

# **REDES NEURONALES Y SISTEMAS DINAMICOS COMPLEJOS**

M<sup>a</sup> Victoria Caballero Pintado y Lourdes Molera Peris

*Universidad de Murcia*

## **RESUMEN**

En el estudio de series temporales los métodos lineales son los adecuados para interpretar una estructura regular siempre que la serie proceda de un sistema dinámico lineal. Sin embargo, cuando el sistema del que procede tiene un comportamiento irregular o complejo se necesitan métodos de análisis no lineales.

Los modelos matemáticos de redes neuronales son una herramienta de gran utilidad en la estimación de funciones no lineales generadoras de datos suficientemente próximos a los que se quieren estudiar. Por ello, las redes neuronales se utilizan en campos tan diversos como Economía, Física o Biología en el intento de predecir el siguiente valor de una serie dada.

En este trabajo se estudian series temporales obtenidas de sistemas dinámicos deterministas no lineales (Lozi, Hénon), series temporales i.i.d. y series temporales deterministas contaminadas con ruido. Se utiliza el test BDS para confirmar que la serie presenta una estructura subyacente estocástica o determinista frente a un comportamiento i.i.d. En segundo lugar, se emplean distintos modelos de redes neuronales (MLP y recurrentes) para predecir el valor futuro de las series consideradas. Por último, se intentarán aplicar estos procedimientos a una serie de datos económicos.

## 1. INTRODUCCIÓN

En los últimos años, las redes neuronales artificiales se han utilizado frecuentemente por su capacidad para aprender a partir de datos. Debido a su flexibilidad, la aplicación de las redes neuronales se ha extendido a las series temporales con el ánimo de conseguir predicciones, con un mínimo error, de un valor futuro de la serie temporal e incluso con el objeto de predecir la volatilidad presente en los datos, cuando la haya.

En este trabajo se comparan los resultados obtenidos aplicando redes neuronales a series temporales procedentes de sistemas dinámicos deterministas conocidos y de sistemas dinámicos deterministas perturbados aditivamente. Por último se aplican estos modelos matemáticos a la serie de rendimientos del Ibex-35. Esta serie, como muchas otras obtenidas de mercados financieros, tiene una apariencia aleatoria y, en los últimos años, ha empezado a estudiarse desde la perspectiva de la no linealidad determinista y estocástica. En este contexto, parece interesante estudiar si las redes neuronales permiten extraer regularidades no lineales presentes en ésta y otras series temporales económicas.

Por ello, en la segunda sección se hace un rápido repaso sobre qué se entiende por red neuronal, su estructura y estimación de sus parámetros. El test BDS, muy utilizado en los últimos años para decidir si una serie presenta alguna estructura, se recuerda en la sección tercera, así como la dimensión de correlación utilizada en el análisis de los sistemas caóticos.

Aplicamos redes neuronales multicapa hacia delante en la sección cuarta. Teniendo en cuenta los resultados obtenidos una vez aplicado el test BDS a las series que tenemos así como calculada su dimensión de correlación, intentamos aproximarnos a los datos utilizando redes neuronales con distintas arquitecturas. Y la sección última la dedicamos a sacar conclusiones de los resultados obtenidos, particularizando en la serie económica que nos interesa y a plantear cuestiones que en un futuro nos interesaría resolver.

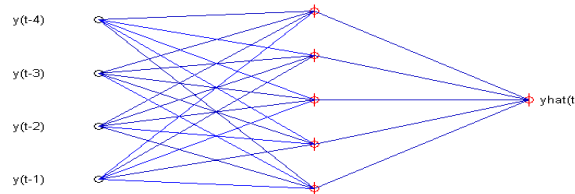
## 2. REDES NEURONALES

Una red neuronal artificial es un modelo matemático inspirado en la estructura de las redes neuronales biológicas y en cómo procesan la información. Estos modelos están formados por unidades simples, llamadas nodos o neuronas, que se conectan entre sí y se disponen en capas. En casi todas las redes neuronales hay una capa de entrada que recibe la información y una capa de salida encargada de transmitir la información procesada al exterior y puede haber una o varias capas ocultas encargadas de establecer relaciones complejas. Las conexiones entre las neuronas indican la dirección y el sentido en el que fluye la información.

Aunque hay muchos tipos de redes neuronales, cuando el objetivo es predecir el valor futuro de una serie temporal, las más utilizadas son la *redes multicapa hacia delante* (MLP), en

las que la información que llega a cada nodo procede de los nodos de la capa anterior, y las *redes recurrentes*, en las que información procesada en la capa oculta o en la capa de salida se toma nuevamente como información en la capa de entrada.

El siguiente gráfico representa una red neuronal con 4 nodos en la capa de entrada que reciben la información de exterior, cinco en la capa oculta y una en la capa de salida.



La información que llega a una neurona  $j$  de la capa oculta es  $\sum_{i=1}^4 \beta_{ij} x_i + \beta_{0j} = \beta'_j x + \beta_{0j}$  donde el vector  $\beta'_j = (\beta_{1j}, \beta_{2j}, \beta_{3j}, \beta_{4j})$  recoge los pesos de las conexiones entre las unidades de la capa de entrada y la neurona  $j$  de la capa oculta. En cada una de las neuronas de la capa oculta se aplica una función no lineal  $\sigma$ , llamada de activación, y  $\sigma(\beta'_j x + \beta_{0j})$  la información que parte de la unidad oculta  $j$  hacia la neurona de salida. La información total que llega a la neurona de salida es

$$\alpha_0 + \sum_{j=1}^5 \alpha_j \sigma(\beta'_j x + \beta_{0j})$$

y el resultado de aplicar una función de activación  $F$  correspondiente a esta neurona es

$$F \left( \alpha_0 + \sum_{j=1}^5 \alpha_j \sigma(\beta'_j x + \beta_{0j}) \right).$$

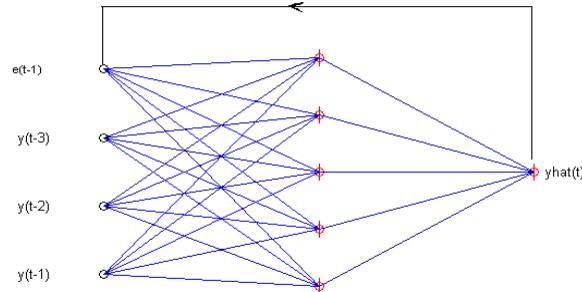
Cuando la función de activación de las neuronas de capa oculta es la logística o la tangente hiperbólica la red recibe el nombre de MLP. En general, la función de activación  $F$  es la identidad, aunque es posible utilizar otra función dependiendo de la naturaleza de los datos, y se considera que todos los nodos de la capa oculta tienen la misma función de activación.

En general una red MLP con  $q$  una capa oculta, una capa de entrada con  $m$  nodos y un nodo en la capa de salida se escribe como

$$h(x, \theta) = F \left( \alpha_0 + \sum_{j=1}^q \alpha_j \sigma(\beta'_j x + \beta_{0j}) \right)$$

donde  $x \in \mathbb{R}^m$  y  $\theta \in \mathbb{R}^{(m+2)q+1}$ .

Hay distintos tipos de redes recurrentes pero en este trabajo se utilizaran aquellas que incorporan como información en la capa de entrada resultados de la capa de salida en iteraciones anteriores y cuya arquitectura puede representarse como en el gráfico siguiente



Mientras que las redes MLP son una extensión no lineal de los modelos econométricos autorregresivos (AR), las redes recurrentes suponen una extensión no lineal de los modelos de medias móviles (MA).

El proceso de aprendizaje de estas redes es supervisado porque para cada  $m$ -úpla de patrones presentados a la red, el resultado obtenido se compara con el resultado deseado del que también se tiene información temporal. La diferencia entre estos valores se utiliza para ajustar los pesos  $\beta_{ij}$  y  $\alpha_k$  de la red utilizando un algoritmo iterativo.

La característica fundamental de estas redes es su capacidad aproximadora de cualquier función medible ( $f : D \subset \mathbb{R}^m \rightarrow \mathbb{R}$ ). Es decir, se puede construir una red neuronal tan próxima a la función  $f$  como se desee con tan sólo aumentar el número de nodos ocultos (Hornik y otros.1989).

## 2.1. Estimación de los parámetros

El objetivo principal en el estudio de series temporales con redes neuronales es construir una red neuronal que permita predecir con el menor error posible, al menos, un dato futuro de la serie. Si la serie temporal respondiera a un modelo autorregresivo de la forma  $y_t = g(y_{t-1}, \dots, y_{t-m}) + \varepsilon_t$  con  $g$  medible y  $\varepsilon$  i.i.d., el teorema de Hornik, Stinchcombe y White asegura la existencia de una red tan próxima como se desee a  $g$  (la distancia  $L^2$  de la red a  $g$  dependerá del número de nodos en la capa oculta, a mayor número de nodos menor error).

Por tanto, fijado el número de nodos de la capa oculta y bajo condiciones de compacidad en el conjunto donde toman valores los parámetros de la red, existe una red  $h_q$  que minimiza la distancia a la función  $g$ . Pero esta red  $h_q$ , no es necesariamente la estimada a partir

de los datos disponibles, denotada por  $h_{q,n}$ , de modo que el verdadero error cometido al tomar  $h_{q,n}$  en sustitución de  $g$ , llamado error de generalización, tiene dos componentes, uno que no depende del número de datos sino de la distancia entre  $g$  y  $h_q$ , llamado de error de aproximación, y el otro cuyo valor se debe al error de estimación debido al número de datos. Así se garantiza dado una serie temporal que responda a  $y_t = g(y_{t-1}, \dots, y_{t-m}) + \varepsilon_t$ , si se dispone de suficientes datos es posible aprender  $g$  utilizando redes neuronales MLP.

Una vez fijada la arquitectura de la red y unos valores iniciales para el vector de parámetros (generalmente aleatorios) se lleva a cabo el entrenamiento de la red, que consiste en la estimación de las conexiones que minimizan el error cuadrático medio. Este proceso de estimación es iterativo y se pueden utilizar diferentes algoritmos, bien usando todos los datos o uno a uno en cada iteración.

Los resultados de White (1989) y Kuan and White (1994) establecen los fundamentos teóricos que permiten obtener estimaciones de los parámetros de una red neuronal.

### 3. TEST BDS

Se han propuesto muchos tests para contrastar si una serie tiene una estructura concreta frente a su ausencia. Sin embargo, muchas veces es más interesante contrastar si una serie temporal es independiente e idénticamente distribuida (i.i.d.) frente a cualquier estructura que pueda tener la serie, como por ejemplo no linealidad determinista o estocástica o no estacionariedad. Brock, Dechert y Scheinkman (1987) desarrollaron este test, denominado BDS, que permite realizar este contraste. Se trata de un test no paramétrico y está basado en el concepto de correlación integral, utilizado en el análisis de sistemas dinámicos caóticos.

Sea  $(\alpha_t)_{t=1}^n$  una serie temporal de longitud  $n$  (que por simplicidad se considera unidimensional). Se llama  $m$ -historia a cada uno de los vectores  $m$ -dimensionales  $\bar{\alpha}_t = (\alpha_t, \alpha_{t+1}, \dots, \alpha_{t+m-1})$  que se pueden formar a partir de la serie.

Para  $m$  y  $\varepsilon > 0$ , la probabilidad de que cualesquiera dos  $m$ -historias estén a una distancia menor que  $\varepsilon$ , cuando  $n \rightarrow +\infty$ , se llama correlación integral. Una estimación de ésta viene dada por

$$C_m(\varepsilon) = \frac{2}{(n-m)(n-m+1)} \sum_{t < s} I_\varepsilon(\bar{\alpha}_t, \bar{\alpha}_s)$$

donde  $I_\varepsilon(\bar{\alpha}_t, \bar{\alpha}_s) = 1$  si  $\|\bar{\alpha}_t - \bar{\alpha}_s\| < \varepsilon$ , siendo  $\|\cdot\|$  la norma del supremo, y  $I_\varepsilon(\bar{\alpha}_t, \bar{\alpha}_s) = 0$  en caso contrario. Si la serie temporal es i.i.d. (la hipótesis nula del test BDS), para cualquier distribución que siga, se tiene que  $C_m(\varepsilon) = [C_1(\varepsilon)]^m$ , frente a cualquier estructura existente. El valor del estadístico de este contraste es

$$\sqrt{n} \frac{C_m(\varepsilon) - [C_1(\varepsilon)]^m}{\sigma_{m,n}(\varepsilon)},$$

donde  $\sigma_{m,n}$  es la varianza estimada de  $C_m(\varepsilon) - [C_1(\varepsilon)]^m$  y cuya distribución asintótica sigue una distribución  $N(0,1)$  para todo  $m$  y  $\varepsilon$ .

Para muestras pequeñas los valores los valores críticos del estadístico no coinciden con los de la  $N(0,1)$  así como cuando se aplica el estadístico a los residuos estimados de un modelo (de Lima,1996). Se han elaborado tablas para los valores críticos del estadístico para distintos tamaños muestrales y valores de  $\varepsilon$ , utilizando simulaciones de Montecarlo (Kanzler,1999).

### 3.1. Dimensión de correlación

Dada una serie temporal de longitud  $n$  y  $m \ll n$ , se construyen las  $m$ -historias, obteniéndose  $n-m+1$  observaciones  $m$ -dimensionales. Se define la dimensión de correlación correspondiente a  $m$  como el valor del límite

$$\lim_{\varepsilon \rightarrow 0^+} \lim_{n \rightarrow \infty} \frac{\ln C_m(\varepsilon)}{\ln(\varepsilon)}.$$

Si la serie temporal tiene una explicación determinista entonces el límite anterior existe y se estabiliza en torno a un valor real  $k$ , para cualquier  $m > k$ . Contrariamente cuando la serie tiene una componente aleatoria, la dimensión de correlación crece con  $m$  (Brock y otros, 1991). El valor del límite anterior puede ser aproximado por la pendiente de la recta ajustada a la nube de puntos  $(\ln(\varepsilon), \ln C_m(\varepsilon))$  para valores pequeños de  $\varepsilon > 0$ .

## 4. DATOS Y METODOLOGIA

Primeramente vamos a establecer a qué series temporales vamos a aproximarnos en este trabajo. Se consideran las siguientes cuatro series temporales:

- Una serie i.i.d. cuya distribución es  $N(0,1)$ .
- Una serie determinista procedente de un sistema dinámico de Hénon.
- La serie determinista de Hénon contaminada con ruido de varianza pequeña.
- La serie de rendimientos del Ibex-35.

En las series generadas se han descartado los primeros datos para evitar una posible dependencia del algoritmo generador o de las condiciones iniciales, quedándonos con 2434 datos que es la longitud de la serie económica de rendimientos del Ibex-35.

#### 4.1. Valores del estadístico BDS<sup>1</sup>

En primer lugar, se ha pasado a las series el test BDS para contrastar si la serie tiene alguna estructura subyacente.

##### a) Serie i.i.d.

Tabla 1: Valores del estadístico BDS para una serie i.i.d.

	<b>m = 2</b>	<b>m = 3</b>	<b>m = 4</b>	<b>m = 5</b>	<b>m = 6</b>	<b>m = 7</b>	<b>m = 8</b>	<b>m = 9</b>	<b>m = 10</b>
$(1/2)\sigma$	-0,1901	-0,388	-0,655	-0,407	-0,464	-0,2434	0,0522	0,274	-2,456
$\sigma$	0,186	-0,296	-0,386	-0,166	-0,456	-0,588	-0,619	-0,719	-0,780
$(3/2)\sigma$	0,106	-0,228	-0,415	-0,514	-0,636	-0,788	-0,865	-0,850	-0,846
$2\sigma$	0,179	-0,0622	-0,372	-0,334	-0,416	-0,562	-0,605	-0,580	0,582

##### b) Serie de Henon

Tabla 2: Valores del estadístico BDS para una serie determinista de Henon.

	<b>m = 2</b>	<b>m = 3</b>	<b>m = 4</b>	<b>m = 5</b>	<b>m = 6</b>	<b>m = 7</b>	<b>m = 8</b>	<b>m = 9</b>	<b>m = 10</b>
$(1/2)\sigma$	182,26	306,59	452,24	745,22	1307,52	24446,92	4759,22	9652,43	20475,8
$\sigma$	81,04	81,043	91,835	98,61	109,94	130,59	156,36	190,27	236,11
$(3/2)\sigma$	37,58	27,56	29,064	25,943	26,38	27,56	28,99	30,45	32,67
$2\sigma$	13,27	4,115	3,096	-1,492	-2,72	-3,716	-3,158	-2,774	-1,881

##### c) Series de Hénon perturbadas con un ruido.

Tabla 3: Valores del estadístico BDS para una serie con componente determinista más un ruido normal de desviación típica un cuarto de la desviación típica de la serie de Hénon.

	<b>m = 2</b>	<b>m = 3</b>	<b>m = 4</b>	<b>M = 5</b>	<b>m = 6</b>	<b>m = 7</b>	<b>m = 8</b>	<b>m = 9</b>	<b>m = 10</b>
$(1/2)\sigma$	117,75	161,26	225,72	322,91	485,5	761,42	1254,8	2161,5	3951,1
$\sigma$	71,18	71,78	86,19	95,37	109,45	129,99	156,28	191,15	239,15
$(3/2)\sigma$	33,22	25,06	26,77	24,61	24,59	25,60	27,12	28,40	30,45
$2\sigma$	9,18	2,39	1,24	-1,87	-2,87	-3,201	-2,38	-1,72	-0,75

##### c) Serie de Hénon perturbada con un ruido de pequeña varianza

Tabla 4: Valores del estadístico BDS para una serie con componente determinista más un ruido normal con desviación típica la mitad de la desviación típica de la serie de Hénon.

	<b>m = 2</b>	<b>m = 3</b>	<b>m = 4</b>	<b>M = 5</b>	<b>m = 6</b>	<b>m = 7</b>	<b>m = 8</b>	<b>m = 9</b>	<b>m = 10</b>
$(1/2)\sigma$	42,56	45,68	51,43	55,35	61,78	70,91	82,77	100,17	125,51
$\sigma$	34,88	34,50	37,14	37,69	39,89	43,35	47,66	53,09	60,53
$(3/2)\sigma$	20,13	17,38	17,05	15,42	14,17	15,04	11,63	16,49	17,63
$2\sigma$	8,54	5,45	3,93	2,018	1,204	0,762	0,863	1,23	1,76

<sup>1</sup> En las tablas se muestran los valores obtenidos por el estadístico BDS para  $\epsilon$  igual a  $1/2\sigma$ ,  $\sigma$ ,  $3/4\sigma$  y  $2\sigma$ , donde  $\sigma$  es la desviación típica de la serie. Los valores críticos son: 1,645 (10%); 1960 (5%); 2,576 (1%).

d) Serie de rendimientos del Ibex-35

Tabla 5: Valores del estadístico BDS para la serie Libex-35 .

	<b>m = 2</b>	<b>m = 3</b>	<b>m = 4</b>	<b>M = 5</b>	<b>m = 6</b>	<b>m = 7</b>	<b>m = 8</b>	<b>m = 9</b>	<b>m = 10</b>
$(1/2)\sigma$	6,40	10,22	13,66	16,85	19,77	22,53	27,33	35,45	47,17
$\sigma$	6,92	10,61	13,88	16,57	18,87	21,19	23,82	27,18	31,61
$(3/2)\sigma$	7,32	19,96	13,95	16,08	17,92	18,92	20,38	21,99	23,79
$2\sigma$	7,61	11,14	13,74	15,51	16,60	17,42	18,32	19,16	19,97

Los valores del estadístico BDS confirman que la primera serie es i.i.d., mientras que en la serie de Hénon hay una estructura subyacente en los datos. Se observa que cuando se aplica el test a la serie de Hénon perturbada con un ruido de pequeña varianza, aunque el estadístico BDS rechaza la hipótesis de i.i.d. indicando estructura subyacente sus valores disminuyen respecto a los valores que toma con la serie determinista no perturbada.

La serie de rendimientos de Ibex-35 tiene 2434 datos puesto que se dispone de información del Ibex-35 al cierre de la sesión desde el 1 de Febrero de 1993 hasta 1 de Febrero de 2003. La serie de rendimientos sobre el Ibex-35 se calcula como la primera diferencia de los logaritmos del índice y se denotará como Libex-35.

$$libex_t = s_t = \ln(p_t / p_{t-1})$$

A esta serie se le ajusta un modelo AR(2) que recoge parte de la estructura dinámica de la media (Font y Baixauli, 2001) y a la serie de sus residuos se les sigue llamando Libex-35. En la tabla 5 se recogen los valores del estadístico BDS para esta serie y confirman que la serie tiene alguna estructura de carácter no lineal.

**4.2. Cálculo de la dimensión de correlación**

Para una serie temporal de Hénon, la dimensión de correlación existe y es un valor comprendido entre 1 y 2, cuya estimación se denota por  $\beta_m$  para  $m$  la dimensión de las historias

$\beta_2$	$\beta_3$	$\beta_4$	$\beta_5$
1,180	1,197	1,223	1,213

Estos valores confirman que la serie no es i.i.d. Sin embargo, para la serie normal el valor de este límite va creciendo con  $m$  y toma valores muy próximos a  $m$ , indicando que sea cual sea la dimensión de las  $m$ -historias construidas a partir de la serie temporal, siempre se llena todo el espacio.

$\beta_2$	$\beta_3$	$\beta_4$	$\beta_5$
1,944	2,939	3,885	4,850

Cuando se trata de la serie de Hénon más ruido se tienen las estimaciones de la dimensión de correlación siguientes

$\beta_2$	$\beta_3$	$\beta_4$	$\beta_5$
1,768	2,513	3,261	3,95

$\beta_2$	$\beta_3$	$\beta_4$	$\beta_5$
1,886	2,815	3,763	4,65

se observa que cuando el ruido añadido tiene mayor varianza la dimensión de correlación aumenta para cada  $m$ . Según Kocenda, cuando el ruido es pequeño, la estimación de la dimensión de correlación confirmaría la existencia de estructura en la serie mientras que en el segundo caso pesaría más el ruido.

Para la serie de rendimientos del Ibex-35, la estimación de la dimensión de correlación para cada  $m$ , indicaría una fuerte componente aleatoria y entraría en contradicción con los resultados del test BDS.

$\beta_2$	$\beta_3$	$\beta_4$	$\beta_5$
1,86	2,78	3,7	4,56

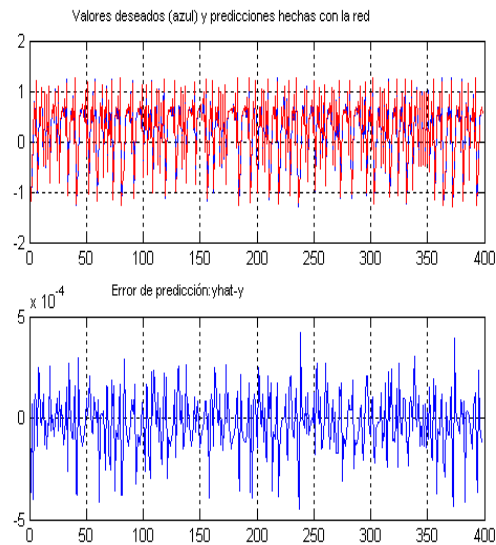
### 4.3. Uso de las redes neuronales para estas series

En el proceso de ajuste de los parámetros de una red neuronal a un conjunto de datos se distinguen dos fases: La fase de entrenamiento, cuyo objetivo es estimar los parámetros y la fase de validación, que consiste en comprobar que sobre datos que no se han utilizado en el proceso de estimación, la red construida se comporta bien. Es decir la red ha aprendido de los datos. Para llevar a cabo este proceso se toman aproximadamente el 80% de los datos para entrenar la red que se proponga y el resto para comprobar cómo generaliza.

Primero se presenta la serie de Hénon a diferentes redes neuronales (redes con distintas arquitecturas), con distinto número de regresores y de nodos en la capa oculta. De los resultados obtenidos, se tiene que son suficientes 5 ó 7 neuronas en la capa oculta y dos nodos en la capa de entrada. El error sobre el conjunto de entrenamiento y el error sobre el conjunto de validación son prácticamente iguales y muy pequeños (del orden de  $10^{-5}$ ), en estos casos se dice que se tiene una estimación del error medio de generalización. En el siguiente gráfico se observa cómo los resultados obtenidos por la red se superponen sobre los datos deseados en el conjunto test. Sin embargo, si una vez estimada una red MLP se toma la serie de los residuos y se le aplica el test BDS sus resultados no confirman que sean i.i.d.<sup>2</sup>

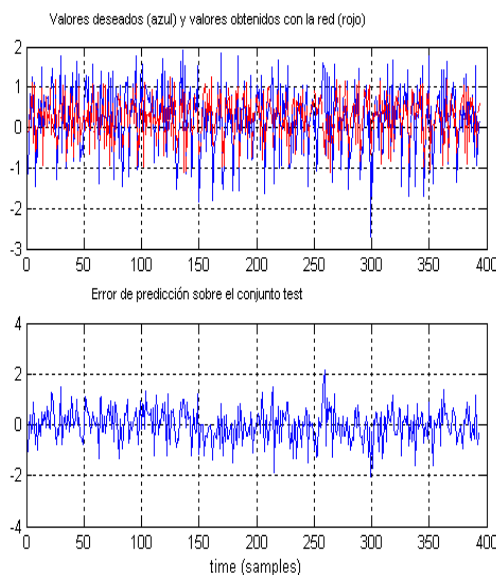
---

<sup>2</sup> No se ha estudiado la distribución asintótica del estimador BDS si se aplica sobre los residuos que se obtienen tras estimar una red neuronal.

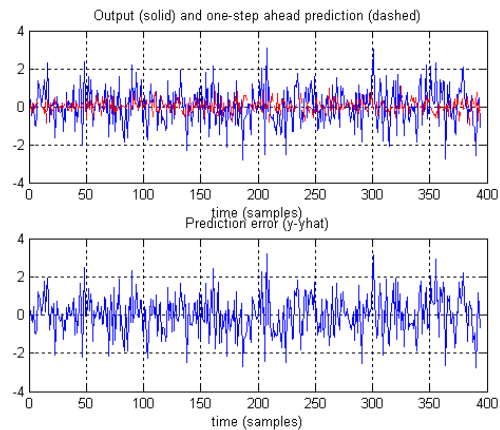


Análogamente se han presentado los datos de las series de Hénon contaminadas con ruidos normales y de los resultados obtenidos se puede destacar que el error tanto sobre el conjunto de entrenamiento como sobre el conjunto test es más pequeño cuando el número de unidades en la capa de entrada es 6 frente a un menor número de regresores, y la diferencia entre ambos errores crece cuando aumenta la varianza del ruido que contamina los datos deterministas. Sin embargo, si se pasa el test BDS sobre los residuos una vez estimados los parámetros de la red no se rechaza la hipótesis de i.i.d. Se podría decir que la red recoge todo el comportamiento no lineal determinista de esta serie.

La correlación entre los valores deseados y los obtenidos a través de la red es 1 cuando la serie de Hénon no está contaminada y va disminuyendo conforme aumenta la varianza del ruido añadido a la serie.

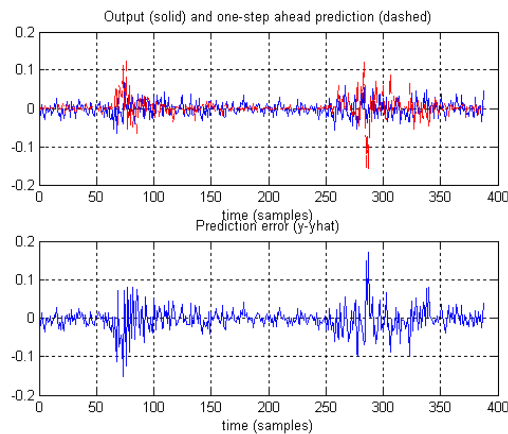


También se ha intentado ver si una red MLP puede aprender el comportamiento aleatorio de una serie i.i.d. Se han utilizado distintas arquitecturas, probando con distinto número de neuronas de entrada y neuronas en la capa oculta. En todos los casos el error sobre el conjunto de entrenamiento y el conjunto test es aproximadamente igual a la varianza de la serie. La red no explica nada de la serie y ello puede observarse en la gráfica siguiente



La serie de rendimientos del Ibex-35 ha sido ampliamente estudiada en la última década. El rechazo de la hipótesis de que los residuos estimados a partir de un modelo AR sean i.i.d. ha llevado a muchos investigadores a considerar que se debe a heterocedasticidad, en particular, no linealidad en varianza (modelos ARCH y GARCH). Pero aunque las redes neuronales se han utilizado también para expresar relaciones funcionales no lineales de la varianza, en este trabajo se querían utilizar para intentar predecir la propia serie. Como para cada  $m$ , el valor de la dimensión de correlación aumenta, se necesitan bastantes retardos de la serie para explicar su comportamiento. Una red neuronal con 12 retardos de la serie como regresores daba mejores resultados que con menos retardos y similares cuando se utilizaban más por tanto, se han elegido 12 neuronas en la capa de entrada. Para la capa oculta, también se ha probado con distinto número de unidades y se ha optado por 13 nodos.

Se observa que el error sobre el conjunto de entrenamiento es mucho más pequeño que sobre el conjunto test por lo que no se puede hablar de una estimación del error medio de generalización. Sobre los residuos estimados correspondientes al conjunto de entrenamiento, el test BDS rechaza i.i.d. tomando valores muy próximos a los valores críticos, que como ya se ha dicho no han sido corregidos. Sin embargo, sobre los residuos estimados correspondientes al conjunto test, los valores del estadístico BDS se alejan de los valores críticos.



Se podría pensar que la red sobreajusta los datos del conjunto de entrenamiento y es incapaz de generalizar a datos no utilizados en éste, pero creemos que aunque es un factor a tener en cuenta, es necesario considerar otros aspectos, como la inclusión de otras variables en la capa de entrada, (porque no sea posible predecir esta serie únicamente con retardos de la misma), la consideración de un horizonte temporal más próximo para las predicciones,...

## 5. CONCLUSIONES

En este trabajo se han aplicado redes neuronales con objeto de predecir valores futuros de series temporales de diferente naturaleza. Se han obtenido buenos resultados cuando se trataba de series deterministas no lineales o estas mismas perturbadas por un pequeño ruido; en cambio, como era de esperar, las redes no han sido capaces de captar el comportamiento i.i.d. Respecto a la serie de rendimientos del Ibex-35, las redes son capaces de reproducir parte de su comportamiento en el conjunto de entrenamiento, pero su capacidad de predicción bastante escasa, quizá debido a la volatilidad de la serie

En el desarrollo del trabajo se han utilizado sólo redes MLP porque, aunque creíamos que el uso de redes recurrentes iba a funcionar mejor, con las datos simulados no se han obtenido mejores resultados que hicieran recomendable su uso.

Creemos que es necesario realizar un estudio más profundo de las relaciones no lineales de los retardos de las serie de rendimientos del Ibex-35 para determinar con mayor precisión el número de retardos a considerar en las redes neuronales y ver si se obtienen mejores resultados. Además, otra perspectiva interesante sería intentar predecir la volatilidad de la serie mediante redes neuronales en lugar de los valores futuros de ésta.

## 6. REFERENCIAS BIBLIOGRÁFICAS

- BROCK, W; DECHERT, W. and SCHEINKMAN, J. (1987). "A test for independence based on the correlation dimension". Working Paper, University of Chicago.
- BROCK, W., HSIEH, D. and LEBARON, B. (1991) "Nonlinear Dynamics, chaos and instability: Statistical theory and economic evidence". MIT press. Cambridge.
- CABALLERO, M.V. y MOL C.-M. Kuan and H. White: "Artificial Neural Networks: An Econometric Perspective," *Econometric Reviews*, 13, 1-92 (1994). ERA, L. (2004). "Redes neuronales perceptrón multicapa" Documento de trabajo nº 45. Facultad de Economía y Empresa. Universidad de Murcia.
- DE LIMA, P. (1996) "Nuisance parameter free properties of correlation integral based statistics". *Econometric Review* 15, 3, 237-259.
- FONT, B. y BAIXAULI, J.S. (2001). "Evidencia sobre eficiencia en el mercado de capitales español" *Revista Europea de Dirección y Economía de la Empresa* 10, 3, 135-166.
- HORNIK, K., STINCHCOMBE, M. and WHITE, H. (1989). "Multilayer feedforward networks are universal approximators". *Neural Networks* 2, 356-366.
- KANZLER, L. (1998). "[BDS: MATLAB module to calculate Brock, Dechert & Scheinkman test for independence](#)," [Statistical Software Components](#) T871803, Boston College Department of Economics.
- KUAN, C.M. and WHITE, H. (1994). "Artificial Neural Networks: An Econometric Perspective," *Econometric Reviews*, 13, 1-92 .
- NORGAARD, M., RAVN, O., POULSEN, N.K. and HANSEN, L.K. (2003). "Neural Networks for modelling and control of dynamic systems" 3<sup>a</sup> edition. Springer-Verlag. London.
- OLMEDA, I. y PEREZ, J. (1995). "Non-linear dynamics and chaos in the spanish stock market". *Investigaciones Económicas* 19, 2, 217-248.
- WHITE, H. (1989). "Some asymptotic results for learning in single hidden-layer feedforward network models" *Journal of the American Statistical Association*, 84, 1003-1013