9)$$

Por simplicidad, se asume que los datos están libres de ruido y todas las diferencias entre el output verdadero y el estimado son ineficiencias.

⁸ Esta situación de congestión puede presentarse por ejemplo en los ámbitos de educación y transporte (Costa y Markellos, 1997).

Para cada escenario se computa la eficiencia técnica mediante OLS y COLS con SPSS, frontera estocástica SFA con FRONTIER 4.1 (Coelli, 1996b), DEA con rendimientos constantes DEAcrs y variables DEAvrs con DEAP 2.1 (Coelli, 1996a), y la red MLP con S-PLUS.

Inicialmente, los datos son divididos en dos conjuntos, entrenamiento (80%) y validación (20%). Sin embargo, se ha detectado un comportamiento casi idéntico en ambos casos por lo que se han considerado conjuntamente. Para cada escenario se han entrenado redes con 3 a 8 neuronas en la capa oculta con un coeficiente de aprendizaje de 0.5 y *weight decay* 0.001, deteniendo el proceso tras 500 iteraciones para evitar el sobreajuste.

La comparación de técnicas se ha llevado a cabo mediante el coeficiente de correlación de Pearson entre las medidas de eficiencia estimadas y reales (tabla 3). En todos los casos excepto uno, la RNA presenta mejores resultados, y alcanza estimaciones robustas a número de unidades y ruido inyectado. La RNA es superior a técnicas tradicionales cuando se somete a la tecnología a ruido moderado y suficientes observaciones. Sin embargo, el análisis DEA con rendimientos variables supera a la RNA cuando existe gran cantidad de ruido y pocas observaciones.

Tabla 3. Coeficientes de correlación de Pearson entre eficiencias reales y estimadas

Efficiency Techniques							
Number of DMUs & Percentage of noise injected	50 DMUs	OLS	COLS	SF	DEAcrs	DEAvrs	MLP BEST
	50(15)	0.180	0.104	0.441	0.297	0.431	0.788
	50(25)	0.230	0.249	0.294	0.119	0.296	0.838
	50(35)	0.464	0.405	0.581	0.419	0.714	0.804
	50(50)	0.584	0.575	0.630	0.378	0.798	0.873
	50(75)	0.608	0.520	0.443	0.473	0.895	0.887
	100 DMUs	OLS	COLS	SF	DEAcrs	DEAvrs	MLP BEST
	100(15)	0.145	0.146	0.096	0.090	0.183	0.897
	100(25)	0.255	0.211	0.239	0.286	0.293	0.751
	100(35)	0.297	0.237	0.332	0.357	0.498	0.919
	100(50)	0.496	0.490	0.321	0.345	0.661	0.951
	100(75)	0.557	0.517	0.474	0.543	0.728	0.855
	200 DMUs	OLS	COLS	SF	DEAcrs	DEAvrs	MLP BEST
	200(15)	0.184	0.205	0.139	0.076	0.249	0.816
	200(25)	0.326	0.322	0.258	0.187	0.439	0.961
	200(35)	0.377	0.329	0.280	0.348	0.479	0.947
	200(50)	0.554	0.557	0.331	0.365	0.686	0.924
	200(75)	0.685	0.705	0.337	0.483	0.794	0.934
	300 DMUs	OLS	COLS	SF	DEAcrs	DEAvrs	MLP BEST
	300(15)	0.214	0.248	0.029	0.026	0.302	0.887
	300(25)	0.374	0.332	0.388	0.280	0.457	0.935
	300(35)	0.447	0.409	0.417	0.316	0.587	0.975
	300(50)	0.606	0.607	0.663	0.319	0.736	0.935
	300(75)	0.759	0.722	0.804	0.541	0.857	0.973

Un caso particular se muestra en la figura 5, para 300 observaciones y un 25% de ruido uniforme en el output real. Se observa que la RNA capta la no linealidad de los datos, y es posible construir una frontera desplazando hacia arriba la curva en el mayor residuo como en COLS. La figura 5 también pone de relieve que la RNA es un buen instrumento para el análisis exploratorio de relaciones no lineales como paso previo al estudio de eficiencia y así evitar problemas de especificación.

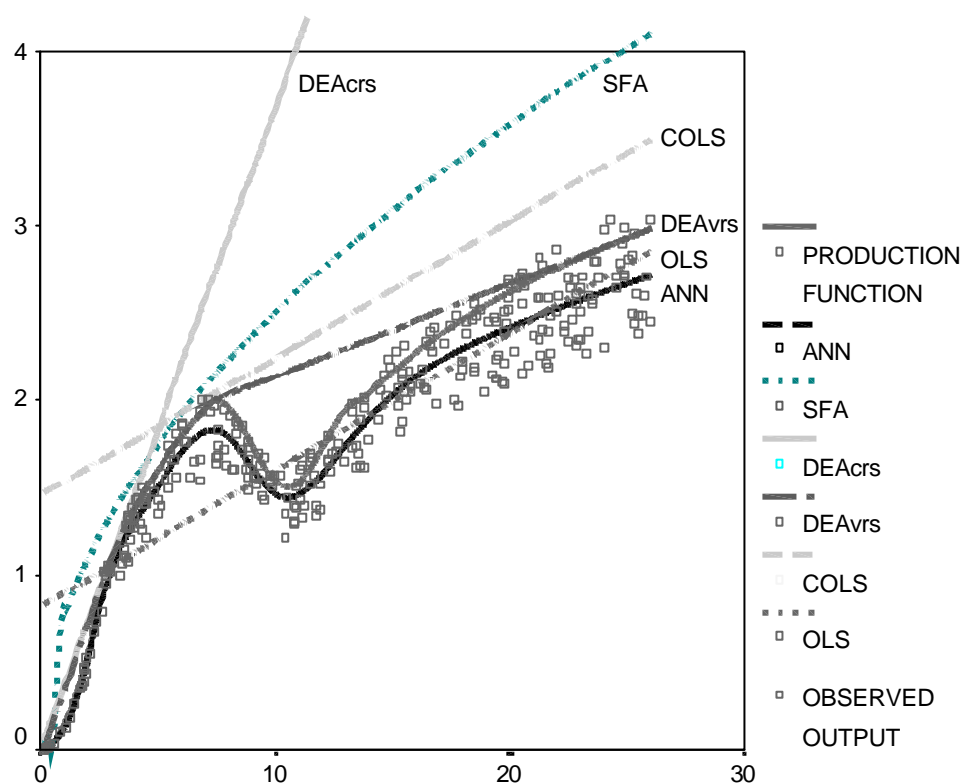


Figura 5. Funciones de producción estimadas

4. Conclusiones

Las RNA son técnicas muy flexibles aplicables en una variedad de campos entre ellos el económico. En este trabajo se han revisado las experiencias de las RNA en el ámbito de la producción y la medición de eficiencia de las que se deduce su importante papel como técnica complementaria a las tradicionales, como las técnicas de fronteras estocásticas y el análisis envolvente de datos (DEA).

En este trabajo se ha simulado una función de producción no lineal con diferentes rendimientos a escala: crecientes, decrecientes y áreas de congestión, y se han comparado técnicas estocásticas, envolvente de datos y redes neuronales. Los resultados de la simulación avalan la relevancia de las redes neuronales en el contexto de eficiencia, conscientes de que ninguna de las técnicas es siempre superior al resto. Las RNA tendrán un papel fundamental cuando se desconozca la forma funcional y así evitar problemas asociados a incorrecta especificación, y también en aquellos casos en que se detecten comportamientos no lineales, siendo recomendable un análisis (contraste) previo de no linealidad.

Referencias

- Aigner, D.J., C.A.K. Lovell y P. Schmidt (1977), "Formulation and estimation of stochastic frontier production function models, *Journal of Econometrics*, 6, 21-37.
- Álvarez, A. (2001), La medición de la eficiencia y la productividad, Ed. Pirámide.
- Bishop, C.M. (1995), *Neural Networks for Pattern Recognition*, Clarendon Press, Oxford.
- Carrol, S. y B. Dickinson (1989), "Construction of Neural Networks using the Rado Transform, *IEEE International Conference on Neural Networks*, 1, 607-611, Washington.
- Charnes A, W. W. Cooper y E. Rhodes (1978), "Measuring the Efficiency of Decision Making Units", *European Journal of Operational Research*, 2, 429-444.
- Coelli, T. (1996a), *A Guide to DEAP Version 2.1: A Data Envelopment Analysis Program*, Centre for Efficiency and Productivity Analysis (CEPA), Working Paper 96/08.
- Coelli, T. (1996b), *A Guide to FRONTIER Version 4.1: A Computer Program for Stochastic Frontier Production and Cost Function Estimation*, Centre for Efficiency and Productivity Analysis (CEPA), Working Paper 96/07.
- Coelli, T., D.S. Prasada Rao y G.E. Battese (1998), *An introduction to efficiency and productivity analysis*, Kluwer Academic Publishers.
- Cheng, B. y D.M. Titterington (1994), "Neural Networks: a Review from a Statistical Perspective", *Statistical Science*, 9 (1), 2-54.
- Costa, A. y R.N. Markellos (1997), "Evaluating Public Transport Efficiency with Neural Network Models", *Transportation Research C*, 5 (5), 301-312.
- Cybenko, G. (1989), "Approximation by Superpositions of a Sigmoidal Function", *Mathematics of Control, Signals and Systems*, 2, 303-314.
- De Veaux, R.D., J. Schumi, J. Schweinsberg y L.H. Ungar (1998), "Prediction Intervals for Neural Networks via Nonlinear Regression", *Technometrics*, 40 (4), 273-282.
- Farë, R., S. Grosskopf y C.A.K. Lovell (1985), *The Measurement of Efficiency of Production*, Kluwer, Boston.

- Farrell, M.J. (1957), "The Measurement of Productive Efficiency", *Journal of the Royal Statistical Society*, 120, 253-281.
- Fleissig, A.R., T. Kastens y D. Terrell (2000), "Evaluating the Semi-nonparametric Fourier, AIM, and Neural Networks Cost Functions", *Economics Letters*, 68, 3, 235-244.
- Fried, H.O., C.A. Lovell y S.S. Schmidt (1993), *The Measurement of Productive Efficiency*, Oxford University Press, Oxford.
- Funahashi, K. (1989), "On the Approximate Realization of Continuous Mappings by Neural Networks", *Neural Networks*, 2, 183-192.
- Geman, S., E. Bienenstock y R. Doursat (1992), "Neural Networks and the Bias/Variance Dilemma", *Neural Computation*, 4, 1-58.
- Guermat, C. y K. Hadri (1999), *Backpropagation Neural Network Vs Translog Model in Stochastic Frontiers: a Monte Carlo Comparison*, Discussion Paper 99/16, University of Exeter.
- Hecht-Nielsen, R. (1989), "Kolmogorov's Mapping Neural Network Existence Theorem", *International Joint Conference on Neural Networks*, 3, 11-14, Washington.
- Hertz, J., A. Krogh y R.G. Palmer (1991), *Introduction to the Theory of Neural Computation*, Addison-Wesley.
- Hornik, K., M. Stinchcombe, y H. White (1989), "Multilayer Feedforward Networks are Universal Approximators", *Neural Networks*, 3, 551-560.
- Hornik, K., M. Stinchcombe, y H. White (1990), "Universal Approximation of an Unknown Mapping and its Derivatives Using Multilayer Feedforward Networks", *Neural Networks*, 3, 551-560.
- Hwang, J.T.G. y A.A. Ding (1997), "Prediction Intervals for Artificial Neural Networks", *Journal of the American Statistical Association*, 92 (438), 748-757.
- Joerding, W., Y. Li, S. Hu y J. Meador (1994), "Approximating production technologies with feedforward neural networks", en J.D. Johnson y A.B. Whinston (eds): *Advances in artificial intelligence in Economics, Finance and Management*, 1, 35-42, JAI Press, London.
- Kuan, C.M. y H. White (1994), "Artificial Neural Networks: an Econometric Perspective", *Econometric Reviews*, 13, 1-91.
- Kuan, C.M. y T. Liu, (1995) "Forecasting Exchange Rates Using Feedforward and Recurrent Neural Networks", *Journal of Applied Econometrics*, 10, 347-364.
- Lee, T.-H., H. White y C.W.J. Granger (1993), "Testing for neglected nonlinearity in time series models. A comparison of neural network methods and alternative test", *Journal of Econometrics*, 56, 269-290.
- McCulloch, W.S. y W. Pitts (1943), "A Logical Calculus of the Ideas Immanent in Nervous Activity", *Bulletin of Mathematical Biophysics*, 5, 115-133.
- Ripley, B.D. (1996), *Pattern Recognition and Neural Networks*, Cambridge University Press.
- Rivals, I. y L. Personnaz (2000), "Construction of Confidence Intervals for Neural Networks Based on Least Squares Estimation", *Neural Networks*, 13, 463-484.

- Rosenblatt, F. (1958), "The Perceptron: a Probabilistic Model for Information Storage and Organization in the Brain", *Psychological Reviews*, 62, 386-408.
- Rumelhart, D., G. Hinton y R. Williams (1986), "Learning Internal Representations by Error Propagation", en D. Rumelhart y J. McClelland, (eds), *Parallel Distributed Processing: Explorations in the Microstructures of Cognition*, 1, 318-362, MIT Press, Cambridge.
- Santín, D. y A. Valiño (2000), "Artificial Neural Networks for Measuring Technical Efficiency in Schools with a Two-Level Model: an Alternative Approach", *II Oviedo Workshop on Efficiency and Productivity*, Oviedo.
- Santín, D. y A. Valiño (2001), "Comparing Performance of Efficiency Techniques in Non-Linear Production Functions", *7th European Workshop On Efficiency And Productivity Analysis*, Oviedo.
- Santín, D., F.J. Delgado y A. Valiño (2001), *Measuring technical efficiency with neural networks: a review*, Efficiency Series Paper, Universidad de Oviedo.
- Scarselli, F. y A. Chung (1998), "Universal Approximation Using Feedforward Neural Networks: a Survey of Some Existing Methods, and Some New Results", *Neural Networks*, 11 (1), 15-37.
- White, H. (1988), "Economic Prediction Using Neural Networks: the Case of IBM Daily Stock Returns", *Proceedings of the IEEE International Conference on Neural Networks*, San Diego, 451-458.
- White, H. (1989a), "Learning in Artificial Neural Networks: a Statistical Perspective", *Neural Computation*, 1, 425-464.
- White, H. (1989b), "Some Asymptotic Results for Learning in Single Hidden-Layer Feedforward Network Models", *Journal of the American Statistical Association*, 84 (408), 1003-1013.
- White, H. (1990), "Connectionist Nonparametric Regression: Multilayer Feedforward Networks Can Learn Arbitrary Mappings", *Neural Networks*, 3, 535-549.
- Zapranis, A. y A-P Refenes (1999), *Principles of Neural Model Identification, Selection and Adequacy. With Applications to Financial Econometrics*, Springer.
- Zhang, Y. y Bartels, R. (1998), "The Effect of Sample Size on the Mean Efficiency in DEA with an Application to Electricity Distribution in Australia, Sweden and New Zealand". *Journal of Productivity Analysis*, 9, 187-204.
- Zhang, G., B.E. Patuwo y M.Y. Hu (1998), "Forecasting with Artificial Neural Networks: the State of the Art", *International Journal of Forecasting*, 14, 35-62.