

# 2

## Introducción a la Neurocomputación

---

### 2.1 Breve reseña histórica de la Neurocomputación

#### a) Los comienzos (década de los cuarenta)

El trabajo de Warren McCulloch y Walter Pitts publicado en 1943 bajo el título:  
*"A logical calculus of the ideas immanent in nervous activity"*

podemos decir que señala el comienzo de la neurocomputación. En este trabajo se muestra que incluso las formas más simples de redes neuronales pueden computar, en principio, funciones aritméticas y lógicas.

Otros investigadores, como Norbert Wiener y John von Neumann (el padre de la computación moderna), realizaron trabajos en los que sugerían que la investigación en el diseño de ordenadores inspirados en el cerebro podría ser interesante.

En 1949, Donald Hebb, psicólogo de la Universidad de McGill, escribió el libro titulado:  
*"The Organization of Behavior"*

donde se diseña la primera ley de aprendizaje para las redes de neuronas artificiales.

#### b) La primera edad de oro (las décadas de los cincuenta y sesenta)

En 1962, Frank Rosenblatt introdujo y desarrolló una amplia gama de redes neuronales artificiales llamadas **perceptrones** que en principio crearon bastante entusiasmo. La regla de aprendizaje del perceptrón utiliza un ajuste iterativo de pesos cuando la respuesta de la unidad es incorrecta, que es más potente que la regla de Hebb.

Ya en 1960, Bernard Widrow y su alumno Marcian Hoff desarrollaron una regla de aprendizaje (que lleva su nombre) que es muy parecida a la regla de aprendizaje del perceptrón y que es la precursora de la regla de retropropagación para redes multicapa. La similitud de los modelos del psicólogo Rosenblatt y de los ingenieros eléctricos Widrow y Hoff pone en evidencia la interdisciplinariedad de las redes neuronales.

En 1969, Marvin Minsky y Seymour Papert, en su libro titulado *"Perceptrons"*, mostraron las limitaciones del perceptrón (con una sola capa) para aprender. Por ejemplo, el perceptrón no puede aprender la función lógica XOR (disyunción exclusiva). Ellos realizaron una campaña para desacreditar la investigación en redes neuronales y desviarla a la "inteligencia artificial".

c) **Los años tranquilos** (década de los setenta)

Los trabajos tempranos de Teuvo Kohonen (1972) de la Universidad de Tecnología de Helsinki desarrollaron las redes neuronales de memoria asociativa. Así mismo, James Anderson (1968, 1972), de la Universidad de Brown, comenzó su investigación en las redes de memoria asociativa. Hay que destacar también, los trabajos de Stephen Grossberg (director del Center for Adaptive Systems de la Universidad de Boston ) y sus muchos colaboradores, como Gail Carpenter, con quien desarrolló una teoría de redes neuronales autoorganizadas, llamada teoría de resonancia adaptativa (1985, 1987 y 1990).

d) **Entusiasmo renovado** (década de los ochenta)

John Hopfield del Instituto Tecnológico de California y premio Nobel en física, junto con David Tank, un investigador de AT&T desarrollaron varias redes neuronales con pesos fijos y activaciones adaptativas que pueden resolver problemas de optimización combinatoria con restricciones, como el problema del viajante (1982, 1984, 1985, 1986 y 1987). Su artículo publicado en la revista *Scientific American* ayudó a que se prestara más atención a las redes neuronales siguiendo el mensaje de un premio Nobel que señalaba la necesidad de estudiar la cognición humana para poder diseñar máquinas que puedan hacer lo que hace el hombre. Estos trabajos, junto con todas las conferencias que dio Hopfield por todo el mundo, contribuyeron a despertar gran interés por la investigación en redes neuronales.

La publicación en 1986 de los dos volúmenes titulados:

*"Parallel Distributed Processing"*

y editados por los psicólogos David Rumelhart de la Universidad de California (en San Diego) y James McClelland de la Universidad de Carnegie-Mellon, produjo una gran explosión en este campo. En ellos se recoge un método general efectivo de entrenamiento de una red neuronal multicapa (*la regla Delta generalizada*) que permite resolver, entre otros, el problema XOR. Este método fue inventado independientemente por varios investigadores; es muy similar a un algoritmo propuesto por Bryson y Ho (1969) en la teoría de control óptimo y a los propuestos por Werbos (1974), Parker (1985) y LeCun (1986), pero hasta entonces era muy poco conocido.

En 1987 se celebró en San Diego la primera conferencia abierta sobre redes neuronales (*IEEE International Conference on Neural Networks*), con más de 1700 participantes, y se formó la *International Neural Network Society* (INNS). En 1988 nace la revista *Neural Networks*; le sigue la revista *Neural Computation* en 1989 y la *IEEE Transaction on Neural Networks* en 1990. Posteriormente han ido apareciendo otras muchas y se han creado Institutos de Investigación y programas de formación en Neurocomputación.

Otro razón del creciente interés por la redes neuronales es la mejora en la capacidad de computación. Para ello se han desarrollado redes neuronales ópticas e implementaciones VLSI. El progreso de la neurocomputación va emparejado con el progreso en el diseño de neurocomputadores.

A pesar de su interesante historia, la teoría de redes neuronales artificiales está todavía en la etapa inicial de su desarrollo.

## 2.2 ¿Qué es una Red Neuronal?

Es un sistema de procesamiento de información que tiene ciertas características de comportamiento inspiradas en el conocimiento actual que tenemos sobre el funcionamiento de las neuronas biológicas. El **cerebro** se puede contemplar como un ordenador complejo, no lineal y paralelo que procesa gran cantidad de información. Su capacidad de organizar las neuronas para realizar ciertas computaciones, como el reconocimiento de objetos, en menos de 0.2 segundos, supone una rapidez de cálculo muy superior al ordenador actual más potente. El sistema visual humano proporciona una *representación* del entorno que le rodea y suministra la información necesaria para interactuar con él.

Las redes de neuronas artificiales son modelos computacionales paralelos que constan de **unidades de proceso** (neuronas) adaptativas y masivamente interconectadas. Son, por tanto, procesadores distribuidos masivamente en paralelo para almacenar conocimiento experimental y hacerlo disponible para su uso. Se basan en los siguientes supuestos:

- 1.- El procesamiento de información ocurre en muchos elementos simples llamados **unidades de proceso (o neuronas)**.
- 2.- Las señales se transmiten entre las neuronas a través de **conexión sináptica**.
- 3.- Cada enlace de conexión entre dos unidades de proceso tiene asociado un peso (llamado **peso sináptico**) que tiene un efecto multiplicador sobre la señal transmitida.
- 4.- Cada unidad de proceso aplica una **función de activación (o transferencia)** a sus señales de entrada para determinar su señal de salida.

Una **red de neuronas artificiales** está caracterizada por su:

- **Arquitectura**: Estructura o patrón de conexiones entre las unidades de proceso
- **Dinámica de la Computación** que nos expresa el valor que toman las unidades de proceso y que se basa en unas **funciones de activación (o de transferencia)** que especifican como se transforman las señales de entrada de la unidad de proceso en la señal de salida.
- **Algoritmo de Entrenamiento o Aprendizaje**: Procedimiento para determinar los pesos de las conexiones

Una característica muy importante de estas redes es su naturaleza *adaptativa*, donde el "**aprendizaje con ejemplos**" sustituye a la "programación" en la resolución de problemas.

Las RNA se parecen al cerebro en dos aspectos:

- La red adquiere el conocimiento mediante un **proceso de aprendizaje**.

- Los pesos de las conexiones (**pesos sinápticos**) entre las unidades de proceso se utilizan para **almacenar el conocimiento**.

El procedimiento utilizado para el proceso de aprendizaje se llama **algoritmo de aprendizaje** y permite modificar los pesos sinápticos de la red neuronal para conseguir un objetivo deseado.

Estas redes se llaman neuronales por estar inspiradas en la neurociencia pero no pretenden ser buenos modelos de neuronas biológicas. Se busca sobre todos su **capacidad computacional**.

Las RNA poseen las siguientes propiedades:

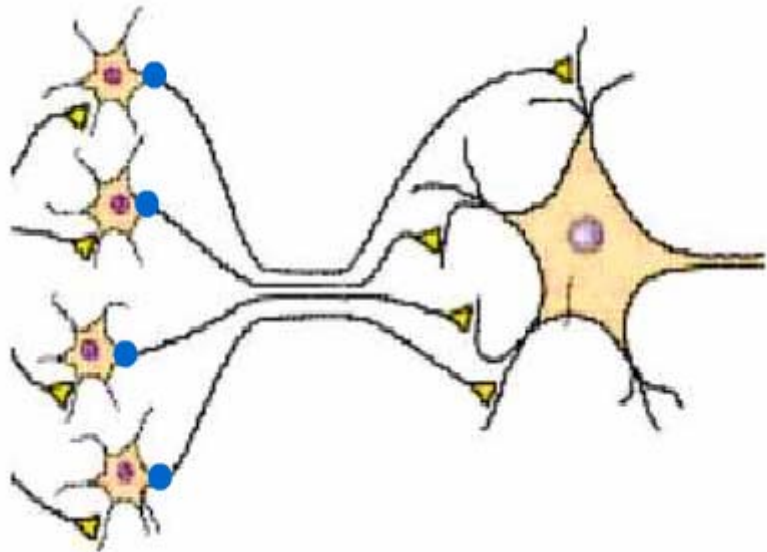
1. **No linealidad**.
- 2.- Permiten la representación de aplicaciones o correspondencias entre las entradas y salidas.
- 3.- **Adaptabilidad**: Acomodan (adaptan) sus pesos sinápticos a los cambios del entorno.
- 4.- **Información contextual**: El conocimiento viene representado por el estado de activación de la red neuronal. Cada neurona está afectada potencialmente por la actividad global de las otras neuronas de la red.
- 5.- **Tolerancia a fallos**: Por una parte, si una neurona o un enlace de la red neuronal son dañados, la respuesta de la red probablemente no quedará afectada. Una persona nace con unos 100 billones de neuronas y a partir de los cuarenta años, o antes, se pierden más de 40.000 neuronas diarias, que no son reemplazadas y, sin embargo, continuamos aprendiendo. Por otra parte, una red es capaz de reconocer señales de entrada diferentes a las señales entrenadas cuando difieren moderadamente. Análogamente, una persona puede reconocer a otra después de muchos años sin verla.
- 6.- **Implementación VLSI**: Al ser las redes masivamente paralelas pueden conseguir rapidez de cómputo en la realización de ciertas tareas. Permiten la implementación usando la tecnología VLSI (very-large-scale-integrated) y conseguir aplicaciones, como el reconocimiento de patrones, procesamiento de señales y control en tiempo real.

## 2.3 Unidades de proceso (neuronas artificiales)

La teoría y modelado de las redes neuronales artificiales está inspirada en la estructura y funcionamiento de los sistemas nerviosos, en los que la neurona es el elemento fundamental de procesamiento. Las unidades de proceso (neuronas artificiales) son las unidades básicas de cómputo en una red de neuronas artificiales y están inspiradas en las neuronas biológicas.

Una de las características que diferencian a las neuronas biológicas del resto de las células vivas, es su capacidad de comunicarse. Entre las estructuras fundamentales de una célula nerviosa (neurona) típica se encuentran las **dendritas**, el **cuerpo** (o soma) de la célula y un único **axón** (figura 1). El axón de muchas neuronas está rodeado por una membrana que se denomina vaina de mielina que se encuentra periódicamente interrumpida por los nodos de Ranvier. Las conexiones que llegan a una neurona procedentes de otras neuronas se producen

en distintos lugares de la misma conocidos con el nombre de **sinapsis**. En términos generales, las dendritas y el cuerpo celular reciben señales de entrada; el cuerpo las combina e integra y emite señales de salida. El axón transporta esas señales a los terminales axónicos, que se encargan de distribuir información a un nuevo conjunto de neuronas. Los impulsos nerviosos que pasan a través de estas neuronas que están conectadas pueden dar lugar a cambios locales en el potencial del cuerpo celular de la neurona receptora. Dichos potenciales pueden ser excitatorios (que hacen disminuir la polarización de la célula) o bien inhibitorios (que incrementan la polarización de la célula). Los potenciales de entrada se suman en el montículo del axón. Si la cantidad de despolarización del montículo del axón es suficiente, se genera un potencial de acción que viaja a lo largo del axón, alejándose del cuerpo celular.



Una **unidad de proceso bipolar** (figura 2) es una función matemática con dominio el conjunto  $n$ -dimensional  $\{-1,1\}^n$  y rango el conjunto  $\{-1,1\}$ , definida por la siguiente expresión:

$$f(x_1, x_2, \dots, x_n) = \begin{cases} 1 & \text{si } w_1x_1 + w_2x_2 + \dots + w_nx_n \geq \theta \\ -1 & \text{si } w_1x_1 + w_2x_2 + \dots + w_nx_n < \theta \end{cases} \quad (1)$$

donde los parámetros  $w_1, w_2, \dots, w_n$ , se llaman **pesos sinápticos** y son los pesos con los que se ponderan los valores de entrada  $x_1, x_2, \dots, x_n$ , o argumentos de la función; la suma ponderada  $u = w_1x_1 + w_2x_2 + \dots + w_nx_n$  se llama **potencial sináptico** y el parámetro  $\theta$  se llama **umbral** o sesgo. También se puede expresar la función  $f$  mediante la función signo, es decir,

$$f(x_1, x_2, \dots, x_n) = \text{sgn}(u - \theta)$$

siendo la función signo,

$$\text{sgn}(x) = \begin{cases} 1 & x \geq 0 \\ -1 & x < 0 \end{cases}$$

y diremos que en este caso la función de transferencia es la función signo. Análogamente, se define una **unidad de proceso binaria** como una función matemática con dominio el conjunto  $n$ -dimensional  $\{0,1\}^n$  y rango el conjunto  $\{0,1\}$ , definida por la siguiente expresión:

$$f(x_1, x_2, \dots, x_n) = \begin{cases} 1 & \text{si } w_1x_1 + w_2x_2 + \dots + w_nx_n \geq \theta \\ 0 & \text{si } w_1x_1 + w_2x_2 + \dots + w_nx_n < \theta \end{cases} \quad (2)$$

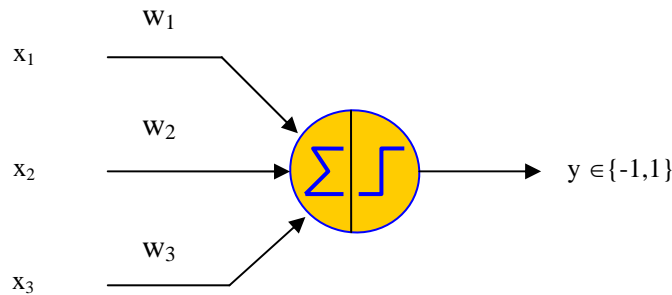


Figura 2. Unidad de proceso binaria.

Una **unidad de proceso logística** es una función de  $R^n$  en  $[0,1]$  dada por la expresión

$$f(x_1, x_2, \dots, x_n) = \frac{1}{1 + \exp\left(-2\beta \sum_{i=1}^n w_i x_i\right)}$$

Una **unidad de proceso tangente hiperbólica** es una función de  $R^n$  en  $[-1,1]$  dada por la expresión

$$f(x_1, x_2, \dots, x_n) = \frac{\exp\left(\beta \sum_{i=1}^n w_i x_i\right) - \exp\left(-\beta \sum_{i=1}^n w_i x_i\right)}{\exp\left(\beta \sum_{i=1}^n w_i x_i\right) + \exp\left(-\beta \sum_{i=1}^n w_i x_i\right)}$$

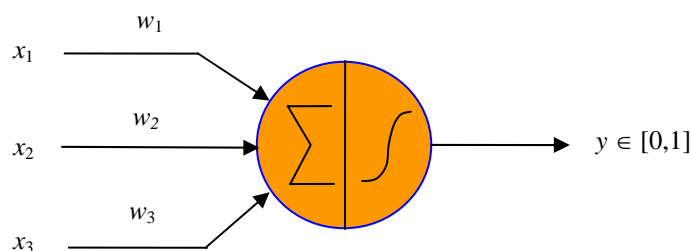


Figura3. Unidad de proceso logística.

Por lo tanto, una unidad de proceso es una función  $f$  de  $R^n$  en  $R$  que viene dada por la composición de dos funciones,  $g(u(x_1, x_2, \dots, x_N))$ . La función  $u$  (llamada *potencial sináptico*) es una función lineal que viene dada por la expresión

$$u(x_1, x_2, \dots, x_N) = w_1x_1 + w_2x_2 + \dots + w_nx_n,$$

y una función  $g$  que es una función monótona no decreciente de  $R$  en  $R$ , llamada *función de transferencia* o *función de activación*.

Las funciones de activación que vamos a utilizar son:

a) La función signo (neuronas discretas)

$$g(x) = \text{sgn}(x) = \begin{cases} 1 & x \geq 0 \\ -1 & x < 0 \end{cases}$$

o la función escalón (neuronas discretas)

$$g(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases}$$

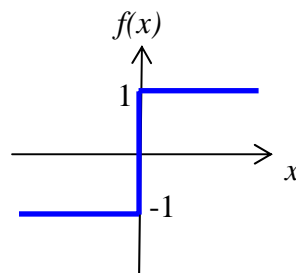


Figura 4. Función escalón.

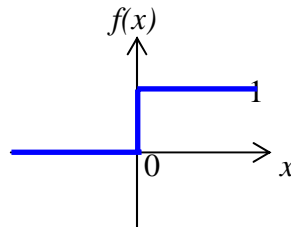


Figura 5. Función escalón.

b) Para unidades de proceso continuas, la función logística,

$$g(x) \equiv \frac{1}{1 + \exp(-2\beta x)}$$

la función tangente hiperbólica

$$g(x) = \tanh(\beta x) = \frac{e^{\beta x} - e^{-\beta x}}{e^{\beta x} + e^{-\beta x}}$$

o la función identidad.

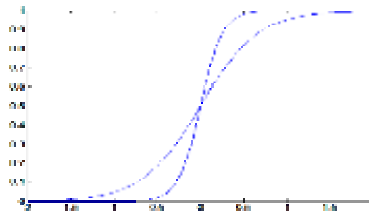


Figura 6. Funciones logísticas.

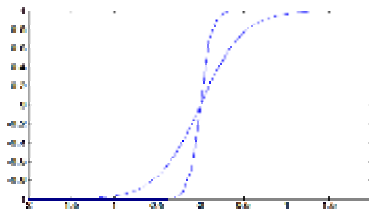


Figura 7. Funciones tangentes hiperbólicas

## 2.4 Características de las unidades de proceso inspiradas en las neuronas biológicas

Algunas características de las unidades de proceso inspiradas en las neuronas biológicas son:

- 1.- La unidad de proceso recibe muchas **señales**.
- 2.- Las señales se pueden modificar por un **peso** en la sinapsis receptora.
- 3.- La unidad de proceso **suma las entradas ponderadas**.
- 4.- En la circunstancias propicias (entradas suficientes) la neurona **transmite una única salida**.
- 5.- La **salida de una unidad de proceso** puede ser la entrada de otras unidades de proceso.
- 6.- El procesamiento de la información es **local**.
- 7.- La **memoria es distribuida**:
  - La memoria a largo plazo reside en los pesos sinápticos.
  - La memoria a corto plazo corresponde a las señales enviadas por las neuronas.
- 8.- Los **pesos sinápticos** se pueden modificar por la experiencia.
- 9.- Los neurotransmisores para la sinapsis pueden ser **excitadores** (pesos positivos) o **inhibidores** (pesos negativos).



## 2.5 Reglas de aprendizaje

Una de las características más importantes de las redes neuronales es su capacidad de aprender interactuando con su entorno o con alguna fuente de información. El aprendizaje de la red es un proceso adaptativo mediante el cual se van modificando los pesos sinápticos de la red para mejorar el comportamiento de la red. De manera general, una red neuronal va a modificar su peso sináptico  $w_{ij}$  correspondiente a la conexión de la neurona  $i$  con la neurona  $j$  mediante una regla de aprendizaje de la forma:

$$w_{ij}(k+1) = w_{ij}(k) + \Delta w_{ij}(k)$$

Es decir, el nuevo valor del peso sináptico se obtiene sumándole una cantidad (modificación) al valor antiguo.

Vamos a distinguir tres paradigmas de aprendizaje:

- **Aprendizaje supervisado**, en el que vamos a disponer de un conjunto de patrones de entrenamiento para los que conocemos perfectamente la salida deseada de la red. Un objetivo para diseñar la regla de aprendizaje supervisada podrá ser minimizar el error cometido entre las salidas (respuestas) de la red y las salidas (respuestas) deseadas. Tendremos así reglas de aprendizaje basadas en la corrección del error, como la regla de retropropagación del error en el caso del Perceptrón y el algoritmo de mínimos cuadrados, muy utilizados en problemas de clasificación y predicción.
- **Aprendizaje no supervisado** (competitivo o autoorganizado), en el que vamos a disponer de un conjunto de patrones de entrenamiento pero no vamos a conocer las salidas deseadas de la red. La red por sí misma buscará su comportamiento más adecuado atendiendo a cierto criterio y encontrará estructuras o prototipos en el conjunto de patrones de entrenamiento. Como ejemplo tenemos la regla de aprendizaje competitivo no supervisado (utilizada en problemas de agrupación de patrones y obtención de prototipos), la regla de Kohonen (utilizada en reconocimiento e identificación de patrones) y la regla de Hebb.
- **Aprendizaje por refuerzo**, basado en un proceso de prueba y error que busca maximizar el valor esperado de una función criterio conocida como una *señal de refuerzo*. La idea de este paradigma surge en la psicología en relación con el estudio del aprendizaje en los animales. Si una acción supone una mejora en el comportamiento entonces la tendencia a producir esta acción se refuerza y en caso contrario se debilita. Por ello, vamos a tener un conjunto de patrones de entrenamiento y sus correspondientes señales evaluativas, que suelen ser valores -1 ó +1 (en lugar de sus respuestas deseadas como en el caso supervisado). Dicha señal evaluativa informa a la unidad entrenada sobre su comportamiento con respecto a la entrada recibida, es decir, evalúa la adecuación de su salida para dicha entrada.