

Autor(es): **Daniel Santín González**

Título: **Detección de alumnos de riesgo y medición de la eficiencia de centros escolares mediante redes neuronales**

Resumen:

En este trabajo, de aproximación a la tesis doctoral que el autor está elaborando, se utilizan las redes neuronales para intentar solucionar empíricamente tres problemas habituales en el campo de la economía de la educación: Detección de alumnos con alto riesgo de fracaso escolar, evaluación de la eficiencia relativa de centros educativos y evaluación de la eficiencia relativa de diferentes planes de estudio.

En una primera parte, se realiza una introducción a los modelos neuronales, con especial atención a los modelos con conexiones hacia delante (feed-forward) con algoritmo de aprendizaje de retropropagación supervisada (backpropagation), utilizados en el ejercicio empírico desarrollado. A continuación, se comentan los problemas planteados en este trabajo y se tratan de resolver mediante un modelo basado en las redes neuronales.

Se han aplicado las redes neuronales para la resolución de los diferentes problemas expuestos en datos individuales de 7454 alumnos de enseñanzas medias de octubre de 1984 (1985) en 256 centros públicos y privados. Los datos incluyen estudiantes que cursan el programa de Reforma de las Enseñanzas Medias (grupo experimental), alumnos de BUP y FP en los mismos centros (grupo de control interno), y alumnos de estos programas en centros no experimentales (grupo de control externo).

Finalmente y a modo de conclusión, se comentan los resultados obtenidos y su relevancia dentro del marco teórico de la economía de la educación.

DETECCIÓN DE ALUMNOS DE RIESGO Y MEDICIÓN DE LA EFICIENCIA DE CENTROS ESCOLARES MEDIANTE REDES NEURONALES

Daniel Santín González

Departamento de Economía Aplicada VI (Hacienda Pública).

Universidad Complutense de Madrid

E-mail: ecap6z9@sis.ucm.es

ÍNDICE

1.- INTRODUCCIÓN.

2.- FUNDAMENTOS DE LAS REDES NEURONALES.

2.1.- Características básicas de las redes neuronales.

2.1.1.- Redes neuronales de tipo biológico.

2.1.2.- La neurona artificial simple.

2.1.3.- Estructura de la red neuronal.

2.1.4.- Mecanismos de entrenamiento-aprendizaje.

2.2.- El Perceptrón multicapa.

2.3.- Funcionamiento del algoritmo Backpropagation.

2.4.- Algunos comentarios prácticos sobre el perceptrón multicapa con backpropagation.

2.5.- Algunas ventajas e inconvenientes de las redes neuronales.

3.- APLICACIÓN EMPÍRICA DE LAS REDES NEURONALES EN ECONOMÍA DE LA EDUCACIÓN: DESCRIPCIÓN DE LOS 3 PROBLEMAS PLANTEADOS Y DE LOS DATOS UTILIZADOS.

3.1.- Detección de alumnos de riesgo mediante redes de neuronas.

3.2.- Medición de la eficiencia relativa de los centros escolares.

3.3.- Medición de la eficiencia relativa de planes de estudios distintos.

4.- CONCLUSIONES.

BIBLIOGRAFÍA

1.- INTRODUCCIÓN

El cerebro humano recibe continuamente, a través de los sentidos, miles de inputs, con el fin de procesarlos en su red de neuronas biológicas y emitir en todo momento la respuesta más adecuada a cada circunstancia para el desarrollo normal de la vida. Una red de neuronas artificiales es una herramienta que nace inspirada en este funcionamiento, con el fin de aprender a asociar vectores inputs a vectores outputs a partir de ejemplos reales. Las redes neuronales están formadas por tanto por gran cantidad de elementos de proceso interconectadas (las neuronas) capaces de aprender mediante el ajuste de los parámetros que componen la

red (pesos sinápticos).

El principal rasgo de esta tecnología es que puede ser y ha sido aplicada a un inmenso conjunto de problemas, pertenecientes a diversas áreas del conocimiento, muchos de los cuales parecían ser demasiado complicados de resolver en modelos teóricos. Las redes neuronales han sido aplicadas con éxito, por ejemplo, en el campo de la inteligencia artificial para el reconocimiento de voz y de escritura, o en el sector militar para el procesamiento de señales (radar, sonar, infrarrojos) para la identificación de objetivos. Asimismo, en el área económica algunas de las aplicaciones más difundidas han sido: en modelos de inversión a fin de optimizar carteras, en el mundo del marketing como instrumento para vender productos a clientes potenciales, en entidades financieras para predecir el riesgo potencial de los créditos que le son solicitados y en la predicción de tipos de cambio de divisas e índices económicos. En general podemos decir que las aplicaciones más desarrolladas en los últimos años han sido todas las relacionadas con el análisis de grandes masas de datos, con vistas a la predicción, clasificación, optimización, aproximación de funciones y la agrupación o el reconocimiento de patrones.

El objetivo de este trabajo es utilizar las redes de neuronas artificiales para la resolución de problemas específicos en el ámbito de la economía de la educación. Para este fin, la primera parte del trabajo será exponer los fundamentos básicos de las redes neuronales. El objetivo de este punto no es llevar a cabo una revisión detallada de la literatura en torno a este tema, sino tratar de introducir, con cierto rigor, al lector no familiarizado con esta técnica en los principales tópicos de la misma y comentando finalmente sus ventajas e inconvenientes respecto a otras técnicas.

A continuación se hará una somera descripción de los datos utilizados y se planteará la resolución de los problemas utilizando para tal fin las redes neuronales. Si bien los tres problemas son similares en cuanto a su resolución cada uno presenta sus propias peculiaridades en cuanto a la interpretación de los resultados obtenidos, es por ello que cada uno será analizado de forma separada.

Llevaremos a cabo la resolución de los tres problemas planteados utilizando principalmente para ello el marco teórico de la Hacienda Pública, donde los conceptos eficiencia y equidad cobran especial relevancia. Por una parte, el conocimiento de la eficiencia con la que están actuando los centros educativos permitiría, por el lado de la oferta, mejorar la asignación de recursos públicos y articular un sistema de incentivos para aquellos centros más eficientes. Desde el punto de vista de la equidad, toda vez que los estados occidentales modernos han incluido la educación primaria y secundaria en el llamado *Estado del Bienestar* atendiendo al principio de equidad categórica, por el cual se exige el consumo de un mínimo educativo para todos los individuos, se debe profundizar en desarrollar los conceptos de equidad vertical y equidad horizontal. Desde el punto de vista de la equidad vertical el sistema debería invertir mayores recursos en aquellos alumnos peor situados tanto económica como aptitudinalmente. La detección de estos alumnos con mayor probabilidad de fracaso constituye el primer problema que hemos planteado (apartado 3.1). Por equidad horizontal entendemos que individuos iguales deberían ser tratados igualmente por el sistema educativo. Haciendo una analogía con el sistema sanitario, nadie entendería que pacientes con una determinada patología resultaran sistemáticamente peor atendidos en un hospital que en otro cercano. La comparación de la eficiencia y la equidad horizontal con la que actúan los centros es el segundo objetivo de este trabajo (apartado 3.2), que implementaremos a partir de un modelo en donde mediremos la eficiencia a través de dos dimensiones distintas.

De igual manera y volviendo al sistema sanitario, si existen dos tratamientos para una patología con unos costes similares y se comprueba que uno de ellos es más efectivo parece lógico abandonar el otro tratamiento. En este trabajo, utilizando de nuevo las redes neuronales, trataremos de aplicar un método que nos permita identificar la eficiencia con la que actúan distintos planes de estudio, a partir del modelo definido, a fin de aplicar el más conveniente a cada situación determinada (apartado 3.3).

2.- FUNDAMENTOS DE LAS REDES NEURONALES

2.1.- Características básicas de las redes neuronales

2.1.1.- Redes neuronales de tipo biológico

Las Redes Neuronales son el resultado de los intentos por reproducir mediante computadoras el funcionamiento del cerebro humano. Se estima que el cerebro está compuesto por más de cien mil millones de neuronas interconectadas de forma variable y compleja. Su peculiar forma de trabajo le convierte en el dispositivo más eficaz para procesar información del mundo real.

Cada neurona recibe impulsos procedentes de otras neuronas (inputs) a través de las *dentritas*, que están conectadas a las salidas de otras neuronas por puntos de contacto especiales llamados *sinapsis*. Las sinapsis, alteran la efectividad con la que la señal es transmitida a través de un parámetro, el peso. El aprendizaje resulta de la modificación de estos pesos, que unido al procesamiento de información de la neurona determinan el mecanismo básico de la memoria. Algunas sinapsis permiten pasar a la señal con facilidad, mientras que otras no. El cuerpo de la neurona recibe todos estos inputs, y emite una señal de salida (output) si la entrada total supera el valor del umbral. Esta salida se transmite a través del axón desde donde se propaga mediante diferencias de potencial a las dentritas de otras neuronas (véase la figura 1).

Bajo esta descripción, las redes neuronales tratan de simular este proceso biológico en un equipo computacional basándose en el siguiente modelo:

La información se procesa en elementos simples llamados **neuronas** (o nodos en las redes neuronales).

Las señales se transmiten entre las neuronas a través de **conexiones**.

Cada conexión tiene asociado un **peso** que multiplica la señal transmitida.

Cada neurona transforma la señal recibida mediante una **función de activación** para determinar la señal de salida.

El aprendizaje se realiza mediante el ajuste de los pesos que ponderan las conexiones entre las neuronas que componen la red. El **algoritmo de aprendizaje** es el método por el que se obtienen los ajustes óptimos evitando tener que supervisar todo el proceso. Este método consta de una serie de pasos sencillos e iterativos.

Para procesar la información las neuronas se organizan en **capas**: la capa de entrada transmite las variables input utilizadas, la capa de salida presenta el resultado y entre medias pueden existir varias capas ocultas que procesan la información.

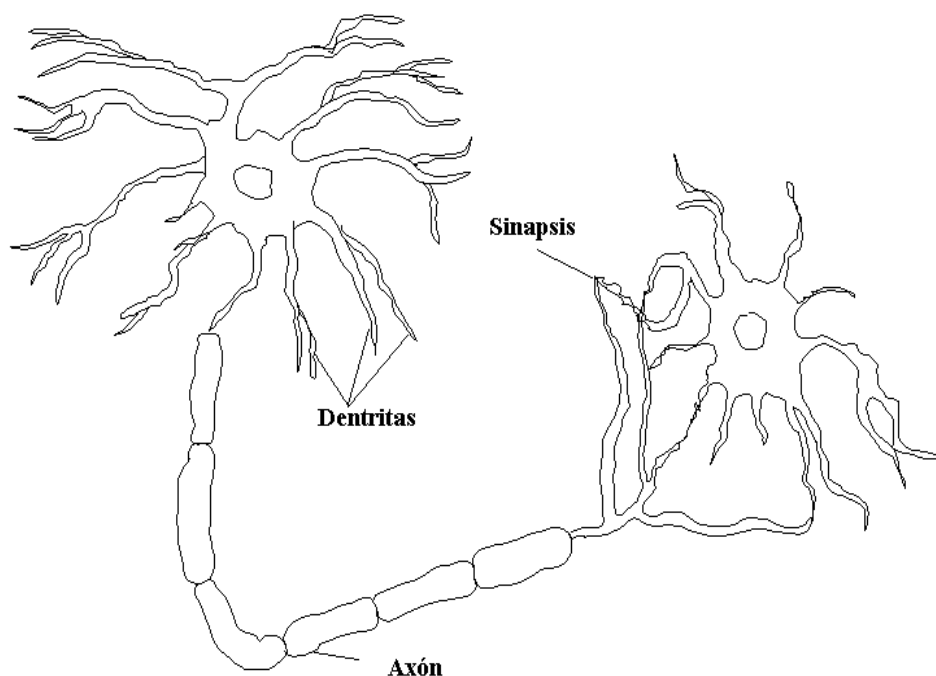


Figura 1. Características básicas de la neurona biológica.

En definitiva, podemos definir una red de neuronas como un conjunto de procesadores sencillos, altamente interconectados, que son capaces de aprender a partir de ejemplos reales. El objetivo de la red es por tanto aprender a asociar datos input a datos output a través de las neuronas $\{X(p), Y(p)\}$, donde $x(p) \in \mathbb{R}^n$, $y(p) \in \mathbb{R}^m$. Esto es, la red debe aprender una función:

$$f: \mathbb{R}^n \rightarrow \mathbb{R}^m$$

$$f(X) = Y$$

a partir de una muestra $\{X(p), Y(p)\}$, $p = 1, 2, \dots, N$

2.1.2.- La neurona artificial simple

La eficiencia de la sinapsis en la transmisión de señales al cuerpo de la neurona puede ser modelada asociando un factor multiplicativo a cada línea de entrada a la neurona. Si la neurona tiene n entradas, podemos escribir:

$$\text{Input total} = \text{peso línea 1} \times \text{input 1} + \dots + \text{peso línea } n \times \text{input } n = \sum_{i=1}^n \omega_i x_i$$

Si esta suma supera el valor umbral, la señal de salida es 1 y, si no 0. Utilizando la notación $c = (x_1, \dots, x_n)$ para el vector de entradas, $w = (w_1, \dots, w_n)$ para el vector de pesos sinápticos y q para el valor del umbral, la activación (salida) de la neurona para la entrada x puede modelarse por:

$$1, \text{ si } \sum_{i=1}^n \omega x_i \geq q \text{ o } W^T \times X - q \geq 0$$

$$g(X) = 0, \text{ si } \sum_{i=1}^n \omega x_i \leq q \text{ o } W^T \times X - q \leq 0$$

El modelo puede ser extendido cambiando la función de activación de la neurona, en este caso la función de Heaviside (función signo), por una función no lineal de activación de la neurona, siendo las más comunes en la literatura la función sigmoideal y la función gaussiana. La existencia del umbral se modela introduciendo entre los inputs un sesgo (bias). Esto equivale a considerar otro input cuyo valor siempre es 1 y cuyo peso asociado representa el umbral (figura 2).

El proceso de la neurona simple puede interpretarse geoméricamente de la siguiente manera: dado un vector de entrada X, se calcula el producto escalar $X \times W$ y se le resta el umbral q . Si el resultado es positivo, la neurona produce una respuesta y si es negativo otra. Así, una neurona con función de activación de Heaviside funciona exactamente igual que un clasificador lineal.

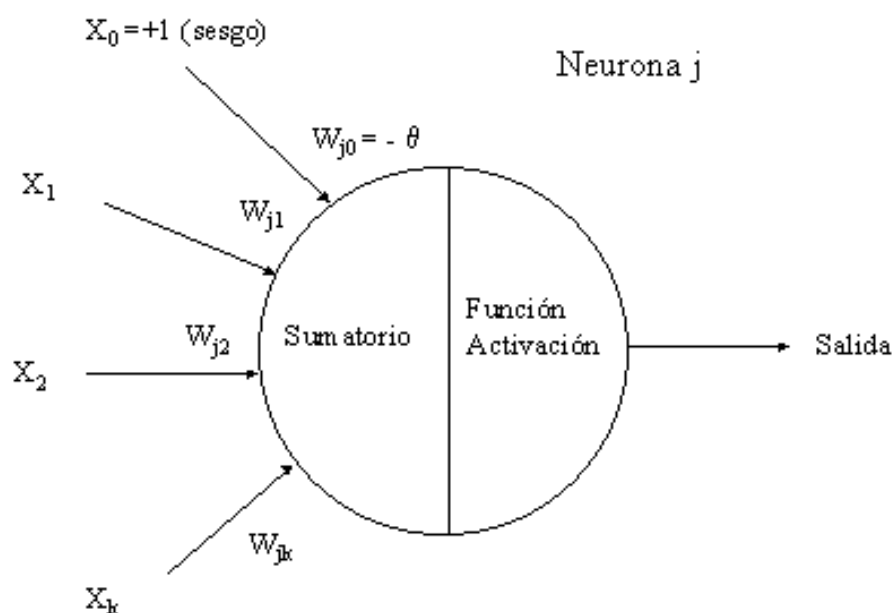


Figura 2. Estructura de la neurona artificial.

2.1.3.- Estructura de la red neuronal.

Como hemos apuntado anteriormente, dentro de la red, las neuronas se encuentran organizadas en capas:

Capa de entrada: Recibe los inputs del exterior. En esta capa normalmente no se procesan las señales recibidas sino que éstas son enviadas a la capa siguiente. El número de inputs que debe utilizarse depende del problema específico que estemos considerando.

Capa(s) oculta(s): Las neuronas de esta(s) capa(s) no tienen contacto con el exterior, transforman la señal recibida y la envían a la capa de salida. Normalmente es suficiente con una

sola capa oculta.

Capa de salida: Una vez la red ha transformado la señal esta capa envía la respuesta al exterior. Las neuronas de esta capa realizan una nueva transformación de la señal recibida por la capa(s) oculta(s).

La forma en que las neuronas están conectadas determinan la estructura de la red. Los modelos de conexión más básicos son los siguientes (figura 3):

Redes de propagación hacia delante (feed-forward). Es la forma más básica de red. Las neuronas de cada nivel sólo están conectadas con las neuronas de los niveles posteriores por lo que la información se propaga hacia delante. Las redes de este tipo más conocidas son: Perceptron, Adaline, Madaline, Linear Adaptive Memory (LAM) y Backpropagation. Todas ellas suelen aplicarse al reconocimiento o clasificación de patrones.

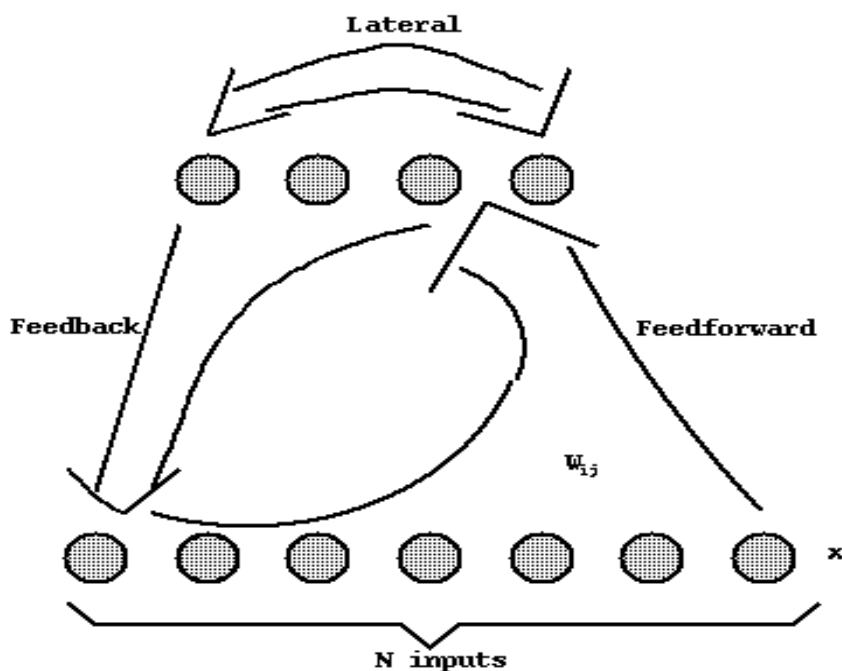


Figura 3. Esquema de las distintas estructuras de redes neuronales.

Redes de propagación hacia atrás (feedback). En estas redes, las neuronas pueden estar conectadas indistintamente con neuronas de niveles previos, posteriores, de su mismo nivel (conexiones laterales) o incluso con ellas mismas (autorrecurrentes). Los modelos más conocidos son: Cognitron, Neocognitron, la red Adaptive Resonance Theory (APT) y la Bidirectional Associative Memory (BAM).

2.1.4. Mecanismos de entrenamiento-aprendizaje.

Existen dos fases en toda aplicación de las redes neuronales: la fase de aprendizaje o entrenamiento y la fase de validación. La fase de entrenamiento, consiste en hacer que la red sea capaz de extraer, a partir de ejemplos, normas generales que la permitan en el futuro responder adecuadamente a patrones nunca vistos anteriormente. Una vez entrenado el modelo, se usará en la fase de validación o de prueba, en la que la red procesa los inputs que constituirán su entrada habitual, normalmente nunca vistos anteriormente por la red, analizándose de esta manera las prestaciones definitivas de la red. Una vez que ésta resulte operativa, la

arquitectura, neuronas, conexiones y pesos quedan fijos y la red está lista para funcionar.

Podemos diferenciar dos reglas básicas de entrenamiento-aprendizaje: supervisado y no supervisado:

Entrenamiento supervisado: Los datos para el entrenamiento están constituidos por un grupo de pares de patrones de entrada-salida deseada. Existe un maestro encargado de comprobar que la salida de la red coincida con la deseada, si finalmente no es así se deben modificar los pesos de las conexiones para intentar que la salida obtenida se aproxime a la deseada. Este tipo de redes han sido las más desarrolladas desde el inicio de estos diseños para la clasificación de patrones.

Entrenamiento no supervisado: En esta ocasión se suministran los inputs de entrada pero la red no recibe ninguna información que le indique cuál es la salida. La red modifica los pesos para poder asociar a la misma unidad de salida los vectores de entrada más similares. Suele decirse que estas redes son capaces de auto-organizarse. Normalmente se habla de dos tipos de entrenamiento no supervisado: el hebbiano y el competitivo/cooperativo: el primero generalmente más orientado a detectar las similitudes de las entradas y el segundo más orientado hacia la clusterización (1).

2.2.- El Perceptrón multicapa

Los orígenes de las Redes Neuronales suelen relacionarse a partir del modelo de neurona básico debido a Mc Culloch y Pitts (1943), en el que proponían, basados en los estudios del funcionamiento de la neurona de Ramón y Cajal, una teoría general del procesamiento de la información basada en redes de elementos binarios, capaces de realizar cálculos similares a los ordenadores digitales, pero con una ejecución paralela y no serial, en donde los pesos hacían el papel que el programa juega en un ordenador convencional. Hebb (1949) propuso una ley de aprendizaje que llegó a ser el ancestro de las modernas técnicas de entrenamiento. Ya en la década de los 50 Rosembat (1958) propuso el perceptrón simple con tan solo dos capas como un sistema de clasificación de patrones, capaz de interpretar patrones tanto abstractos como geométricos. Tras este brillante inicio, la exposición matemática de la naturaleza de los perceptrones por Minsky y Papert (1969) puso en claro el alcance y las limitaciones de estos novedosos modelos de proceso en el campo de la inteligencia artificial. El problema era que este tipo de redes tan solo serían capaces de hacer clasificaciones en conjuntos de datos que fueran linealmente separables ya que la célula de decisión que utiliza es el hiperplano.

Figura 4. Ejemplo de regiones de decisión separables linealmente.

Como es obvio, muchos problemas reales no son linealmente separables lo que llevó al abandono masivo del estudio de las redes neuronales. La figura 5 (A) es un ejemplo de problema XOR. Aunque claramente

existen dos patrones, no existe una recta que pueda dejar los círculos blancos a un lado y los negros a otro. En la figura 4, los dos patrones de datos eran linealmente separables, no así los de la figura 5 (B).

(A) (B)

Figura 5. Dos ejemplos de patrones no separables linealmente.

Para solucionar este problema iba a ser necesario convertir el perceptrón de un solo nivel en un perceptrón multicapa, mediante la introducción de alguna capa oculta junto a un algoritmo de aprendizaje.

Habría que esperar hasta el trabajo de Werbos (1974), quien obtuvo el algoritmo de entrenamiento capaz de obtener los valores óptimos de los pesos de estos sistemas neuronales, y a las contribuciones de Parker (1985) y Le Cun (1985). Estos 3 autores propusieron una arquitectura diferente para las redes neuronales (multicapa) y una regla de aprendizaje supervisado, la regla delta generalizada. Esta regla busca variar los pesos de modo que se reduzca el error cometido por la red lo más rápidamente posible. El trabajo seminal de Rumelhart Y Mc Clelland (1986) popularizó esta regla y extendieron el tema sugiriendo múltiples aplicaciones. Más recientemente, White (1992) ha contribuido a clarificar las relaciones entre las redes neuronales y otras ramas como la estadística y la econometría.

En el perceptrón multicapa los nodos se agrupan por capas conectándose todos los nodos de una capa con los de la siguiente capa y así sucesivamente. Se trata de una red feed-forward en la cual las señales se propagan únicamente hacia delante. Cada neurona suma cada input recibido por su peso asociado.

$$\text{net}_i = w_{i1}I_1 + w_{i2}I_2 + \dots + w_{in}I_n$$

donde net_i es la suma lineal de los inputs recibidos y n denota el número de inputs que estamos considerando. El resultado se hace pasar a continuación a través de una función de transferencia, generalmente sigmoideal.

$$H_i = [1 + \exp(- \text{net}_i)]^{-1}$$

Junto al resto de los resultados de las otras neuronas de la capa oculta, este resultado pasa a la capa de salida. La señal recibida por la neurona de salida k es:

$$\text{net}_k = w_{ki}H_i + \dots + w_{pi}H_p$$

Siendo p el número de neuronas en la capa oculta. Este resultado es de nuevo transformado en la neurona de la capa de salida cuyo resultado es el output definitivo.

$$O_j = [1 + \exp(- \text{net}_j)]^{-1}$$

Siendo j el número de neuronas en la capa de salida. Si denotamos por A y B las matrices que recogen los valores de los pesos de la capa de entrada a la oculta y de la oculta a la de salida respectivamente, O es el vector salida de la red, I el vector input y f la función de transferencia entonces la red puede ser descrita como:

$$O = f [f (I \times A) \times B]$$

2.3.- Funcionamiento del algoritmo Backpropagation

El método de entrenamiento más extendido en la literatura sobre el perceptrón multicapa es el algoritmo de retropropagación supervisada (Backpropagation); Rumelhart y Mc Clelland (1986). El objetivo del proceso de aprendizaje es ajustar los pesos de la red para que los patrones de entrada, introducidos en la red, permitan obtener las salidas deseadas. El aprendizaje, de tipo supervisado, supone la necesidad de disponer de pares de entrenamiento entrada-salida deseada. Este proceso se lleva a cabo mediante la minimización de una función de actuación, que generalmente se corresponde con la suma de errores al cuadrado.

El algoritmo backpropagation consta de dos fases:

1ª Fase: Hacia delante.

Inicialización: Se asigna aleatoriamente a la red un vector de pesos uniformemente distribuidos cuyo rango depende del usuario pero que generalmente se establece entre - 0.5, +0.5.

Propagación: Se presenta a la red un vector de entrada; esta información se va propagando a través de todas las capas (de la forma que vimos en el apartado anterior) hasta obtener el vector salida global de la red.

Comprobación del error cometido: A continuación la red compara la salida obtenida por ella con la deseada.

$$E_t(W) = \frac{1}{2} \sum_{i=1}^r (T_{ij} - O_{ij})^2$$

Siendo T_{ij} el vector de salida deseado en la capa de salida j para el vector de entrenamiento t y O_{ij} el vector de salida obtenido; multiplicar por $\frac{1}{2}$ tan solo se hace para simplificar posteriormente. Una vez calculado el error se tiene que determinar el sentido en que los pesos van a modificarse. La regla general para la modificación de los pesos es:

$$\Delta_t w_{ji} = -\alpha \frac{\partial E_t}{\partial w_{ji}}$$

donde α es el coeficiente de aprendizaje encargado de regular la velocidad de aprendizaje.

2ª Fase: Hacia atrás:

Adaptación del vector de pesos: Los cambios en los pesos se producen en la dirección en la que el error caiga lo más rápidamente posible (gradiente negativo). Estas señales de error se propagan a los nodos de las capas empezando por la de salida y siguiendo con las capas sucesivas. De esta forma, los nodos de cada capa sólo reciben una fracción del error global en función de su aproximada contribución relativa a la obtención de la salida. Los pesos se modificarán en base a la señal de error recibida, de manera que se reduzca el error actual de la red y la salida obtenida se vaya aproximando a la deseada (2).

$$w_{ji}(t+1) = w_{ji}(t) + \Delta_t w_{ji}(t)$$

Vuelta al paso 2: Con los nuevos pesos se introducen un nuevo par vector input-respuesta deseada y el proceso se repite hasta que el algoritmo converge cuando el error es suficientemente pequeño.

2.4.- Algunos comentarios prácticos sobre el perceptron multicapa con backpropagation.

Coefficiente de aprendizaje: Normalmente este coeficiente es fijo y con un valor reducido para asegurar que la red alcanzará una solución. Por otro lado, un valor muy reducido supone incrementar el número de iteraciones necesarias hasta obtener una solución.

Momento: A menudo la ecuación para la modificación de los pesos es modelada de la siguiente forma:

$$\Delta_t w_{ji} = -\alpha \frac{\partial E_t}{\partial w_{ji}} + \Delta_{t-1} w_{ji}$$

donde b es el vector momento. Con esta formulación introducimos un "recuerdo" que introduzca algo de inercia en el sistema ya que en el paso $t+1$ se conserva algo del cambio hecho en el paso t .

Mínimos Locales: El algoritmo backpropagation busca minimizar el error del sistema. Para ello se moverá por la superficie de la función de error dando pasos descendentes. Debido a este método de aprendizaje, el sistema podría eventualmente quedar atrapado en un mínimo local y alcanzar una solución que no fuese óptima, es decir con un error elevado (figura 6). Existen diversos criterios heurísticos diseñados a fin de evitar este problema:

Modificar el número de nodos en la capa oculta.

Comenzar de nuevo el entrenamiento con un conjunto distinto de pesos.

Cambiar los parámetros de aprendizaje.

Añadir capas ocultas.

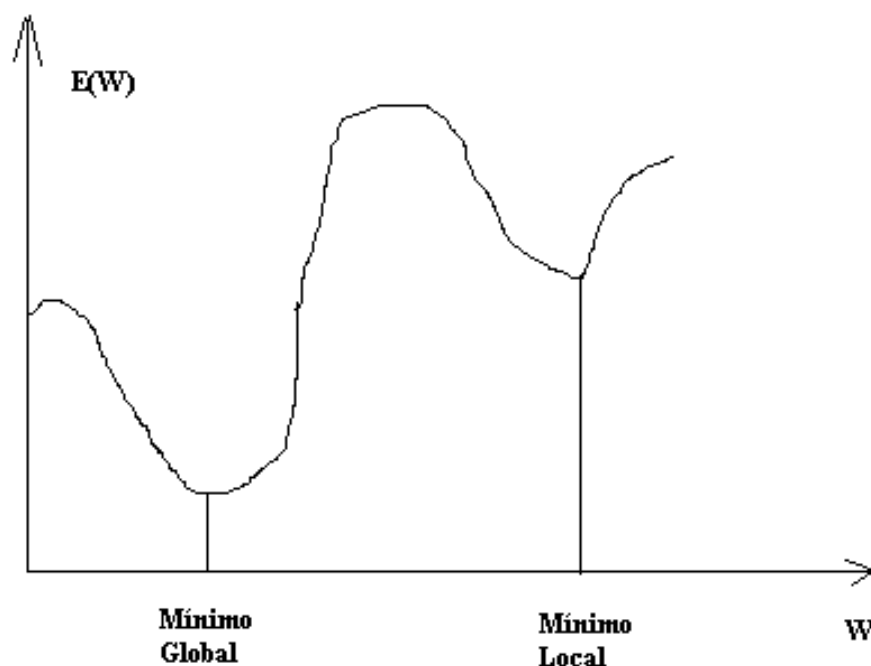


Figura 6: Problemas en el aprendizaje. Mínimo local en la superficie del error.

Arquitectura de la red: La elección de la arquitectura de la red no tiene ninguna regla y es por tanto un proceso de ensayo-error. Respecto al número de capas ocultas, si bien existen resultados teóricos que apuntan a que en redes feed-forward la existencia de una capa oculta es suficiente (véase Hornik (1989 y 1990)); en muchas aplicaciones se usa más de una capa oculta. Además, debe elegirse el número de nodos en la capa oculta teniendo en cuenta que si la red es muy pequeña no será capaz de formar un buen modelo del problema y si es muy grande corre el peligro de sobreaprendizaje, esto es, la red se ajusta a los datos de entrenamiento en vez de al patrón general de comportamiento de los datos produciendo de este modo una mala generalización. Normalmente se empieza con una red pequeña y se van añadiendo nodos en función de los resultados observados.

Entrenamiento y validación del aprendizaje: De la masa de datos disponibles resulta recomendable no entrenar la red con todos los datos, reservando aleatoriamente un conjunto de ellos (en torno a la mitad), para comprobar si el aprendizaje llevado a cabo resulta satisfactorio para datos nunca vistos antes por la red (out-of-sample). Dado el conjunto de datos de entrenamiento, éstos deben incluir vectores de ambas clases. Por otra parte, es deseable que el error de la red una vez entrenada sea parecido para los datos con los que la red aprendió (in-sample) y para los datos nunca vistos pues podremos concluir que la red ha generalizado correctamente.

2.5.- Algunas ventajas e inconvenientes de las redes neuronales.

Tras esta aproximación al mundo de las redes de neuronas parece conveniente resaltar someramente algunas de sus ventajas e inconvenientes respecto a otras técnicas estadísticas y econométricas:

No necesitan ser programadas; son capaces de aprender a partir de ejemplos reales mediante

la adaptación de los pesos.

No hace supuestos acerca de la distribución de las variables independientes y por tanto permite la introducción de variables continuas y categóricas lo cual resulta una importante ventaja frente al análisis discriminante que supone distribución normal multivariante.

La red permite realizar transformaciones no lineales lo que es una ventaja frente a los modelos de regresión y análisis discriminante múltiple tradicionales a la hora de llevar a cabo clasificaciones ya que puede generar superficies de decisión arbitrariamente complejas.

La principal desventaja de las redes de neuronas son por un lado la imposibilidad de determinar cómo se procesa internamente la información y por otro su incapacidad para contrastar el modelo.

El ejercicio empírico que desarrollamos a continuación aplica la técnica de las redes neuronales a los problemas propuestos. Estos problemas podrían ser afrontados con las técnicas de clasificación estadísticas y econométricas tradicionales (análisis discriminante y regresión logística). No es el objetivo de este estudio llevar a cabo comparaciones entre las distintas técnicas capaces de aportar soluciones; existen distintos estudios empíricos que ratifican la mayor fuerza predictiva de las redes (véase Bonilla, M. y Puerta, R. (1997). En general pensamos que para aplicaciones prácticas lo ideal sería aplicar varias técnicas y complementar los resultados de unas y otras de forma que las predicciones resulten lo más eficientes posibles.

3.- Aplicación empírica de las redes neuronales en economía de la educación: Descripción de los 3 problemas planteados y de los datos utilizados.

La carencia de lo que los economistas llamamos *función de producción* en el campo de la economía de la educación, esto es, la falta de un modelo que nos permita determinar objetivos, que para las edades de la educación secundaria se traducen básicamente en objetivos de rendimiento académico, en función de unos determinados recursos tanto materiales, aptitudinales como socioeconómicos, es el principal obstáculo para la investigación en ese campo.

Por otro lado, uno de los objetivos para la mejora de la educación no universitaria (el campo de estudio de este trabajo) es la reducción del fracaso escolar que como sabemos condiciona el nivel de ingresos futuros del individuo. Reduciendo el fracaso escolar, el sector público invierte en capital humano, contribuye a la redistribución de la renta y mejora la eficiencia del sistema en su conjunto al lograr una mejora en los resultados. Como señala San Segundo (1998). "Un análisis de la evidencia empírica existente pone de manifiesto que en España persisten importantes desigualdades educativas asociadas a la zona de residencia, así como al origen socioeconómico familiar(...). En 1996, sólo un 12% de los jóvenes cuyos padres carecen de estudios, acceden a la universidad, mientras que la probabilidad de que el hijo de un titulado superior acceda a la universidad es del 72%. Parece imposible concluir que se ha alcanzado la igualdad de oportunidades". El siguiente trabajo tratará de ofrecer algunas soluciones a esta situación.

El objetivo del primer ejercicio que llevaremos a cabo, de aproximación a este campo de estudio, es intentar predecir el fracaso escolar mediante la construcción de una red de neuronas que identifique a priori (a principio de curso) a los alumnos con alta probabilidad de fracaso en sus resultados al terminar el curso en función de las características socioeconómicas y aptitudinales de cada alumno. Como detallaremos posteriormente la definición de *fracaso* ha sido modelada de distintas formas siempre arbitrarias.

El objetivo del segundo problema planteado es intentar aportar ideas para la medición de la eficiencia relativa con la que actúan los centros de enseñanza a través de la aplicación de redes neuronales en datos reales. Para ello definiremos qué entendemos por eficiencia a partir de los resultados obtenidos por las redes y llevaremos a cabo, a modo de ejemplo, comparaciones entre distintos centros a fin de comentar los resultados obtenidos.

Por último, mediremos la eficiencia relativa en resultados obtenida por distintos planes de estudio. El objetivo de esta última aplicación es desarrollar una herramienta que permita evaluar de forma objetiva futuras reformas de los planes de estudio.

Se utilizan datos referentes a 7454 estudiantes que cursaban 1º (2º) de enseñanzas medias en octubre de 1984 (1985) en 256 centros públicos y privados (3). Los datos provienen de las muestras recogidas para el proyecto de evaluación de la Reforma de las Enseñanzas Medias por el CIDE; incluyen estudiantes que cursan el programa de Reforma de las Enseñanzas Medias (grupo experimental), alumnos de BUP y FP en los mismos centros (grupo de control interno), y alumnos de estos programas en centros no experimentales (grupo de control externo). En total utilizaremos los datos divididos en 6 modalidades:

BUP (Control Externo): 2476 alumnos.

FP (Control Externo): 1830 alumnos.

BUP (Control Interno): 771 alumnos.

FP (Control Interno): 524 alumnos.

Grupo Experimental en centros de BUP: 862 alumnos (4).

Grupo Experimental en centros de FP: 991 alumnos.

Los inputs individuales que hemos manejado para realizar el estudio han sido los siguientes:

CLASE DE CENTRO: FP; BUP

TIPO DE CENTRO: Público; privado laico; privado religioso.

APTITUDES EN 1º: Resultados en los siguientes tests de habilidad.

Aptitudes escolares

Razonamiento mecánico

Aptitudes espaciales

RESULTADOS ESCOLARES EN 1º: Resultados en las siguientes pruebas objetivas

Matemáticas cálculo

Matemáticas aplicaciones

Ortografía

Comprensión lectora

VARIABLES DUMMIES

HOMBRE: Si es hombre o no.

PREESCOLAR: Asistencia o no.

TRABAJA: Si realiza algún trabajo.

REPITECURSO: Si repitió algún curso anterior.

VARIABLES SOCIOECONOMICAS

Habilidad media de la clase.

Nivel socioeconómico medio de la clase

Nivel socioeconómico medio del alumno

La habilidad media de la clase así como el nivel socioeconómico fueron obtenidos llevando a cabo dos análisis de componentes principales, uno con las variables aptitudinales al principio del curso y otro con distintas variables socioeconómicas respectivamente. (Véase Modrego y San Segundo (1988)). Carecemos de inputs escolares tales como el gasto por alumno en cada centro, ratios de alumnos por profesor, infraestructuras escolares, etc. Dada la flexibilidad de la técnica utilizada podríamos, si las hubiéramos tenido, haberlas incluido con el fin de reducir el error de las redes obtenidas.

Transcurrido un curso escolar se vuelven a realizar pruebas a los alumnos cuyos resultados utilizaremos como output en las 7 pruebas nombradas anteriormente, tanto escolares como aptitudinales. En primer lugar si observamos la matriz de correlaciones de las 7 notas vemos que estos resultados parecen estar midiendo a la vez una misma dimensión, el rendimiento académico; ello nos anima a agrupar los 7 resultados obtenidos por cada alumno llevando a cabo para ello un análisis de componentes principales

Tabla 1: Matriz de correlaciones de los resultados obtenidos en las 7 pruebas

	<i>Comlec2</i>	<i>Ortog 2</i>	<i>matcalc 2</i>	<i>Matapl 2</i>	<i>apesp 2</i>	<i>Aptes 2</i>	<i>datrm 2</i>
Comlec 2	1						
Ortog 2	0,28956787	1					
Matcalc 2	0,39946759	0,35167237	1				
Matapl 2	0,41369428	0,2726803	0,60989071	1			
Apesp 2	0,17999404	0,02358337	0,20335777	0,30277622	1		
Aptes 2	0,52861853	0,30649005	0,56201788	0,61618049	0,35588192	1	
Datrm 2	0,21674034	0,01326698	0,22993579	0,3901029	0,51659047	0,39106085	1

Las cargas para cada componente son las siguientes:

Cargas:

Comp. 1 Comp. 2

comlec2 0.372 -0.211

ortog2 0.249 -0.518

matcalc2 0.423 -0.240

mtapl2 0.456 -0.220

apesp2 0.292 0.568

aptes2 0.473 -0.153

datrm2 0.322 0.551

La importancia de cada componente es:

Comp. 1 Comp. 2

Desviación estándar 1.7774418 1.1328493

Proporción de la varianza 0.4513285 0.1833353

Proporción acumulada 0.4513285 0.6346638

El primer factor de este análisis pondera positivamente los 7 tests realizados transcurrido el curso académico y por ello mide el rendimiento conjunto. Por otra parte, la segunda componente, no utilizada en este trabajo, parece decirnos que rendimiento escolar y aptitudinal varían por separado, si bien las aptitudes escolares parecen tener un comportamiento análogo al de los resultados escolares. Las estadísticas descriptivas de la primera componente que hemos asignado para cada alumno es la siguiente.

Min: 23.099 Max: 106.144

1st Qu.: 56.826 Varianza: 148.324

Media: 64.976 Std Dev.: 12.179

Mediana: 65.102 Total N: 7454

3rd Qu.: 73.309

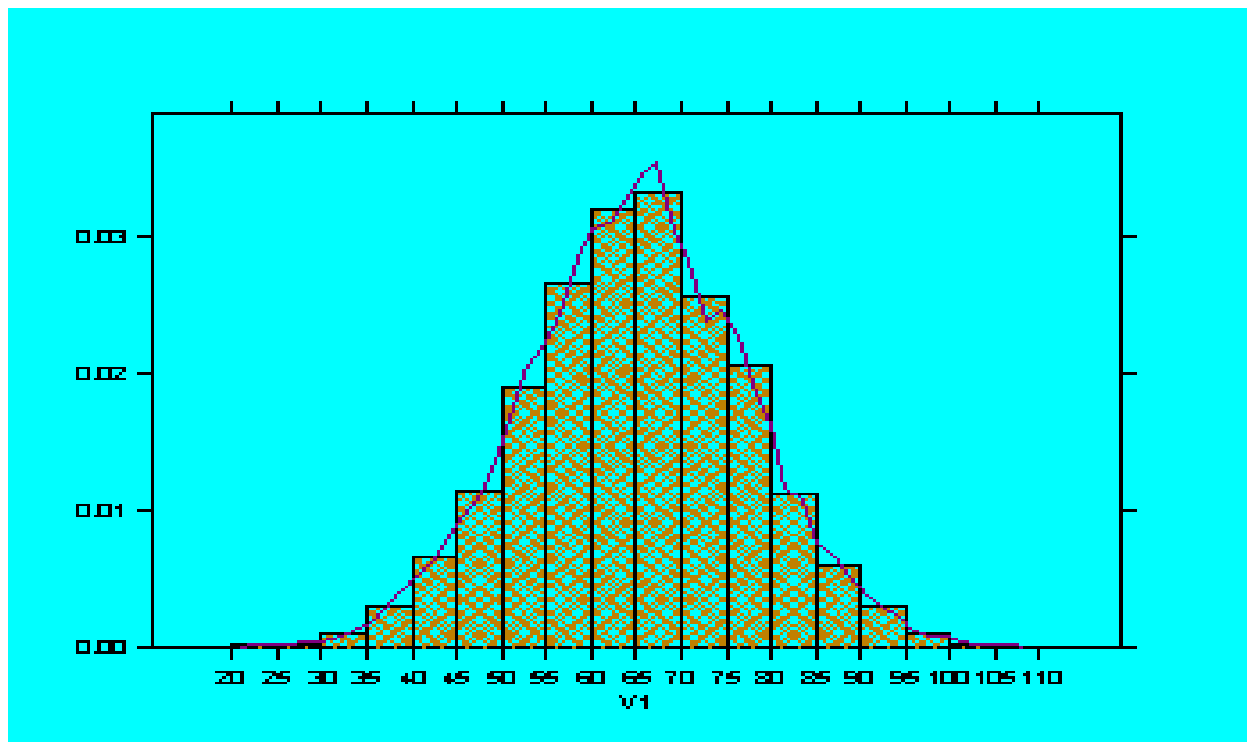


Gráfico 1: Histograma de la componente asociada a cada alumno

A lo largo del trabajo, definiremos el fracaso escolar de 4 maneras diferentes; aquellos alumnos que formen parte del grupo del 25%, 20%, 15% y 10% que peores resultados han obtenido respecto al total de la población en la componente asociada a cada alumno, cada grupo definido de esta forma supone un determinado nivel de exigencia en resultados, que va desde un criterio relativamente estricto a un criterio cada vez más laxo. Como ya hemos apuntado éstas definiciones son totalmente arbitrarias y no tienen ningún fundamento teórico que las apoye. En cualquier caso, el análisis llevado a cabo no resulta afectado por los conceptos de fracaso definidos. Por otra parte, recogemos en la componente asociada a cada alumno los resultados tanto escolares como aptitudinales para con ello medir dos dimensiones del rendimiento académico. De nuevo, podríamos definir el rendimiento del alumno únicamente en función de objetivos escolares o de cualquier otro tipo que sea considerado óptimo.

3.1.- Detección de alumnos de riesgo mediante redes de neuronas: Primera aplicación

En primer lugar, entrenaremos con una muestra aleatoria de los datos 4 redes de neuronas, con el fin de que cada una aprenda a clasificar a los alumnos en los grupos de éxito y fracaso atendiendo a sus características personales y según los distintos criterios de fracaso definidos. Para la construcción de estas 4 redes de neuronas hemos empleado la misma base de datos ordenada por resultados donde hemos ido variando el punto de corte para la definición de los grupos según el criterio de fracaso elegido. En esta base, todos los alumnos son iguales para la red en el sentido de que no existe ninguna variable que codifique ni el tipo de centro ni los estudios que ese alumno va a seguir.

Las siguientes matrices de datos obtenidas recogen los resultados de la clasificación conjunta de los datos de entrenamiento y prueba. Las redes entrenadas tienen una arquitectura feed-forward con una sola capa oculta. El número de neuronas en la capa oculta varía en cada caso (5).

CRITERIO 25% CRITERIO 20% CRITERIO 15% CRITERIO 10%

1 2 1 2 1 2 1 2

1 5200 390 1 5659 305 1 6077 259 1 6527 181

2 531 1333 2 541 949 2 420 698 2 339 407

Los errores cometidos han sido del 12,35%, 11,34%, 9,1% y 6,97% respectivamente. Las notas asociadas a cada punto de corte son 56,826; 54,73; 52,3375 y 49,315 también respectivamente.

La componente (1,1) de la matriz corresponde a alumnos que la red ha clasificado correctamente que no fracasarían y no han fracasado según el criterio definido en cada caso (alumnos que denominaremos del grupo A (aprobados)). La componente (1,2) son alumnos que perteneciendo finalmente al grupo A la red clasificó a priori que pertenecerían al grupo de fracaso (en adelante grupo F (fracaso)); la interpretación es análoga para los elementos (2,2) y (2,1) que representan respectivamente los alumnos que la red clasificó como de alto riesgo y que finalmente fracasaron y los alumnos que por sus características individuales tenían un alto riesgo de fracaso y que finalmente lograron no fracasar. En definitiva, vemos como el resultado de un alumno está fuertemente condicionado por sus aptitudes y sus circunstancias familiares así como por el nivel de sus compañeros de curso. La red nos permite, con el error que hemos señalado, predecir si el alumno fracasará o no para cada definición del fracaso escolar

No debemos perder de vista lo que estamos buscando con esta primera aplicación; por un lado, las políticas educativas se llevan a cabo en todos los centros que dependen de un gobierno (hoy día autonómico en la mayor parte de las regiones) siguiendo el viejo principio de "café para todos". Mismas ratios, mismas dotaciones, mismo gasto y en principio, misma atención para todos los alumnos. Si encontráramos un método que nos permita, con cierta fiabilidad, predecir que alumnos pasarán problemas en el curso académico, el gobierno regional si valora la eficiencia y la redistribución, debería aumentar los recursos sobre estos alumnos a través de los centros, por ejemplo bajando la ratio profesor alumno en aquellos centros (aulas) con más alumnos con posibles futuros problemas, invirtiendo en la contratación de profesores de apoyo...etc.

3.2.- Medición de la eficiencia relativa de los centros escolares.

A continuación vamos a utilizar las redes neuronales para medir la eficiencia de los centros escolares. En cada uno de los 6 grupos de alumnos comentados con anterioridad, llevaremos a cabo componentes principales sobre los resultados obtenidos únicamente por los alumnos de cada grupo, tal como hicimos en el ejercicio anterior para el conjunto total de datos, a fin de comparar posteriormente los centros escolares con otros centros que impartían los mismos estudios que ellos. Posteriormente, para cada uno de estos grupos se han llevado a cabo los siguientes pasos:

En primer lugar se ha diseñado una red de neuronas que permita predecir con el menor error los resultados de cada alumno en función de sus características individuales. Esta clasificación se ha llevado a cabo considerando, de forma arbitraria, que un alumno fracasa si el resultado recogido en su componente se encuentra por debajo del valor del primer cuartil del total del grupo (el estudio sería análogo con los otros criterios definidos). Los resultados obtenidos para cada grupo han sido los siguientes:

BUP C. Externo FP C.Externo BUP C.Interno

1 2 1 2 1 2

1 1732 125 1 1293 79 1 517 61

2 194 425 2 135 323 2 48 145

Los errores de clasificación cometidos son del 12,88%, 11,69% y 14,13%.

Reforma (BUP) FP C.Interno Reforma (FP)

1 2 1 2 1 2

1 571 75 1 371 22 1 693 50

2 46 170 2 33 98 2 77 171

Los errores de clasificación cometidos son del 14,03%, 10,49% y 12,81% .

Una vez obtenidas estas redes el error con la que actúan al predecir el resultado de cada alumno será el mismo para todos los centros que a ese grupo pertenecen con lo que podremos medir la eficiencia relativa de cada centro con respecto a los demás. Pero para ello debemos identificar qué entenderemos por eficiencia en la actuación de un centro, esto lo haremos a través de las siguientes definiciones:

Ineficiencia A: Entenderemos por grado de ineficiencia A el número de alumnos de un centro escolar que siendo clasificados a priori (a principio de curso) en el grupo de éxito pertenecen a posteriori (pasado el curso) al grupo de fracaso respecto del total de alumnos clasificados a priori en el grupo de éxito.

Eficiencia F: Entenderemos por grado de ineficiencia F el número de alumnos de un centro escolar que siendo clasificados a priori (a principio de curso) en el grupo de fracaso pertenecen a posteriori (pasado el curso) al grupo de éxito respecto del total de alumnos clasificados a priori en el grupo de fracaso.

Veamos un ejemplo para entender mejor estos conceptos; para ello vamos a fijarnos en la red obtenida para el grupo de alumnos de BUP control externo. La red, gracias a su proceso de aprendizaje es capaz de predecir a priori y en función de las características individuales de cada alumno y del total que forman el conjunto, el grupo al que con más alta probabilidad pertenecerá cada uno de ellos, en nuestro caso $1732 + 194 = 1926$ la red confía en que pertenecerán al grupo A y $125 + 425 = 550$ al grupo F (leemos en columnas). Esta clasificación a priori la podemos identificar como una "función de producción ad hoc", sin duda rudimentaria debido a que desconocemos en que medida pondera cada uno de los factores y debido a que cometemos un error. Pero este error lleva asociado la eficiencia de los centros escolares; para entender esto debemos leer este resultado de la siguiente forma.

Una vez llevado a cabo el proceso productivo de la educación observamos que 194 de los 1926 alumnos que *deberían* no fracasar han fracasado y que 125 de los 550 "*destinados*" al fracaso han logrado superar el listón del primer cuartil. El 10,07% de los alumnos que han fracasado cuando según la red deberían haber aprobado y el 22,72% de los alumnos que pese a sus peores dotaciones iniciales consiguen aprobar representan las definiciones que hemos apuntado anteriormente .

Rápidamente nos damos cuenta de que en estas definiciones, *ineficiencia A* y *eficiencia F*, se encuentra en cierta medida asociado al error que comete la red en predecir el resultado, esto es, el clasificador utilizado no ha conseguido aprender perfectamente a distinguir los patrones de éxito-fracaso a partir de las variables inputs utilizadas. Pero si una vez elaborada la red la utilizamos para clasificar a cada uno de los centros, tomados individualmente por los alumnos que lo componen, y suponemos que el error se distribuye igualmente para todos, podremos eliminar el error comparando los centros entre sí y midiendo las diferencias relativas de los centros en las dos dimensiones de eficiencia definidas, identificando así a los centros mejores y peores con el fin de que el gestor público pueda actuar de forma distinta sobre ellos en función de sus especiales circunstancias.

Siguiendo con la red de BUP control externo veamos una aplicación práctica:

Del centro 202 de la muestra (centro público) perteneciente al grupo de BUP control externo se tienen los resultados del rendimiento de 46 alumnos. Al introducir la información de éstos en la red, obtenemos la siguiente clasificación (la red no sabe el resultado final que han obtenido).

A – 40 (1,1) + (2,1)

F – 6 (1,2) + (2,2)

Transcurrido el curso académico (suponiendo que ya contáramos con la red) comparamos esta clasificación con los resultados realmente obtenidos, obtenemos así la siguiente matriz.

1 2

1 37 2

2 3 4

(7,5%) (33,33%)

De los 40 alumnos clasificados a priori en el grupo A, la red ha acertado en 37; además observamos dos tipos de errores. Si leemos la primera columna observamos como 3 alumnos que *deberían* no fracasar lo han hecho, esto supone ineficiencia A del centro escolar en el porcentaje que figura en cursiva tal y como ha sido definida. En la columna dos, vemos como dos alumnos del total han conseguido no fracasar lo cual podemos identificar como eficiencia F del centro educativo.

A continuación repetimos este proceso con otro centro de la muestra, en este caso el centro 210, también se trata de un centro público.

A – 36

F – 4

Comparemos ahora la clasificación a priori con la que realmente ha tenido lugar.

1 2

1 31 1

2 5 3

(13,88%) (25%)

La red ha clasificado a los alumnos de este centro utilizando el mismo criterio que para el caso anterior, esto es importante pues ahora, *la diferencia en las dos dimensiones de eficiencia si que podemos identificarla como una medida de la eficiencia relativa con la que uno actúa frente a otro*. En nuestro ejemplo, el centro 202 es un 6,38% menos ineficiente que el 210 (análisis de la primera columna) y un 8,33% más eficiente (segunda columna). Si sumamos ambos porcentajes obtendremos la *distancia en eficiencia total*; en definitiva el centro 202 ha sido un 14,71% más eficiente que el centro 210.

Veamos otros 3 centros (escogidos al azar) pertenecientes a la muestra de centros BUP (control externo). Presentamos la matriz de resultados de cada uno de ellos que debe ser leída de la misma forma que en los casos anteriores.

Centro 300: Centro 313: Centro 281:

(privado laico) (privado religioso) (privado religioso)

1 2 1 2 1 2

1 3 0 1 38 1 1 26 2

2 1 11 2 3 0 2 3 6

(25%) (0%) (7,31%) (100%) (10,34%) (25%)

Como podemos ver existen diferencias significativas en la actuación de unos centros respecto a otros. Con el fin de ilustrar la gran variedad de tipos de centro existentes, en cuanto a su grado de eficiencia, vamos a incluir los resultados en eficiencia de 3 centros más en la red de FP control externo (6).

FP Control Externo

Centro 355: Centro 426: Centro 432:

Privado religioso privado laico privado religioso

1 1 2 1 2

1 37 1 23 2 1 10 0

2 5 2 0 3 2 0 18

(11,90%) (0%) (40%) (0%) (0%)

En el centro 355 vemos como no existe la segunda columna; esto es debido a que la red no encontró a priori alumnos que por sus características tuvieran una alta probabilidad de fracaso. El centro 426 vemos que es un centro muy eficiente ya que no obtiene ineficiencia A y al mismo tiempo obtiene un grado considerable de eficiencia F. Por último, el centro 432 es un ejemplo en el que la red ha clasificado correctamente a priori en los grupos de éxito y fracaso. No obstante su eficiencia es reducida ya que no ha conseguido que ninguno de los alumnos del grupo de fracaso alcanzaran finalmente el éxito.

Con esta forma de medir la eficiencia podemos no sólo medir la eficiencia relativa del centro sino conseguir un mayor grado de desagregación identificando a los profesores que dentro de un centro han impartido clase a distintos grupos. Si bien en centros de primaria y secundaria las actuaciones del profesorado en la impartición del programa de estudios es muy homogénea, debería controlarse la eficiencia relativa de los profesores de un mismo centro a fin de que no se diluyeran responsabilidades y sirviendo como posible guía para la articulación de un sistema de incentivos.

Por otro lado, de las dos dimensiones de eficiencia definidas, parece más deseable que el sistema educativo se esforzara en mayor medida en alcanzar la eficiencia F ya que desde el punto de vista de la equidad, el grupo de fracaso se corresponde en mayor medida con los individuos peor dotados tanto aptitudinal como económicamente. Parece razonable por tanto que el sector público invierta en capital humano en estos individuos en mayor medida que en los demás.

Si bien los comportamientos de cada centro de forma individual son muy heterogéneos, se ha observado a lo largo del trabajo que existen centros donde la segunda columna desaparece (como hemos visto en el

ejemplo anterior) y de esta forma no podríamos medir el grado de eficiencia F. Podemos apuntar algunas soluciones para tratar de corregir este problema: en primer lugar podríamos entrenar de forma paralela al clasificador que estemos utilizando una red con un criterio de fracaso más amplio (subir el listón académico) para así comparar a este grupo de centros con este nuevo clasificador. De igual manera, podríamos diseñar una red capaz de discriminar entre los mejores alumnos y el resto lo que nos permitiría comparar centros de carácter "elitista" entre sí. Una mixtura de estas dos soluciones sería construir una red o clasificador capaz de discriminar entre tres regiones de rendimiento: mayor éxito, región intermedia y fracaso, si bien el inconveniente en este caso es que aumentaríamos el error ya que pasaríamos a tener dos superficies de contacto entre grupos.

3.3.- Medición de la eficiencia relativa de los planes de estudios.

Cuando el sistema educativo piensa introducir una reforma de los planes de estudio existentes parece recomendable evaluar los nuevos planes respecto a los anteriores. De nuevo definiremos el fracaso escolar de 4 maneras diferentes; aquellos alumnos que formen parte del grupo del 25%, 20% 15% y 10% que peores resultados han obtenido respecto al total de la población; a continuación pediremos a estas redes, las mismas que las recogidas en el punto 3.1 de este trabajo, que clasifiquen a los 6 grupos disponibles y compararemos los errores cometidos con el fin de medir la eficiencia relativa de los distintos centros con distintos planes de estudios y en el caso de los centros con grupos de control interno los mismos centros con distintos planes a fin de comprobar la calidad de éstos. El objetivo último de este estudio es determinar la eficiencia relativa con la que actúan los centros agrupados por tipos de estudios y ver como varía ésta en función del criterio de fracaso fijado con un especial interés en comparar las enseñanzas de reforma y los grupos de control interno.

BUP CONTROL EXTERNO

CRITERIO 25%		CRITERIO 20%		CRITERIO 15%		CRITERIO 10%	
1	2	1	2	1	2	1	2
1 2045	80	1 2159	50	1 2257	39	1 2342	12
2 142	209	2 136	131	2 101	79	2 81	41
(6,49)	(27,68)	(5,92)	(27,62)	(4,28)	(33,05)	(3,42)	(22,64)

ESTUDIOS DE REFORMA EN CENTROS DE BUP

CRITERIO 25%			CRITERIO 20%			CRITERIO 15%			CRITERIO 10%		
	1	2		1	2		1	2		1	2
1	658	46	1	713	36	1	745	30	1	783	23
2	52	106	2	63	50	2	55	32	2	43	14
	(7,32)	(30,26)		(8,11)	(41,8)		(6,87)	(48,38)		(5,09)	(62,16)

BUP CONTROL INTERNO

CRITERIO 25%			CRITERIO 20%			CRITERIO 15%			CRITERIO 10%		
	1	2		1	2		1	2		1	2
1	616	36	1	657	24	1	697	15	1	724	11
2	47	72	2	43	47	2	27	32	2	21	15
	(7,08)	(33,33)		(6,14)	(33,8)		(3,72)	(31,91)		(2,81)	(42,3)

Observamos como si bien para el criterio del primer cuartil la eficiencia de los planes de estudio es muy similar, la eficiencia de los estudios de reforma aumenta significativamente a medida que bajamos la exigencia en los resultados debiendo concluir que los planes de reforma ayudan en mayor medida a los alumnos más desfavorecidos. Debido a la falta de datos desconocemos el gasto en educación por alumno en el año que estamos considerando, un gasto por alumno significativamente mayor en los alumnos de reforma podría explicar este resultado.

FP CONTROL EXTERNO

CRITERIO 25%			CRITERIO 20%			CRITERIO 15%			CRITERIO 10%		
	1	2		1	2		1	2		1	2
1	999	128	1	1132	110	1	1274	101	1	1441	79
2	152	551	2	154	434	2	126	329	2	111	199
	(13,20)	(18,85)		(11,97)	(20,22)		(9)	(23,48)		(7,15)	(28,41)

ESTUDIOS DE REFORMA EN CENTROS DE FP

CRITERIO 25%			CRITERIO 20%			CRITERIO 15%			CRITERIO 10%		
	1	2		1	2		1	2		1	2
1	581	66	1	658	58	1	731	46	1	820	35
2	94	250	2	97	178	2	77	137	2	51	85
	(13,92)	(20,88)		(12,84)	(24,57)		(9,52)	(25,13)		(5,85)	(29,16)

FP CONTROL INTERNO

CRITERIO 25%			CRITERIO 20%			CRITERIO 15%			CRITERIO 10%		
	1	2		1	2		1	2		1	2
1	301	34	1	340	27	1	373	28	1	417	21
2	44	145	2	48	109	2	34	89	2	33	53
	(12,75)	(18,99)		(12,37)	(19,85)		(8,35)	(23,93)		(7,33)	(28,37)

En los estudios de FP las distancias son en general mucho menos significativas que en los anteriores. Sin embargo persiste una cierta mejora de los estudios de reforma respecto a los antiguos estudios de FP. También podemos observar como la eficiencia relativa de los centros que imparten FP respecto a los centros de BUP es, en distancia, más reducida.

4.- CONCLUSIONES

Recapitemos las principales conclusiones obtenidas a lo largo del trabajo en el que hemos tratado de aplicar las redes neuronales al campo de la economía de la educación:

Las redes neuronales artificiales pueden ser empleadas en el ámbito de la economía de la educación para predecir el rendimiento académico de los alumnos y detectar así a los alumnos con mayor riesgo de fracaso, a fin de invertir mayores recursos en ellos, mejorando de esta forma la eficiencia y la equidad del sistema educativo en su conjunto.

Asimismo, podemos medir la eficiencia relativa de centros escolares y planes de estudio a partir de los conceptos denominados Ineficiencia A y Eficiencia F que miden dos dimensiones de un mismo concepto: la eficiencia. La eficiencia F parece más deseable desde el punto de vista de la equidad ya que actúa sobre los individuos peor dotados tanto económica como aptitudinalmente.

Las mediciones de eficiencia propuestas podrían obtenerse mediante modelos estadísticos y econométricos. Dada la capacidad de abstracción de la red, aprendizaje no lineal, su capacidad de interpolar, su fácil aplicación (no necesita ser programada) y su completa flexibilidad en la introducción de inputs, hace que los resultados obtenidos sean mejores con la aplicación de las redes neuronales en comparación con otros modelos discriminantes como han probado distintos estudios. En contra tiene que no es posible determinar como se procesa internamente la información presentada a la red. En cualquier caso y para estudios posteriores parecería conveniente utilizar a la vez varios modelos discriminantes (redes, regresión logística, análisis discriminante) que nos permitan ir definiendo funciones productivas educativas ad hoc al contexto que estemos analizando.

A diferencia de los modelos DEA (análisis envolvente de datos), aplicados para medir la eficiencia relativa de centros escolares en el entorno educativo con datos de resultados agregados por centros, en este trabajo hemos utilizado datos individuales que permiten identificar a priori los alumnos en los que radica el objetivo del centro permitiendo de esta forma asignar más recursos a estos alumnos. Otra ventaja frente a las técnicas envolventes es que al trabajar con datos individuales podemos incluir tantas variables inputs como queramos, (significativas eso sí, para lo cual debemos hacer contrastes estadísticos y en cualquier caso conocer detalladamente el problema considerado), sin tener que enfrentarnos a las limitaciones que impone el número de grados de libertad del problema. En el caso de las redes no existe una regla definida, pero algunos heurísticos señalan que debemos tener 20 observaciones por cada variable incluida. En cualquier caso, parece deseable comparar los resultados de eficiencia obtenidos por los centros mediante las dos técnicas a fin de que ambas se complementen y permitan observar posibles inconsistencias en los resultados.

Dada la gran capacidad de generalización de las redes neuronales su degradación es lenta. Por tanto, tras su construcción la red podría ser utilizada en períodos sucesivos para la detección de alumnos de riesgo a partir de las características individuales de éstos.

En España la única prueba objetiva es la prueba de acceso a la Universidad, la selectividad, es por ello que esta prueba parece ser el principal objetivo de futuras investigaciones en cuanto a la medición de la eficiencia de centros, si bien dados los objetivos de este estudio; la mejora en

eficiencia y equidad del sistema educativo parece más deseable intervenir en niveles inferiores donde los resultados están condicionados con mayor fuerza con las variables aptitudinales y socioeconómicas. En cualquier caso se podría realizar la predicción de alumnos de riesgo y la medición de la eficiencia relativa de los centros en varios niveles simultáneamente.

La detección de ineficiencias debe utilizarse para articular un sistema de incentivos de forma que podamos premiar a los profesores más eficientes los cuales podemos identificar a través del modelo propuesto.

La introducción de inputs escolares tales como infraestructuras, gasto por alumno, número de alumnos por unidad, etc. así como el haber contado con las variables socioeconómicas antes de realizar componentes principales (análisis que agrupa variables como: número de hermanos, situación laboral de los padres o estudios de los padres) no considerados en este trabajo, sin duda reducirán en alguna medida el error de la red mejorando de esta manera las predicciones.

BIBLIOGRAFÍA

Andina de la Fuente, D. "Tutorial de redes neuronales". Dpto. de Señales, Sistemas y Radiocomunicaciones. Grupo de circuitos. Universidad Politécnica de Madrid.

Arbib, M.A. (1995): "The Handbook of Brain Theory and Neural Networks". The MIT Press.

Beltratti, A.; Margarita, S.; Terna, P. (1996): "Neural Networks for economic and financial modelling". International Thomson. Computer Press.

Bonilla Musoles, M.; Puerta Medina, R.(1997): "Análisis de las redes neuronales: Aplicación a problemas de predicción y clasificación financiera". Documento de Trabajo. Departamento de Economía Financiera y Matemática. Universidad de Valencia.

Hanushek, E. (1986): "The economics of Schooling". Journal of Economic Literature. Vol 24, nº3, pp. 1141-1171.

Hebb, D.O. (1949): "The organization of behavior". Science Editions, New York, NY.

Hornik, K., Stinchcombe, M. and White, H. (1989): "Multilayer Feed-forward Networks are Universal Approximators". Neural networks 2, pp. 359-66.

Hornik, K., Stinchcombe, M. and White, H. (1990). "Universal Approximation of an Unknown Mapping and its Derivatives using Multilayer Feed-forward Networks". Neural Networks 3, pp. 551-60.

Le Cun, Y. (1985): "Une Procédure d'Apprentissage pour Réseau à Seuil Asymétrique", in Proceedings of Cognitiva.

Martínez de Lejarza, I. (1996): "Redes Neuronales Auto-Organizadas y Clustering: Una Aplicación a la Agrupación Económico-Funcional de Entidades de Población". Documento de Trabajo. Facultat de CC. Econòmiques i Empresariales. Departamento de Economía Aplicada.

Universitat de València.

Mc Culloch, W.S. y Pitts, W.A. (1943): "A logical Calculus of the Ideas Immanent in Nervous Activity". Bulletin of Mathematics and Biophysics, 5, págs. 115-133.

Minsky, M. y Papert, S. (1969): "Perceptrons: An introduction to Computational Geometry". The MIT Press. Cambridge, MA.

Modrego, A.; San Segundo, M.J. (1988): "Ecuaciones de rendimiento escolar para la evaluación de las enseñanzas medias". Revista de Educación. Septiembre-Diciembre 1988. pp 147-179

Parker, D.B. (1985): "Learning Logic". MIT Center for Computational Research in Economics and Management science, T.R. 47.

Pedraja, F. y Salinas, J. (1996): "Eficiencia del gasto público en educación secundaria: una aplicación de la técnica envolvente de datos". Hacienda Pública Española N°138, pp. 87-96.

Roseblatt, F. (1958): "The Perceptron: A probabilistic model for information storage and organization in the brain". Psychological Review 65, págs 386-408.

Roseblatt, F. (1959): "Principles of Neurodynamics". Spartan Books, New York.

Rumelhart, D. E.; Mc Clelland, J.L.; and the PDP Research Group (Eds.) (1986): "Parallel Distributed Processing: Explorations on the Microstructure of Cognition, vol.I. Foundations", MIT Press, Cambridge, MA.

San Segundo, M.J. (1991): "Evaluación del Sistema Educativo a Partir de Datos Individuales". Economía Industrial. Marzo - Abril 1991, pp. 23-37.

San Segundo, M.J. (1998): "Igualdad de oportunidades educativas". Ekonomiaz N° 40 1.º Cuatrimestre.

Santín, Daniel (1998) "Eficiencia relativa de las Comunidades Autónomas en Enseñanza Secundaria: un enfoque socioeconómico" Documento de Trabajo Facultad de Ciencias Económicas y Empresariales. UCM, nº 9822.

Tejedor, F. J.; Caride, J. A. (1988): "Influencia de las variables contextuales en el rendimiento académico". Revista de Educación. Septiembre-Diciembre 1988. pp 112-146.

VV.AA. (1994): "Acerca de la producción educativa: Estudio de los Centros Educativos de la Comunidad Autónoma del País Vasco". Colección estudios y documentos N° 20. Gobierno Vasco. Departamento de educación, universidades e investigación.

Werbos, P.J. (1974): "Beyond Regression: New tools for Prediction and Analysis in the Behavioral Sciences". Unpublished Ph.D. Dissertation, Harvard University. Department of Applied Mathematics.

White, H. (1992): "Artificial Neural Networks: Approximation and Learning Theory". Blackwell, Oxford.

NOTAS AL PIÉ DE PÁGINA

¹ Para una buena introducción a las redes con entrenamiento no supervisado y sus aplicaciones puede acudirse a Martínez de Lejarza y Esparducer, I. (1996).

² Para ver el desarrollo matemático de la propagación hacia atrás del error a través de la regla de la cadena puede acudirse a Bonilla, M. y Puerta, R. (1997).

³ Para un estudio pormenorizado de estos datos puede acudirse a: Modrego y San Segundo (1988) y San Segundo (1991).

⁴ De los 1855 alumnos que figuran en los datos descriptivos, hemos utilizado 1853 debido a que se han detectado 2 alumnos que cursaban el programa de Reforma en un centro no diseñado para tal fin. Ello nos ha llevado a no considerar a estos dos alumnos disponiendo de 7454 y no de los 7456 del estudio original.

⁵ Dado el carácter aplicado de los resultados obtenidos y la construcción de diversas redes para distintos propósitos no vamos a comentar la estructura de cada una de las redes. En general, para todas ellas los porcentajes de error para los datos de entrenamiento y validación son muy similares y el número de neuronas en la capa oculta oscila entre 3 y 8.

⁶ Una vez disponemos de los clasificadores para cada grupo de centros la forma de medir la eficiencia resulta análoga para todos ellos. Dado el carácter metodológico de este trabajo tan solo incluimos algunos centros de los grupos de control externo y para el criterio de fracaso del primer cuartil que como ya hemos visto podría ser variado.