Redes Neuronales: biología, computación o física?

Sergio A. Cannas

27 de Abril de 1999

1 Introducción

El término *"redes neuronales"* ha ido tornándose cada vez mas familiar en los últimos años, especialmente en el ámbito científico-técnico, pero tambien en contextos menos especializados. La primera imágen que se presenta generalmente al neófito, al escuchar mencionar el tema, es la de un "cerebro artificial" y, como tal, algo estrechamente relacionado con la biología (mas especificamente a la neurofisiología) y con la informática. Ambas relaciones son estrictamente correctas pero, no obstante, constituyen una imagen incompleta. Actualmente puede considerarse a tema de las redes neuronales como un área del conocimiento interdisciplinaria por excelencia, contribuyendo a la misma diversas disciplinas tales como biología, informática, matemá tica, psicología y física, entre las mas importantes. Si bien el papel de la mayoría de esta disciplinas es facilmente identificable, el de la física (el cual fue fundamental en las primeras etapas del desarrollo de esta área) suele prestarse a confusión. La primera impresión es este sentido suele ser que las redes neuronales constituyen un intento de modelar en forma completa el sistema nervioso partiendo de primeros principios, es decir, en términos de las leyes de la física y la química microscópicas (enfoque reduccionista). Por el contrario, el enfoque principal de las redes neuronales consiste en intentar emular funciones cognitivas básicas (tales como aprendizaje, memoria asociativa, generalización, etc) mediante modelos simplificados que tomen en cuenta solo las características funcionales esenciales del sistema neurológico, haciendo abstracción de los procesos físicos-químicos que subyacen en las mismas. Estrictamente, son estos modelos lo que se conoce como redes neuronales. El aporte de la física (mas concretamente de la física estadística) es entonces básicamente **metodológico**.

El objetivo principal del presente artículo consiste en presentar, de manera lo menos técnica posible, una visión general de los conceptos básicos de la teoría de las redes neuronales. Por otro lado, se intenta tambien mostrar como los conceptos y la metodología de la física estadística pueden resultar de enorme utilidad en el desarrollo de nuevas áreas del conocimiento como la presente, las cuales por su complejidad requieren necesariamente de un abordaje multidisciplinario.

2 Modelos matemáticos

Desde el punto de vista funcional, el sistema nervioso puede describirse como un *sistema dinámico de alta complejidad*. Se trata entonces de modelar dicho sistema dinámico, trabajando por analogías con sistemas dinámicos conocidos de la física. Cómo?

Todo modelo se supone constituye una *representación simplificada* de cierto sistema real. La principal justificación de dicha simplificación es que permita el tratamiento analítico del problema. No obstante, a menudo el conjunto de simplificaciones es tan drástico que el modelo pierde en gran medida su semejanza con el sistema real que intenta simular. Esto no implica necesariamente que el modelo y sus soluciones se vuelvan inútiles o sin sentido. La física estadística nos muestra numerosos ejemplos en los cuales puede obtenerse información valiosa a partir de modelos ultrasimplificados. Comenzaremos entonces con una descripción breve de un problema paradigmático de la física estadística (estrechamente relacionado con el desarrollo de la teoría que nos interesa).

2.1 Un modelo para el magnetismo

El ejemplo mas característico de lo anteriormente expuesto es el fenómeno conocido como *ferromagnetismo*. Si tomamos un imán permanente y lo calentamos, a cierta temperatura pierde su magnetización permanente. Decimos entonces que el material se encuentra en una fase paramagnética. Si ahora comenzamos a disminuir su temperatura, para un valor muy preciso de la misma (conocido como *temperatura crítica*), el material se magnetiza espontaneamente y permanece así para toda temperatura inferior a la crítica. Decimos entonces que el material se encuentra en una fase ferromagnética. Este tipo de fenómenos se conocen como *transiciones de fase* o *fenómenos críticos* y son completamente análogos a la transformación líquido-vapor en un fluido (por ej., agua) al variar la temperatura.

Desde el punto de vista del modelado, la física microscópica de estos sistemas resulta extremadamente compleja. Estos materiales estan compuestos típicamente de aproximademente 10^{23} átomos ordenados en una estructura regular (cristal). En términos generales las propiedades magnéticas se originan en los electrones de las capas electrónicas incompletas de cada átomo. Cada electrón presenta un momento magnético conocido como spin, de tal manera que puede pensárselo como un imán elemental. El fenómeno macroscópico del ferromagnetismo surge de las fuerzas combinadas de cada electrón dentro del propio átomo junto con las fuerzas producidas por los restantes electrones del material. De esta manera, el estudio de este problema a partir de las ecuaciones de la física atómica involucraría la resolución de como mínimo 10^{23} ecuaciones diferenciales acopladas.

Este fenómeno pudo ser comprendido en profundidad a partir de un modelo ultrasimplificado, el cual puede justificarse a partir de dos observaciones fundamentales:

- En presencia de un pequeño campo magnético externo los imanes elementales asumen en su mayoría sólo dos orientaciones: en la dirección del campo o contraria a él.
- La fuerza efectiva entre dos imanes elementales cercanos tiende a orientarlos paralelos.

El modelo en cuestión, conocido como *modelo de Ising* [3], consiste en reemplazar cada átomo i del cristal por una variable S_i, la cual puede asumir solo dos valores ± 1, simulando las dos orientaciones del imán elemental. Se asigna ademas una energía al sistema, de tal manera que la misma aumenta en una cantidad constante por cada par de imanes que apunten en direcciones opuestas entre sí. Considerando entonces una gran cantidad de estas unidades simples, y aplicando los métodos de la física estadística pueden calcularse en principio todas las funciones termodinámicas de este sistema ficticio y contrastarse con las curvas experimentales. A pesar de su simplicidad, este continúa siendo un problema matemático formidable debido a la gran cantidad de unidades elementales a considerar. No obstante, el mismo pudo resolverse analíticamente, encontrándose que las curvas teóricas no solo describían cualitativamente el fenómeno del ferromagnetismo, sino que graficadas adecuadamente, las mismas reproducían con alta precisión las curvas experimentales de una gran variedad de materiales con diferentes estructuras en las cercanías de la temperatura crítica. Esta característica de que sistemas muy dispares en sus detalles microscópicos presenten el mismo comportamiento macroscópico bajo ciertas circunstancias, dependiendo sólo de unas pocas propiedades elementales en común, se conoce en física como *universalidad* y es una propiedad bastante frecuente en los sistemas complejos.

La fenomenología de estos sistemas es mucho mas rica de lo aquí descripto, pero no nos extenderemos en mas detalles. Solo vamos añadir que este fenómeno es uno de los ejemplos mas típicos de los que se conocen como *fenómenos cooperativos*, esto es, dado un conjunto de unidades funcionales muy simples que interactúan entre sí, es posible que las mismas muestren un comportamiento colectivo nuevo, complejo, si *el número de unidades ensambladas es suficientemente grande*, tal como el ordenamiento espontáneo de los imanes elementales. Una muestra de unos pocos átomos magnéticos no se magnetiza a ninguna temperatura. Así, en la suma de los efectos de una gran cantidad de unidades solo los aspectos funcionales mas básicos son los que sobreviven, dando lugar a la universalidad del comportamiento. Con respecto al ejemplo anterior, en lo que

se relaciona con el comportamiento colectivo, tanto da tener átomos complejos o variables de dos estados: el comportamiento macroscópico para temperaturas cercanas a la crítica es exactamente el mismo en ambos.

2.2 Fenómenos cognitivos

Un segundo ejemplo, que presenta muchas analogías con el anterior, es precisamente el de las redes neuronales (modelos simplificados) y su capacidad de reproducir fenómenos cognitivos básicos como los observados en los seres provistos de un sistema neurológico (sistema por lo general altamente complejo). Las redes neuronales surgen históricamente a partir de la informática, como un intento de reproducir ciertas capacidades cognitivas del cerebro mediante computadoras. Existen una serie de capacidades propias del cerebro sumamente difíciles de reproducir mediante una computadora secuencial. Así, por ejemplo, el reconocimiento visual de lenguaje escrito manuscrito, requiere en principio de una capacidad de almacenamiento de memoria infinita, de manera de guardar todas las posibles formas de las letras. Por otra parte, la recuperación de este tipo de información mediante una búsqueda secuencial sería sumamente lenta. Sin embargo, el cerebro humano, luego de un período de aprendizaje, es capaz de realizar esta tarea de manera inmediata y, obviamente, con una capacidad de almacenamiento finita. Esto indica un mecanismo de almacenamiento y recuperación de información de una naturaleza completamente diferente al de las computadoras, ya que uno no memoriza todas las posibles formas de las letras, sino que de alguna manera se guarda información acerca de la *estructura* del patrón visual (grafía), y el mismo se recupera por asociación entre patrones visuales semejantes. Este tipo de memoria se denomina **asociativa** o **direccionable por contenido**.

Por otra parte, existen tareas que una computadora secuencial realiza de manera mucho mas eficiente que un cerebro, tal como el cálculo de una raiz cuadrada. Vemos así, que los mecanismos internos de funcionamiento de un cerebro y una computadora tradicional deberían ser en principio completamente diferentes.

3 Neuronas formales

Comencemos entonces por intentar modelar las unidades funcionales básicas del sistema nervioso, es decir, las neuronas [4].

Morfológicamente una neurona se diferencia principalmente de otras células en que la membrana celular presenta un conjunto de extensiones en forma ramificada conocidas como *dendritas* y una

única extensión de forma tubular o fibra conocida como *axón*, el cual se ramifica en su extremo; estas últimas ramificaciones se conectan a las dendritas de otras neuronas, siendo estas conexiones conocidas como *synápsis*.

Una neurona posee la capacidad de emitir un impulso eléctrico (estrictamente este mecanismo es electroquímico) el cual se propaga a lo largo del axón, transmitiéndose a otras neuronas a traves de las synapsis con este último. En ausencia de un número importante de estímulos eléctricos aferentes, la membrana se encuentra polarizada eléctricamente, estando el citoplasma (medio interno) cargado negativamente respecto del líquido neural (medio externo), presentando una diferencia de potencial aproximadamente constante de -70 mV. En este caso, decimos que la neurona se encuentra en reposo. La llegada mas o menos simultánea de un cierto número de estímulos electricos puede producir una despolarización local de la membrana, con la consiguiente caída en la diferencia de potencial. Si esta caída supera un cierto valor umbral de aproximadamente 10 mV, se produce una despolarización global que se propaga a lo largo de toda la membrana alcanzando el axón; esta onda de despolarización constituye el impulso eléctrico antes mencionado y se denomina *potencial de acción*. Luego de este evento, la membrana resulta completamente despolarizada y, pasado cierto intervalo de tiempo conocido como *período refractario*, la polarización de membrana se reconstituye. Vemos entonces que una neurona actúa como un dispositivo integrador, "sumando" los impulsos recibidos y emitiendo o no un potencial de acción, según esta suma supere o no un valor umbral.

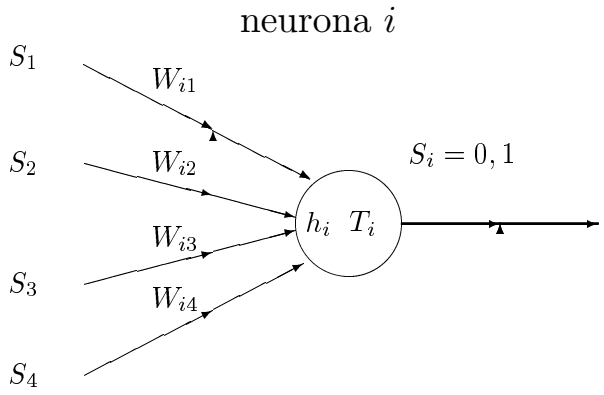

Figura 1: Representación esquemática de una neurona formal

Un aspecto importante a mencionar, es que cada synapsis transmite con diferente intensidad el impulso recibido, pudiendo esta ser de tipo *excitatoria* (disminuye el potencial de membrana) ó *inhibitoria* (aumenta el potencial de membrana hiperpolarizándola).

La característica mas importante, desde el punto de vista funcional, es que las neuronas son dispositivos de dos estados, esto es, en cada instante de tiempo la neurona puede estar inactiva (en reposo) o emitiendo una señal (potencial de acción). Las otras características sobresalientes desde el punto de vista estructural son el alto número de unidades y la alta conectividad entre las mismas.

El primer modelo matemático de una neurona fue introducido por McCulloch y Pitts[5] en 1943, mediante lo que se denominó la **neurona formal**. En este modelo se simula cada neurona mediante una variable que puede asumir en cada instante solo dos valores: 0 y 1, representando los estados desactivado y activado (es decir, emitiendo un potencial de acción) respectivamente. Cada una de estas unidades asume su estado de activación S_i a partir del cálculo de una cantidad h_i, la cual simula un potencial de membrana efectivo como una suma pesada de los estados de activación de las neuronas que realizan synapsis aferentes con esta (ver figura 1):

$$h_i = \sum_{j=1}^{4} W_{ij} S_j$$

donde las variables W_{ij} se denominan **eficacias** o **intensidades sinápticas**. Estas variables pueden tomar valores positivos o negativos, representando synapsis excitatorias o inhibitorias respectivamente.

El mecanismo de funcionamiento de estas unidades es el siguiente: en un dado instante de tiempo t (en alguna escala de eventos discretos) la neurona i computa h_i a partir del estado de activación en el tiempo $t-1$ de todas las neuronas que se conectan aferentemente a la misma. Si h_i supera un cierto valor umbral T_i la neurona se activa, es decir $S_i(t) = 1$ y $S_i(t) = 0$ en caso contrario:

$$S_i = \begin{cases} 1 & si \quad h_i \geq T_i \\ 0 & si \quad h_i < T_i \end{cases}$$

esto es, la función de activación viene dada por un escalón, como se muestra en la figura 2.

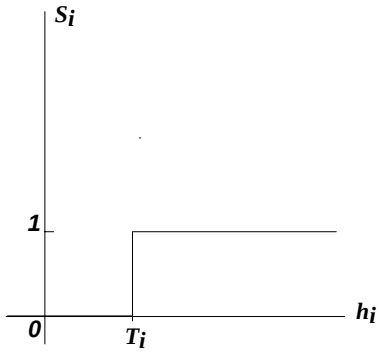

Figura 2: Función de activación en el modelo e McCulloch & Pitts de una neurona formal

El objetivo de McCulloch y Pitts fue mostrar que con dispositivos tan simples como estos, los cuales guardan una analogía a nivel sumamente básico con las neuronas biológicas, era posible construir estructuras que llevaran a cabo operaciones lógicas simples. Sin embargo, en años posteriores se mostró que es posible asociar este modelo a una descripción ultrasimplificada de la electrodinámica de una neurona, mediante una interpretación diferente. Dicha descripción se basa en la observación de que la frecuencia promedio de pulsos emitidos a lo largo del axon V_i, depende de la corriente eléctrica media de entrada a través de la membrana celular, en la forma que se muestra en la figura 3, variando entre cero y una frecuencia máxima de saturación [6].

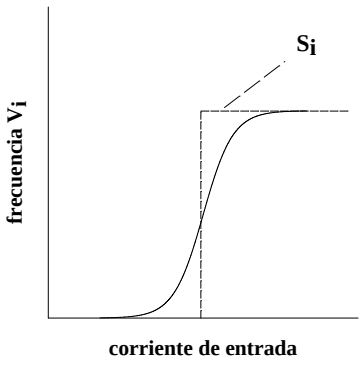

Figura 3: Frecuencia promedio de pulsos emitidos en el axon en función de la corriente promedio de entrada

La corriente de entrada promedio por su parte, proviene fundamentalmente de otras neuronas que hacen sinápsis con ella a través del arbol dendrítico. Así, despreciando otros efectos, la misma puede escribirse como

$$\text{corriente de entrada} = \sum_j W_{ij} V_j .$$

donde ahora la variable W_{ij} es una conductividad efectiva o promedio de la sinápsis entre las neurona i y j. Notemos que la función respuesta de la figura es no lineal, ya que la misma presenta saturación.

Surge entonces la hipótesis de que es esta no-linealidad (es decir, la forma de la función respuesta) la característica mas importante en lo que respecta al manejo de información, tal como era la propiedad de presentar dos orientaciones principales los imanes elementales en el sistema magnético. Asi, se simplifica aún mas el modelo reemplazando la función respuesta por la mas simple posible con forma semejante: un escalón (figura 2).

Podemos entonces reinterpretar la neurona formal de McCulloch y Pitts. Las variables S_i representan entonces las actividades o frecuencias promedio, donde los estados 0 y 1 corresponden a la mínima y máxima frecuencias posibles, en una escala adecuada. La cantidad h_i representa el potencial de membrana promedio, el cual es proporcional a la corriente efectiva a través de la misma. Mas aún, es posible mostrar a traves de simulaciones numéricas que este reemplazo de variables continuas por variables de dos estados no altera cualitativamente el comportamiento de los modelos que se describen a continuación.

4 Perceptrons

El tipo mas simple de estructuras que podemos formar con neuronas formales se conoce como **perceptron simple** [1, 2] y consiste en un conjunto de neuronas dispuestas en dos capas, una de entrada y una de salida como se muestra en la figura 4, donde la información fluye en un solo sentido, de la entrada a la salida. En otras palabras, las neuronas de la capa de salida sólo reciben estímulos aferentes de las correspondientes a la capa de entrada a traves de un conjunto de sinapsis dado. Podemos pensar entonces en dos instantes de tiempo distintos: en el primero las neuronas de la capa de entrada asumen un valor de activación a partir de algun estímulo externo; en el instante siguiente las neuronas de salida computan cada una su potencial post-sináptico a partir de las anteriores y asumen su estado de activación de acuerdo a la regla antes explicada, emitiendo la respuesta al estímulo (ver figura 4).

Ahora bien, tanto la entrada como la salida son conjuntos de ceros y unos. Mediante estos conjuntos puede construirse un *alfabeto* y codificar cualquier tipo de información. Este sistema de codificación se conoce como *binario* y es el utilizado internamente por las computadoras digitales usuales. Así, por ejemplo, si tomamos una ristra grande de ceros y unos podriamos tomarlos de a cinco y codificar números:

00000	00001	00010	00011
0	1	2	3

etc, o bien letras:

10000	10001	10010	10011
A	B	C	D

etc. Así, podemos pensar en estas estructuras como procesadores de información, que producen **asociaciones** entre un cierto conjunto de datos y otro. Por ejemplo, podríamos codificar en la entrada imágenes en blanco y negro, ordenando las neuronas de entrada en una capa bidimensional de tal manera que cada una corresponde a un grano de una foto (o un pixel de una pantalla de computadora) y que la misma toma el valor cero si en la imagen el grano está negro y uno si está blanco. Por otra parte podríamos codificar en la salida letras conformando palabras que describan conceptos asociados con las imágenes. Así, podríamos ingresar una foto de una persona y querer que en la salida aparezca el nombre de esa persona.

Otro ejemplo sería el de suma de números naturales. Codificamos en entrada dos numeros naturales, digamos entre cero y 15 (hacen falta 8 neuronas) y en la salida tambien números entre

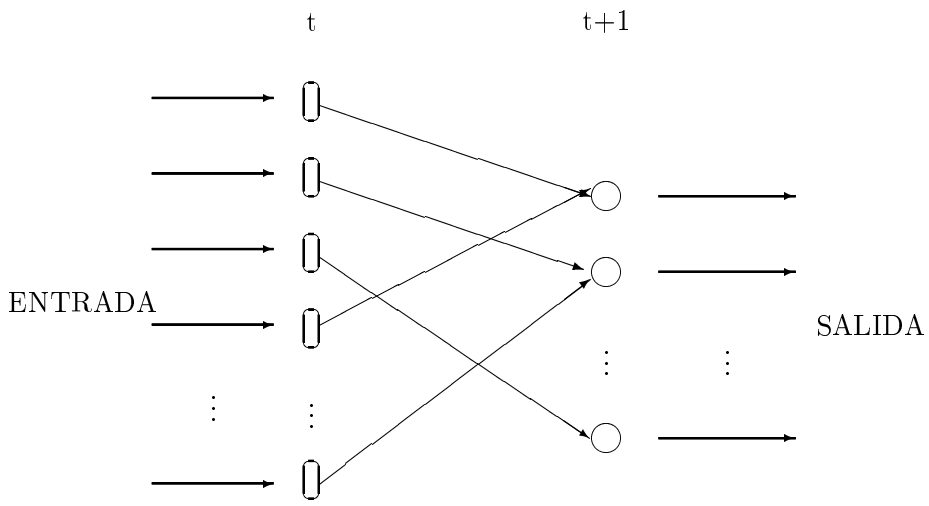

Figura 4: Perceptron simple

cero y 30 con 5 neuronas, y queremos que la salida nos dé el resultado de la suma para cualquier par de números.

Este tipo de estructuras fue introducido en 1962 por F. Rosenblatt [7], el cual las bautizo con el nombre sugestivo de **perceptrons**. Desde el punto de vista computacional, estos dispositivos resultan en principio sumamente poderosos, ya que procesan la información *en paralelo*, es decir, muchas operaciones elementales simultaneamente, a diferencia de las computadoras usuales que funcionan de manera *secuencial*, siendo en consecuencia mucho mas lentas. Como veremos mas adelante, este no es el único aspecto importante de estas redes.

Surge entonces de manera natural la siguiente pregunta: ¿ son capaces estas redes de llevar a cabo cualquier tipo de tarea u asociación? En otras palabras, dado un cierto conjunto de asociaciones entre conjuntos de datos codificables en forma binaria, es posible encontrar un conjunto de intensidades sinápticas y umbrales de activación, tales que la red produce en forma exacta *todas* las asociaciones?

Para el caso de la red antes presentada (*perceptron simple*) la respuesta es **no**. Existe una cierta categoría de asociaciones que resultan imposibles de realizar con esta red [8]. Un ejemplo es precisamente el problema de suma de dos números naturales [9]. No obstante, puede verse que cualquier tipo de problema puede resolverse mediante una generalización de la estructura anterior, conocida como perceptron *multicapa* [10], tal como la que se muestra en la figura 5.

Para cada problema en particular es necesario determinar el número mínimo de capas intermedias necesarias para resolverlo. Así, por ejemplo, en el caso de la suma de dos números naturales es necesario como mínimo una capa intermedia. En el caso de la multiplicación de dos números son necesarias como mínimo dos capas intermedias [11]. No obstante, aun sabiendo que un dado problema (esto es, un dado conjunto de asociaciones) es factible de ser resuelto por una red neuronal, son contados los casos en los cuales se conocen de manera exacta los valores de las intensidades sinápticas $\{W_i\}$ que efectivamente permiten realizar dicha tarea. Es aquí donde entra en consideración el aspecto mas interesante de estos dispositivos y es que los mismos son capaces de **aprender** a realizar una determinada tarea, como por ejemplo sumar dos números.

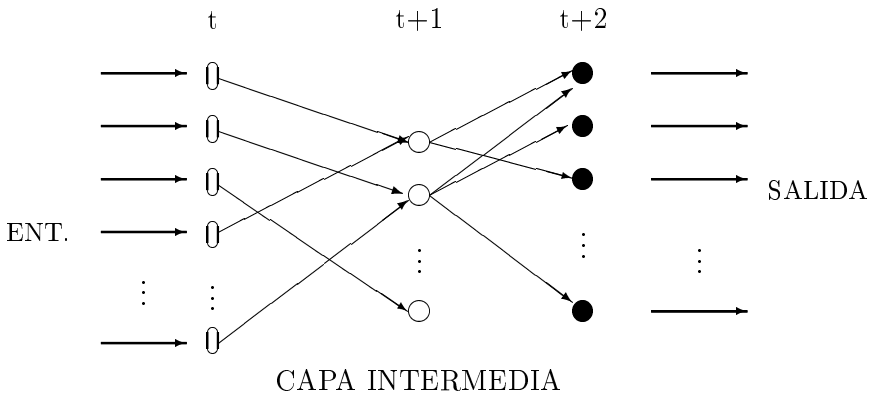

Figura 5: Perceptron multicapa

decimal	binario	S
1	0001	1
2	0010	0
3	0011	1
4	0100	0
5	0101	1

Tabla 1: Ejemplo de función de paridad para una entrada de 5 dígitos binarios

5 Aprendizaje

Todas las teorías de aprendizaje en redes neuronales se basan en alguna medida en la hipótesis de Donald Hebb (1949) [12], según la cual **el aprendizaje tiene lugar mediante modificaciones sinápticas que reflejan las actividades de las neuronas pre y post sinápticas bajo la influencia del estímulo a ser aprendido**.

5.1 Algoritmos

Los modelos de aprendizaje se concretan mediante algoritmos de ajuste de las intensidades sinápticas W_{ij}. Entendemos por algoritmo un cierto conjunto estructurado y secuencial de pasos de cálculo que se repite hasta alcanzar el efecto deseado. Esto algoritmos pueden ser programados en una computadora de manera de simular el proceso automático de modificación sináptica durante el aprendizaje, y se denominan genéricamente **reglas de aprendizaje**.

Tomemos para fijar ideas un ejemplo simple: un perceptron con sólo dos capas (entrada y salida), con un cierto número de neuronas en la capa de entrada que codificaran números naturales (en base binaria) y una única neurona en la capa de salida (ver figura 6), con la cual queremos determinar la paridad del número entrado, es decir, que tome el valor cero si la entrada es un número par y asume el valor uno si es impar (ver tabla 1).

El problema es entonces determinar el conjunto de valores que deben tomar las eficacias sinápticas $\{W_i\}$ con $i = 1, 2, \ldots, N$ para que el perceptron compute correctamente todos los valores de la tabla 1. La regla de aprendizaje mas simple para calcular dichos valores se conoce como *regla del perceptron* [2], y consiste en el siguiente algoritmo:

Partimos de un conjunto de intensidades sinápticas $\{W_i\}$ **arbitrarias**.

1. Se toma un ejemplo particular (esto es, una fila de la tabla 1). Dada la entrada del mismo, en general la salida que se obtendrá no coincidirá con la deseada.

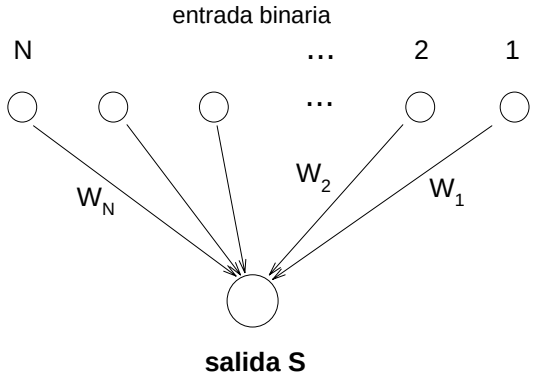

Figura 6:

2. Se toma una neurona de entrada i y se ajusta la sinápis correspondiente W_i de acuerdo con la siguiente regla:

   (a) Si la salida actual coincide con la deseada, W_i se deja inalterada.

   (b) Si la entrada $S_i = 1$ en el ejemplo y la salida actual **no coincide** con la deseada, W_i se modifica de acuerdo con:

entrada S_i	S actual	S deseado	modificación
1	0	1	$W_i \rightarrow W_i + \epsilon$
1	1	0	$W_i \rightarrow W_i - \epsilon$

   donde ϵ es algun número positivo **pequeño**.

   (c) Si la entrada $S_i = 0$, W_i no se modifica

3. Se repiten todos los pasos anteriores para todas las neuronas de entrada con el mismo ejemplo.

4. Se repite todo el proceso anterior para cada uno de los ejemplos.

El conjunto completo de todos estos pasos constituye una **iteración** del algoritmo. Una vez concluido este proceso se vuelven a comparar las salidas para las nuevas intensidades $\{W_i\}$ con las deseadas. Si las mismas coinciden **para todos los ejemplos** se termina el proceso y decimos que el algoritmo convergió o que la red aprendió. En caso contrario se repiten todos los pasos anteriores hasta completar una nueva iteración.

Ahora bien, este proceso termina alguna vez? Para este algoritmo particular existe un teorema, el cual enuncia que, si la solución existe (es decir, si existe al menos un conjunto de valores $\{W_i\}$ para los cuales la red realiza todos los ejemplos) el proceso converge en un numero **finito** de iteraciones [2].

Este algoritmo se generaliza facilmente para el caso en que existen mas de una neurona de salida. Simplemente se repiten dentro una misma iteración todos los pasos para cada una de las neuronas de salida.

Existen muchos otros tipos de algoritmos de aprendizaje semejantes, los cuales permiten entrenar redes multicapa, de manera que puedan aprender cualquier tipo de tarea codificable en base a neuronas formales, para los cuales, si bien no existe un teorema de convergencia como en el caso anterior, puede mostrarse que convergen en un número finito de pasos a una solución con error acotado.

5.2 Aprendizaje supervisado vs. no-supervisado

Este tipo de procedimiento, en el cual se compara permanentemente la salida obtenida durante el mismo con la deseada se denomina **aprendizaje supervisado**, ya que presupone la existencia de un **"maestro"** que determina exactamente cuando la respuesta emitida por el **"alumno"** (esto es, la red) es la correcta o no.

Algoritmos de este tipo resultan bastante útiles desde el punto de vista computacional. Por otra parte, desde el punto de vista biológico son bastante ficticios, ya que resulta difícil discernir cual sería el equivalente se ese supuesto "maestro". Pero existen otros algoritmos de aprendizaje **no-supervisado** que resultan mas interesantes desde el punto de vista de la emulación de sistemas biológicos. Uno de ellos se conoce como método de *premio y penalización*. En este, en lugar de compararse de manera detallada si la salida es la correcta para cada neurona, se introduce una función *"placer" o "displacer"* que cuantifica el exito o error cometido por la red globalmente para todos los ejemplos. Esta funcion se define de tal manera que varia entre cero y uno, correspondiendo estos valores extremos a un fracaso o un exito total respectivamente. A "grosso modo", cada intensidad sináptica es entonces modificada o no, durante una iteración, con una probabilidad que depende de la función placer.

Estudios matemáticos acerca de este algoritmo en problemas particulares muestran que el mismo efectivamente converge a las respuestas correctas, de manera tal que en promedio la función placer tiende siempre a incrementarse. En este caso el tiempo de convergencia, es decir el número de iteraciones necesarias para que el proceso termine, es mayor que en el caso supervisado.

5.3 Generalización

Ahora bien, podría ser que el conjunto de asociaciones a aprender por la red sea demasido grande, de manera tal que el tiempo necesario para que la misma las aprenda todas resulte excesivamente largo. Así por ejemplo, en el problema de suma, el número total de ejemplos (sumas posibles de dos números) crece exponencialmente con la cantidad de dígitos de los números a sumar. Es aquí donde emerge la característica tal vez mas interesante de los perceptrons, que es su *capacidad de generalización* [13] . Es posible ver que con los algoritmos de aprendizaje descriptos, no es necesario enseñarle a red todas la asociaciones posibles entre los conjuntos de datos de entrada y salida; basta con enseñarle (en el sentido antes descripto) sólo un subconjunto mucho menor de asociaciones o ejemplos, conocido como conjunto de *entrenamiento*. Si este conjunto es suficientemente grande (pero siempre mucho menor que el completo) la red reproducirá correctamente asociaciones que *no le fueron enseñadas durante el proceso de aprendizaje*. Así, por ejemplo, en el problema de suma de dos números entre 0 y 15, en el cual el número total de sumas posibles es 256, basta con incluir en el conjunto de entrenamiento aproximadamente 120 sumas elegidas aleatoriamente (50%). Si ahora se ingresa en la entrada cualquier otra suma que no se le enseño la red, esta dará la salida correcta también. Si se aumenta el numero de digitos (y por lo tanto el número de neuronas de entrada y salida) este porcentaje disminuye. Para la suma de números entre 0 y 255 el número total de ejemplos es 64000 y en este caso el numero de ejemplos aleatorios es aproximadamente 6000, es decir un 10% del total posible de ejemplos.

Vemos entonces que la red es capaz de *"inducir"* a partir de un cierto número de ejemplos la *regla* de suma, comportamiento que en algún sentido puede ser catalogado de *"inteligente"*. Por otra parte, el mismo fue obtenido mediante un proceso completamente automático, en el cual en ningún momento se ha suministrado información explícita acerca de la regla de suma o sus propiedades (conmutatividad, asociatividad, existencia del elemento neutro, etc.) Mas aún, esto resulta también muy interesante ya que estas redes serían en principio capaces de resolver problemas para los cuales no se conocen todas las respuestas y sólo se conocen algunos casos particulares. Un ejemplo de esto último es la *predicción de series temporales complejas*. En estos casos se dispone de una cierta serie temporal de datos finita $x(t)$, como podrían ser los registros meteorológicos, de contaminación, etc., y se desea hacer pronósticos acerca de los valores futuros de la serie [14]. Este tipo de datos experimentales combinan una componente aleatoria muy fuerte con una alta no-linealidad, lo cual

torna el proceso de predicción sumamente dificultuoso. En este caso se entrena una red cuya entrada codifica el valor de la serie $x(t)$ y la salida $x(t+1)$. Los pares entrada-salida con los cuales se entrena la red son entonces los pares de valores sucesivos del conjunto conocido. Tests numéricos realizados con series temporales generadas artificialmente con funciones conocidas [1] (es decir, que se conoce la respuesta futura), así como con datos experimentales [14] muestran que efectivamente pueden realizarse predicciones correctas en algunos casos, con una eficiencia mayor que los métodos usuales de regresión lineal.

Un problema en este tipo de aplicaciones consiste en que el numero necesario de capas intermedias varía con la complejidad de la función utilizada, la cual en problemas reales es desconocida. Aplicaciones de este tipo se encuentran actualmente en estudio.

6 Memoria Asociativa

El fenómeno de memoria asociativa puede ser modelado tambien mediante perceptrons. No obstante, resulta mas interesante discutirlo en base a otro tipo de redes neuronales conocido como **"attractor neural networks" (ANN)** o redes con atractores [15], (la razon de este nombre se aclarará mas adelante) ya que las mismas permiten introducir un marco conceptual que posibilita una comprensión mas profunda del fenómeno. La estructura de este tipo de redes, por otra parte, se encuentra mas cercana a las observadas en la mayoría de los sistemas neurológicos.

A diferencia de los perceptrons, que poseen una estructura de capas y en las cuales los estímulos fluyen en una única dirección, en las nuevas redes la información puede fluir en cualquier sentido. Sea W_{ij} la intensidad sináptica entre la neurona presináptica i y la post-sináptica j (ver figura 7). En principio tanto W_{ij} como W_{ji} pueden ser diferentes de cero, es decir, en estos modelos la neurona i puede tanto enviar impulsos a otra neurona j como recibirlos de la misma, a diferencia de los preceptrons en los cuales $W_{ij} \neq 0$ implicaba que $W_{ji} = 0$ (información en un solo sentido).

En este caso la dinámica se modela de una manera un poco diferente. En una dado instante de tiempo inicial t se determina el estado de activación de todas las neuronas $S_i(t)$. Este estado inicial representa un estímulo, el cual puede provenir de otras neuronas sensorias, las cuales realizan solamente sinapsis aferentes con las del sistema en consideración. En el instante siguiente se computa el potencial post-sináptico $h_i(t+1)$ correspondiente a cada neurona i en función *del estado de activación de todas las neuronas restantes* en el *instante anterior*:

$$h_i(t+1) = \sum_j W_{ij} S_j(t)$$

donde las variables W_{ij} tienen el mismo significado de antes. Finalmente, cada neurona asume su nuevo valor de activación al tiempo $t+1$ de acuerdo a la misma regla que el perceptron y todo el ciclo vuelve a repetirse para el instante $t+2$.

Vemos que en este caso no existe diferenciación entre neuronas de entrada y salida. Todas las neuronas formales evolucionan permanentemente, y el estado de todas y cada una interviene en el proceso dinámico en cada instante de tiempo. El estado completo de la red, es decir, el conjunto de valores de activación de todas las neuronas, en un dado instante de tiempo puede representarse en un espacio abstracto, como se muestra en el ejemplo de la figura 8 para el caso de 3 neuronas: S_1, S_2 y S_3.

Vemos que todos los estados posibles de esta red se encuentran en los vértices de un cubo de lado unidad. Para el caso de N neuronas los estados se ubicarán en una generalización del mismo, el cual se denomina un *hipercubo* en N dimensiones, al cual llamaremos el *espacio de estados*. El estado de la red esta representado entonces por un punto en ese espacio discreto abstracto. Bajo la regla de activación antes explicada dicho punto se moverá en el espacio de estados describiendo una trayectoria. Este tipo de red constituye por lo tanto un *sistema dinámico*, donde las trayectorias del mismo estarán determinadas por el estado inicial y el conjunto de intensidades sinápticas y umbrales de activación. Dichas trayectorias alcanzarán al cabo de un cierto tiempo lo que se denomina un

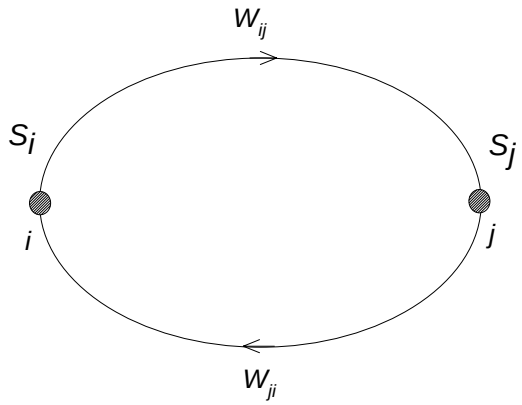

Figura 7:

atractor, esto es, un subconjunto de puntos en el espacio de estados al cual convergen trayectorias provenientes de diferentes estados iniciales lo suficientemente cercanos a algún punto del atractor. El atractor mas simple posible es un único punto en el espacio de estados, el cual se conoce como *punto fijo*. Una vez arribado el sistema a un punto fijo no evoluciona más, con lo cual este corresponde a un estado estacionario. Existen tambien atractores no estacionarios tales como *ciclos límites* en los cuales el sistema recorre periódicamente un conjunto *finito* de puntos, y los atractores *caóticos*, los cuales consisten de un subconjunto infinito de estados (y por lo tanto solo son posibles en un espacio de estados infinito) en los cuales el sistema nunca repite exactamente un mismo estado en su evolución. Si bien todos estos tipos de atractores son de interés en la teoría de las redes neuronales [16], nos concentraremos solamente en el caso de puntos fijos.

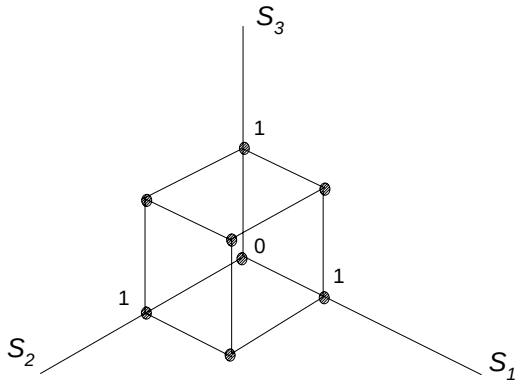

Figura 8: Espacio de estados para 3 neuronas. Los puntos negros indican los únicos estados posibles de la red.

El proceso de manejo de información completo puede entonces esquematizarse como se muestra en la figura 9. Por un lado tenemos un cierto conjunto de neuronas de entrada (las cuales no forman parte del ANN) las cuales envían un estímulo a las del ANN, sin recibir estímulos del mismo. Estas son las que determinan el estado inicial de la red. Por otra parte tenemos un conjunto de neuronas de salida o lectoras, las cuales reciben estímulos de parte del ANN, pero no los reenvían al mismo. Estas tendrían por función monitorear el estado de la red, respondiendo de manera diferente de acuerdo al atractor en el cual se encuentra el ANN y enviando en consecuencia una señal a otro dispositivo para un posterior procesamiento o bien a un sistema motor. En lo que sigue nos vamos a ocupar solamente del ANN, es decir, de la relación entre el estado inicial y el atractor asociado o estado final.

El atractor mas simple posible es un **punto fijo**, es decir un estado en el cual el conjunto de

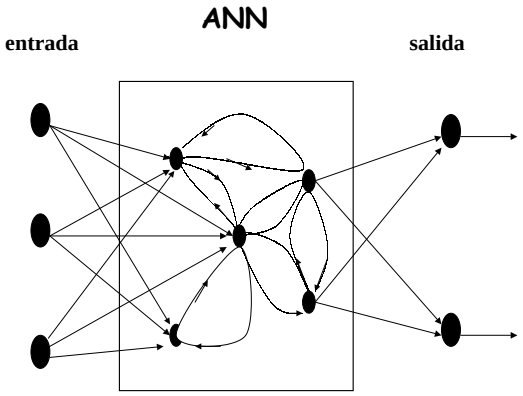

Figura 9: Manejo esquemático de la información en una red con atractores (ANN).

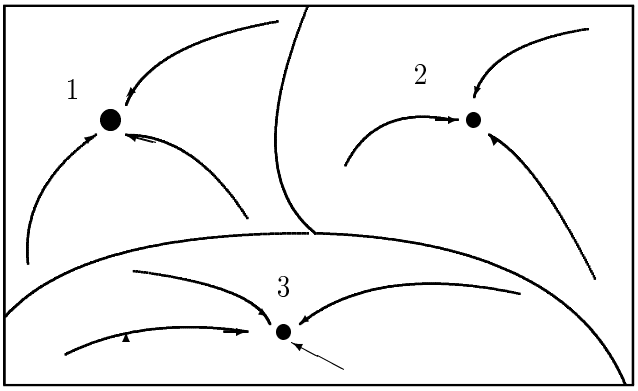

Figura 10: Representación esquemática de las cuencas de atracción asociadas a tres atractores diferentes, representados por los puntos negros 1,2 y 3; las lineas continuas constituyen las fronteras de las cuencas.

actividades de la red no cambia en el tiempo. Si tenemos un conjunto de puntos fijos diferentes, cada uno de ellos posee una **cuenca de atracción**, que se define como la región del espacio de estados cuyos puntos son llevados a traves de la dinámica a dicho punto fijo, tal como se muestra esquemáticamente en la figura 10 para el ejemplo de 3 puntos fijos.

Ahora bien, recordemos que el espacio de estados esta constituido por los vértices de un hipercubo, o en otras palabras, por un conjunto de ceros y unos asociados al estado de activación de cada una de las neuronas que componen la red. Es posible ver que, bajo la acción de la dinámica de activación que hemos presentado, la cuenca de atracción de un punto fijo se encuentra constituida por puntos *cercanos* al punto fijo, donde la distancia al punto se mide en base al *número de neuronas que difieren en su estado de activación*, la cual se conoce técnicamente como *distancia de Hamming*. Pero esto es justamente lo que precisamos para un modelo de memoria asociativa, ya que puntos cercanos en el espacio de estados contendran información semejante al diferir solamente en el estado de unas pocas neuronas. Supongamos para fijar ideas que la red codifica imágenes visuales, es decir, el estado de activación cada neurona indica si un punto de la imágen bidimensional es blanco o negro. Supongamos ademas que queremos almacenar el patrón visual de una letra del alfabeto, por ejemplo el caractér romano para la letra A, y asociar con el mismo letras parecidas que resultan de una deformación del original, por ejemplo, debido a un trazo manuscrito irregular. Cuanto mas parecidas sean al patrón visual original menor sera su distancia de Hamming al mismo.

El problema que se plantea entonces es el siguiente: ¿ existe un conjunto de valores para las intensidades sinápticas, tales que el patron visual original, al cual llamaremos una **memoria**, constituya un punto fijo de la dinámica de la red?

Este problema fue resuelto en los años 80 por John Hopfield [17] y otros [18], quienes introdujeron el modelo (conocido luego como modelo de Hopfield) que se describe a continuación [2].

Antes de continuar conviene introducir el siguiente cambio de variables el cual, si bien es un tecnicismo que no altera el contenido del modelo, simplifica su discusión. Para cada neurona i introducimos una nueva variable de activación σ_i de la siguiente manera:

$$\sigma_i = 2S_i - 1$$

con lo cual si $S_i = 0 \Rightarrow \sigma_i = -1$ y si $S_i = 1 \Rightarrow \sigma_i = 1$. De esta manera el estado $\sigma_i = -1$ representa una frecuencia de activación baja y el $\sigma_i = 1$ continua teniendo el mismo significado de antes. El potencial post-sináptico se calcula ahora como

$$h_i(t+1) = \sum_{j=1}^{4} W_{ij}\sigma_j(t)$$

y el estado de activación de la neurona i al tiempo $t+1$ estará determinado por una función de la forma que se muestra en la figura 11.

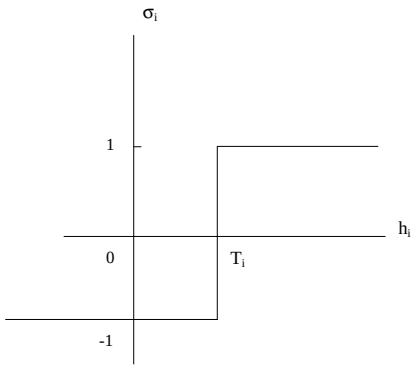

Figura 11: Función de activación para neuronas $\sigma_i = \pm 1$.

La receta para la construcción de las intensidades sinápticas, conocida técnicamente como **regla de Hebb**, es la siguiente:

- Supongamos primero que queremos almacenar sólo una memoria:

  $$(\sigma_1, \sigma_2, \ldots, \sigma_N) = (\xi_1, \xi_2, \ldots, \xi_N)$$

  la cual constituye un conjunto particular de unos y menos unos. En este caso las intensidades sinápticas entre cada par de neuronas i y j se calcula según

  $$W_{ij} = \xi_i \xi_j$$

  Notemos que la sinápsis será exitatoria (+1) o inhibitoria (-1) de acuerdo al estado de **ambas** neuronas i y j en la memoria o patrón, el cual se supone aprendido en alguna etapa previa.

- Supongamos ahora dos memorias

$$(\sigma_1, \sigma_2, \ldots, \sigma_N) = \begin{cases} (\xi_1^1, \xi_2^1, \ldots, \xi_N^1) & (1) \\ (\xi_1^2, \xi_2^2, \ldots, \xi_N^2) & (2) \end{cases}$$

Cada memoria constituye un conjunto particular de unos y menos unos. En este caso

$$W_{ij} = \xi_i^1 \xi_j^1 + \xi_i^2 \xi_j^2$$

y asi sucesivamente al aumentar el número de memorias.

Notemos que con esta prescripción las intensidades sinápticas resultan simétricas $W_{ij} = W_{ji}$. Esta hipótesis, altemente no realista, tiene como objeto simplificar el análisis matemático subsecuente y es posible mostrar que la misma no es fundamental. Esto es, los resultados que se obtienen no se alteran cualitativamente si se levanta esta restricción, por ejemplo, cortando aleatoriamente un cierto número de sinápsis. En particular el modelo asi formulado, con variables que toman valores ± 1 y con sinápsis simetricas resulta *enteramente análogo al modelo magnético mencionado al principio*. Fue esta analogía formal lo que posibilitó a Hopfield y a otros autores posteriores el análisis completo de la dinámica del modelo en base a todo el conocimiento previo de la física de los sistemas magnéticos.

Hopfield mostró que en este caso existe una *función energía* asociada al espacio de estados que depende del conjunto de intensidades sinápticas W_{ij}, la cual determina la dinámica del sistema y que en la analogía con los sistemas magnéticos es precisamente la energía magnética. Así, cada estado neuronal tiene asociado un valor de energía, de manera que la evolución del sistema en el tiempo es tal que el mismo siempre disminuye su energía. Esta evolución puede visualizarse a traves del siguiente esquema.

Supongamos que representamos esquemáticamente el espacio de estados en un plano, de tal manera que cada punto del plano representa un estado colectivo de la red. La función energía puede representarse entonces como una superficie en el espacio tridimensional, con valles y montañas. La proyección en el plano horizontal (piso) de un punto sobre esta superfice corresponde entonces al estado de la red, en tanto que su altura da el valor de la energía. La dinámica del sistema entonces es análoga al movimiento de una bolita que cae bajo la acción de la gravedad por las laderas de estas montañas, pero la cual posee rozamiento de manera que al caer pierde energía y al llegar al fondo de un valle se detiene. De esta manera, los mínimos de la energía o fondos de los valles constituyen los *atractores* de la dinámica, y sus respectivas cuencas de atracción estan determinadas por las laderas de las montañas que rodean a los valles, tal como se muestra en la figura 12.

Hopfield mostró tambien que los mínimos mas profundos (los cuales puede verse que poseen además las mayores cuencas de atracción) corresponden justamente a las **memorias almacenadas segun la regla de Hebb**. Todos los patrones cuya distancia de Hamming no difiera demasiado de la memoria poseerán una energía mayor y por lo tanto se localizarán en las laderas adyacentes, encontrandose por lo tanto en la cuenca de atracción de la memoria.

En la figura 13 se muestra un ejemplo de recuperación asociativa para una red tipo Hopfield con patrones visuales en blanco y negro de 130×180 puntos, en el cual se han almacenado 8 fotos diferentes. La figura muestra los estados inicial, uno intermedio y el final para la recuperación de tres de los patrones, en los cuales se ha suministrado información borrosa en el primero y parcial para los otros dos.

La analogía con los sistemas magnéticos permitió ademas obtener una serie de resultados importantes acerca de diversas propiedades de este modelo, así como diversas generalizaciones del mismo que permitieron estudiar otros fenómenos relacionados. Citemos a título de ejemplo el cálculo de la *capacidad máxima de almacenamiento* de una red con N neuronas. Tal como es de esperar no

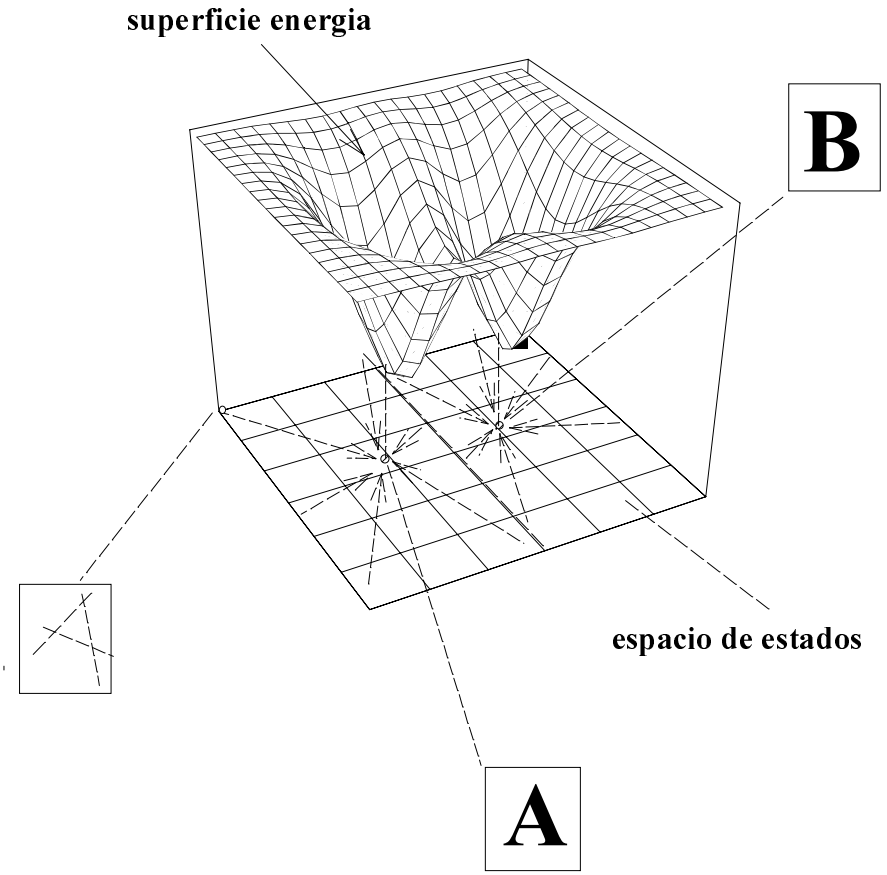

Figura 12: Representación esquemática de la superficie energía asociada al almacenamiento de dos memorias, en este caso, los caracteres romanos A y B.

Figura 13: Recuperación de memoria asociativa en una red tipo Hopfield con ocho fotos almacenadas [1]. Las columnas representan (de izquierda a derecha) los estados inicial, uno intermedio y el final (atractor-memoria). Las entradas corresponden a una imágen borrosa (araña), una cortada por la mitad (botellas) y una en la cual se presenta un cuarto de la foto original (perro).

podemos almacenar memorias indefinidamente sin que se altere la capacidad de recuperar adecuadamente la información. Para el modelo de Hopfield el número máximo de memorias es de un 14% del total de neuronas. Para otros modelos posteriores este porcentaje mejora substancialmente. Así, si tenemos una red de 100 neuronas podremos almacenar hasta 14 memorias sin problemas. Que ocurre si almacenamos 15? En el caso del modelo de Hopfield se demostró que ocurre una catástrofe, ya que la capacidad de recuperación desaparece completamente al superar la capacidad máxima. Esta falencia es debida obviamente a la extrema simplicidad del modelo. Posteriores sofisticaciones del modelo permitieron superar esta falencia. Existe, por ejemplo, un modelo en el cual las memorias en lugar de almacenarse todas simultáneamente mediante la regla de Hebb, se almacenan *en orden cronológico* mediante otra regla de cálculo de las intensidades sinápticas [19]. En este caso, las memorias mas antiguas van perdiendo su capacidad de recuperación, ya que sus cuencas de atracción se van haciendo cada vez mas pequeñas. A su vez, la calidad de la recuperación se deteriora, ya que el atractor correspondiente no se encuentra ya exactamente en la memoria, sino en las cercanías de la misma. Por otra parte, las nuevas memorias poseen las mayores cuencas de atracción y su calidad de recuperación es la mejor. De esta manera, la red nunca pierde su capacidad de memorizar nuevas informaciones, sino que "los viejos recuerdos van dando lugar gradualmente a los nuevos".

Si bien la información en el modelo de Hopfield no fue obtenida a traves de un proceso de aprendizaje como en el caso de los perceptrones, es posible combinar ambos procedimientos y calcular las intensidades sinápticas en los modelos tipo Hopfield mediante algoritmos de aprendizaje supervisado análogos a los anteriores. En este caso lo que se aprende no es un conjunto de asociaciones o pares de preguntas-respuestas, sino directamente el patrón a memorizar, de tal manera que las cuencas de atracción que producen las asociaciones se forman automáticamente [20, 21].

7 Comentarios finales

Un aspecto importante a resaltar es que en todos los modelos presentados el manejo de la información se lleva a cabo a traves de *estados colectivos de la red* (atractores en el caso de las ANN y asociaciones entrada-salida en el caso de los perceptrones). Esto es, la información no se encuentra almacenada de manera detallada y particionada en cada neurona (es decir, cada neurona no almacena un bit de información) sino que cada concepto (memoria u asociación) se encuentra almacenada *simultaneamente* en el *conjunto completo de las intensidades sinápticas*, actuando las neuronas puramente como procesadores de dicha información. Este manejo de la información se conoce como *memoria distribuida* y fue una de los mayores aportes conceptuales de la teoría de las redes neuronales a la neurofisiología, ya que permitió orientar los estudios sobre memoria en una nueva dirección.

El hecho de que la información se encuentre almacenada en forma distribuida tiene algunas consecuencias sumamente importantes. Por un lado, el almacenamiento de memoria (y su consiguiente recuperación) resulta "robusto" frente a eventuales "daños físicos" del sistema. Así por ejemplo, si en el modelo de Hopfield eliminamos algunas neuronas (o sinapsis) al azar (no importa cuales) *a posteriori* de haber almacenado un conjunto de memorias, los atractores resultan cualitativamente inalterados. Esto es, solamente se observará un leve deterioro en la calidad del patrón recuperado (en el caso de una imágen se espera que la reproducción sea mas borrosa) pero el mismo continuará siendo claramente identificable, en tanto el daño no sea masivo.

Por otra parte, una estructura de memoria distribuida resulta consistente con la velocidad de procesamiento de información observada en los sistemas neurológicos, ya que involucra un procesamiento *masivamente paralelo* de la misma. Esto es, grupos de neuronas procesan *simultaneamente* un estímulo, de tal manera que la emisión de una respuesta (o el arribo a un atractor) involucran un intervalo de tiempo pequeño. Esto permite entender la enorme eficiencia de los sistemas neurológicos en funciones tales como la recuperación asociativa de una memoria, en comparación con una computadora digital basada en el procesamiento *secuencial* de un conjunto de instrucciones, ya que las neuronas son dispositivos básicamente *lentos*: la velocidad característica de procesamiento de una neurona es aproximadamente unas 100.000 veces *menor* que la de una memoria RAM de una computadora personal corriente de las actuales.

Todos los modelos presentados aqui pueden ser (y de hecho lo han sido en gran medida hasta el presente) perfeccionados de diversas maneras a fin de, ya sea obtener un manejo mas eficiente de la información (el cual es el interés principal en el área de la informática) o bien mediante la incorporación de ingredientes que permitan aumentar su semejanza con los sistemas biológicos. El principal merito de estos modelos iniciales consistió en que, dada su simplicidad, estos permitieron la comprension de ciertos aspectos esenciales de mecanismos factibles que explican comportamientos cognitivos básicos.

Para finalizar, un comentario sobre la cuestión metodológica aludida en el principio del artículo, la cual puede resumirse en el siguiente aforismo: *si existe un problema complicado que uno no es capaz de resolver, con seguridad existe algun problema relacionado mucho mas simple que uno no comprende.* Así, el modelo de Hopfield nos permitió entender los mecanismos de la memoria asociativa en base a los conceptos de atractor e información distribuida, de la misma manera que el modelo de Ising permitió entender el ferromagnetismo en el marco conceptual de la universalidad. Es esta la contribución que, desde un punto de vista metodológico, pudo hacerse desde la física estadística de los sistemas complejos. Comenzamos con el estudio de sistemas mucho mas simples, aun a costa de alejarnos mucho del problema real, con la esperanza de que los conocimientos adquiridos en los primeros nos permitan en alguna etapa posterior alcanzar una mayor comprensión acerca de la forma en que se desarrollan los comportamientos cognitivos en los sistemas neurológicos[1]

.

[1]La mayoría de los artículos "históricos" aqui citados pueden encontrarse en la excelente recopilación de la Ref.[22], así como numerosos artículos interesantes sobre neurociencias.

Agradecimientos

A Francisco (Pancho) Tamarit, con quien emprendimos juntos (entre tantos otros) este camino fascinante de las neurociencias y a quien estas notas tanto deben de horas de discusión. A Pedro Pury por la crítica metódica (e implacable) al manuscrito original y por el apoyo logístico en la compaginación computacional del mismo. A Mariano Cognigni, cuyas agudas preguntas y cuya notable curiosidad en inumerables charlas nocturnas contribuyeron, no solo a motivar, como fundamentalmente a delinear el presente artículo.

Referencias

[1] *Neural Networks: An Introduction*, B. Muller and J. Reinhardt (Springer Verlag, 1991).

[2] *Introduction to the Theory of Neural Computation*, J. Hertz, A. Krogh and R. G. Palmer (Addison-Wesley Pub. Co., 1991).

[3] *Statistical Mechanics* 2nd. Ed., K. Huang (J. Wiley & Sons, 1987).

[4] *Essentials of Neural Science and Behavior*, E. R. Kandel, J. H. Schwartz and T. M. Jessel Eds. (Appleton & Lange, Stamford, 1995).

[5] *A logical calculus of the ideas immanent in nervous activity* W. S. McCulloch and W. Pitts, Bull. Math. Biophys. **5**, 115 (1943).

[6] *Collective processing and neural states*, J. J. Hopfield, Modeling and Analysis in Biomedicine, pp.369 (World Cientific Publishing Co., 1984).

[7] *Principles of Neurodynamics* (Spartan, New York, 1962).

[8] *Perceptrons: An introduction to computational geometry*, M. Minsky and S. Papert (MIT Press, Cambridge 2nd Ed. 1988).

[9] *Arithmetic Perceptrons*, S. A. Cannas, Neural Computation **7**, 173 (1995).

[10] *Parallel distributed Processing* Vol.1, D. E. Rumelhart and J. L. McClelland (MIT Press, Cambridge, 1988).

[11] *Solving arithmetic problems using feed-forward neural networks*, L. Franco and S. A. Cannas, Neurocomputing **18**, 61 (1998).

[12] *The Organization of Behavior: A Neurophysiological Theory*, D. O. Hebb (Wiley, New York, 1949).

[13] *Neural Networks, a Comprehensive Foundation*, S. Haykin (Macmillan College Publishing Company, 1994).

[14] *Predictive Modular Neural Networks: Applications to Time Series*, V. Petridis and A. Kehagias (Kluwer Academic Publishers, 1998).

[15] *Modeling Brain Function: The world of attractor neural networks*, D. J. Amit (Cambridge University Press, 1989).

[16] *Effects of refractory periods in the dynamics of a diluted neural network*, F. A. Tamarit, D. A. Stariolo, S. A. Cannas and P Serra, Physical Review E **53**, 5146 (1996).

[17] *Neural networks and physical systems with emergent collective computational abilities*, Proc. Natl. Acad. Sci. USA **79**, 2554 (1982).

[18] *Analytic study of the memory capacity of a neural network*, W. A. Little and G. L. Shaw, Math. Biosci. **39**, 281 (1978).

[19] *Basins of attraction of neural networks models*, J. D. Keeler, in *Neural Networks for Computing* pp. 259, American Inst. of Physics Conf. Proc. Vol. 151 (1987). J. S. Denker Ed.

[20] *Learning of correlated patterns in spin-glass networks by local learning rules*, S. Diederich and M. Opper, Phys. Rev. Lett. **58**, 949 (1987).

[21] *The space of interactions in neural networks models*, E. Gardner, J. Phys. A **21** 257 (1988).

[22] *Biology and Computation: A physicist's choice*, H. Gutfreund and G. Toulouse, Advanced Series in Neuroscience Vol. 3 (World Scientific, 1994).