

## GENERALIDADES SOBRE REDES NEURONALES ARTIFICIALES

### INTRODUCCIÓN

Muchos de los desarrollos del hombre se deben a su capacidad para explicar y emular funciones que son realizadas por seres vivos, por ejemplo, se puede citar el radar, el cual surge como una emulación de la forma como un murciélago es capaz de detectar los objetos que están en su camino, sin necesidad de verlos, gracias a la emisión de una onda ultrasónica, su posterior recepción de la señal de eco y procesamiento, con el fin de detectar obstáculos en su vuelo con una rapidez y precisión sorprendentes. Como el mencionado, existen muchos ejemplos más en la naturaleza que han inspirado diversos inventos: el helicóptero, el avión, el submarino, para citar algunos.

Aunque el hombre ha sido capaz de emular funciones de los animales, para él siempre ha sido un sueño poder conocer e imitar, la llamada por muchos la máquina perfecta: el cerebro humano.

Las redes neuronales artificiales (RNA) surgen como un intento para emular el funcionamiento de las neuronas de nuestro cerebro. En este sentido las RNA siguen una tendencia diferente a los enfoques clásicos de la inteligencia artificial que tratan de modelar la inteligencia humana buscando imitar los procesos de razonamiento que ocurren en nuestro cerebro.

En este primer acercamiento, pretendemos introducir los conceptos iniciales y básicos asociados a las RNA. Empezamos con un recorrido por el proceso evolutivo que los científicos han llevado a cabo en la construcción de este apasionante campo del saber. Luego, mostramos como pasar del modelo de la neurona biológica, al modelo de la neurona artificial, que se seguirá utilizando a lo largo de este libro. Establecemos las características más relevantes de las RNA cuando emulan el proceso de aprendizaje que

ocurre en nuestro cerebro. Finalmente, presentamos algunos ejemplos prácticos, donde las RNA han sido aplicadas con éxito.

### BREVE RESEÑA HISTÓRICA

Emular redes neuronales de manera artificial no es un desarrollo reciente, se hicieron algunos intentos antes del advenimiento de los computadores, pero su verdadero desarrollo tuvo lugar cuando las simulaciones por computador fueron factibles por su capacidad de procesamiento y bajo costo. Luego de un periodo inicial de entusiasmo, las redes neuronales cayeron en un periodo de frustración y desprestigio, durante esta etapa el soporte económico y computacional era muy limitado y sólo unos pocos investigadores consiguieron logros de algún nivel de importancia. Estos pioneros fueron capaces de desarrollar una tecnología que superara las limitaciones identificadas en algunas publicaciones de Minsky y Papert en 1969, que sembraron un desencanto y frustración general en la comunidad científica.

Actualmente, las redes neuronales ocupan un sitio preponderante en el ámbito del procesamiento de señales con técnicas adaptativas. Acerquémonos a algunas etapas en las que puede resumirse la historia de las redes neuronales artificiales:

- 1. *El concepto de Neurona:*** A finales del siglo XIX, el científico español Santiago Ramón y Cajal logra describir por primera vez los diferentes tipos de neuronas en forma aislada. Al mismo tiempo plantea que el sistema nervioso estaba constituido por neuronas individuales, las que se comunicaban entre sí a través de contactos funcionales llamados sinapsis (teoría de la neurona). La hipótesis de este investigador se oponía a la de otros científicos de su época que concebía al sistema nervioso como una amplia red de fibras nerviosas conectadas entre sí formando un continuo, en clara analogía con el sistema circulatorio.



### Santiago Ramón y Cajal

(1852-1934). Médico español que obtuvo el Premio Nobel en 1906 por descubrir los mecanismos que gobiernan la morfología y los procesos conectivos de las células nerviosas, una nueva y revolucionaria teoría que se empezó a ser llamada la «doctrina de la neurona» .

Cursó la carrera de Medicina en Zaragoza, a donde toda su familia se trasladó en 1870, allí se centró en sus estudios universitarios con éxito. El año 1875 marcó el inicio del doctorado y de su vocación científica. Ganó la cátedra de Anatomía Descriptiva de la Facultad de Medicina de Valencia en 1883, donde pudo estudiar la epidemia de cólera que azotó la ciudad el año 1885. En 1887 se trasladó a Barcelona para ocupar la cátedra de Histología creada en la Facultad de Medicina de la Universidad de Barcelona. En 1888, definido por Cajal como “mi año cumbre”, descubre los mecanismos que gobiernan la morfología y los procesos conectivos de las células nerviosas de la materia gris del sistema nervioso cerebroespinal.

Su teoría fue aceptada en 1889 en el Congreso de la Sociedad Anatómica Alemana, celebrado en Berlín. Su esquema estructural del sistema nervioso como un aglomerado de unidades independientes y definidas, pasó a conocerse como «doctrina de la neurona» y en ella destaca la ley de la polarización dinámica, modelo capaz de explicar la transmisión unidireccional del impulso nervioso. Su trabajo y su aportación a la neurociencia se verían reconocidos, finalmente, en 1906, con la concesión del Premio Nobel de Fisiología y Medicina.

2. **Primeros intentos:** McCulloch y Pitts (1943) desarrollaron algunos modelos de redes neuronales basados en su conocimiento de neurología, estos modelos se basaban en neuronas simples, consideradas como dispositivos binarios con umbrales fijos. Los resultados de sus modelos fueron funciones lógicas elementales tales como “*a or b*” y “*a and b*”. En 1949, Donald Hebb en su libro *The Organization of Behavior*, presenta el principio del aprendizaje no supervisado, conocido como la Regla de Hebb.
3. **Tecnología emergente y promisoria:** No solo la neurociencia influía en el desarrollo de las redes neuronales, también los físicos y los ingenieros contribuían al progreso de las simulaciones de redes neuronales. Rosenblatt (1958) revitalizó fuertemente el interés y la actividad en esta área cuando diseñó y desarrolló el modelo de red neuronal que denominó Perceptron. El Perceptron tiene dos niveles, una de entrada cuya función es recibir la información del exterior y un nivel de

salida o de procesamiento que es la que se encarga de hacer el procesamiento de los datos entregados por el nivel de entrada para generar así la salida de la red neuronal. Este sistema es capaz de asociar unas entradas dadas a una salida determinada. Otro sistema fue el ADALINE (ADaptive LInear Element) el cual fue desarrollado en 1960 por Widrow y Hoff (de la Universidad de Stanford). El ADALINE fue un dispositivo electrónico analógico hecho de componentes simples, con un método de aprendizaje diferente al del Perceptron, empleando la regla de aprendizaje basada en mínimos cuadrados (LMS –*Least Mean Square*).

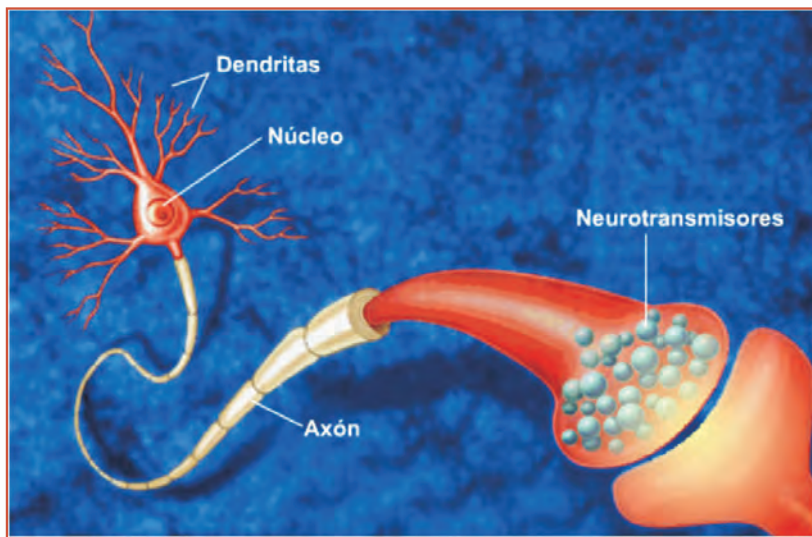
4. **Periodo de frustración y desprestigio:** En 1969 Minsky y Papert presentaron un trabajo, en cual se recalcaron las limitaciones del Perceptron para solucionar problemas complejos, esencialmente aquellos que no son linealmente separables. Estos autores, además generalizaron las limitaciones de un Perceptron monocapa a las redes neuronales multicapa, cuando plantearon: “...nuestro intuitivo juicio es que la extensión (a sistemas multicapa) es una tarea estéril”. El resultado de las afirmaciones de este trabajo fue disminuir el interés de los investigadores en la simulación de redes neuronales; lo que generó un desencanto de los investigadores en el área y trajo como resultado un periodo en el cual prácticamente se dejó de hacer nuevas propuestas.
5. **Innovación:** Aunque el interés por las redes neuronales era mínimo, varios investigadores continuaron trabajando en el desarrollo de métodos computacionales basados en neuromorfología para problemas de identificación y clasificación de patrones. Durante este periodo se generaron varios paradigmas, entre los cuales podemos mencionar a: Steve Grossberg y Gail Carpenter quienes desarrollaron la teoría de la resonancia adaptativa, ART (Adaptive Resonance Theory) (1976), Anderson y Kohonen (1982) quienes desarrollaron técnicas para aprendizaje asociativo, Hopfield (1984) quien desarrolló una red neuronal haciendo un símil energético. Paul Werbos (1982) desarrolló y usó el método de aprendizaje conocido como Backpropagation, destacando que varios años después de popularizarse este método, es actualmente el más utilizado en las arquitecturas multicapa con mayor nivel de aplicación práctica. En esencia una red Back-Propagation es un Perceptron con múltiples capas, con diferentes funciones de activación en las neuronas artificiales y con una regla de aprendizaje más robusta y confiable.
6. **Resurgimiento:** Durante el final de la década del setenta y principios de los ochenta, fue importante el resurgimiento del interés en el campo de las redes neuronales. Varios factores han influenciado este movimiento, tales como la aparición de libros y conferencias que han dado a conocer las bondades de esta técnica a personas de diferen-

tes áreas. Se introdujeron cursos en los programas académicos de las principales universidades europeas y americanas. El financiamiento a proyectos de investigación en redes neuronales en Europa, Estados Unidos y Japón que han dado origen a una gran variedad de aplicaciones comerciales e industriales.

7. **Lo Actual:** Se han realizado progresos muy significativos en el campo de las RNA, lo suficientes como para atraer una gran atención e interés en generar nuevos campos de aplicación. Ya se encuentran comercialmente circuitos integrados basados en RNAs y las aplicaciones desarrolladas resuelven problemas cada vez más complejos. En la actualidad ha surgido un nuevo tipo de máquinas de aprendizaje, como ejemplo las máquinas con vectores de soporte, con grandes capacidades para hacer procesamiento de datos, que aunque no son bioinspiradas como las redes neuronales, pues su aprendizaje se basa en métodos estadísticos, han impactado fuertemente en diversos campos del conocimiento.

#### DE LA NEURONA BIOLÓGICA A LA NEURONA ARTIFICIAL

Una neurona es una célula viva y está constituida por los mismos elementos que conforman las células biológicas. En general una neurona consta de un cuerpo celular más o menos esférico de 5 a 10 micras de diámetro, del que sale una rama principal el axón, y varias ramas más cortas denominadas dendritas. A su vez el axón puede producir ramas en torno a su punto de arranque, y con frecuencia se ramifica extensamente cerca de su extremo. Figura 1.1.



**Fig. 1.1 La neurona biológica**

Una de las características que diferencian a las neuronas del resto de células vivas, es su capacidad de comunicación. En términos generales, las dendritas y el cuerpo celular reciben señales de entrada, el cuerpo celular las combina e integra y emite señales de salida. El axón transporta esas señales a sus terminales, los cuales se encargan de distribuir la información a un nuevo conjunto de neuronas. Por lo general una neurona recibe información de miles de otras neuronas y, a su vez, envía información a miles de neuronas más. Se estima que en el cerebro humano existen del orden de  $10^{15}$  conexiones. El procesamiento de información de esta máquina maravillosa es en esencia paralelo, en la Tabla 1.1 mostramos un análisis comparativo entre un computador secuencial tipo Von Neumann y un sistema biológico neuronal.

**Tabla 1.1. Paralelo entre un computador secuencial y el sistema biológico neuronal**

Característica	Computador secuencial	Sistema biológico neuronal
Unidad de Procesamiento	Compleja Alta velocidad Una sola unidad	Simple Baja velocidad Muchas unidades

<b>Memoria</b>	Separada del procesador Localizada Direccionable aleatoriamente	Integrada dentro del procesador Distribuida Direccionable por contenido
<b>Procesamiento de los Datos</b>	Centralizado Secuencial Instrucciones almacenadas en Programas	Distribuido Paralelo Capacidad de aprendizaje
<b>Confiabilidad</b>	Muy vulnerable ante fallos	Robusto ante fallos
<b>Ambiente de operación</b>	Bien definido	Puede ser ambiguo



**John von Neumann zu Margitta**

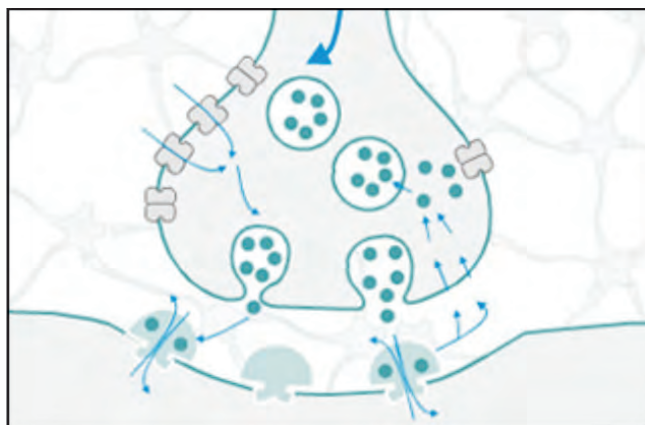
(1903-1957) Matemático húngaro que realizó importantes contribuciones a la Física, Matemáticas, Economía y Ciencias de la Computación. A los 23 años recibió su doctorado en matemáticas de la Universidad de Budapest.

Tras publicar junto a Oskar Morgenstern el libro *Theory of games and economic behavior* ('Teoría de juegos y comportamiento económico'), es considerado como el padre de la Teoría de los Juegos. Además es reconocido su trabajo en el Proyecto Manhattan que dio origen a la bomba atómica.

Es considerado como el pionero del computador digital moderno, y de hecho, una de las arquitecturas de computadores más utilizadas lleva su nombre. Junto con Eckert y Mauchly, en la Universidad de Pennsylvania, desarrollaron el concepto de programa almacenado en memoria, que permitió la ejecución de las instrucciones sin tener que volverlas a escribir. El primer computador en usar este concepto fue el llamado EDVAC (Electronic Discrete-Variable Automatic Computer), desarrollado por Von Neumann, Eckert y Mauchly. Los programas almacenados dieron a los computadores flexibilidad y confiabilidad, haciéndolos más rápidos y menos sujetos a errores que los programas mecánicos.

## La neurona biológica

Las señales que se encuentran en una neurona biológica son de naturaleza eléctrica y química. La señal generada por la neurona y transportada a lo largo del axón es eléctrica, mientras la señal que se transmite entre los terminales del axón de una neurona y las dendritas de las neuronas siguientes es de origen químico. Concretamente se realiza mediante neurotransmisores que fluyen a través de una región especial, llamada sinapsis que está localizada entre los terminales del axón y las dendritas de las neuronas siguientes, tal como vemos en la Figura 1.2.



*Fig. 1.2 Neurotransmisores en una Sinapsis*

Para establecer una similitud directa entre la actividad de una neurona biológica y una artificial, analicemos los siguientes aspectos funcionales:

### **1. Los elementos de proceso (neuronas) reciben las señales de entrada**

Una de las características de las neuronas biológicas, y a la que deben su gran capacidad de procesamiento y realización de tareas de alta complejidad, es que están altamente conectadas con otras neuronas de las cuales reciben un estímulo de algún evento que está ocurriendo o cientos de señales eléctricas con la información aprendida. Esta información al llegar al cuerpo de la neurona, afecta su comportamiento y puede afectar una neurona vecina o algún músculo.

### **2. Las señales pueden ser modificadas por los pesos sinápticos**

La comunicación entre una neurona y otra no es por contacto directo. La comunicación entre neuronas se hace a través de lo que se ha denominado sinapsis. Las sinapsis es un espacio que está ocupado por unas sustancias químicas denominadas neurotransmisores. Estos neurotransmisores son los que se encargan de bloquear o dejar pasar las señales que provienen de las otras neuronas.



**3. Los elementos de proceso suman las entradas afectadas por las sinapsis**

Las neuronas van recibiendo las señales eléctricas provenientes de las otras neuronas con las que tienen contacto. Estas señales se acumulan en el cuerpo de la neurona para definir qué hacer.

**4. Bajo una circunstancia apropiada la neurona transmite una señal de salida**

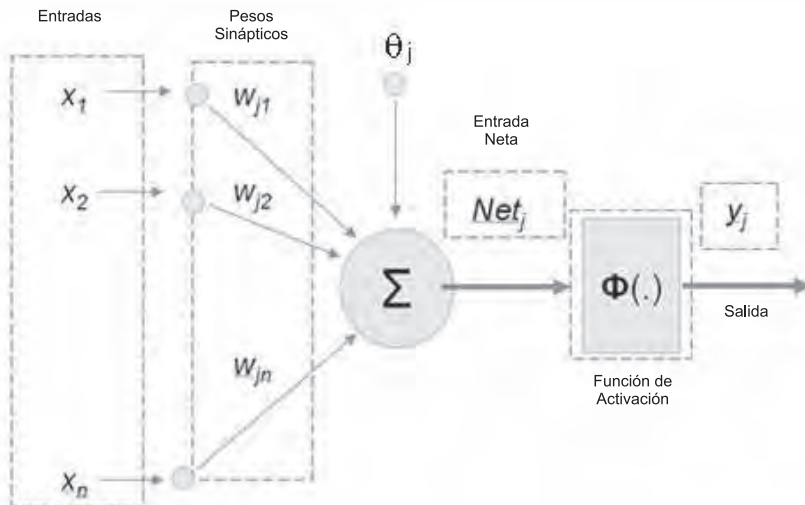
Si el total de la señal eléctrica que recibe la neurona es suficientemente grande, se puede vencer el potencial de acción, lo cual permite que la neurona se active o por el contrario permanezca inactiva.

**5. La salida del elemento de proceso puede ir a muchas neuronas**

Al activarse una neurona, está en capacidad de transmitir un impulso eléctrico a las neuronas con las cuales tiene contacto. Este nuevo impulso, por ejemplo, actúa como entrada para otras neuronas o como estímulo en algún músculo.

**La neurona artificial**

A partir de los aspectos funcionales de la neurona biológica, vamos a proponer un modelo de neurona artificial como el ilustrado en la figura 1.3.



**Fig. 1.3 Modelo de Neurona Artificial**

Como en el caso de la neurona biológica, la neurona artificial recibe unas entradas de estímulo que pueden provenir del sistema sensorial externo o de otras neuronas con las cuales posee conexión. Para el caso del modelo que proponemos en la ilustración, la información que recibe la neurona la definimos con el vector de entradas  $X = [x_1, x_2, \dots, x_n]$ .

La información recibida por la neurona es modificada por un vector  $w$

de pesos sinápticos cuyo papel es el de emular la sinapsis existente entre las neuronas biológicas. Estos valores se pueden asimilar a ganancias que pueden atenuar o amplificar los valores que se desean propagar hacia la neurona. El parámetro  $\theta_j$  se conoce como el *bias* o umbral de una neurona, cuya importancia veremos más adelante.

Los diferentes valores que recibe la neurona, modificados por los pesos sinápticos, los sumamos para producir lo que hemos denominado la entrada neta. Esta entrada neta es la que va a determinar si la neurona se activa o no.

La activación o no de la neurona depende de lo que llamaremos Función de Activación. La entrada neta la evaluamos en esta función y obtenemos la salida de la red. Si, por ejemplo, esta función la definimos como un escalón unitario, la salida será 1 si la entrada neta es mayor que cero, en caso contrario, la salida será 0.

Aunque no hay un comportamiento biológico que indique la presencia de algo parecido en las neuronas del cerebro, el uso de la función de activación es un artificio para poder aplicar las RNA a una gran diversidad de problemas reales. De acuerdo a lo mencionado la salida  $y_j$  de la neurona se genera al evaluar la neta en la función de activación.

Podemos propagar la salida de la neurona hacia otras neuronas o puede ser la salida de la red, que de acuerdo a la aplicación tendrá una interpretación para el usuario.

### Procesamiento matemático en la neurona artificial

En una neurona artificial el cálculo de la entrada neta se puede representar con la ecuación 1.1, y en forma vectorial se representa con la ecuación 1.2

$$Net_j = \sum_{i=1}^N x_i w_{ji} + \theta_j \quad [1.1]$$

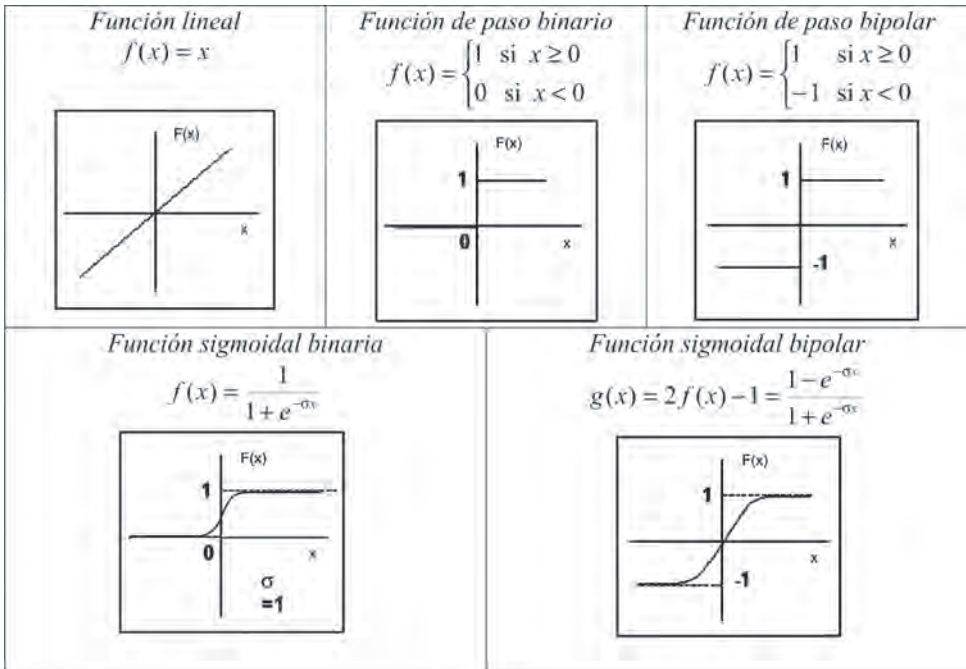
$$Net_j = w_1 x_{j1} + w_2 x_{j2} + \dots + w_i x_{ji} + \dots + w_N x_{jN} + \theta_j \quad [1.2]$$

$$Net_j = \mathbf{w}^T \mathbf{X}_j + \theta_j$$

La salida de la neurona artificial está determinada por una función de activación (*Fact*), tal como se aprecia en la ecuación 1.3.

$$y_j = Fact_j(Net_j) \quad [1.3]$$

La función de activación generalmente es del tipo escalón, lineal o sigmoïdal según se presenta en la Figura 1.4.



**RED NEURONAL ARTIFICIAL**

La neurona artificial por si sola posee una baja capacidad de procesamiento y su nivel de aplicabilidad es bajo, su verdadero potencial radica en la interconexión de las mismas, tal como sucede en el cerebro humano. Esto ha motivado a diferentes investigadores a proponer diversas estructuras para conectar neuronas entre si, dando lugar a las redes neuronales artificiales. En la literatura encontramos múltiples definiciones, de las cuales queremos destacar las siguientes, que se ajustan muy bien al concepto de red que seguiremos a lo largo de este libro.

La Agencia de Investigación de Proyectos Avanzados de Defensa (DARPA), define una red neuronal artificial como un sistema compuesto de muchos elementos simples de procesamiento los cuales operan en paralelo y cuya función es determinada por la estructura de la red y el peso de las conexiones, donde el procesamiento se realiza en cada uno de los nodos o elementos de cómputo.

Según Haykin, una red neuronal es un procesador paralelo masivamente distribuido que tiene una facilidad natural para el almacenamiento de conocimiento obtenido de la experiencia para luego hacerlo utilizable. Se parece al cerebro en dos aspectos:

1. El conocimiento es obtenido por la red a través de un proceso de aprendizaje.
2. Las conexiones entre las neuronas, conocidas como pesos sinápticos, son utilizadas para almacenar dicho conocimiento.

Kohonen, las define como redes de elementos simples (usualmente adaptativos) masivamente interconectados en paralelo y con organización jerárquica, las cuales intentan interactuar con los objetos del mundo real del mismo modo que lo hace el sistema nervioso biológico.

En síntesis se puede considerar que una red neuronal artificial es un sistema de procesamiento de información que intenta emular el comportamiento con las redes neuronales biológicas. Las redes neuronales artificiales han sido desarrolladas como generalizaciones de modelos matemáticos del conocimiento humano o de la biología neuronal, con base en las siguientes consideraciones:

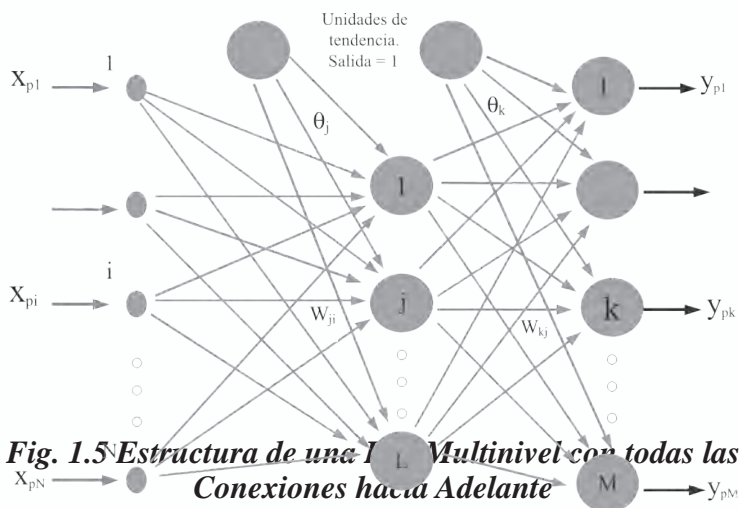
1. El procesamiento de información se realiza en muchos elementos simples llamados neuronas.
2. Las señales son pasadas entre neuronas a través de enlaces de conexión.
3. Cada enlace de conexión tiene un peso asociado, el cual, en una red neuronal típica, multiplica la señal transmitida.
4. Cada neurona aplica una función de activación (usualmente no lineal) a las entradas de la red (suma de las señales de entrada pesadas) para determinar su señal de salida.

La distribución de las neuronas dentro de una red neuronal artificial se realiza formando niveles de un número de neuronas determinado. Si un conjunto de neuronas artificiales reciben simultáneamente el mismo tipo de información, lo denominaremos capa. En una red podemos diferenciar tres tipos de niveles:

- **Entrada:** Es el conjunto de neuronas que recibe directamente la información proveniente de las fuentes externas de la red.
- **Oculto:** Corresponde a un conjunto de neuronas internas a la red y no tiene contacto directo con el exterior. El número de niveles ocultos puede estar entre cero y un número elevado. En general las neuronas de cada nivel oculto comparten el mismo tipo de información, por lo que formalmente se denominan Capas Ocultas. Las neuronas de las capas ocultas pueden estar interconectadas de diferentes maneras, lo que determina, junto con su número, las distintas arquitecturas de redes neuronales.

- **Salida:** Es el conjunto de neuronas que transfieren la información que la red ha procesado hacia el exterior.

En la figura 1.5, se puede apreciar la estructura de capas de una red neuronal artificial con varios niveles.



**Fig. 1.5 Estructura de una Red Neuronal Artificial Multinevel con todas las Conexiones hacia Adelante**

### ARQUITECTURAS DE REDES NEURONALES ARTIFICIALES

La arquitectura de una red neuronal artificial es la forma como se organizan las neuronas en su interior y está estrechamente ligada al algoritmo de aprendizaje usado para entrenar la red. Dependiendo del número de capas, definimos las redes como monocapa y multicapa; y si tomamos como elemento de clasificación la forma como fluye la información, definimos las redes como Feedforward y Recurrentes. En este libro conservamos el Anglismo Feedforward, para las redes cuya información fluye en un solo sentido desde el nivel de entrada hacia la capa de salida.

#### Redes monocapa

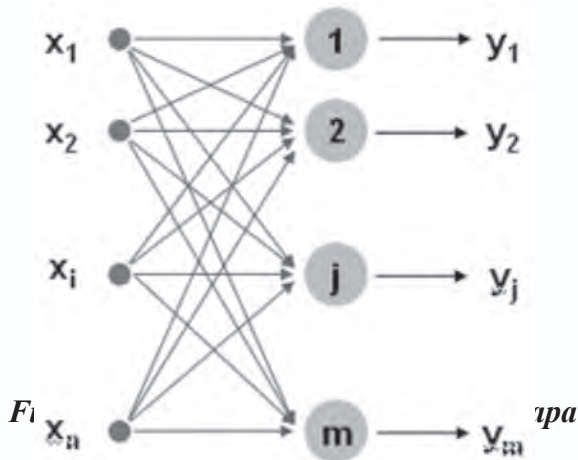
En la figura 1.6, observamos dos niveles de neuronas, el primero asociado al vector de entrada, pero no hay ningún tipo de procesamiento de estos datos, por esta razón no se considera formalmente como una capa, únicamente sirve de interfaz entre los datos de entrada y las siguientes capas de neuronas.

Este primer nivel tiene la misma dimensión del vector de entrada, la

información entra al mismo nivel y los datos son transferidos al siguiente nivel, modificados por los pesos sinápticos. Como las neuronas de este nivel reciben el mismo tipo de información lo denominamos capa y, a su vez, corresponde a la salida de la red, la llamaremos Capa de Salida. Notemos que en esta arquitectura solo disponemos de una capa de procesamiento, de ahí su nombre arquitectura monocapa.

Observemos que hay conectividad total entre el nivel de entrada y la capa de salida, pues todas las neuronas de entrada están conectadas con todas las neuronas de salida, por ejemplo, la neurona de entrada *i-ésima* se conecta a las *m* neuronas de salida.

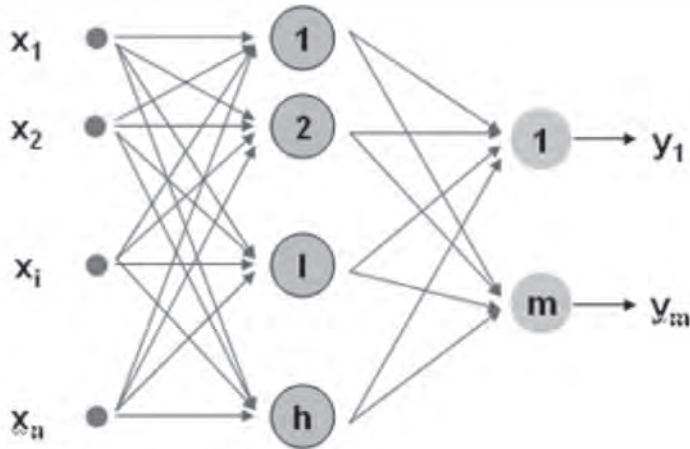
La capa de salida tiene *m* neuronas, por lo que luego del entrenamiento, la red neuronal establecerá una relación matemática de un espacio *n* dimensional a uno *m* dimensional.



### Redes multicapa

En este caso, la red tiene un nivel de entrada con *n* neuronas y una capa de salida de *m* neuronas; cuyo comportamiento es similar al que describimos en la red monocapa. La diferencia sustancial, es que incluimos una nueva capa intermedia entre la entrada y la salida, a esta capa la denominaremos Capa Oculta, que está conformada por *h* neuronas.

Como en el caso anterior, la información fluye en una única dirección, de la entrada a la capa oculta y finalmente, a la capa de salida, además existe conectividad total entre ellas. En este ejemplo, presentamos una única capa oculta, pero una red puede tener más de una capa intermedia. ¿Por qué el nombre de oculta? Simplemente porque ésta no tiene contacto con los datos que modelan el mundo real, es decir, los datos de entrada y salida.



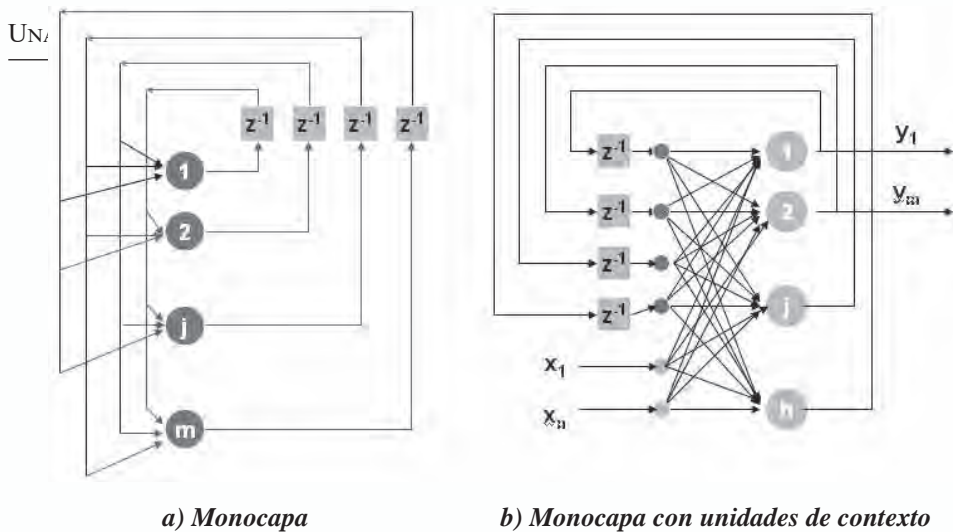
### Redes feedforward

En este tipo de red neuronal artificial, la información fluye en un único sentido desde las neuronas de entrada a la capa o capas de procesamiento, para los casos de redes monocapa y multicapa, respectivamente; hasta llegar a la capa de salida de la red neuronal. En la figura 1.6, la información fluye desde la neurona *i-ésima* de entrada hacia la *j-ésima* de salida. En la figura 1.7, la información proviene de la neurona *i-ésima*, pasa por la neurona *l-ésima* y finaliza en las neuronas de salida.

### Redes recurrentes

En este tipo de red neuronal, la información no siempre fluye en un sentido, puesto que puede realimentarse hacia capas anteriores a través de conexiones sinápticas. Este tipo de red neuronal puede ser monocapa o multicapa. En este caso, presentamos dos ejemplos de redes neuronales recurrentes monocapa. La salida de cada neurona es pasada por una unidad de retardo y, luego, llevada hacia todas las neuronas, excepto a sí misma. Observe en la figura 1.8.a, que hay conexión desde la neurona *j-ésima*, hacia las *m* neuronas, menos a la misma neurona *j-ésima*.

En la figura 1.8.b, hemos definido un grupo de neuronas que luego de recibir la información que proviene de las unidades de retardo, llevan su señal de estímulo hacia todas las neuronas de salida. A este conjunto de neuronas, se le suele llamar neuronas de contexto.



**Fig. 1.8 Redes Neuronales Recurrentes**

**EL APRENDIZAJE EN LAS REDES NEURONALES ARTIFICIALES**

El concepto de aprendizaje lo asociamos normalmente, en nuestra vida cotidiana, al proceso de formación que llevamos a cabo en las aulas de clase; adicionalmente, lo podemos asociar al resultado que nos dejan las diversas experiencias que tomamos de a nuestro diario vivir y a la manera como éstas nos condicionan frente a los diferentes estímulos que recibimos del entorno. La Real Academia de la Lengua Española define el aprendizaje como: “*Adquirir el conocimiento de algo por medio del estudio o de la experiencia*”.

Otro punto de vista, es mirar el aprendizaje como un proceso que nos permite apropiarnos un conocimiento, alguna habilidad, construir actitudes o valores, gracias al estudio, la experiencia o la enseñanza. Este proceso da origen a cambios permanentes, que son susceptibles de ser medidos y que, generalmente, modifican el comportamiento del individuo.

Biológicamente, se acepta que la información memorizada en el cerebro está más relacionada con los valores sinápticos de las conexiones entre las neuronas que con ellas mismas; es decir, el conocimiento se encuentra en las sinapsis. En el caso de las redes neuronales artificiales, se puede considerar que el conocimiento se encuentra representado en los pesos de las conexiones entre las neuronas. Todo proceso de aprendizaje implica un cierto cambio en éstas. En realidad se puede decir que se aprende modificando los pesos sinápticos de la red neuronal artificial.

El aprendizaje para la red neuronal artificial, es la habilidad para aprender del entorno y mejorar su desempeño, es un proceso interactivo que permite ajustar los pesos sinápticos. Según Mendel McClare, el aprendizaje en las redes neuronales artificiales, se puede definir como: “*Un proceso mediante el cual los parámetros libres de una red neuronal artificial son*



*adaptados a través de un proceso de estimulación del ambiente en el cual está embebida la red. El tipo de aprendizaje está determinado por la forma como se cambian los parámetros en el proceso”.*

Teniendo en cuenta lo mencionado en la ecuación 1.4, planteamos una expresión general para modelar el proceso de aprendizaje en las redes neuronales artificiales.

$$w(t+1) = w(t) + \Delta w(t) \quad [1.4]$$

Donde,

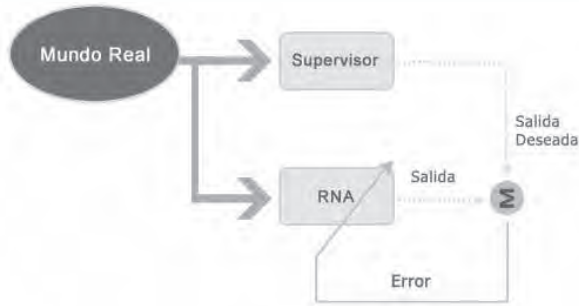
- $w(t+1)$  : Valor actualizado del peso sináptico
- $w(t)$  : Valor actual del peso sináptico
- $\Delta w(t)$  : Variación del peso sináptico

Es importante precisar, que la variación del peso sináptico depende del algoritmo o regla de aprendizaje que se esté utilizando para modificar los pesos sinápticos de la red neuronal artificial. En general, en este libro estudiaremos dos tipos de aprendizaje: el Supervisado y el No-Supervisado.

### **Aprendizaje supervisado**

El aprendizaje Supervisado, cuyo esquema presentamos en la de la figura 1.9, se caracteriza porque el proceso de entrenamiento es controlado por un agente externo llamado *supervisor o maestro*. En la figura, el mundo real corresponde al problema a modelar que se representa mediante un conjunto de datos de entrada y salida. El Supervisor opera como un “maestro” que guía el aprendizaje en la red, y conoce las salidas deseadas correspondientes a las respectivas entradas.

Los datos de entrada se le presentan al supervisor y a la red de neuronal simultáneamente, el supervisor propone la salida deseada para ser comparada con la salida de la red neuronal artificial. Para que la labor del supervisor sea exitosa se define en él, un error de entrenamiento como la diferencia entre la salida deseada o esperada y la salida que produce la RNA (Ecuación 1.5).



$$error = d - y \quad [1.5]$$

Donde:

- $y$  : Salida de la red neuronal artificial
- $d$  : Salida deseada

Generalmente, en los procesos de aprendizaje se usa un conjunto de patrones de entrenamiento, que está conformado por un conjunto de vectores de entrada  $\mathbf{X}$  y su correspondiente conjunto de vectores de salida  $\mathbf{D}$ .

$$\mathbf{X} = \{x_p, x_2, \dots, x_p, \dots, x_p\} \text{ Conjunto de vectores de entrada}$$

$$\mathbf{D} = \{d_p, d_2, \dots, d_p, \dots, d_p\} \text{ Conjunto de vectores de salida}$$

Donde cada patrón de entrenamiento está constituido por la pareja ordenada de vectores  $\{x_p, d_p\}$ , los cuales los podemos definir así:

$$\mathbf{x}_p = \{x_{p1}, x_{p2}, \dots, x_{pN}\} \text{ Elementos de la entrada de un patrón de entrenamiento}$$

$$\mathbf{d}_p = \{d_{p1}, d_{p2}, \dots, d_{pM}\} \text{ Elementos de la salida de un patrón de entrenamiento}$$

Por lo regular, la red debe aprender todo el conjunto de patrones de entrenamiento y, por esta razón, no se puede entrenar con un error local, sino que el aprendizaje se hace en términos de un *error global*  $E_p$ , que lo calculamos con base en la ecuación 1.6 y define el error cuadrático promedio. Se debe entender el error global, como el error que produce la red en sus diferentes neuronas de salida antes todos los patrones de aprendizaje que se estén utilizando para el proceso de entrenamiento.

$$E_p = \frac{1}{2P} \sum_{p=1}^P \sum_{j=1}^M (d_{pj} - y_{pj})^2 \quad [1.6]$$

Donde,

$M$  : Número de neuronas en la capa de salida

$P$  : Números de patrones de entrenamiento

### Aprendizaje no-supervisado

En este caso el vector de datos que describe el problema, se le presenta directamente a la red, pero ahora ya no hay un supervisor o maestro que guía el aprendizaje. En este caso los pesos de la red se calculan en función de la caracterización que se haga de la entrada que la red neuronal artificial esté recibiendo, de acuerdo a un objetivo específico que nos permite obtener el conocimiento que queremos representar con la red. El esquema de este tipo de aprendizaje lo presentamos en la figura 1.10.



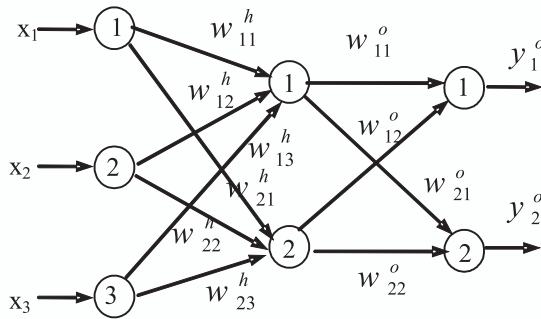
**Fig. 1.10 Aprendizaje No-Supervisado**

### EJEMPLO DE PROCESAMIENTO DE LA INFORMACIÓN EN UNA RED NEURONAL

El procesamiento de la información que realiza una red neuronal artificial, generalmente, consiste en recibir la información de entrada e ir propagándola por toda la estructura de la red neuronal hasta generar la salida de la misma

El vector de entrada lo denotaremos por  $x$ , y está conformado por los diferentes valores que va recibir la red neuronal. Este vector se propagará por las diferentes neuronas y capas que la conforman hasta generar la salida de la red neuronal.

En este ejemplo presentaremos el procesamiento necesario para generar la salida de la red neuronal mostrada en la figura 1.11. El vector de entrada esta compuesto por tres componentes  $x = [x_1, x_2, x_3]$ . El objetivo es generar la expresión para el vector de salida  $y^o = [y_1^o, y_2^o]$



**Fig. 1.11 Estructura de conexión entre neuronas**

En el anterior esquema la entrada neta de la neurona de la capa oculta es:

$$Net_j^h = \sum_{i=1}^3 w_{ji}^h x_i \tag{1.7}$$

$$Net_j^h = w_{j1}^h x_1 + w_{j2}^h x_2 + w_{j3}^h x_3$$

y su salida:

$$y_j^h = Fact(Net_j^h) \tag{1.8}$$

donde,

$$y_1^h = Fact(w_{11}^h x_1 + w_{12}^h x_2 + w_{13}^h x_3)$$

$$y_2^h = Fact(w_{21}^h x_1 + w_{22}^h x_2 + w_{23}^h x_3)$$

Empleando las ecuaciones 1.7 y 1.8, la salida se puede expresar como:

$$Net_1^o = w_{11}^o y_1^h + w_{12}^o y_2^h \quad y_1^o = Fact(Net_1^o) \tag{1.9}$$

$$Net_2^o = w_{21}^o y_1^h + w_{22}^o y_2^h \quad y_2^o = Fact(Net_2^o) \tag{1.10}$$

**NIVEL DE APLICACIÓN**

Las redes neuronales artificiales, con su inherente capacidad para extraer el conocimiento a partir de datos complejos e imprecisos, sin requerir un modelo *a priori* del problema, se han convertido en una herramienta útil para el procesamiento de series de tiempo, o para el reconocimiento o cla-

sificación de patrones. Esta capacidad está asociada al aprendizaje que la red hace del problema, convirtiéndose en una técnica emergente que supera ampliamente a las técnicas computacionales tradicionales o determinísticas en la solución de problemas complejos. Una red entrenada puede ser vista como un experto en el manejo de la información que se le ha dado para analizar, este experto puede ser utilizado para proporcionar proyecciones ante nuevas situaciones de interés.

Entre las características y ventajas representativas de las redes neuronales artificiales tenemos:

- *Capacidad de aprendizaje.* Las redes neuronales artificiales son capaces de aprender a partir de un conjunto de ejemplos que sean representativos o intenten modelar el problema a solucionar. Ésta es una de las características más poderosas, pues permite que la red neuronal se adapte a nuevas situaciones siempre y cuando exista la información necesaria para realizar el proceso de entrenamiento. Como en el caso de nosotros los seres humanos, la calidad de la información o datos que le entreguemos a la red, va a ser fundamental para garantizar la calidad del aprendizaje, en otras palabras, nosotros garantizamos que “la red aprende lo que le enseñemos”.
- *Capacidad de generalización.* Cuando se usan las redes neuronales artificiales se necesita que ellas sean capaces de extender su respuesta adecuada a eventos o datos que no han sido utilizados en la fase de aprendizaje, a esta característica la denominaremos: generalización. Esta propiedad de nosotros los humanos, nos permite extender un conocimiento adquirido bajo ciertas circunstancias, a situaciones si-milares pero no exactas, también la tienen las redes neuronales, si los datos son representativos del problema. Lo que buscamos es que la red aprenda la relación entre la entrada y la salida usando algunos ejemplos y, luego, a la red ya entrenada, si le suministramos un conjunto de entradas del mismo contexto del problema, la salida generada debe ser la correcta.
- *Capacidad para extraer características esenciales los datos.* Algunos tipos de redes neuronales artificiales, son capaces de extraer la información relevante que hay en ciertos datos y rechazar aquella información que no sea importante. Esta característica se ha aplicado con mucho éxito en problemas asociados a la minería de datos.
- *Capacidad de asociación.* Esta es una característica muy importante de la inteligencia de nosotros los seres humanos e incluso, podemos atrevernos a decir que es más elevada que la capacidad de aprender a partir de ejemplos. ¿Qué entendemos por asociación? Cuando establecemos una relación entre dos conceptos. Por ejemplo, si decimos “aprendizaje en redes neuronales”, para todos nosotros, esperamos

que inmediatamente se genere una asociación al concepto de modificación de pesos sinápticos. Si decimos esta misma frase en un auditorio que no conozca de Redes Neuronales, muy seguramente no le dirá nada o no la asociarán a un conocimiento. Más adelante aprenderemos sobre un tipo de aprendizaje que trata de emular esta característica y sobre la red, ya entrenada, que es capaz de asociar.

- *Capacidad de agrupación.* Hemos venido evolucionando en las diferentes formas de aprender que presenta el ser humano, iniciamos con un aprendizaje basado en la repetición de una serie de ejemplos de entrada y salida que modelan el conocimiento al aprender, esto podría asociarse a la memorización, pero les garantizamos que las redes neuronales artificiales son capaces de generalizar. Luego vimos que al igual que los seres humanos, las redes neuronales artificiales son capaces de separar la información relevante de la irrelevante en un conjunto de datos. Pasamos al conocimiento adquirido por asociación y ahora presentamos la facilidad que tenemos para agrupar conocimiento. Por ejemplo, tenemos los libros de nuestra comunidad y queremos organizarlos en una biblioteca... ¿Qué pasa si pedimos a diferentes grupos que los organicen? Cada uno de ellos buscará algún criterio para hacerlo; pues bien, veremos cómo las redes neuronales artificiales emulando nuestra capacidad, podrán, igualmente solucionar problemas de asociación por sectores o grupos de información, algo que la literatura se conoce como *clustering* o agrupamiento.
- *Aprendizaje Adaptativo.* Hemos visto que las redes neuronales artificiales, al igual que las naturales, aprenden a realizar algunas tareas gracias a un proceso de entrenamiento utilizando ejemplos lo suficientemente ilustrativos para ello, por lo que no se requiere de modelos a priori, del sistema o problema que queremos solucionar. Pero el nivel de emulación no se limita a este importante hecho, sino que las redes neuronales artificiales son capaces de modificar su propia estructura de pesos sinápticos para adaptarse a nuevas situaciones que se presenten en el sistema y modifiquen su desempeño. Visto desde otra perspectiva, una misma arquitectura de red neuronal puede aprender diferentes problemas modificando su matriz de pesos para adaptarse a las diferentes tareas que le pedimos sean solucionadas.
- *Auto-Organización.* En el caso concreto de los modelos de redes neuronales artificiales propuestos por Teuvo Kohonen, su arquitectura se puede modificar completamente dependiendo de la forma como estén organizados los datos o del objetivo a cumplir. De alguna manera, la red neuronal trata de seguir (imitar) la estructura de los datos que le presentemos a su entrada sin necesidad de que le impongamos una salida predeterminada.
- *Tolerancia a fallos.* Las redes neuronales artificiales son tolerables

a fallos en los datos (ruido, distorsiones, datos incompletos) debido a que la red no almacena el conocimiento de una manera localizada, sino que él mismo se encuentra distribuido en toda la estructura de la red.

- *Operación en Tiempo Real.* Las redes neuronales deben su fortaleza a su capacidad para procesar información en paralelo, en implementaciones hardware de redes neuronales artificiales se puede realizar este tipo de procesamiento, lo que permite un cálculo de la información a muy altas velocidades, suficientes para cumplir las restricciones de tiempo real, de la gran mayoría de las aplicaciones en las cuales se utilizan las redes neuronales artificiales. De igual manera, en implementaciones software, hemos visto que el procesamiento de una neurona es simple y de rápida ejecución; aunque, en este tipo de uso no podemos garantizar el paralelismo, de igual manera por su simplicidad, podemos asegurar que la ejecución total de una red, se hace en un tiempo lo suficientemente corto como para garantizar las restricciones de tiempo real para aplicaciones típicas, como de filtrado, control, identificación de sistemas, entre otras.

Los computadores digitales actuales, superan al hombre en su capacidad de cálculo numérico y el manejo de símbolos relacionales. Sin embargo, el hombre puede solucionar problemas mucho más complejos de percepción (por ejemplo, reconocer a un amigo entre un tumulto desde un simple vistazo de su cara o al escuchar su voz, incluso por el modo de caminar; definir la habitabilidad de un aula a partir de sensaciones de temperatura, ruido, humedad, iluminación, etc.) a muy altas velocidades y sin necesidad de concebir un complejo modelo matemático o computacional. La respuesta está en la arquitectura del sistema neuronal biológico que es completamente diferente a la arquitectura del computador tradicional von Neumann, como se indicó en la Tabla 1.1. Estas diferencias, le brindan un factor de diferenciación y aumento del desempeño a las redes neuronales artificiales en una gran variedad de campos de aplicación como mencionados a continuación:

- *Reconocimiento y Clasificación de Patrones.* Esta es una de las grandes fortalezas de las Redes Neuronales Artificiales y, al igual, que las naturales, tienen una especial facilidad para solucionar problemas de reconocimiento y clasificación de patrones, en especial se puede encontrar una gran cantidad de problemas resueltos utilizando las redes tipo Perceptron Multicapa.
- *Categorización de Patrones (“clustering”).* Otra área donde las redes neuronales han sido utilizadas con éxito es el agrupamiento o clustering. Este tipo de aplicación es muy útil en minería de datos donde se necesitan encontrar grupos de patrones con características similares.

- *Procesamiento de señales.* Otro campo de aplicación de las redes neuronales es el procesamiento de señales como, por ejemplo, la voz. En este caso la red se puede usar para realizar alguna parte del proceso de extracción de características o como una etapa clasificadora
- *Optimización.* Aunque la principal fortaleza de las redes neuronales no es la solución de este tipo de problemas y existen otras técnicas de inteligencia computacional para ello, como los Algoritmos Genéticos, se han utilizado algunos modelos de RNA, como las redes neuronales de Hopfield y Kohonen, para resolver problemas clásicos de optimización en ingeniería, v.gr. el problema del Agente Viajero.
- *Control.* En las industrias hay procesos donde han sido usadas con éxito las redes neuronales para el modelado y control de los mismos.
- *Medicina.* Las redes neuronales se han usado para la extracción de características, a partir de una imagen o como un clasificador para determinar la pertenencia a una clase del patrón representado por una imagen.
- *Gestión financiera.* En este caso se ha aprovechado la capacidad de las redes neuronales para modelar series de tiempo, lo cual nos permite utilizarlas para hacer predicciones de la misma. En este contexto las redes neuronales se han usado, por ejemplo, para modelar el comportamiento de variables económicas, como el valor de la acción de una empresa.
- *Robótica.* La robótica móvil es un área donde se puede sacar partido de la capacidad de aprendizaje de las redes neuronales. Básicamente, para ayudarle al robot a tomar decisiones en ambientes cambiantes. En este momento se ha dado solución a problemas que se caracterizan por tener altos niveles de incertidumbre que con métodos tradicionales jamás se habría obtenido. Soluciones tan novedosas e interesantes como la reconstrucción craneofacial para la identificación de hombres, música neurocomputacional, sistemas de detección de virus en computadores, identificación de usuarios en cajeros automáticos desde la imagen del iris de los ojos, reconocimiento de emisores en comunicaciones, diagnóstico de hepatitis, recuperación de telecomunicaciones ante fallas en el software, interpretación de palabras chinas, detección de minas submarinas, análisis de texturas, reconocimiento de objetos tridimensionales, reconocimiento de textos manuscritos, entre otros.