

# BOLETÍN MATEMÁTICO

Instituto de Investigaciones de Matemática Aplicada  
Facultad de Ciencias Económicas y Empresariales

17

ABRIL 2009

REGISTRO DE LA PROPIEDAD INTELECTUAL ISSN 0329-0255



UNIVERSIDAD DE MORÓN

**Autoridades de la  
Facultad de Ciencias  
Económicas y  
Empresariales**

**Decano**

Dr. Jorge Raúl Lemos

**Vicedecano**

Dr. Jorge Emilio Salvel

**Secretario Académico**

Dr. Osvaldo Luis Perillo

**Secretaria Adjunta**

Dra. Amanda Raquel Llistosella

**Director de Estudios y Coordinación**

Dr. Vicente Filleti

**Directora de Investigaciones y Enseñanza Experimental**

Prof. Elvira Venturo

**Consejeros del Honorable Consejo Académico  
de la Facultad de Ciencias Económicas y Empresariales**

Dr. Miguel Gregorio Skubic

Dra. Norma Beatriz Irigoyen

Dr. Raúl Roque Orellano

Dr. Fernando Appesseche

Dr. Sergio Andrés Ghedin

Dr. Domingo José Mazza

**Representante de Profesores ante el H. C. S.**

Dra. Alicia I. de Montagut de Rodriguez

**Directores de Carrera:**

Dr. Raúl Roque Orellano (Contador Público)

Dr. Miguel Gregorio Skubic (Licenciatura en Administración)

Lic. Domingo José Mazza (Licenciatura en Economía)

Lic. Carlos Alberto Ferreras (Licenciatura en Comercialización /

Técnico Superior en Comercialización)

Lic. Luis Fabián García Nicora (Licenciatura en Recursos Humanos /

Analista Universitario en Recursos Humanos)

A cargo del Señor Decano (Licenciatura en Relaciones Públicas /

Analista Universitario en Relaciones Públicas)

Dra. Amanda Raquel Llistosella (Licenciatura en Seguros / Técnico Superior en Seguros)

Lic. Marcelo Emilio Mirón (Técnica en Comercialización Minorista)

Lic. Germán Avelino Kraus (Licenciatura en Comercio Internacional)

**Directores de Institutos de Investigación:**

• Instituto de Investigaciones Contables

Dr. Isaac Aizik Senderovich

• Instituto de Investigaciones Económicas

Dr. Vicente H. Monteverde

• Instituto de Investigaciones Administrativas

A Designar

• Instituto de Investigaciones de Matemática Aplicada

Ing. Luino Edelfio Vilches

• Instituto de Investigaciones Tributarias

Dr. Juan Ferrari Herrero

Subdirector: Dr. Alfredo Destuniano

• Instituto de Metodología Jurídica Aplicada en las Ciencias Económicas

Dr. Eduardo Mario Favier Dubois

• Instituto de Investigaciones de la Pequeña y Mediana Empresa

Dr. Horacio Armando Irigoyen

• Instituto de Investigaciones de Humanidades y Ciencias Sociales Aplicadas

a las Ciencias Económicas y Empresariales

Prof. Elvira Venturo

**Directores de Departamentos Pedagógicos**

• Área Pedagógica de Administración

Dr. Jorge Eduardo Marcos

• Área Pedagógica de Contabilidad

Dr. Sergio Daniel Arguissain

• Área Pedagógica de Economía

Dr. Vicente Filleti

• Área Pedagógica de Humanidades

Prof. Elvira Venturo

• Área Pedagógica Jurídica

Dr. Eduardo Mario Favier Dubois

Subdirectora: Dra. Amanda Raquel Llistosella

• Área Pedagógica de Matemática

Ing. Martín Adler

• Área Pedagógica de Comercialización

Dr. Fernando Appesseche

Instituto de Investigaciones de Matemática  
Aplicada de la Facultad de Ciencias  
Económicas y Empresariales de la UM



# BOLETÍN MATEMÁTICO

Año 11 - Nº 17 - Abril 2009

Registro de la Propiedad Intelectual: ISSN. 0329-0255

## ÍNDICE

|                                                                                                                                                                                                            | Pág. |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------|
| <b>Métodos para estimar el componente cíclico en series temporales con aplicaciones al Producto Bruto Interno por Habitante en Argentina - (Segunda Parte)</b><br>Autor: Dr. José Luis Iparraguirre D`Elía | 5    |
| <b>Convergencia del Perceptron</b><br>Autores: Profesores Lisandro Curia y Francisco Paz                                                                                                                   | 31   |

Las opiniones vertidas en los trabajos que se publican en este Boletín son de exclusiva responsabilidad de sus autores.

## STAFF

### Director

Ing. Luinor E. Vilches  
[lvilches@unimoron.edu.ar](mailto:lvilches@unimoron.edu.ar)

### Redacción

Profesores de la Universidad de Morón  
y colaboradores especiales

### Producido por la Oficina de Medios UM

#### Editor:

Lic. Alejandro Gavric

#### Diseño Gráfico:

DCV. Sandra Luján

#### Coordinación:

Lic. Marcela Golía

#### Corrección:

Prof. Susana Lamaison

### Impreso en los Talleres Gráficos UM

### Año 11 Número 17

Registro de la Propiedad  
Intelectual ISSN 0329-0255

### Universidad de Morón

Cabildo 134 - (B1708JPD) Morón  
(011) 5627-2000 (líneas rotativas)  
Fax: 5627-2002

E-mail: [webmaster@unimoron.edu.ar](mailto:webmaster@unimoron.edu.ar)

Internet: [www.unimoron.edu.ar](http://www.unimoron.edu.ar)

LOS CONTENIDOS DE TODOS LOS NÚMEROS DEL BOLETÍN  
ESTÁN INSTALADOS EN LA PÁGINA WEB DE LA  
UNIVERSIDAD DE MORÓN:

[www.unimoron.edu.ar](http://www.unimoron.edu.ar) → Facultades → Facultad de  
Ciencias Económicas y Empresariales → Publicaciones

Universidad de Morón  
Cabildo 134 - (B1708JPD) Morón  
54-011-5627-2000 (líneas rotativas) - Fax: 5627-2002  
[e-mail: webmaster@unimoron.edu.ar](mailto:webmaster@unimoron.edu.ar) - Internet: [www.unimoron.edu.ar](http://www.unimoron.edu.ar)

Se autoriza la reproducción parcial o total de los artículos publicados en este Boletín con la condición de que se mencionen su fuente y sus autores.

# MÉTODOS PARA ESTIMAR EL COMPONENTE CÍCLICO EN SERIES TEMPORALES, CON APLICACIONES AL PRODUCTO BRUTO INTERNO POR HABITANTE EN ARGENTINA

## Segunda Parte (\*)

(\*) Ver primera parte en el Boletín Matemático N° 16, octubre 2008.

(\*\*) Investigador de la Facultad de Ciencias Económicas y Empresariales de la Universidad de Morón y Economista Senior en el Economic Research Institute of Northern Ireland, Belfast, Irlanda del Norte, Reino Unido.

Por el Dr. José Luis Iparraguirre D`Elía (\*\*)

### *2.e. Filtro de Christiano-Fitzgerald*

El filtro de Baxter-King asume, como vimos, que la tendencia responde a un promedio finito con ponderaciones simétricas. En cambio, el filtro desarrollado por Christiano y Fitzgerald (2003) utiliza un promedio con ponderaciones cambiantes no simétricas. Dado que este filtro permite que los ponderadores varíen según cada observación, se considera un filtro más flexible que el de Baxter-King.

Asimismo, recordemos que para estimar el ciclo según el filtro de Baxter-King, perdimos una misma cantidad de observaciones al comienzo y al final de la serie (en nuestro caso, 3). En cambio, el filtro de Christiano-Fitzgerald es asimétrico, es decir que puede computarse para los valores extremos de la serie. Esta asimetría se obtiene utilizando toda la información disponible en cada período (tanto hacia adelante, como correspondiente a los períodos anteriores) para proceder al filtrado de la serie. El algoritmo del promedio ponderado en sí se supone que sigue un modelo de camino aleatorio. El componente cíclico según el filtro de Christiano-Fitzgerald se obtiene a partir de la siguiente expresión:

$$c_t = B_0 y_t + B_1 y_{t+1} + \dots + B_{T-1-t} y_{T-1} + \tilde{B}_{T-t} y_T + B_1 y_{t-1} + \dots + B_{t-2} y_2 + \tilde{B}_{t-1} y_1$$

donde

$$B_j = \frac{\sin(jb) - \sin(ja)}{\pi j}, \quad j \geq 1$$

$$B_0 = \frac{b-a}{\pi}$$

$$a = \frac{2\pi}{p_u}$$

$$b = \frac{2\pi}{p_l}$$

$p_u$  es el período de oscilación inferior,  $p_l$  el período de oscilación superior y  $2 \leq p_u < p_l < \infty$ . (Las series mayores a  $p_l$  y menores a  $p_u$  son capturadas por el término  $c_t$ ). Para datos anuales, los valores sugeridos de  $p_u$  y  $p_l$  son 2 y 8, respectivamente. Finalmente,  $\tilde{B}_k$  son funciones lineales de los ponderadores  $B_j$ :

$$\tilde{B}_k = -\frac{1}{2} B_0 - \sum_{j=1}^{k-1} B_j$$

La **Figura 8** presenta el componente cíclico del PBI por habitante de la República Argentina entre 1950 y 2006, según el filtro de Christiano-Fitzgerald.

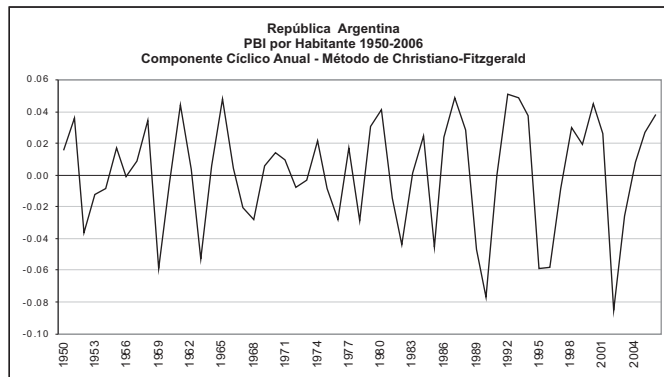


Figura 8

## 2.f. Método de Beveridge-Nelson

Mencionamos en la introducción que la visión tradicional de una serie de tiempo es que la misma puede descomponerse en elementos ortogonales no observables –principalmente, el tendencial, el cíclico, el estacional y el término de error–. Nelson y Plossner (1982) argumentaron que si la serie de tiempo es no estacionaria –como por ejemplo, es el caso de la mayoría de las series macroeconómicas–, es decir, si presenta raíces unitarias en su representación autorregresiva, el componente tendencial no puede ser considerado determinístico sino un proceso estocástico que define el comportamiento de largo plazo de la variable, mientras que el componente cíclico solamente es transitorio, es decir que las fluctuaciones se disipan con el tiempo. Los cambios observados en la serie son resultado de ambos componentes.

Qué persistentes sean los efectos de las perturbaciones depende de la importancia relativa en el comportamiento de la serie de cada uno de los dos componentes: la tendencia estocástica de largo plazo y el ciclo de corto plazo. Hemos aplicado el método de ratio de varianzas de Cochrane (1988) (ver Apéndice) y hallamos que el ratio tiende a un valor constante ( $\sim 0.65$ ), con lo que la serie de PBI de la República Argentina por habitante entre 1950 y 2006 es estacionaria y siempre regresaría a su valor tendencial de largo plazo si es afectada por choques temporarios. El valor obtenido indica, además, que la varianza del componente tendencial estocástico explica alrededor de dos tercios del total de la varianza de la serie, con lo que el elemento temporario solamente explicaría el tercio restante.

Beveridge y Nelson (1981) propusieron un método de descomposición de series de tiempo univariadas y multivariadas (aunque solamente trataremos las primeras en este artículo) en componentes permanentes y transitorios –ambos estocásticos– que permite asimismo la existencia de correlación entre ambos. El componente permanente se modela como un camino aleatorio con tendencia o deriva (*random walk with drift*). La tendencia

estocástica se define como la predicción futura en el límite (es decir, hacia el infinito) del valor de la variable menos cualquier componente determinístico –en otros términos, el valor en el período actual más el impacto a largo plazo de las desviaciones transitorias de la serie de su valor de estado estacionario. El componente transitorio, por su parte, que es lo que se define como ciclo, surge de la diferencia entre la serie original y la tendencia.

Este enfoque tiene implicancias económicas importantes, mas excede el objeto de este artículo discurrir sobre ellas. Solamente cabe resaltar que el método plantea que el componente permanente de la serie temporal es estocástico, con lo que todas las fluctuaciones se imputan a la tendencia de largo plazo, mientras que los movimientos transitorios registran cambios de corta duración sucedidos a lo largo del tiempo<sup>1</sup>. Un fenómeno económico que podría explicar el carácter estocástico de la tendencia de una serie temporal de producto bruto es el carácter aleatorio de innovaciones con efectos permanentes en la actividad económica (por ejemplo, el ferrocarril, los superconductores, la nanotecnología, etc).

El método consiste en hallar el modelo ARMA (p,q) que mejor ajuste la serie de tiempo en cuestión. La estimación del filtro sigue un procedimiento en dos etapas: primeramente, se aplica un filtro de diferencias primeras a la serie para tornarla estacionaria<sup>2</sup> y eliminar, así, la existencia de raíces unitarias<sup>3</sup> en la serie original, y en segundo término, se estima la tendencia como un promedio móvil autorregresivo ARMA(p,q). El componente irregular se sustrae de la serie original<sup>4</sup>.

Un modelo ARMA (p,q) se compone de un componente autorregresivo (AR) de  $p$  períodos de rezago de la serie de tiempo  $y_t$  que se quiere modelar, y un promedio móvil (MA, *moving average*) de  $q$  períodos de los errores estocásticos. El modelo puede representarse como:

(1) Ver el trabajo reciente de Oh *et al.* (2008) para un tratamiento más extenso de la descomposición de Beveridge y Nelson.

(2) Es decir que la serie de PBI es integrada de orden 1, lo que significa que en cualquier período  $t$ , el valor del PBI ( $y_t$ ) es igual a un valor inicial  $y_0$  más la suma de todas las perturbaciones aleatorias entre el período inicial y el período  $t$ . Es decir que los residuos son de memoria larga; los shocks aleatorios tienen efectos perdurables en el tiempo. Toda serie integrada de orden 1 se puede transformar en una serie integrada de orden cero aplicando diferencias primeras (ver Enders, 1995).

(3) El orden de integración de una serie está asociado a la cantidad de raíces unitarias de la misma –una serie integrada de orden 1 posee una raíz unitaria–. Ver Libanio (2005) para una introducción didáctica a la existencia de raíces unitarias en series temporales macroeconómicas. Sosa Escudero (1997) presenta un estudio pormenorizado para el PBI de la República Argentina. La referencia clásica es Nelson y Posner (1982). Cualquier texto de econometría (por ej., Stock y Watson, 2007) o de series temporales (como Enders, *op. cit.*)



$$y_t = \phi_0 + \sum_{i=1}^p \phi_i y_{t-1} + \sum_{j=1}^q \theta_j \varepsilon_{t-1} \quad (10a)$$

donde,

$\{\varepsilon_t\}$  es ruido blanco, es decir que

$$E(\varepsilon_t) = 0;$$

$$\text{Var}(\varepsilon_t) = \sigma^2; \text{ y}$$

$$\forall i \neq j, \text{Cov}(\varepsilon_i, \varepsilon_j) = 0$$

Así, por ejemplo, un modelo ARMA (0,1) se representa por  $y_t = \phi_0 + \varepsilon_{t-1} + \theta_1 \varepsilon_{t-1}$ ,

mientras que un modelo ARMA (1,0) se representa por

$$y_t = \phi_0 + \phi_1 y_{t-1} + \varepsilon_t$$

Para aplicar este método, en principio calculamos las diferencias primeras de los logaritmos del PBI de la República Argentina entre 1950 y 2006, con lo que obtenemos las tasas anuales de crecimiento del PBI. La diferenciación torna a la serie estacionaria, con lo que podemos rechazar la hipótesis de la existencia de raíces

4) Oh *et al* (2008) califican a este método de extracción del ciclo siguiendo la descomposición de Beveridge-Nelson como el “de la fuerza bruta”.

(5) Estadístico Dickey-Fuller Aumentado para la serie de primeras diferencias sin ordenada ni tendencia:

-6.072200

Valores Críticos:

1% = -2.607686

5% = -1.946878

10% = -1.612999

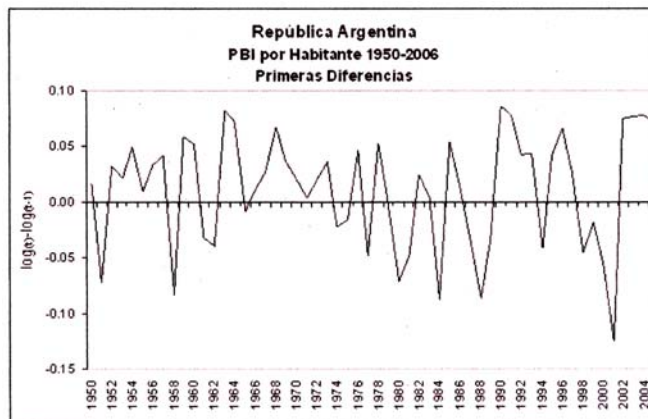


Figura 9

Luego, ajustamos esta serie según un modelo ARMA (p, q) y encontramos que el modelo ARMA (1,1) la representa adecuadamente<sup>6</sup>. Obtenemos la siguiente ecuación (estadístico t entre paréntesis):

$$\Delta y_t = -0.6413 \Delta y_{t-1} + 0.9618 \varepsilon_{t-1} \quad (10b)$$

(-5.97)                      (39.05)

Reemplazando los valores de  $y_t$  por el logaritmo del PBI en el período  $t$ , obtenemos los valores residuales en la ecuación (10b). La diferencia entre las primeras diferencias logarítmicas de la Figura 8 y la curva de ajuste según la ecuación (10b) corresponde al componente cíclico, cuya representación

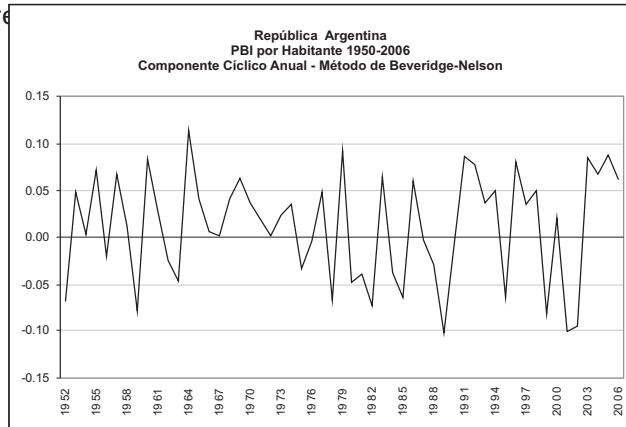


Figura 10

*2.g. Método de Componentes no Observados de Harvey-Clark y el Filtro Kalman*

El método ideado conjuntamente por Harvey (1985) y Clark (1987) supone que la tendencia responde a una evolución suavizada, por cuanto los choques macroeconómicos provocan desviaciones distribuidas según una curva normal. Con este método, la serie se descompone en un componente de tendencia local lineal según un modelo de camino aleatorio con tendencia o deriva

**((6))** Tras estudiar los estadísticos AIC, BIC y el Ljung-Box Q de este modelo y los modelos alternativos ARMA(1,0) y ARMA(1,2) el correlograma indicaba la presencia de un proceso AR(1).

y un componente estocástico que corresponde al ciclo.

La serie puede definirse como:

$$y_t = m \cdot v_t + u_t, \quad (11)$$

donde

$$v_t = a \cdot v_{t-1} + \varepsilon_t \quad (12)$$

es la tendencia y

$$u_t = \phi_1 u_{t-1} + \phi_2 u_{t-2} + \kappa_t + \theta \kappa_{t-1}, \quad (13)$$

es el componente cíclico (los términos  $\varepsilon_t$  y  $\kappa_t$  se suponen normalmente distribuidos).

En el sistema de ecuaciones anterior, la única variable observable es  $y_t$  –de allí que el método se conozca como de 'componentes no observados'. Para resolver este sistema, se recurre a un algoritmo llamado filtro de Kalman<sup>7</sup>. El algoritmo se compone de una ecuación de medida (ecuación 11) y una ecuación de transición (ecuación 12), también conocidos como componente de observación y estado, respectivamente. El filtro permite estimar la variable no observada  $v_t$  definida en la ecuación 12 (conocida como vector de estado) a partir de observaciones pasadas de la misma y de la variable observada  $y_t$ . Estas dos ecuaciones forman lo que se conoce como una "representación en el espacio de los estados" (*state-space representation*). El procedimiento es iterativo y requiere establecer un valor inicial de la variable no observada  $v_t$ , es decir,  $v_0$ , para con ello predecir el valor de  $v_t$ . (el superíndice  $p$  en la ecuación siguiente denota que es un valor predicho).

7) Para exposiciones didácticas sobre el filtro de Kalman, ver Ribeiro (2000) y Arnold *et al.* (2008); un tratamiento exhaustivo puede hallarse en Hamilton (1994). Ver también la temprana presentación en castellano de Terceiro Lomba (1978).

$$v_1^p = a \cdot v_0 + \xi_t \quad (14)$$

Reemplazando (14) en (11) obtenemos:

$$y_1^p = m \cdot v_1^p + u_1 = m \cdot (a \cdot v_0 + \xi_t) + u_1 \quad (15)$$

Ahora bien, la ecuación 15 estima o predice un valor para la variable observable en el momento 1, que puede o no coincidir con el verdadero suceso, es decir con el valor de la variable en dicho momento. Definamos el error de predicción como:

$$y_1^e = y_1 - y_1^p \quad (16)$$

Este error de predicción puede incorporarse en la predicción de la variable no observada para ajustar dicha estimación. Este valor estimado ajustado se acercará al valor observado de la variable comparado con el valor estimado en la ecuación (14); en otras palabras, habrá una 'ganancia' por incluir el error de predicción dentro del modelo. El filtro Kalman es el valor de la derivada parcial de la varianza del error ajustado que permite minimizar dicha varianza:

$$v_1^{aj} = a \cdot v_1^p + k_1 \cdot y_1^e, \quad (17)$$

donde  $k_1$  es la ganancia

La varianza del valor ajustado de la variable no observable es igual a:

$$VAR(v_1^{aj}) = VAR(v_1^p) \cdot (a - k_1 \cdot m_1)^2 + k_1^2 \cdot r_1 \quad (18)$$

de donde,

$$k_1 = \frac{VAR(v_1^{aj}) \cdot m}{[VAR(v_1^{aj}) \cdot m^2 + r_1]} = \frac{COV(v_1^p, y_1^p)}{VAR(y_1^p)} \quad (19)$$

Recordando la definición del coeficiente de regresión lineal simple como el cociente entre la covarianza de la variable dependiente y la independiente y la varianza de la variable independiente, la expresión (19) indica que la ganancia de Kalman equivale al coeficiente de regresión lineal entre el valor predicho de la variable no observable y la variable observada. El proceso iterativo lleva a utilizar el valor ajustado obtenido en (17) en la ecuación (14) y obtener los valores para los siguientes períodos.

Hemos aplicado el filtro de Kalman para estimar los parámetros no observados de la serie de Producto Bruto Interno por Habitante de la República Argentina entre 1950 y 2006, según un sistema similar al (11)-(13), pero donde la serie sigue un proceso ARIMA(1,1,1), según los resultados obtenidos en la sección anterior. La **Figura 11** presenta los resultados.

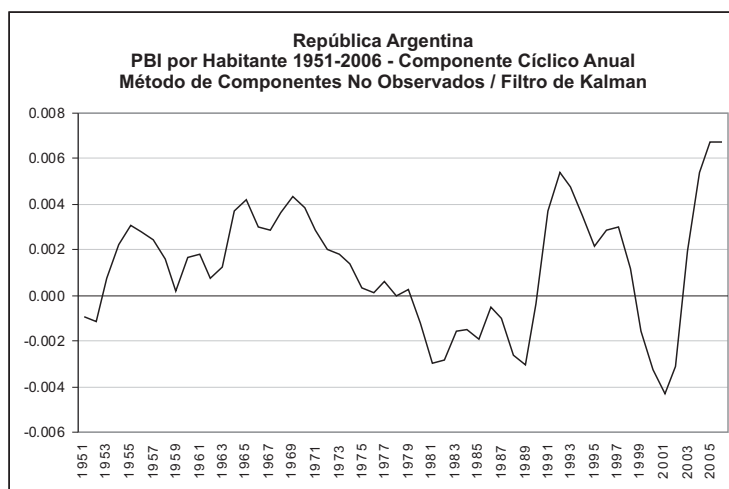


Figura 11

## 2.h. Algunos conceptos de Análisis Espectral

En econometría se acostumbra a analizar series de tiempo, es decir sucesos de una variable acaecidos a lo largo de un lapso temporal, y el estudio de los ciclos corresponde al análisis de las variaciones o fluctuaciones de los valores que las variables adoptan en el transcurso del tiempo. Sin embargo, podemos estudiar el mismo fenómeno económico desde otra perspectiva, no centrándonos en el transcurso del tiempo sino en las características de las fluctuaciones en sí: su frecuencia y amplitud, por ejemplo. Supongamos que asumimos la existencia de ciclos de Juglar de 8 años de duración en una serie de 40 años, con lo que tendremos 5 ciclos de Juglar en total. Dado este supuesto, no nos interesará saber cuánto duran, pues por la definición adoptada, duran 8 años cada uno. Pero sí nos interesará conocer su amplitud, es decir hasta cuánto ascendió la

variable durante la fase de expansión y cuánto cayó hasta que la crisis alcanzó su piso. El análisis espectral propone este cambio de enfoque en una serie: pasar del tiempo al ciclo o, en términos más técnicos, pasar del dominio temporal al dominio de frecuencias. Otro ejemplo es remover el componente estacional en una serie de tiempo: corresponde al componente de periodicidad igual a 12 meses.

Recordemos que algunos de los componentes de un ciclo son:

- Período: distancia entre dos picos o dos pisos.
- Frecuencia: inversa del período. Mide la cantidad de ciclos completos por intervalo de tiempo.
- Amplitud: distancia entre el valor máximo superior (pico) o inferior (piso) y el valor en torno al cual fluctúa el ciclo. Cuanto mayor sea la amplitud, mayor será la variabilidad de la serie; una propiedad importante es que la varianza del componente cíclico puro (es decir, despojado de elementos aleatorios o de error) es igual al cuadrado de la amplitud.

Otro aspecto relevante es el desfase: cuando se comparan dos o más ciclos, es importante estimar el desfase entre las series, es decir, es desplazamiento en el tiempo de ambas. Por ejemplo, si dos ciclos son idénticos en cuanto a los componentes antes mencionados, pero uno presenta su pico en marzo y el otro en mayo, existe un desfase de 2 meses entre ambas series.

Una serie de tiempo puede expresarse como una combinación de sinusoides (i.e. senos y cosenos) de las diversas ondas en que puede descomponerse la misma –esto se conoce como transformada de Fourier–. En general,

$$X_t = \alpha + \sum_{j=1}^{[N/2]} [A_j \cos(2\pi f_j t) + B_j \sin(2\pi f_j t)], \quad (20)$$

donde A y B son las amplitudes de las sinusoides,  $[N/2]$  indica el número entero mayor o igual a  $N/2$  con N igual al

tamaño de la muestra, y  $f_j (= j/N)$  denota la frecuencia.

Utilizando el teorema de Euler, según el cual

$$e^{int} = \cos(nt) + i \sin(nt), \quad (21)$$

donde  $i = \sqrt{-1}$ ,

podemos descomponer una serie de tiempo  $X_t$  en:

$$X_t = \sum_{t=-\infty}^{t=+\infty} c_t e^{int} \quad (22)$$

En esta representación, los coeficientes  $c_t$  se conocen como coeficientes de Fourier y surgen de resolver la siguiente expresión:

$$c_t = \frac{1}{2\pi} \int_{-\pi}^{+\pi} X_t e^{-int} dt \quad (23)$$

El espectro requiere suponer que las amplitudes A y B en (20) son variables aleatorias de distribución normal [ $N \sim (0, \sigma^2)$ ]. La media igual a cero requiere, entonces, trabajar con series normalizadas en torno a la media. Por otra parte, puede demostrarse que la varianza de la serie anterior es igual a la suma de las varianzas de los componentes sinusoidales asociados a las distintas frecuencias; en términos matemáticos:

$$\sigma_{X_t}^2 = E\{X_t - \alpha\}^2 = \sum_{j=1}^{[N/2]} (\sigma_j^2) \quad (24)$$

A partir de (20) y (21), la varianza puede expresarse en función de las frecuencias. El espectro equivale a la varianza de la serie a cada frecuencia, con lo que es inmediato calcular la contribución de cada frecuencia a la variabilidad total de la serie –lo que se conoce como el espectro de la serie a cada frecuencia–. Si el espectro de una serie alcanza su máximo en una determinada frecuencia  $\omega_0$ , entonces el componente de dicha frecuencia en la descomposición de Fourier es el que explica la mayor parte

de la varianza de la serie.

Existen diversos métodos de estimación del espectro de una serie, amén de cuestiones técnicas que exceden el objeto del presente trabajo. Hemos aplicado el método de la transformada discreta de Fourier a la serie anual de Producto Bruto Interno por Habitante de la República Argentina entre 1950 y 2006; la **Figura 12** presenta el periodograma en el espacio de frecuencias. El periodograma no corresponde al ciclo sino que es la representación gráfica de la siguiente función:

$$\sqrt{A^2 + B^2} = \frac{2\pi}{f} \quad (25)$$

Es decir, el periodograma es una representación gráfica de la densidad espectral de la serie.

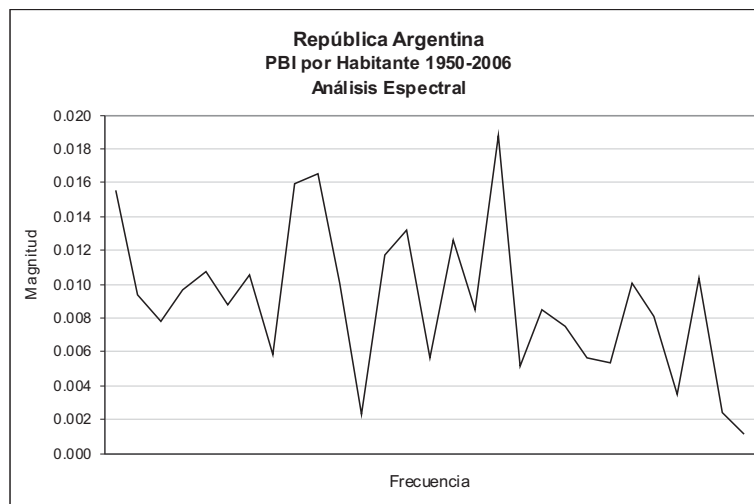


Figura 12

### 2.i. Transformada Wavelet

Supongamos que tenemos la siguiente serie de cuatro números: {14 8 5 13}. En esa serie, existen dos pares ordenados contiguos: [14 8] y [5 13]. Los promedios de dichos pares ordenados son  $(14 + 8)/2 = 11$  y  $(5 + 13)/2 = 9$ . Estos dos promedios forman otro par ordenado, cuyo



promedio a su vez es:  $(11 + 9)/2 = 10$ . Cada promedio ha generado un valor de pérdida, es decir que se ha perdido información (también conocido como coeficiente de detalle). El primer promedio, por ejemplo, tiene un valor perdido igual a 3, porque  $11 + (+3) = 14$  y  $11 - (+3) = 8$ . El segundo promedio tiene un coeficiente de detalle igual a -4 y el promedio general, una pérdida de 1. La transformada wavelet (en castellano, también conocida como de ondina u ondeleta) de la serie original es la secuencia  $\{10 \ 3 \ -4 \ 1\}$ , es decir, el promedio general de la serie de pares ordenados contiguos y los valores de pérdida en orden creciente de resolución<sup>8</sup>.

Dijimos en la sección anterior que una transformada de Fourier supone que la frecuencia a lo largo de la serie es constante. La transformada wavelet permite variar la frecuencia a lo largo de la serie, con lo que posibilita localizar el proceso bajo estudio tanto en el tiempo como según la frecuencia, es decir que hace posible distinguir fenómenos recurrentes y singularidades en el tiempo (sucesos no repetitivos). Por ejemplo, dentro de una serie como la que estamos tomando de ejemplo, del PBI por habitante de la República Argentina entre 1950 y 2006, ciertas caracterizaciones de las fluctuaciones económicas nos llevarían a distinguir ciclos de distintas duraciones (a veces conocidos como ciclos de Juglar, Kitchin, Kuznets, Kondratieff, etc.). La transformada wavelet podría aplicarse para tal caracterización. Con datos trimestrales, como las series de los componentes de la demanda y oferta global entre 1993-Q1 y 2008-Q1, podemos descomponerlas en:

$$x_t = x_t^0 + \sum_{j=0}^3 y_t^j, \quad (26)$$

donde  $x_t^0$  denota ciclos largos mayores a 32 períodos;

$y_t^0$  denota ciclos de entre 16 y 32 trimestres;

$y_t^1$  denota ciclos de entre 8 y 16 trimestres;

$y_t^2$  denota ciclos de entre 4 y 8 trimestres; y

$y_t^3$  denota ciclos de periodicidad menor a 4 trimestres.

(8) Para una introducción a aplicaciones de las wavelets en economía, ver Gencay *et al.* (2002), Ramsey (2002) y Cowley (2005).

Existen diversos métodos para estimar la transformada wavelet, e incluso las wavelets pueden ser discretas o continuas. Una de las wavelets más sencillas y utilizadas es la que se conoce como wavelet de Haar, que es una variante de una función de pulso; por ejemplo:

$$W_t^{Haar} = \begin{cases} 1 & 0 \leq t < 1 \\ -1 & -1 \leq t < 0 \\ 0 & \text{para todo otro valor de } t \end{cases} \quad (27)$$

La siguiente figura presenta la wavelet de Haar definida en (27)

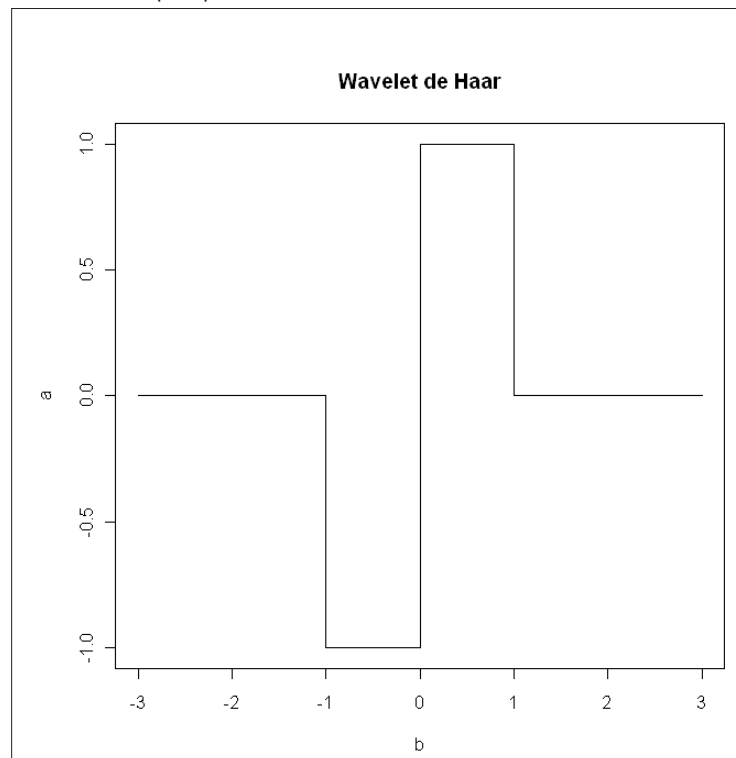


Figura 13

La escala de una wavelet es el ancho del intervalo temporal. En la wavelet (27), la escala es igual a 1 –hay dos intervalos, uno entre [-1,0] y otro entre [0,1].

Podemos calcular wavelets a distintas escalas y centradas en distintos períodos. Esto es lo que se conoce como una wavelet madre, la cual, utilizando una terminología habitual en estudios de wavelets, se dice que es trasladada en el eje temporal por un parámetro de ubicación que determina el período en el cual se centra y dilatada por un coeficiente de escala que denota el ancho del intervalo.

Otra wavelet bastante usual es la de Morlet, que corresponde a una función Gaussiana con un exponente de decaimiento exponencial:

$$W_t^{morlet} = e^{-0.5*\left(\frac{t}{a}\right)^2} e^{i2\pi ft} \quad (28)$$

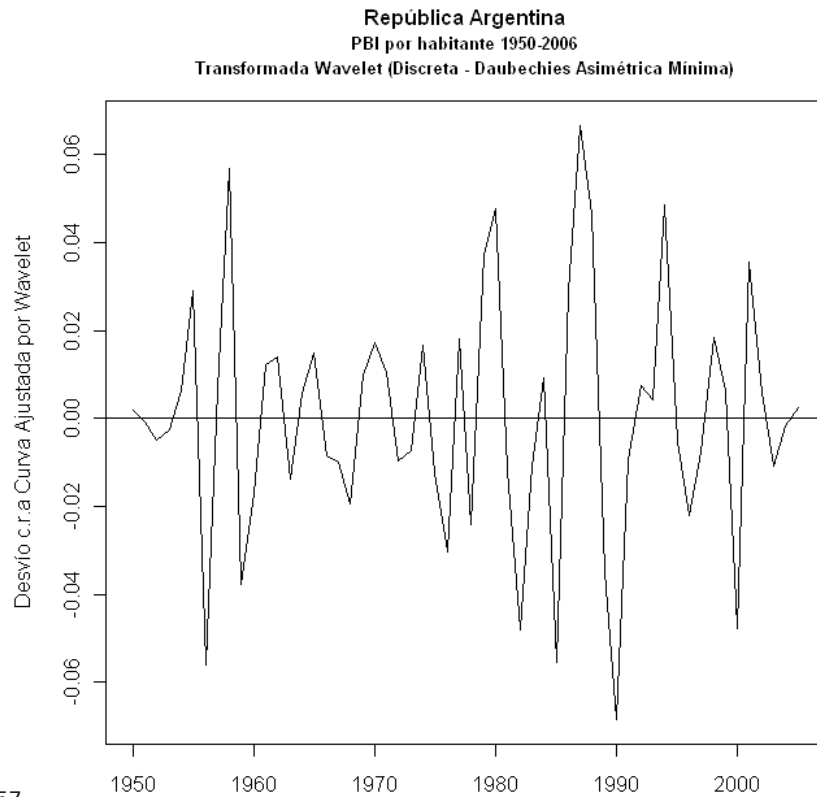
donde  $a$  es el coeficiente de escala, y  $f$  es la frecuencia. Ambos coeficientes están relacionados según  $a = f_0/f$ , donde  $f_0$  regula la forma de la wavelet. A mayores valores de  $a$  la wavelet se dilata, lo que implica menor frecuencia, y viceversa.

Finalmente, la familia de wavelet de Daubechies define una transformada discreta con momentos decrecientes, lo cual las torna muy adecuadas para series macroeconómicas, pues usualmente presentan cambios suaves, pero, por ejemplo, no demasiado adecuadas para series financieras o bursátiles, las cuales presentan cambios más bruscos.

La transformación wavelet surge de multiplicar la serie original por la función wavelet madre elegida (Haar, Morlet, Daubechies, etc.) para diferentes valores paramétricos de escala y a diferentes valores de centrado. Las wavelets resultantes son, entonces, versiones

trasladadas temporalmente y comprimidas o expandidas de la wavelet madre. El cómputo conlleva calcular todas las wavelets centradas en cada uno de los períodos de tiempo a lo largo de los cuales se extiende la serie y para cada uno de los valores deseados de escala. El camino inverso permite descomponer una serie de tiempo en wavelets de diferentes escalas y centradas en diferentes períodos.

La **figura 14** presenta el ciclo resultante de aplicar el método de transformada wavelet de Daubechies asimétrica mínima con definición de umbral (*thresholding*) universal (Nason, 2008) al logaritmo de la serie anual de PBI por habitante entre 1950 y 2006<sup>9</sup>



(9) La serie original cuenta con 57 observaciones. Fue extendida con 7 ceros para llegar a 64 observaciones, la potencia de 2 más cercana, utilizamos el paquete WaveThresh del programa R. (Nason, 2009).

**Figura 14**

### 3. Comparación entre las distintas estimaciones

En este trabajo hemos presentado algunos de los más habituales métodos para estimar componentes cíclicos en series temporales. Esta sección, en aras de completar la presentación de los métodos, compara los resultados obtenidos sobre la serie de PBI por habitante de la República Argentina entre 1950 y 2006.

Suelen cotejarse los desvíos estándar de las series cíclicas o los coeficientes de correlación entre las mismas para comparar las estimaciones; sin embargo, entendemos que un procedimiento más riguroso requiere el uso de técnicas estadísticas no paramétricas de comparación entre series, como el estadístico de Kolmogorov-Smirnov o el de Lilliefors. Presentamos en la tabla 1 los desvíos estándar y los estadísticos de asimetría y kurtosis de cada serie (con tests de significatividad respectivos) y en la tabla 2 los coeficientes de correlación pareados, para coincidir en la presentación con otros autores, pero estamos más interesados en los resultados incluidos en la tabla 3, que surgen de aplicar el estadístico de Kolmogorov-Smirnov a cada par de series de ciclos obtenidas.

La **tabla 1** muestra que el ciclo resultante de aplicar el método de Tendencia Lineal es el de mayor variabilidad y que el obtenido por medio del filtro Kalman es el que menos desvío presenta en torno al valor medio. En igual sentido, todos los ciclos resultan simétricos al 5 por ciento de significatividad, pero puede rechazarse la hipótesis nula de que el ciclo resultante del filtro Baxter-King sea simétrico al 6 por ciento de significatividad y el resultante del filtro Hodrick-Prescott, al 7.5 por ciento. En cuanto a la kurtosis, solamente los ciclos obtenidos por medio del filtro Hodrick-Prescott y la descomposición de Beveridge-Nelson presentan fakes significativamente distintos al de una distribución normal (al 5 por ciento de significatividad). El primero, con una distribución leptocúrtica, mientras que el segundo, platocúrtica.

| Método                                         | Desvío estándar | Asimetría |                    | Kurtosis |                        |
|------------------------------------------------|-----------------|-----------|--------------------|----------|------------------------|
|                                                |                 | Coef.     | Test de D'Agostino | Coef.    | Test de Anscombe-Glynn |
| Tendencia lineal                               | 0.0985          | 0.0893    | 0.8416             | 2.2930   | 0.1990                 |
| Tendencia cúbica                               | 0.0760          | -0.3816   | 0.4038             | 2.5961   | 0.6662                 |
| Tendencia lineal con quiebres                  | 0.0515          | -0.7545   | 0.1197             | 4.0524   | 0.0825                 |
| Filtro Hodrick-Prescott                        | 0.0494          | -0.8837   | 0.0755             | 4.4733   | 0.0355                 |
| Filtro Hodrick-Prescott de doble paso de banda | 0.0124          | -0.5984   | 0.2045             | 3.4684   | 0.2750                 |
| Filtro Baxter-King                             | 0.0321          | -0.9582   | 0.0576             | 4.0804   | 0.0779                 |
| Filtro Christiano-Fitzgerald                   | 0.0343          | -0.5977   | 0.2051             | 2.6060   | 0.6836                 |
| Descomposición de Beveridge-Nelson             | 0.0570          | -0.3205   | 0.4802             | 2.0703   | 0.0330                 |
| Componentes no observados / Filtro Kalman      | 0.0026          | -0.0186   | 0.9667             | 2.3801   | 0.3121                 |
| Wavelet                                        | 0.0278          | -0.0595   | 0.8939             | 3.3035   | 0.3848                 |

La **tabla 2** indica que hay ciclos que están más correlacionados entre sí que con respecto a aquellos obtenidos por otros métodos. Esto no debe llamar la atención por cuanto algunos de estos métodos, como vimos, intentan detectar o filtrar fluctuaciones a distintas frecuencias. La mayor correlación se obtuvo entre el ciclo obtenido por medio del filtro de Hodrick-Prescott y por el método de tendencia lineal con quiebres (puede entenderse, en términos muy generales, que el de HP es una generalización no lineal de una tendencia lineal con n quiebres).

| Método                               | Tendencia Lineal   | Tendencia Cúbica             | Tendencia Lineal con Quiebres | Hodrick Prescott | HP con paso de banda |
|--------------------------------------|--------------------|------------------------------|-------------------------------|------------------|----------------------|
| <b>Tendencia Lineal</b>              | 1.000              |                              |                               |                  |                      |
| <b>Tendencia Cúbica</b>              | 0.772              | 1.000                        |                               |                  |                      |
| <b>Tendencia Lineal con Quiebres</b> | 0.616              | 0.755                        | 1.000                         |                  |                      |
| <b>Hodrick-Prescott</b>              | 0.633              | 0.817                        | 0.923                         | 1.000            |                      |
| <b>HP con paso de banda</b>          | 0.482              | 0.623                        | 0.860                         | 0.890            | 1.000                |
| <b>Baxter-King</b>                   | 0.519              | 0.622                        | 0.819                         | 0.855            | 0.862                |
| <b>Christiano-Fitzgerald</b>         | 0.364              | 0.498                        | 0.746                         | 0.739            | 0.859                |
| <b>Beveridge-Nelson</b>              | 0.195              | 0.225                        | 0.398                         | 0.374            | 0.306                |
| <b>Kalman</b>                        | 0.181              | 0.170                        | 0.373                         | 0.320            | 0.256                |
| <b>Wavelet</b>                       | -0.038             | -0.049                       | -0.117                        | -0.077           | 0.006                |
|                                      | <b>Baxter King</b> | <b>Christiano Fitzgerald</b> | <b>Beveridge Nelson</b>       | <b>Kalman</b>    | <b>Wavelet</b>       |
| <b>Tendencia Lineal</b>              |                    |                              |                               |                  |                      |
| <b>Tendencia Cúbica</b>              |                    |                              |                               |                  |                      |
| <b>Tendencia Lineal con Quiebres</b> |                    |                              |                               |                  |                      |
| <b>Hodrick-Prescott</b>              |                    |                              |                               |                  |                      |
| <b>HP con paso de banda</b>          |                    |                              |                               |                  |                      |
| <b>Baxter-King</b>                   | 1.000              |                              |                               |                  |                      |
| <b>Christiano-Fitzgerald</b>         | 0.892              | 1.000                        |                               |                  |                      |
| <b>Beveridge-Nelson</b>              | 0.400              | 0.420                        | 1.000                         |                  |                      |
| <b>Kalman</b>                        | 0.213              | 0.238                        | 0.679                         | 1.000            |                      |
| <b>Wavelet</b>                       | -0.161             | -0.165                       | -0.540                        | -0.191           | 1.000                |

El estadístico de Kolmogorov-Smirnov, cuyos resultados se presentan en la **tabla 3**, permite rechazar o no la hipótesis de que ambas series pareadas son estadísticamente equivalentes. La tercera columna de la tabla (columna  $p$ ) contiene los valores de significatividad del estadístico de Kolmogorov-Smirnov incluido en la columna  $ks$ . Valores de  $p$  mayores a 0.05 no permiten rechazar la hipótesis nula de que ambas series de ciclos sean iguales. De las 45 comparaciones, en 26 se rechaza dicha hipótesis; es decir que para las comparaciones restantes, no podemos rechazar la hipótesis de que los ciclos obtenidos se traten de series temporales estadísticamente equivalentes.

**Tabla 3**  
Comparación entre Ciclos  
Estadístico de Kolmogorov-Smirnov

| Ciclos                                                                        | ks     | $p$    |
|-------------------------------------------------------------------------------|--------|--------|
| Tendencia Lineal - Tendencia Cúbica                                           | 0.1754 | 0.3442 |
| Tendencia Lineal - Tendencia Lineal con Quiebres                              | 0.2105 | 0.1598 |
| Tendencia Lineal - Filtro de Hodrick-Prescott                                 | 0.2281 | 0.1031 |
| Tendencia Lineal - Filtro de Hodrick-Prescottband                             | 0.4211 | 0.0001 |
| Tendencia Lineal - Filtro de Baxter-King                                      | 0.2982 | 0.0126 |
| Tendencia Lineal - Filtro de Christiano-Fitzgerald                            | 0.2807 | 0.0224 |
| Tendencia Lineal - Descomposición de Beveridge-Nelson                         | 0.1930 | 0.2390 |
| Tendencia Lineal - Filtro de Kalman / Modelo de Espacio-Estado                | 0.5263 | 0.0000 |
| Tendencia Lineal - Método de Wavelets                                         | 0.2982 | 0.0126 |
| Tendencia Cúbica - Tendencia Lineal con Quiebres                              | 0.2281 | 0.1031 |
| Tendencia Cúbica - Filtro de Hodrick-Prescott                                 | 0.2281 | 0.1031 |
| Tendencia Cúbica - Filtro de Hodrick-Prescottband                             | 0.4035 | 0.0002 |
| Tendencia Cúbica - Filtro de Baxter-King                                      | 0.3158 | 0.0068 |
| Tendencia Cúbica - Filtro de Christiano-Fitzgerald                            | 0.2982 | 0.0126 |
| Tendencia Cúbica - Descomposición de Beveridge-Nelson                         | 0.1754 | 0.3442 |
| Tendencia Cúbica - Filtro de Kalman / Modelo de Espacio-Estado                | 0.5088 | 0.0000 |
| Tendencia Cúbica - Método de Wavelets                                         | 0.2807 | 0.0224 |
| Tendencia Lineal con Quiebres - Filtro de Hodrick-Prescott                    | 0.1053 | 0.9103 |
| Tendencia Lineal con Quiebres - Filtro de Hodrick-Prescottband                | 0.3684 | 0.0009 |
| Tendencia Lineal con Quiebres - Filtro de Baxter-King                         | 0.1930 | 0.2390 |
| Tendencia Lineal con Quiebres - Filtro de Christiano-Fitzgerald               | 0.1579 | 0.4761 |
| Tendencia Lineal con Quiebres - Descomposición de Beveridge-Nelson            | 0.1930 | 0.2390 |
| Tendencia Lineal con Quiebres - Filtro de Kalman / Modelo de Espacio-Estado   | 0.5614 | 0.0000 |
| Tendencia Lineal con Quiebres - Método de Wavelets                            | 0.2281 | 0.1031 |
| Filtro de Hodrick-Prescott - Filtro de Hodrick-Prescottband                   | 0.3860 | 0.0004 |
| Filtro de Hodrick-Prescott - Filtro de Baxter-King                            | 0.1579 | 0.4761 |
| Filtro de Hodrick-Prescott - Filtro de Christiano-Fitzgerald                  | 0.1053 | 0.9103 |
| Filtro de Hodrick-Prescott - Descomposición de Beveridge-Nelson               | 0.2807 | 0.0224 |
| Filtro de Hodrick-Prescott - Filtro de Kalman / Modelo de Espacio-Estado      | 0.5263 | 0.0000 |
| Filtro de Hodrick-Prescott - Método de Wavelets                               | 0.2456 | 0.0642 |
| Filtro de Hodrick-Prescottband - Filtro de Baxter-King                        | 0.2807 | 0.0224 |
| Filtro de Hodrick-Prescottband - Filtro de Christiano-Fitzgerald              | 0.3333 | 0.0036 |
| Filtro de Hodrick-Prescottband - Descomposición de Beveridge-Nelson           | 0.4737 | 0.0000 |
| Filtro de Hodrick-Prescottband - Filtro de Kalman / Modelo de Espacio-Estado  | 0.3509 | 0.0018 |
| Filtro de Hodrick-Prescottband - Método de Wavelets                           | 0.1754 | 0.3442 |
| Filtro de Baxter-King - Filtro de Christiano-Fitzgerald                       | 0.1053 | 0.9103 |
| Filtro de Baxter-King - Descomposición de Beveridge-Nelson                    | 0.3333 | 0.0036 |
| Filtro de Baxter-King - Filtro de Kalman / Modelo de Espacio-Estado           | 0.4211 | 0.0001 |
| Filtro de Baxter-King - Método de Wavelets                                    | 0.1404 | 0.6284 |
| Filtro de Christiano-Fitzgerald - Descomposición de Beveridge-Nelson          | 0.2807 | 0.0224 |
| Filtro de Christiano-Fitzgerald - Filtro de Kalman / Modelo de Espacio-Estado | 0.4737 | 0.0000 |
| Filtro de Christiano-Fitzgerald - Método de Wavelets                          | 0.1930 | 0.2390 |
| Descomposición de Beveridge-Nelson - Filtro de Kalman                         | 0.5439 | 0.0000 |
| Descomposición de Beveridge-Nelson - Método de Wavelets                       | 0.3509 | 0.0018 |
| Filtro de Kalman / Modelo de Espacio-Estado - Método de Wavelets              | 0.4386 | 0.0000 |

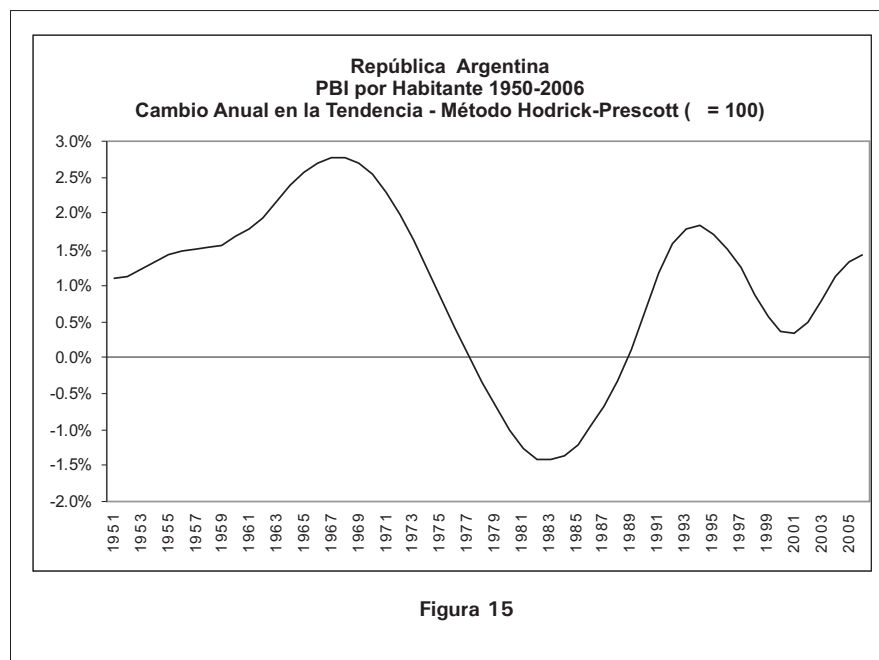
**Nota:** Los valores de  $p$  marcados con *bastardilla* denotan significatividad estadística menor al 5 por ciento. Es decir que se rechaza la hipótesis nula de igualdad entre ambas series temporales de ciclos respectivos.

El presente trabajo tuvo como principal objetivo el describir brevemente las principales técnicas univariadas de extracción del componente cíclico en una serie de tiempo. Los resultados de los tres cuadros anteriores plantean dudas acerca de la bondad de ajuste de muchos de los métodos utilizados, pero excedería el alcance del artículo el comentar acerca de esto. En la literatura especializada, algunos autores sostienen que “diferentes métodos de extracción de la tendencia son ventanas alternativas que miran a las variables macroeconómicas desde perspectivas diferentes” (Canova, 1993, p.3), con lo que proponen cierto agnosticismo en cuanto a la técnica a utilizar, mientras que otros autores prefieren un método en particular por sobre el resto.

La respuesta final acerca de la bondad relativa de cada método no descansa en técnicas estadísticas, sino en un conocimiento de la realidad histórica y en cuestiones de teoría económica –caso contrario se caería en los errores de “medición sin analizar los datos originales” y de “medición sin teoría”. Una forma de considerar ambos aspectos es teniendo en cuenta las series de tendencia de largo plazo resultantes de la extracción del componente cíclico, y en particular, las tasas de crecimiento anuales de la tendencia. Para una discusión en mayor profundidad acerca de la importancia de combinar el juicio histórico con las técnicas estadísticas de filtrado de las series económicas, ver Ball y Word (1996).

En la **figura 15** presentamos la variación anual de las tendencias de largo plazo del PBI por habitante de la República Argentina entre 1950 y 2006 según el método de Hodrick-Prescott. Este filtro sugiere que habría habido un pico en el crecimiento de la tendencia de largo plazo en 1968 equivalente al 2.78 por ciento anual y un piso en 1983, cuando la tendencia anual habría caído un 1.4 por ciento. Podemos asociar el primer año con la culminación del proceso de sustitución de importaciones y el segundo con la liberación financiera y apertura importadora iniciado en la segunda mitad de la década de 1970. El filtro de Hodrick-Prescott parece recoger, asimismo, un crecimiento en la tendencia a comienzos de la década de 1990 y un decaimiento en la misma tras la crisis de 2001, pero de menor impacto en la tendencia de largo plazo.





#### 4. Comentarios Finales

La importancia de detectar fluctuaciones a lo largo del tiempo en las variables económicas ha llevado a la profusión de técnicas estadísticas de extracción del componente cíclico de las series. Este trabajo ha presentado las técnicas univariadas más habituales, las cuales son, a su vez, de aplicación más frecuente que los enfoques multivariados.

A los efectos de guiar la presentación, hemos utilizado la serie de Producto Bruto Interno por habitante de la República Argentina entre 1950 y 2006, y otras series auxiliares cuando la exposición, por motivos didácticos, lo sugería. Confiamos en que ambas partes de este artículo lleven a los lectores a profundizar en el tema y a aplicar estas y otras técnicas a otras series económicas, porque, como afirmó Schumpeter: "Analizar ... los ciclos significa, nada más ni nada menos, que analizar el proceso económico de la era capitalista. Los ciclos no son, como las amígdalas, cosas separables que puedan tratarse por sí mismas, sino que son, como el latido de un corazón, parte de la esencia del organismo que los presenta".

## 5. Referencias

AHUMADA, H. y GAREGNANI, M. (1999). "*Hodrick-Prescott Filter in Practice*", en: *Económica*, Universidad Nacional de La Plata, Vol. XLV, N° 4. La Plata, Argentina.

ARNOLD, T., BERTUS, M. y GODBEY, J. (2008). "*A Simplified Approach to Understanding the Kalman Filter Technique*", en: *The Engineering Economist*, Vol. 53, pp. 140-155.

BACCINO, O. (2006). "*Comportamientos Cíclicos en la Actividad Económica de la Argentina 1993-2006*". Trabajo presentado en la XLI Reunión Anual de la Asociación Argentina de Economía Política, Salta. Asociación Argentina de Economía Política. Buenos Aires.

BACKUS, D., KEHOE, P. y KYDLAND, E. (1992). "*International real business cycle*", en: *Journal of Political Economy*, Vol. 100, N° 4, pp. 745- 775.

BALL, M. y WOOD, A. (1996). "*Trend growth in post-1850 British economic history: the Kalman filter and historical judgment*", en: *The Statistician*, Vol. 45, N° 2, pp. 143-152.

BARRIOS, S., BRÜLHART, M., ELLIOT, R. y SENSIER, M. (2002). "*A Tale of Two Cycles: Co-fluctuations between UK Regions and the Euro Zone*", Discussion Paper 003, September 2002, Centre for Growth & Business Cycle Research, Universidad de Manchester.

BEVERIDGE, S. y NELSON, C. (1981). "*A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the business cycle*", en: *Journal of Monetary Economics*, Vol. 7, pp. 151-174.

BUCH, C. (2002). "*Business Cycle Volatility and Globalization: A Survey*". Kiel Working Paper N° 1107. Kiel Institute of World Economics: Kiel, Alemania.

BURNS, A. y MITCHELL, W. (1946). *Measuring Business Cycles*. National Bureau of Economic Research. Nueva York.

CANOVA, F. (1993). "*Detrending and business cycle facts*". Discussion Paper 782. Centre for Economic Policy Research. Londres.

CHRISTIANO, L. y FITZGERALD, T. (2003). "*The bandpass filter*", en: *International Economic Review*, Vol. 44, N° 2, pp. 435-465.

CLARK, P. (1987). "The Cyclical Component of U.S. Economic Activity", en: *The Quarterly Journal of Economics*, Vol. 102, N° 4, pp. 797-814.

COCHRANE, J. (1988). "How big is the random walk in GNP?", en: *Journal of Political Economy*, Vol. 96, N° 5, pp. 893-920.

COWLEY, P. (2005). "An intuitive guide to wavelets for economists". Bank of Finland Research Discussion Papers N° 1. Bank of Finland.

FRISCH, R. (1933). "Propagation Problems and Impulse Problems in Dynamic Economics", en: *Economic Essays in Honour of Gustav Cassel*. Allen & Unwin. Londres.

GENCAY, R.; SELCUK, F. y WHITCHER, B. (2002). *An Introduction to Wavelets and Other Filtering Methods in Finance and Economics*. Academic Press, Nueva York.

GÓMEZ, V. y BENGOCHEA, P. (1998). "El Filtro de Ciclo-Tendencia en la Contabilidad Nacional Trimestral frente a los Filtros Basados en Modelos". Documento de Trabajo D-98004. Dirección General de Análisis y Programación Presupuestaria. Ministerio de Economía y Hacienda. España.

HAMILTON, J. (1994). *Time Series Analysis*. Editorial Princeton University Press. Princeton. Nueva Jersey.

HARDING, D. y PAGAN, A. (2005). "A Suggested Framework for Classifying the Modes of Cycle Research", en: *Journal of Applied Econometrics*, Vol. 20, N° 2, pp. 151-160.

HARVEY, A. (1985). "[Trends and Cycles in Macroeconomic Time Series](#)", en: *Journal of Business & Economic Statistics*, Vol. 3, N° 3, pp. 216-27.

HARVEY, A. y TRIMBUR, T. (2003). "General model-based filters for extracting trends and cycles in economic time series", en: *Review of Economics and Statistics*, Vol. 85, pp. 244-55.

HESS, G. y SHIN, K. (1997). "International and Intranational Business Cycles", en: *Oxford Review of Economic Policy*, Vol. 13, N° 13, pp. 93-109.

HM TREASURY (2005). *Evidence on the UK Economic Cycle*. HM Treasury: Londres.

HODRICK, R. y PRESCOTT, E. (1980). "Post-war U.S. business cycles: An empirical investigation." Discussion paper N° 451, Carnegie-Mellon University. (Published in the Journal of Money, Credit and Banking, Vol. 29, N° 1, Feb. 1997, pp. 1-16).

JORRAT, J. (2004). Comentario al Trabajo del Lic. Luis A. Trajtenberg: "Modelización de Ciclos y Tendencias en Series de Tiempo Macroeconómicas Argentina: 1980-2000". Presentado en la XXXIX Reunión Anual de la Asociación Argentina de Economía Política, Buenos Aires.

LIBANIO, G. (2005). "Unit roots in macroeconomic time series: theory, implications, and evidence", en: *Nova Economia Belo Horizonte*, Vol. 15, N° 3, pp. 145-176.

MORLEY, J. (2002). "A state-space approach to calculating the Beveridge-Nelson decomposition", en: *Economic Letters*, Vol. 75, pp. 123-127.

NASON, G.P. (1998) *WaveThresh3 Software*. Department of Mathematics, Universidad de Bristol, Bristol, Reino Unido.

NASON, G. (2008). *Wavelet Methods in Statistics with R*. Use R! Series. Springer. Berlin.

NELSON, C. y PLOSSER, C. (1982). "Trends and random walks in macroeconomic time series: some evidence and implications", en: *Journal of Monetary Economics*, Vol. 10, pp. 139-169.

OH, K., ZIVOT, E. y CREAL, D. (2008). "The relationship between the Beveridge-Nelson decomposition and other permanent-transitory decompositions that are popular in economics", en: *Journal of Econometrics*, Vol. 146, pp. 207-219.

PEDERSEN, T. (2003). *Alternative Linear and Non-Linear Detrending Techniques: A Comparative Analysis based on Euro-Zone Data*. European Communities: Luxemburgo.

RAMSEY, J. (2002). "Wavelets in Economics and Finance: Past and Future". Economic Research Reports 2002-02. C.V. Starr Center for Applied Economics. Universidad de Nueva York.

RAVN, M. y UHULIG, H. (1997). "On Adjusting the HP-Filter for the frequency of Observations", en: Tilburg University Working Paper, Tilburg, Alemania.

RIBEIRO, M. (2000). *"Introduction to Kalman Filtering. A set of two lectures"*. Instituto Superior Técnico / Instituto de Sistemas e Robótica. Lisboa, Portugal.

SCHUMPETER, J. (1939). *"Business Cycles. A Theoretical, Historical, and Statistical Analysis of the Capitalist Process"*, McGraw-Hill, Nueva York.

SICHEL, D. E. (1993). *"Business cycle asymmetry: A deeper look"*, en: *Economic Inquiry*, Vol. XXXI, pp. 224–236.

SOSA ESCUDERO, W. (1997). *"Testing for unit-roots and trend-breaks in Argentine real GDP"*, en: *Económica*, Vol. 43, pp. 123-142.

STOCK, J. y WATSON, M. (2007). *Introduction to Econometrics*. Ed. Addison-Wesley. Nueva York.

STURZENEGGER, A. y MOYA, R. (2003). *"Economic Cycles"*, en: Della Paolera, G. y Taylor, A. (eds.): *A New Economic History of Argentina*, Cambridge University Press. Cambridge.

TERCEIRO LOMBA, J. (1978). *"Aplicaciones de filtro de Kalman en economía"*, en: *Estadística Española*, Vol. 80, pp. 25-38.

TINBERGEN, J. (1939). *Statistical Testing of Business-Cycle Theories: A Method and Its Application to Investment Activity*. League of Nations. Ginebra.

TRAJTENBERG, L. (2004). *"Modelización de Ciclos y Tendencias en Series de tiempo Macroeconómicas. Argentina: 1980-2000"*. Trabajo presentado en la XXXIX Reunión Anual de la Asociación Argentina de Economía Política, Buenos Aires. Asociación Argentina de Economía Política. Buenos Aires.

## Apéndice

### Resultados de la Estimación de Ratios de Varianza del PBI por Habitante

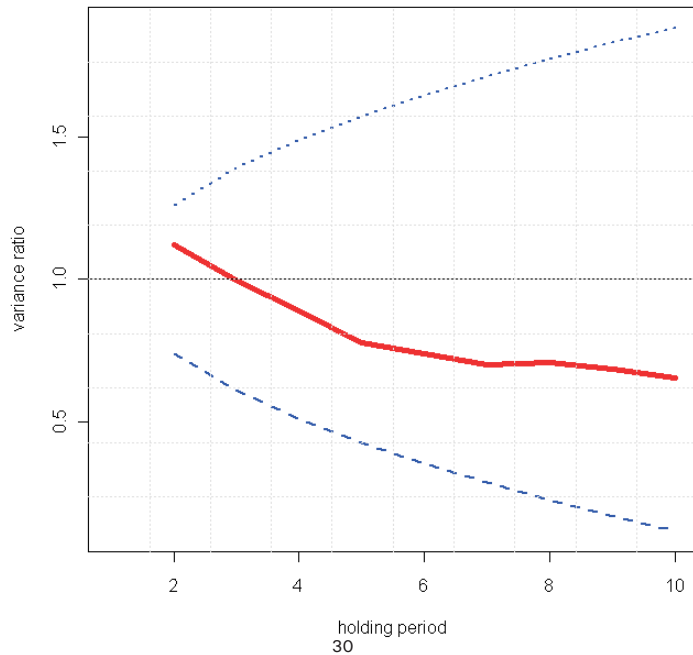
Cochrane (1988) propone estimar las k-diferencias de la serie de PBI por habitante. La primera diferencia de una serie de tiempo  $y_t$  es  $y_t - y_{t-1}$ ; la segunda diferencia es  $y_t - y_{t-2}$  y la k-diferencia es  $y_t - y_{t-k}$ . En nuestro caso, la serie se extiende entre 1950 y 2006. Hemos calculado las 10 primeras diferencias.

Luego se calcula la varianza de cada serie y se divide la misma por la varianza de la serie de primeras diferencias multiplicada por 1/k. En otros términos,

$$V_k = \frac{1}{k} \frac{\text{var}(y_t - y_{t-k})}{\text{var}(y_t - y_{t-1})}$$

Si  $V_k$  tiende a un valor constante a medida que k aumenta, la serie es estacionaria; si, en cambio, crece en función de k, la serie es un camino aleatorio. La figura a continuación muestra los resultados.  $V_k$  tiende a un valor constante a medida que k aumenta, con lo que la serie es estacionaria.

Variance Ratios and 95% confidence band



# CONVERGENCIA DEL PERCEPTRON

(\*) Del Departamento de Matemática de la Facultad de Economía y Administración de la Universidad Nacional del Comahue.

e-mail: [lcuria@uncoma.edu.ar](mailto:lcuria@uncoma.edu.ar)

(\*\*) Del Departamento de Matemática de la Facultad de Economía y Administración de la Universidad Nacional del Comahue y del Departamento de Electrotecnia de la Facultad de Ingeniería de la Universidad Nacional del Comahue.

e-mail: [pancho.pazbustillo@gmail.com](mailto:pancho.pazbustillo@gmail.com)

Por los Profesores Lisandro Curia (\*) y Francisco Paz (\*\*)

## Resumen

Las redes neuronales artificiales (RNA) vienen desarrollándose desde hace aproximadamente sesenta años. A medida que la ciencia fue avanzando, uno de los desafíos más relevantes a los que se enfrentó el ser humano es el de desarrollar sistemas inteligentes capaces de realizar tareas específicas emulando las habilidades del cerebro humano. Una de las características más significativas de las redes neuronales es su capacidad para aprender a partir de la información relevante presente en su entorno. El propósito de este trabajo es presentar una red elemental formada por una única neurona: el Perceptron simple; resaltando sus características principales y haciendo hincapié en su estructura y principio de funcionamiento cuando se quiere representar las funciones lógicas NOT, AND, OR.

Dado que comúnmente se tiende a considerar a las redes neuronales como una "caja negra", en esta publicación se desarrollará el teorema de convergencia del Perceptron, mostrando así su capacidad para resolver problemas de clasificación de patrones linealmente separables. Finalmente, se pondrán de manifiesto las limitaciones que presenta el Perceptron cuando se trata de resolver la función lógica XOR.

## Introducción

Los sistemas de cómputo, que desarrollan su tarea formando secuencias de operaciones, muestran ser eficientes cuando se trata de resolver determinados problemas matemáticos, en la creación de grandes bases de datos o incluso en el control de las funciones de

instrumentos de la industria y electrodomésticos. Sin embargo, este tipo de sistemas presenta una gran desventaja cuando se requiere que interpreten situaciones que evolucionan de manera estocástica con el transcurso del tiempo. Estas situaciones se presentan, por ejemplo, cuando no se sabe qué va a ocurrir con los valores futuros de una serie de tiempo, o bien cuando se conoce la expresión analítica que nos permite hallar la respuesta en función del tiempo, pero que por algún motivo, puede estar contaminada por algún ruido. En este caso, se espera que el sistema pueda copiar la componente efectiva de ruido que se mezcla con la señal y finalmente obtener el valor limpio o restituido de la señal original; este es el caso de los comúnmente llamados filtros adaptivos.

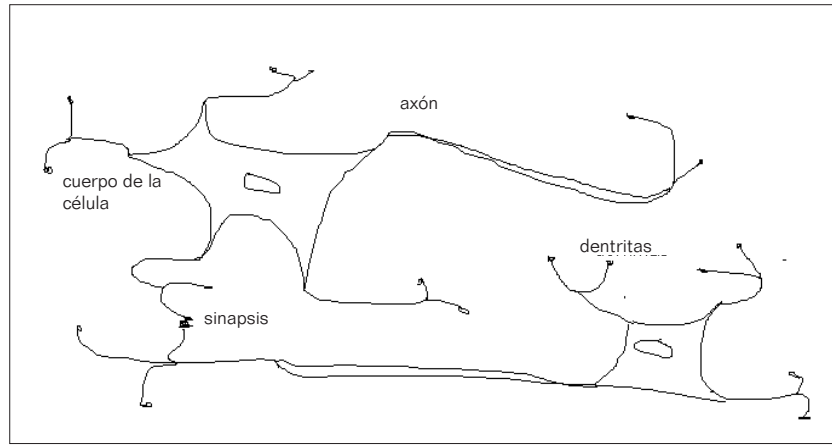
En gran medida la tarea con redes neuronales artificiales se basa en el desarrollo de programas de cómputo que logren aprender de la experiencia. Estos sistemas deben ser capaces de aprender y adaptarse a características de funcionamiento de forma similar a la manera de operar que tiene el ser humano. Con la finalidad de evolucionar y ajustarse a los cambios, los sistemas implementados sobre la base de redes neuronales están desarrollados de manera que puedan aprender de ejemplos que se les presentan y obtener características generales del entorno que permitan al usuario disponer de información confiable de modo rápido y efectivo.

### **Estructura de una neurona artificial**

Una neurona biológica es una célula que presenta las siguientes características: un cuerpo aproximadamente esférico de 5 a 10 micras de diámetro, el axón y las dendritas. La Figura 1 ilustra dos neuronas biológicas, donde se destacan las siguientes particularidades:

- Las dendritas, que constituyen el camino de ingreso de las señales que son procesadas en el cuerpo de la neurona.
- El axón, por donde sale la señal que se generó en la neurona.
- La sinapsis que constituye la zona de contacto entre las dos neuronas.





**Figura 1.** Conexión entre neuronas

Por otro lado, tratando de emular el funcionamiento de una neurona biológica, en la **Figura 2** se representa un modelo de una neurona artificial: los valores  $x_i$  representan las componentes de un vector de entradas de  $r$  componentes, los pesos o intensidades de las conexiones están dados por  $w_{ij}$ , mientras que la salida de la neurona se representa con la letra  $a_j$ .

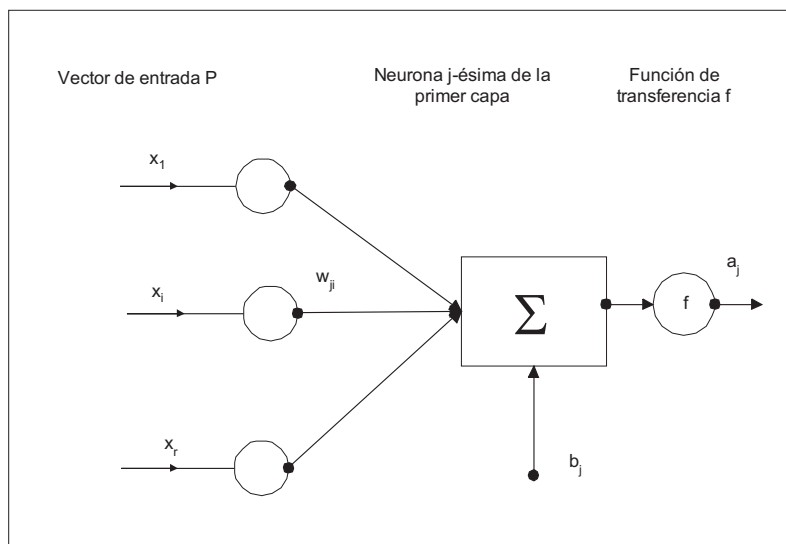
Una neurona artificial puede definirse como un procesador elemental o dispositivo de cálculo que a partir de una entrada, produce una única salida. Una red neuronal artificial es un modelo artificial que se basa en la conexión de varios procesadores elementales que tienen por finalidad realizar una función común.

Las características que permiten a su vez identificar una red neuronal son las siguientes:

1. El conocimiento de la red es adquirido a través de un procedimiento denominado aprendizaje.
2. Este conocimiento radica en la modificación de los pesos sinápticos provenientes de la regla de aprendizaje específica de la red.

Según este criterio, la red neuronal es un modelo computacional capaz de adaptarse y aprender a organizar y generalizar la información basándose en el procesamiento de las señales que ingresan en la red.

En la **Figura 2** también puede verse cómo el sumador  $\Sigma$  se encarga de efectuar las sumas ponderadas de todas las entradas en la red. La función de transferencia  $f$  recibe el valor ponderado y produce la salida de la neurona. Normalmente se incluye un umbral o ganancia  $b$ , que permite controlar el estado a partir del cual la neurona se activará.



**Figura 2.** Red neuronal simple

En el año 1943, Warren McCulloch y Walter Pitts desarrollaron el primer modelo de red neuronal. El funcionamiento de esta red, consistía en sumar las entradas ponderadas y comparar este valor con un umbral  $b$ . En el caso que las sumas multiplicadas por los pesos adoptados supere el valor  $b$ , la neurona se activará produciendo una salida  $a = 1$ ; en el caso contrario la salida será 0 y la neurona no se activará. Inicialmente se esperaba que un conjunto de neuronas de McCulloch-Pitts pudiera servir para modelar cualquier función lógica. Veremos qué ocurre al modelar con una neurona las funciones lógicas NOT, AND y OR. De acuerdo con lo mencionado, la neurona con estas características responde a un modelo tal que la salida  $a$  puede expresarse matemáticamente mediante la ecuación (1).

$$a_j = \begin{cases} 1 & \sum w_{ji}x_i > b \\ 0 & \text{en los casos restantes} \end{cases} \quad (1)$$

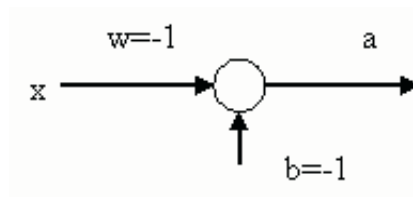
### La representación de funciones lógicas

Las neuronas de McCulloch-Pitts pueden considerarse como las primeras estructuras de redes neuronales artificiales (Isasi, 2004 y Veelentuf, 1995). Con su representación puede modelarse el funcionamiento simplificado de una neurona biológica, considerando que una neurona artificial admite solo dos estados posibles: encendido (1) o apagado (0).

Como paso siguiente, se representarán las funciones lógicas NOT, OR y AND mediante neuronas de McCulloch-Pitts.

#### La función NOT

La función lógica NOT representada en la **Figura 3**, puede modelarse con una neurona, con una sola entrada, peso  $W=-1$  y umbral o ganancia  $b=-1$ .



**Figura 3.** Representación de la función NOT

De acuerdo con la representación anterior, la salida de la neurona estará dada por la ecuación (2).

$$a = \begin{cases} 1 & \text{si } x = 0, \text{ resulta } wx > -1 \\ 0 & \text{si } x = 1, \text{ resulta } wx \not> -1 \end{cases} \quad (2)$$

### La función OR

En este caso la neurona tiene dos entradas y el umbral toma valor 0. La **Figura 4** muestra esta situación. El **Cuadro 1** representa las salidas y las sumas ponderadas; también puede apreciarse el significado del umbral  $b_1$ .

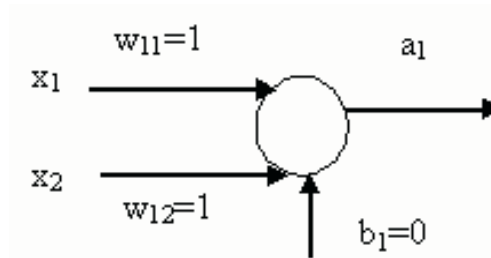


Figura 4. Representación de la función OR

| $x_1$ | $x_2$ | $\sum w_i x_i$ | $a_1$ |
|-------|-------|----------------|-------|
| 1     | 1     | 2              | 1     |
| 1     | 0     | 1              | 1     |
| 0     | 1     | 1              | 1     |
| 0     | 0     | 0              | 0     |

Cuadro 1. Salidas posibles de la red correspondientes a la función OR

### La función AND

En este caso la neurona tiene dos entradas y el umbral toma valor 1. La **Figura 5** muestra esta situación; el **Cuadro 2** representa las salidas y las sumas ponderadas.

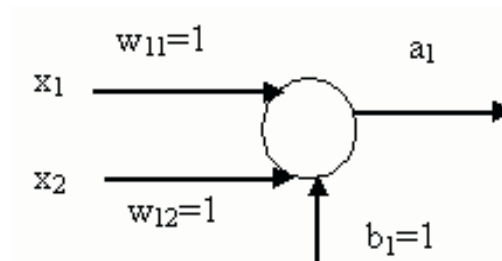


Figura 5. Representación de la función AND

| $x_1$ | $x_2$ | $\sum w_i x_i$ | $a_1$ |
|-------|-------|----------------|-------|
| 1     | 1     | 2              | 1     |
| 1     | 0     | 1              | 0     |
| 0     | 1     | 1              | 0     |
| 0     | 0     | 0              | 0     |

**Cuadro 2.** Salidas posibles de la red correspondiente a la función AND

## El Perceptron

El Perceptron constituye el primer modelo de red neuronal presentado y fue desarrollado por Rosenblatt en el año 1958. Inicialmente se presentó como un modelo capaz de clasificar patrones partiendo de un conjunto de ejemplos. Sin embargo al disponer sólo de una capa de entrada y otra de salida, tiene la posibilidad de operación muy limitada, pudiendo resolver problemas que sean linealmente separables. El caso más conocido es la incapacidad que muestra el Perceptron para resolver el problema OR exclusivo (XOR). Para entrenar al Perceptron, se recurre a un conjunto de ejemplos que constituyen lo que se denomina patrones de entrenamiento. A partir de ellos, el Perceptron debe establecer regiones donde queden separados los patrones de una y otra clase. Esta red simple, una vez entrenada y habiéndose establecido la característica de decisión, será capaz de clasificar cualquier nuevo patrón que se le presente.

La arquitectura del Perceptron es simple: está constituida por la capa de entrada con valores  $x_i$  y una sola capa de salida que puede presentar varias neuronas. En este caso particular, la salida se representa con  $a_1$  y la ganancia con  $b_1$ . La salida  $a_j$  de cada neurona de la capa de salida estará determinada por la ecuación (3)

$$a_j = f\left(\sum_{i=1}^n w_{ji} x_i + b_j\right) \quad (3)$$

En el caso del Perceptron, la salida  $a_j$  queda establecida al aplicarle la función de transferencia escalón a las sumas ponderadas de las entradas en la neurona. La función binaria  $f$  permite discriminar las clases en la que se quiere clasificar los patrones. Por ejemplo, si estos se quieren clasificar en dos categorías, cada patrón  $p = [x_1, x_2]$ , el plano  $x_1 x_2$  resultará dividido en dos semiplanos de acuerdo con la recta de decisión representada por la ecuación (4)

$$w_{11} x_1 + w_{12} x_2 + b_1 = 0 \quad (4)$$

Es decir, para el caso de dos dimensiones, las clases quedarán separadas por esta recta, a un lado se ubicarán los patrones de una clase y del otro los patrones correspondientes a la otra clase. Esta situación se representa en la **Figura 6**.

Es posible simplificar la expresión anterior al hacer que la sumatoria comience con  $i=0$ . Para ello se considera  $w_{00} = b_1$  y  $x_0 = -1$ ; entonces resulta el estado de la neurona dado por la ecuación (5)

$$a_j = f\left(\sum_{i=0}^n w_{0i} x_i\right) \quad (5)$$

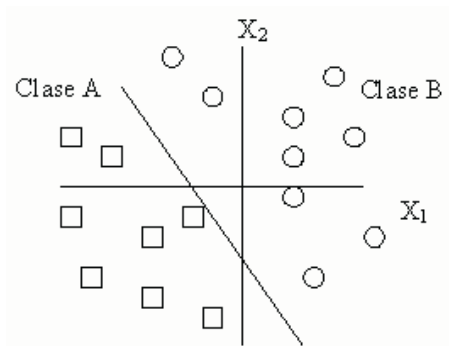


Figura 6. Patrones de clases diferentes

### Algoritmo de entrenamiento del Perceptron

El Perceptron tiene asociado un algoritmo de aprendizaje mediante el cual las salidas de la red son evaluadas comparándolas con las salidas que se espera obtener para cada patrón, **Figura 7**. Este aprendizaje se

llama supervisado y puede resumirse en los siguientes pasos (Veelenturf, 1995).

*Paso 1.* Inicializar pesos y ganancia con valores aleatorios.

*Paso 2.* Presentar el patrón  $p_1 = [x_1, \dots, x_r]$  y la salida esperada  $t_1$ .

*Paso 3.* Calcular la salida  $a_1$  mediante la ecuación (6)

$$a_j = f\left(\sum_{i=1}^n w_{1i} x_i + b_1\right) \quad (6)$$

donde  $f$  es la función de Heaviside.

*Paso 4.* Calcular el error  $e_1 = t_1 - a_1$  y actualizar pesos y ganancia.

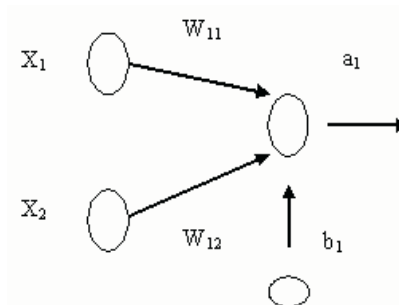


Figura 7. Modelo neuronal para el Perceptron simple

Si la red retorna la salida correcta, el error es cero y los valores de pesos y ganancia no sufren cambios.

Si la salida de la red no coincide con la esperada, los pesos y la ganancia se corrigen de acuerdo con las ecuaciones (7)

$$\begin{aligned} w^{Nuevo} &= w^{Anterior} + \varepsilon e p^T \\ b^{Nuevo} &= b^{Anterior} + \varepsilon e \end{aligned} \quad (7)$$

Donde  $\varepsilon$  simboliza la tasa que representa el ritmo de aprendizaje de la red.

Gráficamente esta situación puede entenderse como sigue: consideremos el Perceptron con salidas posibles 1 o -1, donde el vector de pesos es  $w$ , la salida de la neurona es  $a_j$  y la salida esperada es  $t_j$ . la salida de la neurona estará dada por la ecuación (8):

$$a_j = f\left(\sum_{i=0}^n w_{0i} x_i\right) \quad (8)$$

Durante el entrenamiento de la red, pueden presentarse las siguientes alternativas:

a) la salida de la neurona es  $a_j = -1$  y la esperada es  $t_j = 1$ . En este caso el producto escalar  $w \cdot x$  es negativo, y al aplicarle la función de transferencia  $f$  a un valor negativo, la salida será -1. Por lo tanto, debe procurarse que el ángulo  $\theta$  formado por los vectores  $w$  y  $x$  sea  $\theta < \pi / 2$ . Una posibilidad, tal como ilustra la **Figura 8**, consiste en adicionar a  $w$  el vector  $\alpha x$ . De aquí surge la ecuación (9), para el nuevo valor del peso:

$$w^{Nuevo} = w^{Anterior} + \alpha e p^T \quad (9)$$

b) Si la salida de la neurona es  $a_j = 1$  y la esperada es  $t_j = -1$ , el producto escalar  $w \cdot x$  es positivo, tal como se ilustra en la Figura 8. Razonando de manera similar al caso anterior, se obtiene la ecuación (10) que permite actualizar los pesos:

$$w^{Nuevo} = w^{Anterior} - \alpha e p^T \quad (10)$$

c) Cuando  $a_j = t_j$ , no es necesario hacer modificaciones en los pesos puesto que la red realizó correctamente la clasificación del patrón que se le presentó.

*Paso 5.* Volver al paso 2 hasta que la red haya clasificado todos los patrones.



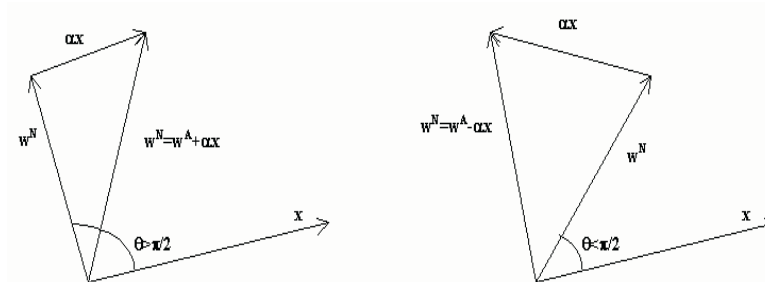


Figura 8. Actualización de pesos durante el entrenamiento de la red

## Teorema de Convergencia del Perceptron

Consideremos un Perceptron simple sobre el que se basará el siguiente estudio de convergencia. Se dispone de un conjunto de  $n$  patrones de entrenamiento junto con las correspondientes salidas esperadas. Cada patrón del conjunto de entrenamiento tiene  $r$  componentes que constituyen las entradas al Perceptron. Con estas condiciones se enuncia el teorema siguiente.

### Teorema

Sea el conjunto de entrenamiento, junto con sus salidas deseadas  $t$ :  $\{(p_1, t_1), (p_2, t_2), \dots, (p_m, t_m), \dots, (p_n, t_n)\}$ . Suponiendo que el mismo es linealmente separable, entonces el Perceptron simple encuentra una solución en un número finito de iteraciones; es decir, la salida de la red coincide con la salida deseada para cada patrón del conjunto de entrenamiento.

### Demostración

Dado que los patrones de entrenamiento son linealmente separables existe un conjunto de pesos  $\hat{W} = \{\hat{w}_{11}, \hat{w}_{12}, \dots, \hat{w}_{1r}\}$  y la correspondiente ganancia  $\hat{b}_1$ , como se representa en la Figura 2, tales que:

$$\sum_{i=1}^r \hat{w}_{1i} x_i > \hat{b}_1 \text{ para los patrones de la clase A}$$

$$\sum_{i=1}^r \hat{w}_{1i} x_i < \hat{b}_1 \text{ para los patrones de la clase B}$$

Nótese que los  $\hat{b}_1$  pueden considerarse como un " $\hat{w}_{1r+1}$ " asociado a una entrada  $x_{r+1} = 1$ . El símbolo " $\hat{\cdot}$ " sobre los pesos indica que estos valores corresponden al gradiente en el hiperplano que resuelve el problema de clasificar los patrones en las clases A y B. Por ejemplo, si consideramos que el patrón de entrada tiene dos componentes  $x_1$  y  $x_2$ , y dado que el problema es linealmente separable, podemos encontrar una recta que separe los patrones en dos clases A y B, tal como se ilustra en la Figura 6. En este caso la red contiene solo una neurona y por lo tanto  $j = 1$ .

Si se considera la iteración  $k$ -ésima, donde aun no han sido clasificados correctamente todos los patrones, resulta que se deben ajustar los pesos de la red pues la salida obtenida no coincide con la esperada. Este ajuste se realiza de acuerdo con la regla de aprendizaje del Perceptron:

$$w_{1i}^{k+1} = w_{1i}^k + \alpha e x_{mi}$$

$$e = t_m - a_m$$

Restando a ambos miembros  $\hat{w}_{1i}$ , elevando al cuadrado y extendiendo la suma para los  $r+1$  pesos, se obtiene la expresión del error cuadrático medio que deriva en la siguiente expresión:

$$\sum_{i=1}^{r+1} (w_{1i}^{k+1} - \hat{w}_{1i})^2 = \sum_{i=1}^{r+1} (w_{1i}^k + \alpha e x_{mi} - \hat{w}_{1i})^2 \quad (11)$$

Teniendo en cuenta la expresión del error y desarrollando:

$$\sum_{i=1}^{r+1} (w_{1i}^{k+1} - \hat{w}_{1i})^2 = \sum_{i=1}^{r+1} (w_{1i}^k - \hat{w}_{1i})^2 + \alpha^2 [t_m - a_m]^2 \sum_{i=1}^{r+1} x_{mi} +$$

$$+ 2\alpha [t_m - a_m] \left( \sum_{i=1}^{r+1} w_{1i}^k x_{mi} \right) - 2\alpha [t_m - a_m] \left( \sum_{i=1}^{r+1} \hat{w}_{1i} x_{mi} \right)$$

Inicialmente, consideremos el caso en que  $\sum_{i=1}^{r+1} w_{1i}^k x_{mi} > 0$ . Entonces resulta  $a_m = 1$  y dado que la salida de la red es incorrecta, debe ser  $t_m = 0$ , con lo que  $[t_m - a_m](\sum_{i=1}^{r+1} w_{1i}^k x_{mi}) < 0$ ; y si  $\sum_{i=1}^{r+1} w_{1i}^k x_{mi} < 0$  entonces  $a_m = 0$  y  $t_m = 1$ , con lo que  $[t_m - a_m](\sum_{i=1}^{r+1} w_{1i}^k x_{mi}) < 0$ , entonces es posible escribir este término como:

$$-\left| \sum_{i=1}^{r+1} w_{1i}^k x_{mi} \right|$$

De igual forma y siguiendo un análisis similar al anterior, resulta  $[t_m - a_m](\sum_{i=1}^{r+1} w_{1i}^k x_{mi}) > 0$ , puesto que si  $\sum_{i=1}^{r+1} \hat{w}_{1i} x_{mi} > 0$  entonces  $t_m = 1$  y la salida errónea es  $a_m = 0$ ; y si  $\sum_{i=1}^{r+1} \hat{w}_{1i} x_{mi} < 0$  entonces  $t_m = 0$  y la salida errónea es  $a_m = 1$ . La sumatoria también puede escribirse como

$$\left| \sum_{i=1}^{r+1} \hat{w}_{1i} x_{mi} \right|$$

Volviendo a la ecuación (11) y haciendo las sustituciones correspondientes se obtiene

$$\sum_{i=1}^{r+1} (w_{1i}^{k+1} - \hat{w}_{1i})^2 \leq \sum_{i=1}^{r+1} (w_{1i}^k - \hat{w}_{1i})^2 + \alpha^2 \sum_{i=1}^{r+1} x_{mi}^2 - 2\alpha \left| \sum_{i=1}^{r+1} \hat{w}_{1i} x_{mi} \right| \quad (12)$$

donde se ha prescindido del término

$$-2\alpha \left| \sum_{i=1}^{r+1} w_{1i}^k x_{mi} \right|$$

A los fines de simplificar la notación y hacer más claro el análisis, se harán las siguientes sustituciones:

$$\begin{aligned}
 D^{k+1} &= \sum_{i=1}^{r+1} (w_{1i}^{k+1} - \hat{w}_{1i})^2 \\
 D^k &= \sum_{i=1}^{r+1} (w_{1i}^k - \hat{w}_{1i})^2 \\
 T &= \min_{1 \leq k \leq p} \left\{ \sum_{i=1}^{r+1} \hat{w}_{1i} x_{mi} \right\}
 \end{aligned}$$

Como  $\sum_{i=1}^{r+1} x_{mi}^2 \leq r+1$  (esta condición surge de suponer que las entradas toman valores  $[0,1]$ ), esta condición se logra sin pérdida de generalidad para cualquier conjunto de  $\mathbb{R}^r$  mapeándolo a otro de igual dimensión pero con valores de  $[0,1]^r$ .

Finalmente la desigualdad (M+1) puede ser escrita de la forma

$$\begin{aligned}
 D^{k+1} &\leq D^k + 2\alpha^2(r+1) - 2\alpha T \\
 D^{k+1} &\leq D^k + 2\alpha[\alpha(r+1) - T]
 \end{aligned}$$

Eligiendo un valor de  $\alpha$  tal que  $\alpha \leq \frac{T}{r+1}$ , entonces  $D^{k+1} \leq D^k$ , lo que significa que en cada iteración  $D^k$  disminuye al menos en la cantidad  $2\alpha[\alpha(r+1) - T]$ . Si el número de iteraciones fuese infinito, entonces se llegaría a alcanzar el absurdo donde  $D^{k+1} < 0$ , que evidentemente no puede ocurrir.

En consecuencia, este teorema garantiza que en un número finito de pasos, el problema de clasificación de los patrones linealmente separables en las clases A y B queda resuelto.

### Ejemplo

Supóngase que se quiere que el Perceptron clasifique los siguientes patrones:

$$p_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, t_1 = -1; \quad p_2 = \begin{bmatrix} 2 \\ -1 \end{bmatrix}, t_2 = -1; \quad p_3 = \begin{bmatrix} -1 \\ 0 \end{bmatrix}, t_3 = 1; \quad p_4 = \begin{bmatrix} -2 \\ -1 \end{bmatrix}, t_4 = 1$$

Dado que las salidas esperadas toman solo dos valores posibles -1 o 1, la función de transferencia utilizada será un limitador fuerte simétrico. La red consiste en dos entradas y una sola salida; los pesos iniciales y la ganancia serán:

$$w = [-0.5, 0.5], \quad b = 0.5.$$

Inicialmente la recta característica se obtiene de la ecuación (11)

$$w_{11} x_1 + w_{12} x_2 + b_1 = 0 \quad (13)$$

$$\text{Haciendo } x_1 = 0, \text{ resulta } x_2 = \frac{b_1}{w_{12}} = -1$$

$$\text{Si } x_2 = 0, \text{ resulta } x_1 = \frac{b_1}{w_{11}} = 1$$

La **Figura 9** muestra que  $p_1$  no es clasificado correctamente.

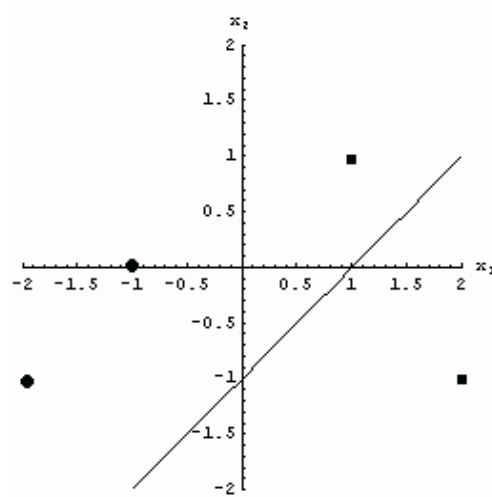


Figura 9. Las clases no son separadas correctamente

*Iteración 1*

Se presenta el patrón  $p_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$  junto con la salida esperada  $t_1 = -1$ .

La salida  $a_1$  será:

$$a_1 = f\left(\sum_{i=0}^n w_{1i} x_i + b_1\right) = f(0.5) = 1$$

$$\text{El error } e_1 = t_1 - a_1 = -1 - 1 = -2$$

El paso siguiente consiste en actualizar los pesos y la ganancia:

$$w^{Nuevo} = w^{Anterior} + ep^T = [-2.5, -1.5]$$

$$b^{Nuevo} = b^{Anterior} + e = -1.5$$

*Iteración 2*

Se presenta el patrón  $p_2 = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$  junto con la salida esperada  $t_1 = -1$

Ahora se trabaja con los pesos y ganancias corregidos. La salida de la red será:

$$a_1 = f\left(\sum_{i=0}^n w_{1i} x_i + b_1\right) = f(-.5) = -1$$

$$\text{El error correspondiente resulta ser } e_1 = t_1 - a_1 = -1 - (-1) = 0$$

El patrón ha sido bien clasificado. La recta característica se obtiene como antes:

$$\text{Haciendo } x_1 = 0, \text{ resulta } x_2 = \frac{b_1}{w_{12}} = \frac{1.5}{-1.5} = -1$$

$$\text{Si } x_2 = 0, \text{ resulta } x_1 = \frac{1.5}{-2.5} = -\frac{3}{5}$$

La **Figura 10** muestra esta recta y puede verse que  $p_2$  está clasificado correctamente.

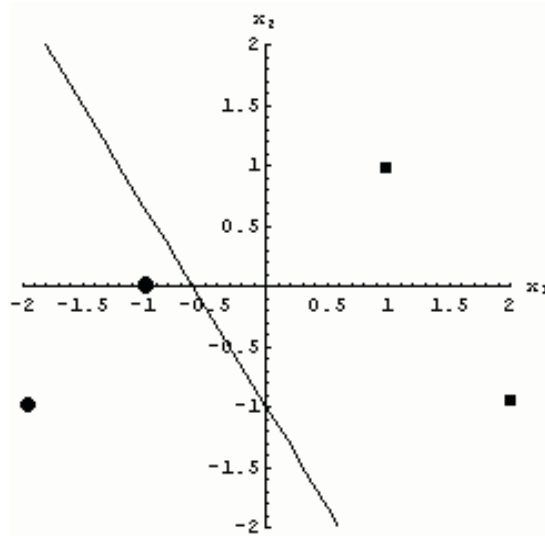
Dado que el error es cero los pesos y la ganancia para la nueva iteración no sufren modificaciones es decir:

$$w^{Nuevo} = w^{Anterior} + ep^T = [-2.5, -1.5]$$

$$b^{Nuevo} = b^{Anterior} + e = -1.5$$

Este procedimiento se repite para los patrones restantes, pero como puede verse, a partir de la segunda iteración todos los patrones ya han sido bien clasificados, por lo tanto en las siguientes iteraciones el error será nulo y la recta característica seguirá ocupando esta posición, tal como se ilustra en la Figura 10. De ahora en adelante, cualquier nuevo patrón que se ingrese en la red será clasificado de acuerdo con la característica de decisión final dada por la ecuación (12)

$$-2.5x_1 - 0.6x_2 - 1.5 = 0 \quad (14)$$



**Figura 10.** Luego del entrenamiento los patrones son clasificados correctamente

### Problema OR exclusivo (XOR). Limitaciones del Perceptron

Como se evidencia en el teorema de convergencia del Perceptron, éste presenta sus limitaciones a la hora de resolver un problema no separable linealmente. A continuación, y luego de haber resuelto los problemas de clasificación anteriormente descritos, se plantea un nuevo desafío: resolver el problema de la función XOR. Inicialmente se pensó que el tipo de red empleado en los ejemplos anteriores sería capaz de comportarse como un sistema clasificador general y capaz de aproximar cualquier función basándose en un conjunto de muestras que permitieran describir adecuadamente el problema.

El problema XOR se muestra en el **Cuadro 3**. A los fines de proceder sólo operativamente y analizar cómo se comporta el Perceptron al implementar el algoritmo de aprendizaje descrito anteriormente, se verá como éste se comporta cuando se intenta resolver el problema no separable linealmente de la función lógica XOR. Evidentemente, tal como pone de manifiesto el Cuadro 3, y a la luz del teorema de convergencia anterior, se espera que el Perceptron no sea capaz de separar este problema.

| $x_1$ | $x_2$ | $a_1$ |
|-------|-------|-------|
| 1     | 1     | 0     |
| 1     | 0     | 1     |
| 0     | 1     | 1     |
| 0     | 0     | 0     |

**Cuadro 3.** Representación de la función XOR

A continuación se verá cómo se comporta el Perceptron simple a la hora de intentar clasificar la función XOR por los patrones anteriores.

$$p_1 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, t_1 = 1; \quad p_2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, t_2 = 1; \quad p_3 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, t_3 = 0; \quad p_4 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, t_4 = 0$$



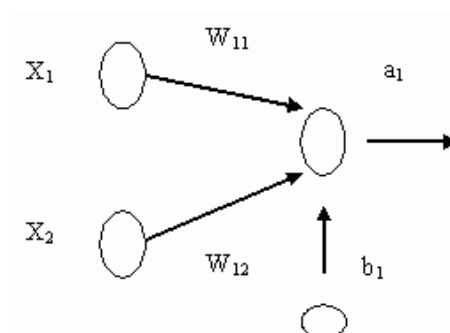


Figura 11. Perceptron simple

Para simplificar el problema sin perder generalización en su resolución, se supone que la ganancia sea  $b = 0$ . Dado que las salidas esperadas toman sólo dos valores posibles 0 o 1, la función de transferencia utilizada será un limitador fuerte. En primer lugar se adoptan los valores iniciales de los pesos de las conexiones:  $w = [0.5, -0.5]$

Se presenta a la red el patrón  $p_1 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$  y la salida esperada  $t = 1$ .

El producto escalar resultante es

$$wp^T = [0 \ 1] \begin{bmatrix} 0,5 \\ -0,5 \end{bmatrix} = -0,5 \text{ y la salida } a \text{ es}$$

$$a = f(-0,5) = 0$$

El error es  $e = t - a = 1$  y los nuevos pesos

$$w^{Nuevo} = w^{Anterior} + ep^T = \begin{bmatrix} 0,5 \\ 0,5 \end{bmatrix}$$

Ahora, la red recibe como entrada el patrón

$$p_2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \text{ y la salida esperada } t = 1.$$

El producto escalar resultante es

$$w^N p^T = [0,5 \ 0,5] \begin{bmatrix} 1 \\ 0 \end{bmatrix} = 0,5, \text{ y la salida resulta}$$

$$a = f(0,5) = 1$$

El error es  $e = t - a = 0$  y los nuevos pesos

$$w^{Nuevo} = w^{Anterior} + ep^T = \begin{bmatrix} 0,5 \\ 0,5 \end{bmatrix}$$

Para el patrón siguiente, se tiene  $p_3 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$  y la salida esperada  $t = 0$ .

El producto escalar resultante es

$$w^N p^T = \begin{bmatrix} 0,5 & 0,5 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \end{bmatrix} = 0, \text{ y la salida resulta}$$

$$a = f(0) = 1.$$

El error es  $e = t - a = -1$  y los nuevos pesos

$$w^{Nuevo} = w^{Anterior} + ep^T = \begin{bmatrix} 0,5 \\ 0,5 \end{bmatrix}$$

Finalmente, ingresando el último patrón, resulta  $p_4 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ ; la salida esperada es  $t = 0$ .

El producto escalar resultante es

$$w^N p^T = \begin{bmatrix} 0,5 & 0,5 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 1 \text{ y la salida es}$$

$$a = f(1) = 1.$$

El error es  $e = t - a = -1$  y los nuevos pesos

$$w^{Nuevo} = w^{Anterior} + ep^T = \begin{bmatrix} -0,5 \\ -0,5 \end{bmatrix}.$$

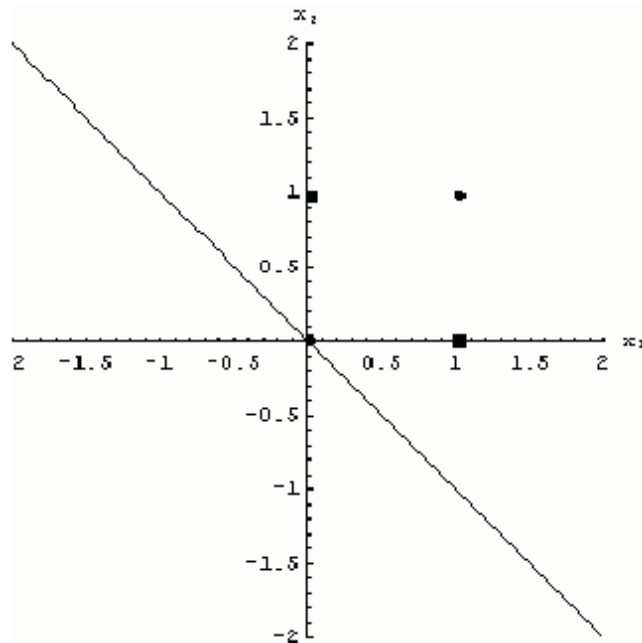


Figura 12. La recta de decisión no separa los patrones en clases

Por lo tanto la recta de decisión, una vez que la red ha sido entrenada con los cuatro patrones es:  $x_2 = -x_1$ . La **Figura 12** muestra que con esta red los patrones no han sido correctamente clasificados. Se puede observar que éste no es un problema linealmente separable, es decir no existe una recta única que separe los patrones de una clase de los de la otra. Incluso si se hubiera adoptado una ganancia, lo único que se hubiera logrado sería desplazar esta recta verticalmente, pero la red seguiría clasificando incorrectamente los patrones que se le presentaron.

Usualmente, para resolver problemas no separables linealmente se emplean redes del tipo Perceptron Multicapa (MLP) a las cuales se le agrega el mecanismo de aprendizaje Backpropagation. Aunque para resolver el problema específico de la función XOR, el mecanismo de trabajo puede simplificarse notablemente, puesto que bastará trabajar con dos neuronas tipo Perceptron con funciones de transferencia de Heaviside, éstas producen dos rectas características que separan el espacio en tres regiones, de las cuales la región central corresponde a un grupo y las externas a otro.

Asimismo, en muchos casos, en problemas que no son linealmente separables, puede existir una franja borrosa que contenga patrones de una y otra clase. Para resolver este problema, se recurre a la lógica difusa y concretamente para los casos que se trabajan en el presente artículo se recurre al denominado Fuzzy Perceptron. Estos temas requieren otra extensión para ser tratados y se espera poder divulgarlos en futuras entregas de este Boletín Matemático.

---

#### **Bibliografía**

1. FREEMAN, J., SKAPURA, D. (1991), *Neural Networks Algorithms Applications and programming Techniques*. Addison Wesley. Nueva York.
2. HAGAN, M. DEMUTH, H., BEALE, M. (1996), *Neural Network Design*, PWS Publishing Company.
3. HAYKIN, S. (1994), *Neural Networks, A Comprehensive Foundation*, Prentice Hall, Nueva Jersey.
4. HILERA, J., MARTÍNEZ, V. (1995), *Redes Neuronales Artificiales. Fundamentos, Modelos y Aplicaciones*, Ra-Ma, Madrid, España.
5. ISASI VIÑUELA, P. (2004), GALVÁN LEÓN, I.,(2004) *Redes de Neuronas Artificiales, Un Enfoque Práctico*. Pearson, Prentice Hall, Nueva York.
6. MARTÍN del BRÍO, B. y SANZ MOLINA, A. (2001) *Redes Neuronales y Sistemas Difusos*. Alfaomega, Ra-Ma, Madrid, España.
7. MATHLAB. (1993), *Mathlab Neural Network Toolbox, The Mathworks*.

---

### INVITACIÓN

Se invita a los Señores Profesores y Docentes en general, a presentar sus trabajos para incluirlos en los próximos números de este Boletín. Los mismos pueden presentarse a nombre de esta dirección en el aula 311 del Edificio Central de la Universidad.

Por cualquier consulta, comunicarse al teléfono 4659-2417.

Ing. Luinor E. Vilches

---

AUTORIDADES DE LA UNIVERSIDAD DE MORÓN

**Rector**  
Dr. Héctor Norberto Porto Lemma

**Secretario General**  
Dr. José María Baños

**Prosecretaria General**  
Arq. Marcela Kral

**Secretario Académico**  
Dr. Eduardo Néstor Cozza

**Secretario Administrativo**  
Dr. Jorge Marcos

**Secretario de Ciencia y Tecnología**  
Dr. Domingo Liotta

**Secretario Ejecutivo de Anexos y Subsedes**  
Dr. Carlos Humberto Pedrini

**Director de la Oficina de Comunicaciones  
y Relaciones Institucionales**  
Lic. Alejandro Gavric

**Director de la Oficina de Control de Gestión**  
Arq. Oscar Aníbal Borrachia

**Decano Facultad de Agronomía  
y Cs. Agroalimentarias**  
Ing. Agrónomo Antonio Ramón Angrisani

**Decano Facultad de Arquitectura,  
Diseño, Arte y Urbanismo**  
Arq. Oscar Aníbal Borrachia

**Decano Facultad de Ciencias  
Económicas y Empresariales**  
Dr. Jorge Raúl Lemos

**Decano Facultad de Ciencias Exactas,  
Químicas y Naturales**  
Dr. Aquiles Carlos Ferranti

**Decano Facultad de Derecho,  
Ciencias Políticas y Sociales**  
Dr. Bruno Oscar Corbo

**Decano Facultad de Ciencias Aplicadas  
al Turismo y la Población**  
Lic. Alejandro F. Gavric

**Decano Facultad de Filosofía,  
Cs. de la Educación y Humanidades**  
Lic. Roberto Mario Paterno

**Decano Facultad de Informática,  
Cs. de la Comunicación y Técnicas Especiales**  
Ing. Hugo René Padovani

**Decano Facultad de Ingeniería**  
Ing. Oscar Nuñez

**Decano Facultad de Medicina**  
Dr. Domingo Liotta