

25
zej



UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

FACULTAD DE CIENCIAS

"DESARROLLO Y COMPROBACION EXPERIMENTAL
DE UNA RED NEURONAL PARA RECONOCER
AUTOMATICAMENTE LA ACTIVACION VENTRI-
CULAR DENTRO DEL ELECTROCARDIOGRAMA"

T E S I S

QUE PARA OBTENER EL TITULO DE
F I S I C O
P R E S E N T A:

JUAN ANTONIO HERNANDEZ GONZALEZ



México, D. F.

DIRECTOR DE TESIS:

DRA. ROSALIA RIDAURA SANZ



1997

**TESIS CON
FALLA DE ORIGEN**

FACULTAD DE CIENCIAS
SECCION ESCOLAR



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.



UNIVERSIDAD NACIONAL
AUTÓNOMA DE
MÉXICO

M. en C. Virginia Abrin Batule
Jefe de la División de Estudios Profesionales de la
Facultad de Ciencias
Presente

Comunicamos a usted que hemos revisado el trabajo de Tesis: "DESARROLLO Y COMPROBACION EXPERIMENTAL DE UNA RED NEURONAL PARA RECONOCER AUTOMATICAMENTE LA ACTIVACION VENTRICULAR DENTRO DEL ELECTROCARDIOGRAMA".

realizado por Juan Antonio Hernández González
con número de cuenta 8153275-3 . pasante de la carrera de Física.
Dicho trabajo cuenta con nuestro voto aprobatorio.

Atentamente

Director de Tesis

Propietario

Propietario

Propietario

Suplente

Suplente

DRA. ROSALIA RIDAURA SANZ

DR. LUIS BOJORQUEZ TAPIA

ING. GENARO RODRIGUEZ ROSSINI

DRA. HORTENSIA GERTRUDIS GONZALEZ GOMEZ

DR. PEDRO MIRAMONTES VIDAL

Consejo Departamental de Física
FACULTAD DE CIENCIAS
UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
DR. ROBERTO ALEJANDRO RUELAS MAYORGA

AGRADECIMIENTOS.

Para el Dr. Carlos García Moreira del Laboratorio de Biofísica de la facultad de Ciencias de la U.N.A.M. que aunque desgraciadamente ya no se encuentra con nosotros agradezco la confianza y el apoyo que me brindó en este trabajo.

Para la Dra. Rosalía Ridaura Sanz por aceptar dirigir este trabajo, dándome valiosos consejos y aportaciones durante la realización del mismo.

Al Investigador Oscar Infante Vázquez, al Ingeniero Genaro Rodríguez Rossini y en general al Departamento de Diseño y Desarrollo de Instrumentación del Instituto Nacional de Cardiología " Ignacio Chávez " por toda la ayuda y facilidades que recibí de ellos.

A los Dr. Pedro Miramontes Vidal y Dr. Luis Bojórquez Tapia por su gran paciencia, apoyo y comprensión.

Para familiares, compañeros y amigos que de alguna forma contribuyeron a la realización de este trabajo, para todos ellos, muchas gracias.

Así mismo, dedico este trabajo con mucho cariño a mi pequeño hijo Calheg Azael Hernández y a mi gran esposa Angelina Muñoz.

INDICE

Resumen.

Agradecimientos

Capítulo 1. ANTECEDENTES FÍSICOS DE LAS REDES NEURONALES ARTIFICIALES.

1. LA FÍSICA DE LAS REDES NEURONALES.	2
1.1 Teoría del orden y desorden.	3
1.1.1 Enrejados estadísticos (Modelo de Ising).	3
1.1.2 Ferromagnetismo y antiferromagnetismo.	11
1.1.3 Modelo poblacional; interacción competitiva.	13
1.2 Equilibrio en dinámicas no locales.	19
1.2.1 El experimento de Bénard (Turbulencia).	19
1.2.2 Sistemas dinámicos atractores. (Formulación cuantitativa).	23
1.2.3 Significado de auto-organización.	29

Capítulo 2. METODOLOGIA DE LAS REDES NEURONALES ARTIFICIALES.

2. CONCEPTOS NEURONALES BÁSICOS.	30
2.1 Procesamiento de datos en una red neuronal.	31
2.1.1 Introducción.	31
2.1.2 La neurona biológica. Unidad básica del sistema nervioso.	32
2.1.3 Desarrollo de auto-organización neuronal. (Postulado neurofisiológico de Hebb).	37
2.2 Cálculos colectivos y redes neuronales.	40
2.2.1 Relación entre dinámica de atractores y arquitectura de redes neuronales artificiales.	40
2.2.2 Un enrejado enfocado como una capa neuronal artificial en donde se promueve la auto-organización.	43
2.2.3 Redes multicapas de alimentación progresiva.	47

Capítulo 3 EL PERCEPTRÓN.

3. CARACTERIZACIÓN CONEXIONISTA DEL APRENDIZAJE	49
3.1 Configurando una red neuronal multicapas de alimentación progresiva.	50
3.1.1. La neurona artificial. El perceptrón.	50
3.1.2. Estructurando la red artificial.	51
3.2 Propagación. Descripción general.	53
3.2.1. El manejo y propagación de la información.	53
3.2.2. Entrenamiento.	58
3.3 Retropropagación. Descripción general a bloques.	59
3.3.1 La regla delta.	59
3.3.2 El gradiente descendente del error.	63
3.3.3 Enseñanza - aprendizaje. Entrenamiento hebbiano. (Basado en ejemplos y ejercicios).	68

Capítulo 4 AUTOMATIZACIÓN DE LA RED AJUSTABLE.

4. AUTOMATIZACIÓN.	69
4.1 Proceso de aprendizaje de la red neuronal multicapas.	70
4.1.1 El programa entrenador de una red neuronal.	70
4.1.2 Ejemplos simples de tareas de reconocimiento y clasificación.	73
4.1.2.1 La X-OR y la separabilidad no lineal.	73
4.1.2.2 La escuela de Capercucita.	77
4.2 La red ya entrenada (su eficiencia y productividad).	79
4.2.1 La generalización.	80
4.2.2 Evaluación del trabajo realizado por la red.	82

Capítulo 5 ENSAYO DEL MÉTODO.

5. APLICACIÓN.	84
5.1 Utilización de una red neuronal artificial como ayuda en la interpretación del electrocardiograma.	84
5.1.1 El sistema cardiovascular.	85
5.1.2 El electrocardiograma. (Huella de la dinámica del corazón).	87
5.2 Planteamiento general del sistema electrocardiográfico.	92
5.2.1 El amplificador.	93
5.2.2 El convertidor analógico - digital.	94
5.2.3 El detector de QRS.	95
5.3 Ensayo del método.	96
5.3.1 El programa neuronal reconocedor del complejo QRS.	96
5.3.2 Selección y almacenamiento de datos.	97
5.3.3 Etapa de enseñanza y entrenamiento.	98
5.3.4 Pruebas de generalización y evaluación.	99
APÉNDICE A. Interacción aditiva a primeros vecinos en un plano.	104
APÉNDICE B. Conteo del número de interacciones a primeros vecinos en un plano.	108
APÉNDICE C. Intervalo de λ en el doble ciclo de la ecuación logística $X_{i+1} = \lambda X_i(1 - X_i)$	109
APÉNDICE D. Listado del programa neuronal.	112
BIBLIOGRAFÍA.	122

DESARROLLO Y COMPROBACIÓN EXPERIMENTAL DE UNA RED NEURONAL PARA RECONOCER AUTOMÁTICAMENTE LA ACTIVACIÓN VENTRICULAR DENTRO DEL ELECTROCARDIOGRAMA.

RESUMEN

Se utiliza la metodología de las redes neuronales para encontrar una solución empírica al problema físico de reconocer una señal eléctrica, emitida periódicamente por un generador multipolo, de configuración variable, que se mueve en el seno de un conductor de volumen heterogéneo, como es el caso de la activación cardíaca.

Desde una plataforma física se abordan las redes neuronales artificiales. Haciendo uso de la teoría del orden y desorden se introducen algunos ejemplos de la mecánica estadística, en el campo del electromagnetismo (los puntos críticos) y en la dinámica de fluidos (las turbulencias y estabilidad), fundamentos que llevan a formular propiedades análogas a las encontradas en los sistemas biológicos como memoria asociativa, aprendizaje autoconducido (experiencia), auto - organización. Dentro de este marco se construye un simulador neuronal que utiliza el algoritmo de aprendizaje de la retropropagación y se evalúa su habilidad para reconocer la actividad ventricular en el E.C.G. Se inicia el proceso de entrenamiento de la Red Neuronal utilizando una gran colección de electrocardiogramas digitalizados y calificados por un especialista (en más de 500 pacientes). Se estudia asimismo, el progreso de aprendizaje y se formulan las bases para su comparación con los métodos tradicionales y su implementación en los monitores cardiológicos .

CAPITULO UNO

ANTECEDENTES FÍSICOS DE LAS REDES NEURONALES ARTIFICIALES

1 LA FÍSICA DE LAS REDES NEURONALES

1.1 Teoría del orden y desorden.

- 1.1.1 Enrejados estadísticos (Modelo de Ising).
- 1.1.2 Ferromagnetismo y antiferromagnetismo.
- 1.1.3 Modelo poblacional; interacción competitiva.

1.2 Equilibrio en dinámicas no lineales.

- 1.2.1 El experimento de Bénard (Turbulencia).
- 1.2.2 Sistemas dinámicos atractores.
Formulación cuantitativa.
- 1.2.3 Significado de auto-organización.

CAPITULO UNO

**ANTECEDENTES FÍSICOS
DE LAS
REDES NEURONALES ARTIFICIALES**

1. LA FISICA DE LAS REDES NEURONALES

A continuación se exhiben los tipos de dificultades y problemas que se enfrentan cuando se estudia a los sistemas neuronales (un sistema con una multiplicidad de elementos relacionados por interacciones simultáneas) y cómo la física ha abordado estos problemas, generando una nueva rama que ahora se conoce como física de las complejidades.

Se empezará por introducir la teoría del orden y desorden apoyados por los enrejados estadísticos ocupados por entes, que presentan alguna dependencia e interacción entre sí. Se dan algunos ejemplos de cómo pueden ser las interacciones entre los pobladores dentro de estos enrejados (sólo mencionamos, de entre una gran posibilidad de casos, las interacciones de alcance a primeros vecinos); así mismo se introducirá una formalidad matemática para su tratamiento. Los temas de ferromagnetismo y puntos críticos se presentan como casos concretos de la utilización de estos modelos, haciendo ver cómo lo observado puede ser el resultado de los movimientos e interacciones de tipo competitivo por alcanzar un cierto orden (estados que adopta la población). Se introduce el punto de vista de la mecánica estadística como manera de cuantificar estos estados de orden, estableciendo en particular la ecuación de la energía (1.14).

El comportamiento de los elementos dentro del marco físico de los modelos de Ising trae consigo propiedades emergentes análogas a las de algunos sistemas biológicos (auto-organización y memoria), lo que se muestra a través de un tratamiento basado en las celdas de Bénard, con el cual se ve, cómo el sistema trata de llegar, a través de competencia y organización entre sus mismos elementos (pobladores), a adquirir ciertos estados de equilibrio coherente, los que se van a relacionar posteriormente como respuestas, reacciones o blancos esperados en una red neuronal. Con dinámicas de conducta muy sencillas de estos sistemas se exhiben los posibles flujos, trayectorias y bifurcaciones que pueden tomar, para encontrar su estabilidad (atractores o blancos destinos). Estos estados van a responder o depender tanto de las condiciones iniciales del sistema como de la intensidad de la perturbación (información).

1.1 TEORÍA DEL ORDEN Y DESORDEN

1.1.1 ENREJADOS ESTADÍSTICOS

Los enrejados estadísticos estudian a los sistemas considerando las posiciones e interacciones de sus elementos o partículas, imaginándolas dentro de celdas que forman un gran enrejado.

Con un simple modelo estadístico, frecuentemente referido como el modelo de Ising, se puede llevar a los sistemas complejos a un sencillo pero riguroso análisis matemático que nos conduzca al entendimiento de la naturaleza física del problema. El modelo consiste en una rejilla regular de sitios en una, dos o tres dimensiones, en la cual cada sitio puede ser ocupado de una manera A o de una manera B, donde el significado de las dos clases de ocupación depende del problema físico particular que se esté tratando.

Se considera que existe una energía potencial de interacción E_{AA} , E_{BB} o E_{AB} entre pares de sitios de primeros vecinos, dependiendo de la manera de ocupación del par. Como ejemplo introductorio consideraremos una mezcla hecha con dos tipos de átomos, como la retícula mostrada en la fig. 1, en la que habrá algún arreglo preferencial de la manera en la cual los sitios del enrejado son ocupados por el tipo de átomos A y el tipo B.

Esencialmente dos tipos de arreglos reticulares son posibles y ellos pueden ser entendidos desde un punto de vista físico como sigue:

Demos a V_{AB} la interpretación de la interacción potencial entre átomos diferentes y V_{AA} y V_{BB} el potencial entre átomos iguales. Si $V_{AB} > (V_{AA} + V_{BB})/2$, entonces los átomos A prefieren estar cerca de los átomos A e igualmente para los átomos B. Así si estuviéramos en un sólido encontraríamos dos regiones distintas, una de átomos A solamente y otra de átomos B solamente. Si $V_{AB} < (V_{AA} + V_{BB})/2$, los átomos A y B podrían tender a alternarse ordenadamente en posiciones, a través del sólido. Si el sólido es calentado gradualmente, entonces los átomos A y B cambiarán posiciones en una forma azarosa hasta cerca de la temperatura de Curie de fundición, en donde se vuelven a alcanzar agrupaciones o colocaciones con distintas formaciones ordenadas (orden que será definido más tarde).

1 LA FISICA DE LAS REDES NEURONALES

A continuación se exhiben los tipos de dificultades y problemas que se enfrentan cuando se estudia a los sistemas neuronales (un sistema con una multiplicidad de elementos relacionados por interacciones simultáneas) y cómo la física ha abordado estos problemas, generando una nueva rama que ahora se conoce como física de las complejidades.

Se empezará por introducir la teoría del orden y desorden apoyados por los enrejados estadísticos ocupados por entes, que presentan alguna dependencia e interacción entre sí. Se dan algunos ejemplos de cómo pueden ser las interacciones entre los pobladores dentro de estos enrejados (sólo mencionamos, de entre una gran posibilidad de casos, las interacciones de alcance a primeros vecinos); así mismo se introducirá una formalidad matemática para su tratamiento. Los temas de ferromagnetismo y puntos críticos se presentan como casos concretos de la utilización de estos modelos, haciendo ver cómo lo observado puede ser el resultado de los movimientos e interacciones de tipo competitivo por alcanzar un cierto orden (estados que adopta la población). Se introduce el punto de vista de la mecánica estadística como manera de cuantificar estos estados de orden, estableciendo en particular la ecuación de la energía (1.14).

El comportamiento de los elementos dentro del marco físico de los modelos de Ising trae consigo propiedades emergentes análogas a las de algunos sistemas biológicos (auto-organización y memoria), lo que se muestra a través de un tratamiento basado en las celdas de Bénard, con el cual se ve, cómo el sistema trata de llegar, a través de competencia y organización entre sus mismos elementos (pobladores), a adquirir ciertos estados de equilibrio coherente, los que se van a relacionar posteriormente como respuestas, reacciones o blancos esperados en una red neuronal. Con dinámicas de conducta muy sencillas de estos sistemas se exhiben los posibles flujos, trayectorias y bifurcaciones que pueden tomar, para encontrar su estabilidad (atractores o blancos destinos). Estos estados van a responder o depender tanto de las condiciones iniciales del sistema como de la intensidad de la perturbación (información).



$$2(N_A - N_B) = 0$$

(a)

N_A = Número de átomos del tipo A
 N_B = Número de átomos del tipo B



$$N_A > N_B$$

$$2(N_A - N_B) > 0$$

(b)

Fig. 1. (a) Enrejado con disposición alternada. (b) Enrejado con aglomeraciones o agrupación poblacional.

1.1 TEORÍA DEL ORDEN Y DESORDEN

1.1.1 ENREJADOS ESTADÍSTICOS

Los enrejados estadísticos estudian a los sistemas considerando las posiciones e interacciones de sus elementos o partículas, imaginándolas dentro de celdas que forman un gran enrejado.

Con un simple modelo estadístico, frecuentemente referido como el modelo de Ising, se puede llevar a los sistemas complejos a un sencillo pero riguroso análisis matemático que nos conduzca al entendimiento de la naturaleza física del problema. El modelo consiste en una rejilla regular de sitios en una, dos o tres dimensiones, en la cual cada sitio puede ser ocupado de una manera A o de una manera B, donde el significado de las dos clases de ocupación depende del problema físico particular que se esté tratando.

Se considera que existe una energía potencial de interacción E_{AEA} , E_{AEB} o E_{BEA} entre pares de sitios de primeros vecinos, dependiendo de la manera de ocupación del par. Como ejemplo introductorio consideraremos una mezcla hecha con dos tipos de átomos, como la retícula mostrada en la fig. 1, en la que habrá algún arreglo preferencial de la manera en la cual los sitios del enrejado son ocupados por el tipo de átomos A y el tipo B.

Esencialmente dos tipos de arreglos reticulares son posibles y ellos pueden ser entendidos desde un punto de vista físico como sigue:

Demos a V_{AB} la interpretación de la interacción potencial entre átomos diferentes y V_{AA} y V_{BB} el potencial entre átomos iguales. Si $V_{AB} > (V_{AA} + V_{BB})/2$, entonces los átomos A prefieren estar cerca de los átomos A e igualmente para los átomos B. Así si estuviéramos en un sólido encontraríamos dos regiones distintas, una de átomos A solamente y otra de átomos B solamente. Si $V_{AB} < (V_{AA} + V_{BB})/2$, los átomos A y B podrían tender a alternarse ordenadamente en posiciones, a través del sólido. Si el sólido es calentado gradualmente, entonces los átomos A y B cambiarán posiciones en una forma azarosa hasta cerca de la temperatura de Curie de fundición, en donde se vuelven a alcanzar agrupaciones o colocaciones con distintas formaciones ordenadas (orden que será definido más tarde).

Es interesante calcular la función de partición de uno de tales sistemas :

$$e^{-\beta F} = \sum_{\text{arreglo de átomos}} \left[\exp \left(-\sum_{\text{(lazos ij)}} \beta V_{ij} \right) \right] \quad (1.1)$$

donde β y F son constantes que están referidas a la probabilidad de encontrar al sistema con una energía dada (E_n).

Definamos los estados de excitabilidad:

$$\begin{aligned} \epsilon_i &= +1 \text{ para el átomo A} \\ &= -1 \text{ para el átomo B} \end{aligned}$$

N_A, N_B = Número de átomo del tipo A y B respectivamente.

$$\begin{aligned} \text{Así } \epsilon_i \epsilon_j &= +1 \text{ Si es el par AA} \\ &= +1 \text{ Si es el par BB} \\ &= -1 \text{ Si es el par AB} \end{aligned}$$

Considerando sólo interacciones de primeros vecinos, la interacción energética entre los átomos puede escribirse en la forma:

$$V_{ij} = a \epsilon_i \epsilon_j + b \left(\frac{\epsilon_i + \epsilon_j}{2} \right) + c \quad (1.2)$$

donde a, b y c son constantes.

Así la exponencial en la ecuación 1.1 es:

$$\exp \left(-\beta a \sum_{ij} \epsilon_i \epsilon_j \right) \exp \left(-\beta b \sum_{ij} \frac{\epsilon_i + \epsilon_j}{2} \right) \exp \left(-\beta c \right)$$

donde la última exponencial $e^{-\beta c}$ es una constante.

En la función de partición cada átomo contribuye un cierto número de veces. El número de veces que un átomo contribuye, es igual al número de lazos que éste haga, esto es, el número de primeros vecinos que tenga. Por ejemplo en el caso de una redicula cuadrada de dos dimensiones, cada uno de los átomos tiene cuatro primeros vecinos.

El término $\sum_j \frac{\epsilon_i + \epsilon_j}{2}$, como se demuestra en el apéndice A, se puede expresar como:

$$\sum_j \frac{\epsilon_i + \epsilon_j}{2} = 2(N_A - N_B)$$

En general, si N_A y N_B son constantes predeterminadas, este término se puede ignorar porque no depende del arreglo o disposición de los átomos en el enrejado. El problema es entonces calcular, en la función de partición, la cantidad:

$$\sum_{\text{arreglo de átomos}} [\exp(-\beta \sum_{ij} \epsilon_i \epsilon_j)] = \sum_{\text{arreglo de átomos}} [\exp(-H \sum_{ij} \epsilon_i \epsilon_j)] \tag{1.3}$$

donde $H = \beta a$

Como es de notar, el modelo que se describe puede ser tomado como un modelo de ferromagnetismo, interpretando $\epsilon_i = +1$ como un espín hacia arriba y $\epsilon_i = -1$ como un espín hacia abajo; de hecho los puntos más importantes en esta teoría han sido desarrollados y presentados usando el lenguaje del ferromagnetismo.

En seguida presentaremos un problema resuelto por vez primera por Onsager, como un ejemplo particular de cómo es evaluada la función de partición (1.3). Se tratará de manera general y no con el interés original de encontrar la respuesta analítica a los puntos críticos en el fenómeno de transición de fase magnética, sino viendo más bien la adaptación de esta herramienta a sistemas donde existe una multiplicidad de elementos interactuando simultáneamente.

EL PROBLEMA OSANGER *

Considérese una red de celdas en dos dimensiones de N celdas, donde cada celda puede ser ocupada de dos maneras distintas (por ejemplo un dipolo, con su eje perpendicular al plano reticular; este elemento puede tener sólo dos orientaciones posibles designadas por \uparrow o \downarrow), esto es, el i -ésimo sitio lleva un número ϵ_i el cual puede ser $+1$ o -1 , A o B , \uparrow o \downarrow , etc. La energía de interacción de los primeros vecinos es $H\epsilon_i\epsilon_j$ en unidades de KT . Un estado particular del sistema es denotado al dar las ϵ_i de cada sitio. El problema es calcular la energía libre del sistema [1,2].

De la función de partición

$$Q = \sum_{\text{Todos los estados}} e^{-\beta E_k}$$

donde E_k es la suma resultante de las interacciones energéticas de primeros vecinos en un estado particular,

$$Q = \sum_{\text{Todos los estados}} e^{-\sum H\epsilon_i\epsilon_j} = \sum_{\text{Todos los estados}} \prod e^{-H\epsilon_i\epsilon_j}$$

Por tanto $-\beta E_k = -\sum H\epsilon_i\epsilon_j$ es la interacción total de energía del sistema de todos los pares de primeros vecinos en la k -ésima configuración y $-\sum H\epsilon_i\epsilon_j$ y $\prod e^{-H\epsilon_i\epsilon_j}$ significan, la suma y el producto sobre pares de puntos i y j que son adyacentes en el reticulado.

Es fácil ver que si el signo de H se cambia, Q permanece sin cambio (ya que el signo de H depende de los valores que tomen las variables ϵ_i , ϵ_j). Por tanto, y por conveniencia escribimos:

$$Q = \sum_{\text{Todos los estados}} \left| \prod e^{H\epsilon_i\epsilon_j} \right| \quad (1.4)$$

* Conocido con este nombre desde que en 1944 el físico alemán Onsager resolvió el problema de transición de fase ferromagnética, determinando analíticamente los puntos críticos en los cristales. El desarrollo expuesto aquí no es el utilizado por Onsager ya que el tratamiento usado por él, es extremadamente complicado y engorroso, por lo que se prefirió ilustrar una versión más sencilla desprendida de ideas de Feynman.

Como:

$$\frac{e^H + e^{-H}}{2} = \text{Cosh } H \qquad \frac{e^H - e^{-H}}{2} = \text{Senh } H$$

y puesto que $\epsilon_i \epsilon_j = \pm 1$ podemos escribir la siguiente ecuación:

$$e^H + e^{-H} + \epsilon_i \epsilon_j (e^H - e^{-H}) = 2 e^{\epsilon_i \epsilon_j H}$$

Por lo tanto

$$\begin{aligned} \mathbf{c}^{\text{H}\epsilon_j} &= \text{Cosh } H + \epsilon_i \epsilon_j \text{ Sinh } H \\ &= \text{Cosh } H + \epsilon_i \epsilon_j T \text{ Cosh } H \\ &= (1 + \epsilon_i \epsilon_j T) \text{ Cosh } H \end{aligned} \qquad (1.5)$$

donde $T = \text{Tanh } H$.

Sustituyendo (1.4) en (1.5) se encuentra

$$Q = \sum_{\text{Todos los estados}} \prod (1 + \epsilon_i \epsilon_j T) \text{ Cosh } H \qquad (1.6)$$

Dado que para una red de N celdas hay $2N$ parejas $(\epsilon_i \epsilon_j)$, (excepto para $N < 4$) de primeros vecinos interactuando (Ver apéndice B) y teniendo tantas veces (Cosh H) como parejas, se puede escribir:

$$Q = \text{Cosh}^{2N} H \sum_{\text{Todos los estados}} \prod (1 + \epsilon_i \epsilon_j T) = 2^N Q' \text{Cosh}^{2N} H \qquad (1.7)$$

donde

$$Q' = \frac{1}{2^N} \sum_{\text{Todos los estados}} \prod (1 + \epsilon_i \epsilon_j T) \qquad (1.8)$$

expandiendo el producto

$$Q' = \frac{1}{2^N} \sum_{\text{Todos los estados}} (1 + \epsilon_i \epsilon_j T) (1 + \epsilon_k \epsilon_l T) (1 + \epsilon_m \epsilon_n T) \dots$$

y

$$2^N Q' = \sum_{\text{Todos los estados}} [1 + T \sum \epsilon_i \epsilon_j + T^2 \sum (\epsilon_i \epsilon_j)(\epsilon_k \epsilon_l) + T^3 \sum (\epsilon_i \epsilon_j)(\epsilon_k \epsilon_l)(\epsilon_m \epsilon_n) + \sum T^4 (\epsilon_i \epsilon_j)(\epsilon_k \epsilon_l)(\epsilon_m \epsilon_n)(\epsilon_o \epsilon_p) + \dots] \quad (1.9)$$

La suma sobre todos los estados pretende tener todos los posibles estados, es decir $\sum_{\text{Todos los estados}}$ se refiere al número total de posibles configuraciones que el enrejado puede tener. Siendo que cada sitio de las N celdas puede ser ocupado de dos maneras diferentes, tenemos que el número total de combinaciones posibles que el enrejado puede adoptar es de 2^N . Lo que está entre paréntesis en la Ec. 1.9, se refiere a las interacciones de las ϵ 's cuando el enrejado está en un estado dado. El coeficiente de T^4 es una suma de productos correspondientes a las diferentes selecciones de L . Las ϵ 's en cada uno de estos términos ocurren por pares correspondiendo a los primeros vecinos y como se puede observar, cada par no aparece dos veces en un mismo producto, por lo que puede decirse que no se repite ningún lazo, pues asociado a cada par $\epsilon_i \epsilon_j$, está un lazo que conecta el i -ésimo punto con el j -ésimo punto (es decir i y j son vecinos cercanos conectados por la interacción $\epsilon_i \epsilon_j$, interpretada gráficamente como un lazo). Así a cada término o producto formado por las parejas de ϵ 's, le asociamos una gráfica o conjunto de L lazos con longitud L . Y ya que no hay pareja $\epsilon_i \epsilon_j$ que aparezca dos veces en un término dado, decimos en consecuencia que no aparece el mismo lazo dos veces en una gráfica dada. Por ejemplo, la fig. 1.2a muestra la gráfica de tamaño $L = 10$ asociada con el término:

$$(\epsilon_1 \epsilon_2)(\epsilon_3 \epsilon_4)(\epsilon_5 \epsilon_6)(\epsilon_1 \epsilon_5)(\epsilon_2 \epsilon_6)(\epsilon_3 \epsilon_7)(\epsilon_4 \epsilon_8)(\epsilon_7 \epsilon_9)(\epsilon_{10} \epsilon_{11})(\epsilon_{10} \epsilon_{14})$$

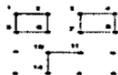


Fig. 1.2a.

Fig. 1.2 a Una gráfica de longitud 10 formada por un conjunto de 7 lazos.

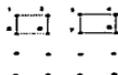


Fig. 1.2b.

Fig. 1.2 b Una gráfica de longitud 8 formada por un conjunto de 7 lazos.

Después de sumar o considerar todos los $\epsilon_i \neq 1$ en la Ec. 1.9, los términos que contienen potencias impares de ϵ_i desaparecen y no contribuyen, quedando sólo los términos que contienen potencias en ϵ_i de 0, 2, o 4. Por ejemplo la suma de los términos de la fig. 1.2a nos da cero, obsérvese:

$$\begin{aligned} \sum (\epsilon_1 \epsilon_2)(\epsilon_3 \epsilon_4)(\epsilon_5 \epsilon_6)(\epsilon_7 \epsilon_8)(\epsilon_9 \epsilon_{10})(\epsilon_{11} \epsilon_{12})(\epsilon_{13} \epsilon_{14}) = \\ = \sum (\epsilon_1)^2 (\epsilon_2)^2 (\epsilon_3)^2 (\epsilon_4)^2 (\epsilon_5)^2 (\epsilon_6)^2 (\epsilon_7)^2 (\epsilon_8)^2 (\epsilon_9)^2 (\epsilon_{10})^2 (\epsilon_{11})^2 (\epsilon_{12})^2 (\epsilon_{13})^2 (\epsilon_{14})^2 \end{aligned}$$

Nótese cómo las primeras nueve variables están elevadas al cuadrado y las dos últimas no, por tanto no importa el valor que estas nueve variables adopten (± 1) darán un resultado en su producto de 1. No así el valor final del término completo, el que está determinado por los valores que tome ϵ_{11} y ϵ_{14} en cada estado.

$$(1) (\epsilon_{11}) (\epsilon_{14})$$

$$(1) (1) (1) = 1$$

$$(1) (1) (-1) = -1$$

$$(1) (-1) (1) = -1$$

$$(1) (-1) (-1) = 1.$$

$$+ 0$$

Sumando todas las posibles configuraciones que los lazos puedan tener, nos da un valor neto igual a cero por lo que la contribución del término correspondiente a la gráfica de la fig. 1.2a en la ecuación 1.9 es cero.

Ahora veamos el término asociado con la gráfica de la fig. 1.2b; siempre en cualquier estado dará 1 y el número de 1's es igual al número de combinaciones posibles que las variables de este término puedan tener, teniendo por tanto 2^4 unos dentro de la sumatoria.

$$\sum (\epsilon_1 \epsilon_2)(\epsilon_3 \epsilon_4)(\epsilon_5 \epsilon_6)(\epsilon_7 \epsilon_8)(\epsilon_9 \epsilon_{10})(\epsilon_{11} \epsilon_{12})(\epsilon_{13} \epsilon_{14})(\epsilon_{15} \epsilon_{16})(\epsilon_{17} \epsilon_{18})(\epsilon_{19} \epsilon_{20}) = \sum (\epsilon_1)^2 (\epsilon_2)^2 (\epsilon_3)^2 (\epsilon_4)^2 (\epsilon_5)^2 (\epsilon_6)^2 (\epsilon_7)^2 (\epsilon_8)^2 (\epsilon_9)^2 (\epsilon_{10})^2 (\epsilon_{11})^2 (\epsilon_{12})^2 (\epsilon_{13})^2 (\epsilon_{14})^2 (\epsilon_{15})^2 (\epsilon_{16})^2 (\epsilon_{17})^2 (\epsilon_{18})^2 (\epsilon_{19})^2 (\epsilon_{20})^2 = 1$$

Por lo tanto, es claro que sólo términos donde cada una de las ϵ_i aparecen un número par de veces contribuirán a Q' . Esto es equivalente a decir que sólo cuentan o contribuyen en Q' las gráficas en las cuales, cada punto reticular, tiene un número cero o par de lazos que salen procedentes de él (0, 2 o 4). En otras palabras, la contribución gráfica debe ser superposición de

simples poligonos cerrados que no tienen lados en común. Tales gráficas son llamadas gráficas cerradas. En la fig. 1.3 se dan dos ejemplos de gráficas cerradas. Nótese que la gráfica de la figura 1.2a no es cerrada y no contribuye a Q' ; la fig. 1.3c es otro ejemplo de una gráfica no cerrada, la cual tiene una contribución cero.

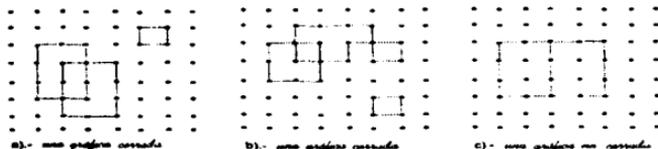


Fig. 1.3 Representaciones gráficas de algunos productos en la ecuación 1.9. a y b son ejemplos de gráficas cerradas mientras c es un ejemplo de una gráfica no cerrada.

Así cada término en la Ec. (1.9) debe corresponder a una gráfica cerrada, pues, para las interacciones de nuestro interés cada ϵ_i ocurre a una potencia de 0, 2 o 4. Por tanto simplemente consideramos todos los términos T^L , donde L es el número de lazos y además, ya que la sumatoria de todos los estados sobre las ϵ_i nos da un factor 2^N , que va a ser cancelado con el 2^N en el lado izquierdo de la Ec. (1.9). Así podemos decir que cada gráfica cerrada de longitud L contribuye con un término T^L en Q' escribiendo:

$$Q' = \sum_L \varphi(L) T^L = \sum_{\substack{\text{gráficas} \\ \text{cerradas}}} T^L \quad (1.11)$$

donde $\varphi(L)$ es el número de gráficas cerradas de tamaño L que pueden ser dibujadas sobre la retícula. Nótese que $\varphi(L) = 0$, a menos que L sea par, así el signo de T (y por tanto el signo de H) es inmaterial como podría esperarse.

1.1.2 FERROMAGNETISMO Y ANTIFERROMAGNETISMO

Uno de los primeros logros en la teoría cuántica del estado sólido durante los años 20's, fue el desarrollo de una teoría de materiales magnéticos; desde entonces a la fecha se han venido realizando diferentes trabajos, de donde han surgido las ideas y conceptos que se presentan a continuación, los cuales han sido base fundamental para innumerables estudios dentro y fuera de esta disciplina.

Un ferromagneto es un material, como el hierro, en el cual puede haber un momento magnético permanente. La situación es complicada por el hecho de que los materiales magnéticos están formados por pequeños dominios, cada uno de los cuales lleva un momento magnético relativamente grande. La energía más baja se obtiene cuando estos dominios o regiones magnéticas están alineados en tal forma que sus momentos magnéticos se cancelan, de manera que el todo tiene un momento magnético muy pequeño. Un magnetismo permanente se obtiene cuando el campo dominante ha forzado a los momentos a tomar una alineación paralela en una u otra posición, teniendo que cada región magnética de dominio individual se integra al todo, adaptándose en una interacción de magnetización de conjunto.

En un simple dominio magnético de tamaño fijo, cuando la temperatura se eleva, el momento magnético decrece deslizándose a cero conforme se alcanza la temperatura crítica, la cual para el hierro es alrededor de 1050°K . Cuando a cualquier material se le aplica un campo magnético, se le induce un momento magnético y a la razón del momento magnético sobre el campo aplicado, se le conoce como susceptibilidad magnética. Justo arriba de la temperatura crítica, la susceptibilidad magnética es muy grande, de hecho ésta llega a ser infinitamente grande cuando nos aproximamos, por el lado de las temperaturas más altas, hacia la temperatura crítica.

Como en todos los átomos, en los pertenecientes a los materiales aislantes del magnetismo se puede hablar de un momento magnético (nos referimos a los metales de transición, elementos en el centro de la tabla periódica tales como el Cr). También ellos tienen un momento angular conocido como "Espin", el cual es, por lo usual, considerablemente más grande que el espín básico de un solo electrón. La pura y simple interacción entre estos momentos magnéticos es muy débil y es sólo importante a temperaturas tan bajas como el punto de ebullición del helio (4°K). La teoría cuántica predice que en este punto, la interacción de los momentos magnéticos podría llegar a ser mucho muy fuerte [3,4]. Bajo algunas circunstancias, estas interacciones pueden ser favorecidas por alineaciones opuestas de espines vecinos y en algunos otros por alineaciones paralelas en una misma dirección. Tales interacciones son conocidas con el

nombre de "interacción de intercambio" (entre espines), por la forma en la cual éstas son manejadas. Cuando la alineación paralela en una misma dirección es favorecida, los espines vecinos tienden a adoptar la configuración de más baja energía, dentro de un pequeño dominio, que corresponde a tener todos los espines apuntando en una misma dirección y podrá surgir como consecuencia de muchos momentos magnéticos interactuando.

La configuración energética más baja es una en la cual un dominio magnético se coloca dentro de una temperatura muy baja. En esta circunstancia un incremento en la temperatura provocará fluctuaciones de baja energía y por lo tanto, no todos los espines estarán apuntando en la misma dirección. Esto fue argumentado, alrededor de fines de la centuria pasada, por P. Curie y R. Weiss, a través de su ley [5]. Así si la proporción de espines apuntando en la dirección del campo magnético (a una temperatura T_c) es la menor de todas, es porque la tendencia de los espines para apuntar en esta dirección es muy baja. Este argumento permite predecir, que el campo magnético deberá, de hecho, deslizarse uniformemente a cero hasta que se alcance una temperatura definida, en la cual la susceptibilidad magnética podría diverger conforme la temperatura se aproxime a la temperatura crítica, por el lado de arriba.

Se ha mencionado también que el intercambio entre muchos acoples, puede ser más favorecido por alineaciones opuestas de espines vecinos, que por las alineaciones paralelas en una misma dirección. Así uno podría esperar encontrar materiales cuya configuración, a bajas temperaturas, tengan a todos los espines vecinos opuestos, uno en contra de otro. Este es, de hecho, el caso más frecuente en la mayoría de los materiales, por eso son conocidos como antiferromagnéticos. Los aislantes antiferromagnéticos son mucho más comunes que los aislantes ferromagnéticos; no hay en ellos momento magnético en conjunto, así que sus propiedades magnéticas son mucho menos espectaculares y además no hay producción de campo desmagnetizador, por lo que la complejidad de la estructura del dominio magnético no surge de inmediato. El orden antiferromagnético puede ser seguido con el comportamiento de la susceptibilidad magnética. En las temperaturas críticas la susceptibilidad magnética en estas dos direcciones perpendiculares, llegan a ser idénticas.

Las estructuras cristalinas de un sólido magnético producen fuerzas adicionales sobre el momento magnético de los átomos; éstas podrán tender a producir alineaciones en alguna dirección definida del espacio. Si esta tendencia es fuerte, uno puede obtener una situación con una gran probabilidad, de que los espines estén apuntando en una de las dos direcciones permitidas, a lo largo de un eje particular. Esta situación suele ser explorada por el modelo de Ising.

En 1925 E. Ising publicó un artículo en el que discutía un sistema simple ferromagnético, el cual tomó su nombre. Hablaba de un conjunto de átomos, cada uno con un momento magnético o espín correspondiente. En tal modelo cada espín atómico puede estar en uno de dos estados, hacia arriba o hacia abajo de acuerdo a las interacciones entre vecinos. Si no hay campo magnético aplicado, los dos estados posibles, de todos los espines hacia arriba o todos los espines hacia abajo, tendrán la misma energía, pero si hay un campo aplicado, aún uno muy débil, éste podrá favorecer a una dirección más que la otra. Ising estudió el modelo para una dimensión, en el cual los espines están todos arreglados en una fila con solamente dos vecinos, uno a cada lado. Durante los siguientes 60 años, grandes esfuerzos fueron realizados, sobre todo en estudios de transición de fase magnética, con el modelo de Ising en más de una dimensión, surgiendo conceptos como el de longitud de correlación. Esta longitud es la distancia sobre la cual, los efectos interactivos en una expansión de desorden, llega a ser muy grande, dentro de la vecindad de un punto crítico. Por ejemplo, en un sistema ferromagnético, justo sobre la temperatura crítica, si un espín es forzado a apuntar en una dirección particular, entonces todos los otros espines vecinos dentro de una distancia igual a la longitud de correlación serán alineados con éste, es decir es como si ellos, en una cierta extensión, imitaran el comportamiento de su vecino (esto es conocido como auto-organización). En la temperatura crítica la longitud de correlación llega a ser infinita; pero debajo de la temperatura crítica, hay de nuevo una longitud de correlación finita la cual da la distancia en la cual las desviaciones ocurren.

1.1.3 MODELO POBLACIONAL.

INTERACCIÓN COMPETITIVA

Los fenómenos de cooperación y competencia que experimentan grupos poblacionales que pueden ser tratados con el modelo de Ising son tan abundantes y familiares en nuestra vida diaria, que muchas veces pasan desapercibidos; mencionaremos, por citar uno de ellos, el efecto cooperativo de silencio producido en un gran auditorio justo al empezar una conferencia o un concierto, o el efecto contrario donde el nivel de ruido puede incrementarse dramáticamente, como en una fiesta de cocktail. Pero sin temor a equivocarnos los primeros ejemplos representativos que más se han estudiado y de los cuales ha surgido gran parte del modelo, que sin lugar a dudas ha marcado historia en el desarrollo de la física teórica en los últimos 50 años, los encontramos dentro del estudio del fenómeno de punto crítico, al cual nos

referiremos. Un punto crítico se presenta cuando un sólido se funde, o en el momento en que se inicia la ebullición de un líquido y nos marca la llamada transición de fase; por ejemplo, para el H_2O la fase líquida se mantiene favorecida en una atmósfera de presión y a temperaturas menores de 100°C . A los 100°C el agua se evapora y para temperaturas superiores a los 100°C , el H_2O existe en su fase gaseosa. El cambio de fase que ocurre a $T_b = 100^\circ\text{C}$ es acompañado por un cambio abrupto en la densidad del fluido. El volumen para una masa dada en la fase de vapor, es 1600 veces más grande que para la misma masa en su estado líquido y su densidad correspondiente es más pequeña. La temperatura de ebullición T_b depende de la presión p . En presiones superiores a la atmosférica, la temperatura de ebullición es mayor que 100°C y el salto, en lo que se refiere a la densidad del fluido que ocurre con la ebullición, es menor que cuando el cambio de fase toma lugar a 100°C . Esta tendencia continúa gradualmente conforme la presión se incrementa. La diferencia en densidades entre las fases líquidas y gaseosas a ambos lados de la temperatura de ebullición $T_b(p)$, decrece suavemente hasta una presión P_c (218 atmósferas), para la cual $T_b = 374^\circ\text{C}$. En este punto la diferencia desaparece del todo. Los parámetros P_c y $T_c = T_b(P_c)$ juntos, localizan el punto crítico que determina la transición de fase conocida como de primer orden.

Algo semejante podemos decir para materiales como el hierro, donde a temperatura ambiente existe en su fase magnética, identificada por su momento magnético, que se manifiesta por la capacidad de atraer otros materiales ferrosos y su disposición a orientarse por él mismo, hasta que su momento magnético es alineado con un campo magnético externo. Cuando la temperatura es incrementada, el momento magnético del hierro decrece suavemente hasta que a una temperatura $T_c = 770^\circ\text{C}$ se desvanece por completo. Este comportamiento es conocido como transición de fase de segundo orden y ha generado gran interés por el reconocimiento de que este problema posee una gran generalidad y es igual a muchos otros problemas que en la ciencia se presentan. Así por ejemplo el fenómeno es parecido a los encontrados en el movimiento turbulento de los fluidos, en la actividad neuronal de los seres vivos o la interacción en la confrontación de los quarks de la física teórica. Además de encontrar similitudes tan marcadas entre todos estos sistemas estudiados en diferentes ramas de la ciencia, la actividad que ha generado ha llevado al surgimiento de una nueva y muy poderosa técnica y sólo por esto bastaría para que valiera la pena estudiarlo.

En la física teórica el ataque a un problema de interés, tradicionalmente comienza (y algunas veces termina), con un intento de identificar y entender el modelo más simple, que exhiba los mismos aspectos esenciales que el problema en cuestión. El modelo más simple fue introducido por Lenz, no obstante que es ahora invariablemente conocido como el modelo de Ising. A pesar que el

dispositivo es un modelo de ferromagnetismo (como ya se ha mencionado en la sección pasada), nosotros hemos de describirlo en términos generales, más que en sólo específicamente magnéticos. Así el modelo considera un ordenamiento regular de puntos en el espacio (una rejilla), donde el número de puntos N es muy grande. La dimensión del arreglo espacial, es una elección muy significativa en el sentido de la complejidad operativa que puede acarrear; pero la esencia física del modelo se encuentra manifestada, en un arreglo en dos dimensiones y lo que ocurre en un plano se puede utilizar con propósitos ilustrativos. La naturaleza precisa de un patrón o estado, es plasmada por la forma en que los sitios son ocupados en la rejilla considerada cuadrada, la simetría y forma de la rejilla es más una cuestión de gusto y conveniencia que otra cosa. Los sitios de las rejillas están etiquetados por un índice $i = 1, 2, \dots, N$ y asociamos a cada sitio i una variable S_i . Cada variable puede adoptar entre uno u otro de dos valores posibles, $+1$ o -1 . Variables vecinas ejercen una interacción o influencia unas con otras. Esta influencia es expresada a través de asociar, con cada par de sitios vecinos i y j , una energía $-J_i S_j$, donde J es alguna constante positiva. La interacción es más baja con $-J$ y más alta con $+J$, acorde a si las variables vecinas tienen el mismo valor, o valores opuestos. Esto completa la definición del modelo.

En el modelo, el tratar de explicar lo que significa y lo qué es un punto crítico, es algo abstracto; aunque la siguiente definición es formalmente adecuada, conceptualmente posee un poco de ininteligibilidad: un punto crítico, es determinado por los términos de la gráfica de una curva cerrada, la cual existe en un espacio de propiedades (normalmente de temperaturas) controlando el estado del sistema y su determinación señala el hecho de que, en este punto, la distinción entre dos fases se desvanece. Este desvanecimiento, en el caso del ferromagnetismo, se observa a medida que nos vamos aproximando a T_c , pues antes de alcanzar este punto se puede detectar un momento magnético neto por unidad de volumen, que nos permite percatarnos de las fases que coexisten a temperaturas menores de T_c , ya que ellas son distinguibles debido a la orientación del momento magnético. Así, a medida que vamos alcanzando T_c se va perdiendo tal distinción, junto con la magnetización. En el caso del modelo de Ising, dos fases coexisten en las proximidades de la temperatura crítica especificada: una fase con una mayoría de variables que tienen un valor $+1$ y otra con una mayoría que tiene el valor -1 . A medida que nos vamos aproximando al punto crítico, las diferentes medidas de las dos poblaciones se contraen reorganizándose y sobre T_c las poblaciones son esencialmente iguales.

En este estudio, es útil tener una medida cuantitativa de la diferencia entre las fases unidas en el punto crítico: éste es el papel que juega el parámetro de orden, Q . En el caso del fluido, el parámetro de orden es tomado como la

diferencia entre las densidades de las fases líquidas y gaseosas. En el ferromagnetismo es tomado como la magnetización y en el modelo de Ising es definido por la fracción o razón de las variables $+1$ sobre las variables -1 . Claramente, así definido el parámetro de orden, se desvanece cuando nos aproximamos al punto crítico, a lo largo de la línea de la coexistencia de fases. Como su nombre sugiere, el parámetro de orden puede ser pensado como una medida de tipo metódico, que localiza a un sistema que está en un estado, por debajo de una temperatura crítica. En el caso del magnetismo una configuración espinorial estable $\{S_i, 1 \leq i \leq N\}$ corresponde o determina una magnetización específica dada.

Nuestra próxima tarea es dar algún sentido a los principios que metodizan los procesos de ordenamiento. Para esto, debemos apelar al resultado fundamental de la física estadística. La probabilidad p_μ de que un sistema físico a temperatura T , tenga un arreglo microscópico particular, etiquetado con μ y energía E_μ , es:

$$p_\mu = \frac{e^{-E_\mu/KT}}{Z} \quad (1.12)$$

donde Z es una función de partición, que puede expresarse como:

$$Z = \sum_{\mu} e^{-E_\mu/KT} \quad (1.13)$$

K = constante de Boltzmann y T = temperatura

ya que los sistemas deben siempre estar en algún arreglo determinado, y la suma de las probabilidades p_μ debe ser la unidad, donde la suma se extiende sobre todos los posibles arreglos microscópicos. En la utilización de esta ecuación es generalmente correcto suponer que el sistema físico se desarrolla rápidamente a través de todos los arreglos permitidos. El valor observado de alguna propiedad física, es vista como el promedio de todos los arreglos μ , pesados con la apropiada probabilidad p_μ , en cada configuración μ . Un arreglo μ es definido por un conjunto particular de valores $S_1^{(\mu)}, S_2^{(\mu)}, \dots, S_N^{(\mu)}$ que toman las variables en el estado μ ; la configuración energética E_μ es entonces dada por la interacción total de las energías involucradas. Por lo que:

$$E_\mu = \sum_{\langle ij \rangle} -J_{ij} S_i^{(\mu)} S_j^{(\mu)} \quad (1.14)$$

donde la suma se extiende sobre todos los pares de sitios adyacentes. La probabilidad p_{μ} (ecuación 1.12) es, de esta forma, la probabilidad de encontrar las N variables $\{S_i, 1 \leq i \leq N\}$, con valores asignados que correspondan a un estado de equilibrio energético E_{μ} . La suma sobre μ , en la ecuación 1.13, se extiende sobre el conjunto completo de estados μ , de las posibles asignaciones en que encontremos a cada una de las N variables.

El parámetro de orden puede entonces ser dado por:

$$Q = \sum_{\mu} Q_{\mu} P_{\mu} \quad (1.15a)$$

donde

$$Q_{\mu} = \frac{1}{N} \sum_i S_i^{(\mu)} \quad (1.15b)$$

da la proporción fraccional entre las poblaciones de las variables con $S_i^{(\mu)} = +1$ y $S_i^{(\mu)} = -1$ en el arreglo μ .

Las sumas de las ecuaciones (1.13, 1.14, 1.15) no son fáciles de evaluar, no obstante, es útil enfrentar esta dificultad, pues importantes e interesantes propiedades surgen de ello.

En el límite en el cual la temperatura T se reduce a un mínimo (en la escala J/K), la ecuación (1.14) del sistema, se hallará probablemente en su estado de mínima energía. Hay dos de tales arreglos: uno en el cual todas las variables S_i son $+1$ y otro en el que todas las variables S_i son -1 . Aunque estén totalmente opuestos estos arreglos, sus ordenaciones correspondientes tienen la misma energía y en consecuencia son igualmente probables. Un sistema en uno de tales estados, en algún instante, es extremadamente difícil que pueda encontrar una forma de ir al otro, pero un argumento importante a la excepción de esta regla general, es que un sistema se desarrolla rápidamente a través de todos sus arreglos y por tanto, el estar en un estado de mínima energía, significa que debe haber pasado a través de arreglos de más alta energía (los cuales tienen una probabilidad muy pequeña, de hecho totalmente insignificante, cuando estamos cerca del límite de baja T). Por tanto, en $T=0$, el sistema permanece localizado en alguna de estas configuraciones totalmente ordenadas, en la cual el parámetro de orden Q (ecuaciones 1.15 a, b) necesariamente tiene la magnitud de 1.

Ahora considere el límite de alta temperatura. Observemos que la diferencia fraccional (Ec. 1.15b) para alguna configuración μ ($T \neq 0$), tendría un valor intermedio entre cero y uno. Cuando se aumenta la T , lo que primeramente ocurre es que aparecen grupos o poblaciones con límites distinguibles sobre el enrejado, que van a contribuir en el parámetro de orden (llevado en la suma de

la Ec. 1.15a) en virtud de su energía E_{μ} , que no es aún lo suficientemente grande como para compensar el hecho de que los arreglos no tengan todavía una diferencia fraccional en sus poblaciones Q_{μ} ; así éstos tienen determinado algún valor intermedio de orden, bien definido. A medida que nos empezamos a acercar a las temperaturas altas, los grupos o poblaciones se van haciendo muy numerosas y su orden cada vez más pequeño. Una pequeña reflexión, nos indica que estos estados son mucho muy abundantes en altas temperaturas; esencialmente en T_c , estos arreglos son ya totalmente desordenados dominando la suma de la ecuación (1.15a) y el parámetro de orden se vuelve cero.

En el transcurso del desarrollo, el sistema pasa a través de un gran número de arreglos energéticos entrópicos E_{μ} , donde el sistema se ve forzado a elegir uno, de entre un número macroscópico de diferentes conjuntos de arreglos microscópicos (un abanico de posibilidades que compiten por tener un reacomodo). En sistemas de espines de vidrio, ejemplificado por el $\text{Eu}_2\text{Sn}_2\text{S}_7$ y en concentraciones suficientemente altas de impurezas magnéticas, se observa una competencia entre un orden ferro o antiferromagnético, entrando y saliendo de estas transiciones, reorganizándose entre configuraciones débiles dentro de un gran número de estados metaestables. Estas son llamadas fases variables, o de espines de vidrio magnetizado. Una detallada discusión teórica de la manera en que ocurre este proceso, se encuentra más allá del horizonte de este trabajo [4]; pero una herramienta útil con la que se puede adquirir cierta experiencia sobre estos procesos de ordenamiento, es apelar a las técnicas de simulación computacional del modelo de Ising. En estas simulaciones normalmente se presenta un arreglo microscópico inicial del enrejado y el sistema genera un arreglo consecuente a través de aplicar un algoritmo (un conjunto de instrucciones) que lo van llevando a su estado final [6].

1.2 EQUILIBRIO EN DINÁMICAS NO LINEALES

1.2.1 EL EXPERIMENTO DE BÉNARD

(Turbulencia - Lenguaje de las Permutaciones, Fluctuaciones y Estabilidad)

El siguiente experimento simple, realizado por primera vez por Bénard, permite ilustrar la física de las perturbaciones, interacción competitiva y auto-organización.

Imagine una capa de fluido (digamos agua) limitada por dos placas paralelas horizontales, cuya dimensión longitudinal es mucho más grande que el ancho de las placas, separadas una distancia d . Dejemos libre al sistema y por él mismo, el fluido tenderá rápidamente a un estado en el cual, estadísticamente hablando, todas sus partes serán idénticas. Así, un diminuto observador

plantado en la base, viendo su medio ambiente, será incapaz de decir si él está dentro del pequeño volumen V_A , o del pequeño volumen V_B (ver Fig. 1.4). Es decir, todos los volúmenes que uno pueda definir arbitrariamente dentro del fluido, serán indistinguibles, ya que no existe ninguna diferencia entre ellos y el conocimiento del estado de uno cualquiera podría ser suficiente para conocer el estado de todos los demás, independientemente de su forma y de su tamaño.

En otras palabras, desde el punto de vista de nuestro observador no hay diferencia en la posición que él ocupa, ya que ahí no existe forma intrínseca que le pueda ayudar a percibir la noción del espacio. La homogeneidad de este sistema se extiende por supuesto a todas sus propiedades y en particular a su temperatura, que será la misma en todas partes del fluido e igual a la temperatura de las placas limitantes. El sistema se encuentra

caracterizado en un particular estado de equilibrio en el cual se cumple que, si T_1 y T_2 son las temperaturas de las placas 1 y 2, respectivamente, entonces:

$$\Delta T_c = T_2 - T_1 = 0 \quad (1.16)$$

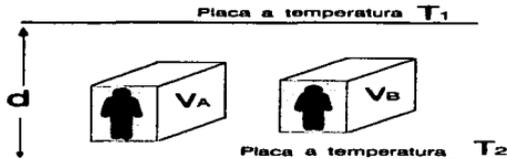


fig. 1.4 Un diminuto observador contemplando los estados elementales de volumen V_A , V_B en una capa de líquido en equilibrio, encontrará que son indistinguibles y concluye que el fluido muestra traslación invariante, a lo largo de la dirección horizontal.

Si alguien coloca un dedo sobre una de las placas por un momento, la temperatura en esta parte de la placa podrá ser momentáneamente modificada. Un incidente como éste, que modifica localmente algunas de las propiedades del sistema, es llamado una **perturbación**. Para nuestro sistema en equilibrio, esta perturbación de la temperatura, podrá no tener influencia, ya que la temperatura rápidamente llegará de nuevo a ser uniforme e igual a sus valores iniciales. En otras palabras, la perturbación desaparece y el sistema se recupera de ella. Cuando un sistema está en un estado tal, que la acción de la perturbación sobre éste, desaparece más o menos rápidamente en el tiempo, se dice que el estado es "estable".

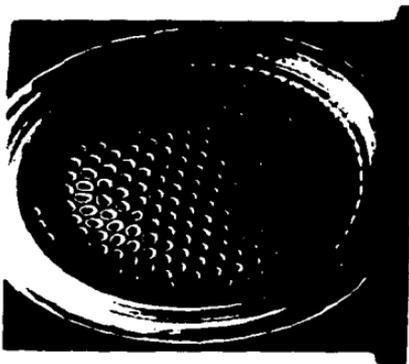
Desde el punto de vista de nuestro observador diminuto, no solamente la homogeneidad del fluido hace imposible que él desarrolle una concepción intrínseca del espacio, sino que además, la estabilidad del estado de equilibrio hace que todos los instantes sean idénticos. Por consiguiente tampoco se desarrolla en él una concepción intrínseca del tiempo.

Si el sistema se empieza a calentar desde abajo, la temperatura T_2 de la placa baja, será más alta que T_1 y la ecuación que expresa la condición de equilibrio (1.16) será violada ($\Delta T \neq 0$). En otras palabras, mediante la aplicación de una perturbación, no permitimos que el sistema alcance el equilibrio. Cuando la perturbación es débil (una ΔT pequeña), nuestro sistema irá cambiando hasta adoptar un estado simple y único, el cual está caracterizado por un proceso particular de transporte de calor desde la placa baja a la alta, en donde el calor será cedido al mundo externo. La diferencia en los posibles estados de estabilidad será la temperatura y en consecuencia, la densidad y la presión que no serán uniformes, pues variarán, prácticamente dentro de un "estilo lineal", desde regiones cálidas (abajo) a las frías (arriba). Este fenómeno es conocido como **conducción térmica**. En este nuevo estado que el sistema se ha buscado para responder o reaccionar a la perturbación, la estabilidad será precisada otra vez y el comportamiento eventualmente será descrito en forma tan "simple" como el que caracteriza al de equilibrio.

Al ir cambiando el sistema desde el equilibrio, una y otra vez, por medio de incrementos de ΔT , se observa que, de repente, en un paso de ΔT , que se llama crítico (ΔT_c), la materia comienza a realizar movimientos volumétricos turbulentos (fig. 1.5a); este movimiento está lejos de ser azaroso; pues el fluido se ha estructurado en una serie de pequeñas "celdas" (ver fig. 1.5b) conocidas como celdas de Bénard.

Considerando que el fluido cercano a la placa de abajo está caracterizado por una densidad menor que el más cercano a la placa superior tendremos, en consecuencia, el surgimiento de un gradiente de densidad el cual se opone a

la fuerza de gravedad. Esta configuración es potencialmente inestable, ya que si un pequeño volumen de fluido cercano a la placa inferior, es débilmente desplazado hacia arriba por una perturbación, aparece ahora en una región más fría y por tanto, siendo más densa, experimentará una fuerza de Arquímedes hacia arriba, la cual tenderá a intensificar el movimiento ascendente, aún más (volúmenes próximos que vienen atrás). Si por otro lado, una pequeña gotita inicialmente cercana al plato superior es desplazada hacia abajo, podrá penetrar a un ambiente de baja densidad, hundiéndose y aumentando la densidad inicial del fluido en la parte de abajo. Así se tiene una forma clásica de explicar, en principio, como el fluido puede generar corrientes ascendentes y descendentes ordenadas, como las observadas en el experimento. La razón por la cual estas corrientes no aparecen tan pronto como ΔT no sea estrictamente cero, podría deberse a que el efecto de inestabilidad está conectado con el efecto de estabilidad atribuido a la viscosidad del fluido, la cual genera una fricción interna opuesta al movimiento, (en pequeñas diferencias de ΔT la viscosidad gana y permanece una conducción térmica uniforme); además la conducción térmica tiende a "afectar" la diferencia de temperatura por causa de su desplazamiento, tanto del medio ambiente interno como externo.



(a)



(b)

Fig. 1.5 (a) Vista superior de convección en celdas de Bénard. (b) Vista esquemática mostrando el sentido en la rotación de los flujos. Note la dirección opuesta en la rotación en dos celdas adyacentes (L) = Rotación de lado izquierdo (R) = Rotación de lado derecho.

A partir de un umbral crítico T_c , se tiene que, siguiendo un pequeño volumen de fluido a través de su camino, éste va primero hacia arriba, después se mueve a lo largo del plato 1 y luego se dirige hacia abajo y se mueve a lo largo del plato 2 y va otra vez hacia arriba, etc. Es decir, el movimiento de la materia está estructurado en celdas, adoptando sucesivamente una rotación derecha (R) seguida de una rotación izquierda (L). Con esto, se puede decir que nos encontramos otra vez con el modelo de Ising.

Pero regresemos una vez más al diminuto observador. A este nivel, su universo ha sido totalmente transformado. Por ejemplo, él puede ahora percibir donde está y donde no está, por observar simplemente el sentido de rotación en la celda que ocupa. Por otra parte, por el conteo del número de celdas que atraviesa, él puede adquirir alguna noción eficiente de espacio. Es decir, emerge la noción de espacio en un sistema en el cual, hasta entonces, no podía ser percibido de ninguna manera, a esto se le nombra "rompimiento de la simetría".

Tal vez, la característica que debe ser enfatizada en esta transición repentina del comportamiento, es el haber pasado desde lo simple a lo complejo y observar un orden y coherencia de conjunto en este sistema. Cuando ΔT mantiene la temperatura abajo del valor crítico T_c , la homogeneidad del fluido en la dirección horizontal fue interpretada en cada una de sus diferentes partes independientemente de las otras. Carecía totalmente de importancia permutar los pequeños volúmenes V_A y V_B (fig. 1.4). En contraste, cuando se alcanza el umbral T_c , las cosas ocurren como si cada elemento del volumen estuviera "observando" el comportamiento de sus vecinos y lo estuviera "tomando" en cuenta para jugar su papel adecuadamente y participar en el patrón de ordenamiento total. Esto recuerda el concepto de correlación entre partes distantes del sistema (ya mencionado al tratar los espines, en el caso ferromagnético).

El experimento es perfectamente reproducible y realizándolo en las mismas condiciones experimentales, se puede ver siempre que el patrón de convección aparece en el mismo valor umbral T_c . Una vez que la dirección de rotación es establecida, ésta permanecerá así en cada celda, hasta que alguna perturbación suficientemente grande la obligue de nuevo a cambiar. No obstante que se tiene la posibilidad de dos situaciones cualitativamente diferentes que pueden originarse justamente después del umbral crítico T_c , sólo una aparece.

Tan pronto como una ΔT excede apenas delicadamente la T_c , las celdas podrán aparecer, por tanto, este fenómeno está sujeto a un determinismo estricto. En contraste, la dirección de rotación de la celda es impredecible e incontrolable. La forma particular de la perturbación que puede haber prevalecido justo al momento del experimento, decidirá si una celda dada

será derecha o izquierda. Llegamos así a una cooperación notable entre la oportunidad y el determinismo, la cual es una dualidad familiar en biología desde la era de Darwin (selección evolutiva, por mutación natural), que ha estado hasta ahora limitada a las ciencias biológicas.

Lejos del equilibrio - cuando la perturbación es suficientemente fuerte - el sistema puede ajustarse a su medio ambiente de varias maneras. El que varias soluciones sean posibles para los mismos valores de los parámetros, hacen, en consecuencia, que sólo por probabilidades se pueda decidir cuáles de estas soluciones se alcanzarán. El hecho de que entre varias alternativas sólo una haya sido retenida, confiere al sistema de una dimensión histórica, de alguna clase de "memoria" que le recuerde un evento pasado que tuvo lugar en un momento crítico, el cual afectará su evolución posterior.

La posibilidad para describir, a través de estos conceptos fundamentales, el comportamiento de ambos, seres vivos y sistemas físicos, no es algo ordinario, pues ni en el mejor desarrollo de hace unos años, pudo haber sido pronosticado.

1.2.2 SISTEMAS DINÁMICOS ATRACTORES

A lo largo de este capítulo nos hemos estado refiriendo a estados de equilibrio, a la estabilidad, las perturbaciones y las fluctuaciones, de una forma poco operacional. Ahora entraremos a un punto de vista más dinámico, apuntando a establecer la relación entre estos conceptos y las relaciones que gobiernan su evolución en el tiempo.

Considere un sistema que se desarrolla en un ambiente con el que está interactuando, a través de cambios de ciertas propiedades, las cuales llamaremos flujos (ver fig. 1.6); como resultado de estos intercambios, las variables propias del sistema $\{X_i\}$ describen el estado instantáneo en que se encuentra al transcurrir el tiempo. Estos estados se van desarrollando hacia valores que están típicamente caracterizados por un estado particular del medio ambiente $\{X_{i,e}\}$. La evolución en el tiempo de tal sistema puede ser determinada por un conjunto de coordenadas generalizadas. Así, un punto en este espacio, representa la condición instantánea del sistema. Este estado espacial, puede

ser ya sea continuo (una partícula moviéndose con amortiguamiento o fricción en un potencial con varios mínimos) o discreto (como en el caso de N espines de Ising). El sistema puede ser descrito por muchas coordenadas X_1, X_2, \dots, X_n , que pueden presentarse como las componentes de un vector de "estado" X , que tiene los puntos límites de arriba estables X^* . El sistema físico $\{X_i\}$ puede denotar temperatura, velocidad hidrodinámica, polarización eléctrica, etc. y en un sentido más general, pensando en el dibujo de la fig. 1.6, éste puede ser aplicado fuera del dominio exclusivo de la ciencia física. Como por ejemplo, a una sociedad humana, donde las X_i pueden representar la población de trabajadores ejerciendo diferentes clases de actividad económica o como en la dinámica de comunidades estudiada en ecología, donde la evolución en el tiempo de los estados sucesionales, dependen de los estados previos, etc.; pero sea cual sea la interpretación detallada de X_i , su evolución frecuentemente puede ser dada, desarrollándola en la siguiente forma general.

$$\frac{dX_i}{dt} = F_i(X_1, \dots, X_n; \lambda_1, \dots, \lambda_m), \quad (i = 1, \dots, n) \quad (1.17)$$

en la cual, F_i denota la ley, razón o costo y $\lambda_1 \dots \lambda_m$ son un conjunto de parámetros presentados en el problema, los cuales pueden ser modificados por el mundo externo. Nosotros llamamos a estas cantidades parámetros de control. Un rasgo característico importante, encontrado en la basta mayoría de los ejemplos de la naturaleza, es que las F 's a las que nos referimos, son complicadas funciones no lineales de X 's. Por ejemplo en un fluido en movimiento, éste tiene que ver con el hecho del transporte de sus propiedades tales como la energía, que es llevada por su propio movimiento y en una población animal o humana, la no linealidad puede reflejar el proceso de comunicación, competencia, crecimiento o el cambio de información.

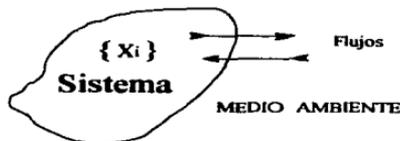
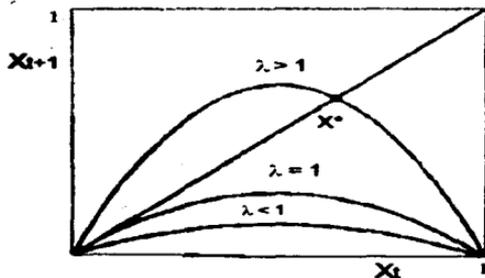


Fig. 1.6 Representación esquemática de un sistema abierto comunicándose con el medio ambiente a través de intercambios de propiedades tales como masa, energía, momento o información. La cantidad transportada por unidad de superficie y unidad de tiempo es el flujo de las correspondientes propiedades a través del sistema.



$$\begin{aligned}
 &0 < 1 - X_{t+1} < 1 \\
 &0 < 1 - \lambda X_t(1 - X_t) < 1 \\
 &0 < 1 - \lambda \frac{1}{2}(1 - \frac{1}{2}) < 1 \\
 &0 < 1 - \frac{1}{4}\lambda < 1 \\
 &\Rightarrow \\
 &\lambda < 4 \\
 &\therefore \lambda \in [0, 4]
 \end{aligned}$$

Fig. 1.7 Gráfica de la ecuación parabólica $X_{t+1} = \lambda X_t(1 - X_t)$ para tres diferentes valores de λ . Note como se eleva y contrae conforme λ toma valores comprendidos en el intervalo $[0, 4]$.

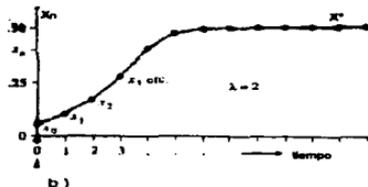
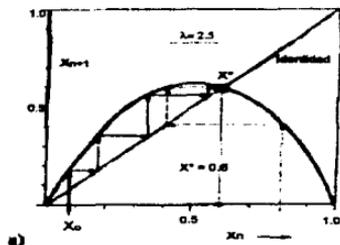


Fig. 1.8 Curva logística. a).- Construcción geométrica para llegar a la solución X^* cuando $\lambda = 2.5$ b).- Ilustración gráfica de la aproximación asintótica al estado estable X^* con $\lambda = 2$.

Para ilustrar algunas de las propiedades generales de los sistemas no lineales que hasta aquí se han venido manejando y aclarar un poco más los conceptos en torno a ellos, tomaremos uno de los ejemplos más sencillos que la ecología nos brinda: El crecimiento logístico de una población desarrollándose en un ambiente dado. Este es modelado a través de ecuaciones en diferencias finitas, que se definen como entes matemáticos que tienen la forma $X_{i+1} = f(X_i)$; esto es, una sucesión de estados posibles que comienzan desde alguna condición inicial y se desenvuelven en el tiempo, donde el estado actual del sistema depende de la historia anterior. Muy concretamente hablaremos de la ecuación $X_{i+1} = \lambda X_i (1 - X_i)$, para $0 \leq X_i \leq 1$, donde λ juega el papel de parámetro de control, el cual puede tomar valores entre $0 \leq \lambda \leq 4$.

Este ejemplo aparentemente ajeno, exhibe propiedades genéricas, que pueden ser relacionadas tanto con el experimento de Bénard como con los cambios de transición de fase ocurridos en los puntos críticos que hemos discutido anteriormente, lo que haremos notar más adelante.

Observe el punto $X^* = f(0) = 0$; $X^* = \lambda 0 (1 - 0) = 0$; es decir no importa cuanto valga λ la función siempre estará en el origen, esto es si X^* es un punto fijo entonces debe cumplir $X^* = f(X^*)$.

Vea como el origen es un punto fijo para toda λ . ¿Pero habrá algún otro punto fijo más?; si los hubiese, estos puntos fijos deben satisfacer la condición $X^* = f(X^*) = \lambda X^* (1 - X^*) \Rightarrow X^* = 1 - 1/\lambda$ donde X^* debe pertenecer a $X_i \in [0, 1]$ y por lo tanto $0 < 1 - 1/\lambda < 1$. Así que el nuevo punto fijo X^* permisible, se presenta sólo cuando λ es mayor que uno, $\lambda \geq 1$; con $\lambda \in [1, 4]$. Veamos este resultado en un análisis gráfico (Fig. 1.7). Para $\lambda < 1$ la parábola yace por abajo de la diagonal de la identidad y el origen es el único punto fijo. Conforme λ se va incrementando, la parábola se va haciendo más alta, llegando a ser tangente con la diagonal en $\lambda = 1$. Para $\lambda > 1$ la parábola intercepta la diagonal en un segundo punto fijo $X^* = 1 - 1/\lambda$ y el origen pierde estabilidad.

¿Qué estabilidad pueden tener estos puntos fijos que hemos encontrado? Para averiguarlo situémonos un momento en el origen y supongamos que en un cierto momento se presenta una fluctuación provocada por una pequeña perturbación que nos saca un poquitito del origen (tal vez la emigración de unos cuantos animales que han sobrevivido después de haber cruzado una barrera geográfica y por tanto pueden llegar a ser los primeros colonizadores en una cierta área).

Desde un punto de vista gráfico y apoyados en la fig. 1.8.a donde $\lambda = 2.5 > 1$, observamos que el punto con coordenadas $(X_0, f(x_0))$, en vez de regresar al

estado de equilibrio localizado en el origen, se aleja rápidamente de él (es lo que se conoce como punto de equilibrio inestable). Es decir a partir de un estado inicial que no es el origen, la evolución del sistema se desarrolla aproximándose monótonamente a un estado de mayor estabilidad, conforme $t \rightarrow \infty \Rightarrow X(t) \rightarrow X^*$. En estas condiciones se garantiza que cualquier trayectoria que comience a una distancia dada de X^* converge a X^* conforme $t \rightarrow \infty$.

La razón de cambio, o la rapidez con que dos estados cercanos dados, se acercan o se alejan entre sí, nos lo dá la derivada, es decir la estabilidad de un punto fijo depende del multiplicador $f'(x^*) = \lambda - 2\lambda x^*$. Así pues, para $f'(0) = \lambda$, de la discusión anterior, el origen es estable siempre y cuando $\lambda < 1$ e inestable para $\lambda > 1$. En relación a los otros puntos fijos tenemos que $f'(x^*) = \lambda - 2\lambda(1 - 1/\lambda)$; $f'(x^*) = 2 - \lambda$, por tanto el punto fijo $X^* = 1 - 1/\lambda$ es estable para $-1 < (2 - \lambda) < 1$, es decir para $1 < \lambda < 3$ y es inestable para $\lambda > 3$.

Así podemos decir que el punto fijo X^* es un estado de equilibrio

- i.)- **Atractor**, si sucede $|f'(x^*)| < 1$
- ii.)- **Repulsor**, si $|f'(x^*)| > 1$
- iii.)- **Indiferente o Metaestable**, si $|f'(x^*)| = 1$
- iv.)- **Super Atractor** si $|f'(x^*)| = 0$

y en una forma más general definiremos un atractor como un conjunto cerrado A con las siguientes propiedades:

- 1.- A es un conjunto invariante: cualquier $X(t)$ que arriba a A , permanece en A por todo el tiempo.
- 2.- A atrae a un conjunto abierto U contenido en A tal que si $X_0 \in U$, entonces la distancia desde $X(t)$ a A tiende a cero cuando $t \rightarrow \infty$. Esto significa que A atrae a todas las trayectorias que arrancan lo suficientemente cerca del atractor. Lo más grande que la tal U puede alcanzar es llamado la base o cuenca de atracción de A .
- 3.- A es minimal, es decir, no hay subconjunto propio de A tal que satisfaga las condiciones 1 y 2; por lo tanto no puede haber un atractor dentro de otro atractor.



Fig. 1.9 Secuencia o cascada de bifurcaciones. El periodo se va doblando y con ello el número de estados estables antes de alcanzar un régimen caótico.

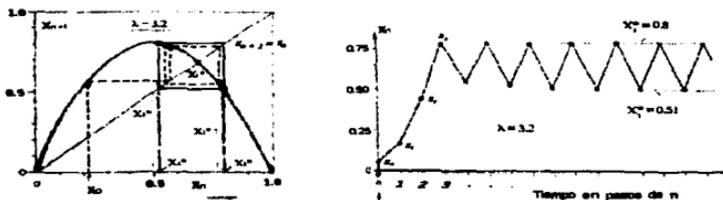


Fig. 1.10 Si λ es incrementado más allá de un valor crítico $\lambda_1 > 3$ entonces el valor de los estados sucesivos $\{X_n\}$, salta periódicamente entre dos valores, antes de rebasar otro valor crítico de transición.

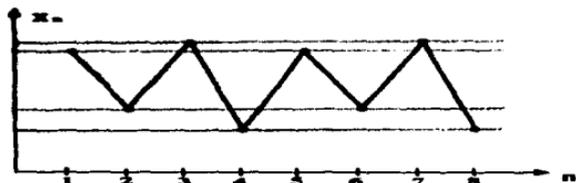


Fig. 1.11 Salto periódico de X_n a través de cuatro valores.

En la Fig. 1.9 vemos que X^* se va bifurcando; primero sólo tenemos el origen como único arribo posible, alcanzado después de un cierto tiempo, cuando $\lambda < 1$, es decir en el intervalo $\lambda \in [0, 1]$. Aquí el último estado alcanzado es lo que se llama (en el caso ecológico) extinción. Ahora bien en $\lambda = 1$ una bifurcación transcítica ocurre; pues si λ adquiere un valor más allá de 1, aparece en escena X_1^* , donde para estos valores de λ , el origen ha pasado a ser un punto fijo de equilibrio inestable y el punto fijo X_1^* que se ha originado, ha nacido como un punto de equilibrio atractor estable, en el caso ecológico adaptación natural. Si vamos deslizando λ a través del intervalo $\lambda \in [1, 3]$, la escena se mantiene cualitativamente más o menos igual pero en el límite de transición, cuando λ se hace ligeramente mayor que 3 aparecen dos nuevos puntos fijos más X_2^* y X_3^* con equilibrio estable, mientras X_1^* pierde estabilidad y se convierte en un punto fijo de equilibrio inestable repulsor. A la bifurcación que ha resultado se le llama una bifurcación excitada de doble ciclo. (ver fig. 1.10, así como el apéndice C).

Para conocer el intervalo de la estabilidad de los puntos de equilibrio que han surgido $X_2^* = p$ y $X_3^* = q$ calcularemos el multiplicador:

$$\frac{d}{dx} (f(f(x)))_{x=p} = f'(f(p)) f'(p) = f'(q) f'(p)$$

y lo resolveremos para $|f'(q) f'(p)| < 1$

Note que el mismo multiplicador es obtenido en $X = q$, esto es por la simetría del término final de arriba. Por tanto, cuando p y q se ramifican en una bifurcación, ellos lo deben de hacer simultáneamente.

Las diferenciales para p y q son:

$$f'(q) = \lambda - 2\lambda q \qquad f'(p) = \lambda - 2\lambda p$$

por lo que

$$\begin{aligned} f'(q) f'(p) &= (\lambda - 2\lambda q)(\lambda - 2\lambda p) \\ &= \lambda(1 - 2q)\lambda(1 - 2p) \\ &= \lambda^2 [1 - 2(p+q) + 4pq] \end{aligned}$$

y como la suma y el producto son:

$$p + q = \frac{\lambda + 1}{\lambda} \qquad \text{y} \qquad pq = \frac{\lambda + 1}{\lambda^2} \qquad (\text{ver apéndice C})$$

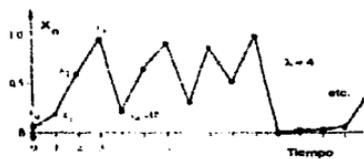
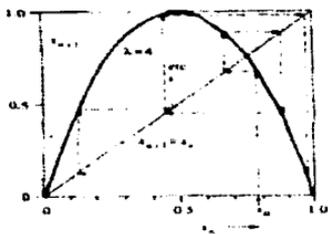


Fig. 1.13 Caos.

$$\begin{aligned} f'(q) f'(p) &= \lambda^2 \left[1 - \frac{2(\lambda + 1)}{\lambda} + \frac{4(\lambda + 1)}{\lambda^2} \right] \\ &= \lambda^2 + 2\lambda(\lambda + 1) + 4(\lambda + 1) \\ &= 4 + 2\lambda - \lambda^2 \end{aligned}$$

Así el doble ciclo es linealmente estable para $|4 + 2\lambda - \lambda^2| < 1$ es decir para $3 < \lambda < 1 + \sqrt{6}$.

Los subsiguientes resultados son difíciles de sacar de manera analítica, por tanto, para conocer el comportamiento en la región cuando λ crece aún más, generalmente se apela a los métodos numéricos y a argumentos gráficos. De hecho, los resultados obtenidos hasta aquí se pueden confirmar fácilmente en forma numérica con una calculadora de bolsillo.

Así, si incrementamos λ para alcanzar un nuevo valor crítico λ_2 , los puntos X_n terminan en una secuencia la cual es sólo repetida después de 4 pasos (fig. 1.11). Vemos que se ha duplicado el periodo, pues con respecto al periodo dos anterior, el valor ha sido doblado.

Si vamos incrementando λ más y más, en una secuencia de valores críticos λ_k (donde k es un paso de tiempo), el periodo se dobla en cada ocasión, siendo así como se va construyendo la fig. 1.9, la cual muestra una secuencia de bifurcaciones con una regla de evolución de doble periodo.

El periodo se dobla en una secuencia de valores críticos $\lambda_1, \lambda_2, \dots$ donde, más allá de un valor crítico λ_c , el desenvolvimiento llega a ser más y más irregular hasta convertirse en caótico (fig. 1.9 y 1.12). Cada bifurcación, ilustra las ramificaciones posibles o alternativas de soluciones para la ecuación no lineal. Físiamente, ello denota una transición de fase desde un estado de equilibrio, a un nuevo y posible estado de equilibrio. Y si el equilibrio es entendido como un estado de simetría, entonces la transición de fase significa un rompimiento de simetría, la cual es causada por fuerzas de fluctuación. Un sistema con un orden total, tiene la más alta de las simetrías, todo homogéneo, todo igual, estructuras idénticas que se repiten como las de un cristal. Sin embargo este orden se puede perder a través de bifurcaciones (transiciones). Se dice que la alta simetría que corresponde a un alto orden se pierde en cada transición de fase, adquiriendo estados los cuales tienen menos simetría y menos orden (por ejemplo, el paso de sólido a líquido y a vapor en el caso del agua ya mencionado). Otro ejemplo lo encontramos en el caso del experimento de Bénard donde la capa de fluido calentado desde abajo llega a ser inestable y el estado de convección estacionario desarrolla círculos (fig. 1.5).

Este caso significa una transición de fase con rompimiento de simetría porque una fluctuación pequeñísima causa que el sistema adopte alguna de las dos direcciones posibles. El ejemplo muestra que la transición de fase y el rompimiento de simetría son causados por un cambio de parámetros externos que conducen eventualmente a un nuevo patrón espacio-temporal de carácter macroscópico de simetría y de orden emergente inesperado.

1.2.3 SIGNIFICADO DE AUTO-ORGANIZACIÓN.

Una de las propiedades que emergen de las secciones anteriores es la auto-organización, la cual descansa sobre la habilidad de los sistemas dinámicos no lineales, contruidos por millones y millones de elementos interactuando para crear y sostener (cuando ellos están alejados del equilibrio), estados de la materia que exhiben regularidades, las cuales serían extremadamente difíciles de estudiar bajo condiciones individuales. Estos sistemas con muchas interacciones que se desenvuelven hacia estados críticos, se han venido estudiando a través de considerar sus elementos separadamente y con mecanismos microscópicos individuales y casi siempre utilizando técnicas lineales, asumiendo que la respuesta de estos sistemas, es proporcional a la perturbación.

Intuitivamente, un ejemplo lo podríamos encontrar pensando en un grupo de trabajadores, o más exactamente en el comportamiento organizado de ellos, si cada trabajador actúa en una forma bien definida bajo unas órdenes externas dadas por el jefe. Se entiende que el comportamiento así regulado resulta en una acción conjunta para producir algo. Si en el mismo proceso no hay órdenes externas dadas a los trabajadores y trabajan juntos por algún tipo de entendimiento mutuo, cada uno hace su trabajo y así colabora en el resultado de producir algo, estaríamos hablando de auto-organización. Las órdenes dadas por el jefe, son la causa de las acciones subsecuentes de los trabajadores.

La auto-organización es un comportamiento conjunto, de características emergentes generales, poco o nada dependiente de la individualidad de la mecánica microscópica, donde las características globales del sistema no pueden ser entendidas por analizar sólo las partes separadamente. Son sistemas no lineales donde eventos mínimos provocan reacciones grandes. Su dinámica busca alcanzar estados estables motores, que surgen de la turbulencia global del comportamiento cooperativo de miles de unidades actuando en una armonía y en un orden.

CAPITULO DOS

Metodología de las redes neuronales artificiales.

CONCEPTOS NEURONALES BÁSICOS

2.1 Procesamiento de datos en una red neuronal.

- 2.1.1 Introducción.
- 2.1.2 La neurona biológica. Unidad básica del sistema nervioso.
- 2.1.3 Desarrollo de auto - organización neuronal.
(Postulado neurofisiológico de Hebb).

2.2 Cálculos colectivos y redes neuronales.

- 2.2.1 Relación entre dinámica de atractores y arquitectura de redes neuronales artificiales.
- 2.2.2 Un enrejado enfocado como una capa neuronal artificial donde se promueve la auto - organización.
- 2.2.3 Redes multicapas de alimentación progresiva.

CAPITULO DOS

CONCEPTOS NEURONALES BÁSICOS

2 CONCEPTOS NEURONALES BASICOS

En este capítulo se establece la analogía entre los conceptos físicos expuestos en el capítulo uno y las redes neuronales, estableciendo así los modelos que permiten entender su funcionamiento a primera aproximación y nos guían en el camino hacia la construcción de una red artificial y en particular la que se conoce como de multicapas de alimentación progresiva.

Se muestra cómo, utilizando el postulado de Hebb dentro de nuestro marco físico, emerge una potencialidad de cálculo en dichos modelos, expresada por la capacidad de aprender y por la forma en que adquieren habilidades parecidas a las mentales, tales como memoria asociativa y reconocimiento de patrones.

La relación con las redes de alimentación progresiva se establece definiendo una neurona artificial, que se sustituye o se coloca dentro de una celda del modelo de Ising; con estas unidades, que se llaman perceptrones, se forman capas, cúmulos o aglomeraciones neuronales, que se interconectan entre sí para interactuar y formar arquitecturas multicapas donde la información va pasando progresivamente de capa en capa.

2.1 PROCESAMIENTO DE DATOS EN UNA RED NEURONAL.

2.1.1 INTRODUCCIÓN

Es sabido que el más eficiente sistema en procedimientos de análisis y procesos de cómputo es el cerebro humano. Pero sin ir demasiado lejos, tenemos que sistemas biológicos simples ejecutan manipulaciones de señales más allá de las capacidades alcanzadas por nuestros más modernos y sofisticados sistemas.

Los procesos de control y los mecanismos de transmisión de información que los seres vivos utilizan en sus actividades cotidianas, superan a las técnicas más avanzadas creadas por el hombre. Por tal motivo, muchas investigaciones motivadas por tal eficiencia, tratan de imitar, de alguna manera, los sistemas biológicos, para resolver una variedad de problemas complejos que la naturaleza nos presenta cotidianamente. En este sentido, la física contribuye desarrollando métodos con los cuales se puedan construir herramientas para encontrar soluciones empíricas a estos problemas, en forma análoga como lo haría un sistema biológico y al mismo tiempo, al crear artificialmente algunas de las propiedades de la mente (reconocer formas, recordar, aprender, asociar, clasificar patrones, evocar imágenes, ver, hablar, etc.), se ha empezado a tener un conocimiento teórico formal del funcionamiento de estos procesos y aunque se está aún muy lejos de entenderlos por completo, los intentos de imitarlos han arrojado valiosos conocimientos, que han tenido gran aceptación en lo que se conoce hoy en día como neurociencias, generando una retroalimentación entre la biología y la física.

La célula es la estructura básica de cualquier ser viviente y es la célula neuronal la unidad básica del tejido nervioso ó neuronal. Como cualquier célula, posee una membrana, la que separa el medio interior del exterior. Ésta juega un papel de fundamental importancia en una de las actividades de la neurona, la actividad eléctrica. La membrana posee ciertos estados físicos que le permiten controlar la carga eléctrica existente entre el interior y exterior de ella y establece corrientes eléctricas (flujo de iones) que causan cambios en el potencial a través de ella, los que son observados como pulsos de voltaje y llamados ~~potenciales de acción~~. Los pulsos o disparos llevan información, la cual se puede ver como una secuencia de pulsos (un tren), con determinada frecuencia, definida por el tiempo en que el potencial de membrana de la neurona transmisora es capaz de regenerarse. Los pulsos viajan a lo largo de una fibra larga llamada ~~axón~~, distribuyéndose a través de ramificaciones, para alcanzar otras neuronas.

Las neuronas también actúan como receptores de información, recibiendo los pulsos que le entran por los sitios denominados ~~sinapsis~~, donde se realiza el contacto entre neuronas. Esta información se incorpora a la que proviene de diferentes sinapsis, procedentes de otras neuronas. Dentro de la neurona la

información es procesada, resultando en la producción o no, de una respuesta o reacción que la neurona transmitirá a sus compañeras, repitiéndose el proceso en tal forma, que la información se propague por la red neuronal, desarrollándose así un trabajo conjunto, auto-organizado, del procesamiento de la información para dar una respuesta global, sobre la solución de un problema o la ejecución de una tarea.

La frecuencia de disparo de una neurona está determinada en primera aproximación, por el efecto acumulativo de un gran número de sinapsis excitatorias e inhibitorias al cuerpo celular en un corto intervalo de tiempo. Es decir la combinación de entradas excitatorias e inhibitorias podrán variar la razón de disparo, a través del arribo de un pequeño número de neurotransmisores excitatorios e inhibitorios que afectan la frecuencia de disparo de la neurona. Parecido a como desarrollan el trabajo las neuronas en los sistemas biológicos, las redes artificiales, al imitar estos procesos, emplean configuraciones o circuitos con arquitecturas que ejecutan procesamientos de datos en paralelo y por tanto no sufren del problema secuencial, que frecuentemente es un cuello de botella en las computadoras convencionales. Otro punto importante que vale la pena señalar, es el que se refiere explícitamente a los algoritmos o a los métodos rígidos con los cuales tradicionalmente se resuelven los problemas. Opuesto a esto, las redes neuronales artificiales tienen la habilidad de aprender a resolver los problemas con base en la experiencia, como una función de aproximación dependiente de las excitaciones de entrada.

2.1.2 LA NEURONA BIOLÓGICA. UNIDAD BÁSICA DEL SISTEMA NERVIOSO.

Cuando a Frank Rosenblatt, creador del Perceptrón, unidad utilizada en casi cualquier red neuronal artificial (R.N.A.), se le preguntaba como es una neurona, él solía responder: ¿acaso hay alguien que sepa como es un electrón? Pues bien, aunque nadie sabe lo que es en realidad una neurona, el conocimiento que se tiene de ella es bastante amplio, pero sin embargo en esta ocasión, sólo haremos una breve reseña de la célula nerviosa, con el fin de familiarizarse con algunos aspectos biológicos, principalmente los referidos a sus propiedades electroquímicas, su dinámica y sus interconexiones (sinapsis), ya que muchas veces, a pesar de realizar el análisis matemático de la dinámica de algún modelo neuronal en forma correcta, no se alcanza a interpretar el significado y alcance de las ideas físicas involucradas, o se pasan por alto; caso concreto fue el sucedido a N. Wiener en 1948 quien desarrolló, de forma

excelente y clara, a través de teoría de grupos y con mecánica estadística [7], la dinámica colectiva de un gran número de unidades relativamente simples; su trabajo fue de poco interés científico hasta que fue retomado 30 años más tarde por John Hopfield [8], quien conjugó estas ideas en una interpretación biológica, dándole formalidad matemática a términos biológicos, resaltando cualidades comunes, relacionando términos aparentemente distintos y efectuando analogías y equivalencias que enriquecen ambos puntos de vista [9].

Así, empezaremos refiriéndonos a la maravillosamente sorprendente acción compleja de la ~~membrana~~ a la que se le atribuye la responsabilidad de la actividad celular, para producir y transmitir tanto señales químicas, como eléctricas. Las propiedades de la membrana suelen encontrarse principalmente, dentro de cualquiera de las tres clases de los siguientes mecanismos o estructuras características: bombeo, canales y receptores, las que iremos tratando de manera superficial en el momento adecuado, a medida que vayamos avanzando.

Dado que tanto el medio extracelular como intracelular son ricos en iones (Na^+ , Cl^- , K^+ , Ca^{++} , entre otros) y que éstos se encuentran en distintas concentraciones, se produce una diferencia de voltaje, entre el interior y exterior de la célula, conocido como **Potencial de Membrana**, que en estado de reposo es típicamente una diferencia de voltaje de 60 - 90 mV, presentado entre las paredes interiores y exteriores de la membrana celular (el interior negativo con respecto al exterior). En este estado de reposo de la neurona, la membrana celular es muy poco permeable al Na^+ ; sin embargo, el coeficiente de permeabilidad de la membrana al Na^+ (P_{Na}) puede incrementarse y esto sucede cuando la célula neuronal está descargándose o produciendo su potencial de acción. Si la permeabilidad de la membrana cambia, entonces la caída de voltaje (V), entre el interior y exterior de la célula cambia. Durante el estado de reposo la neurona, presenta: $P_{\text{Na}} \approx 0$ y $V \approx -75$ mV, pero en el pico del potencial de acción $V \approx +50$ mV y la P_{Na} está elevada.

Durante cualquier potencial de acción que se produzca, hay un flujo temporal de Na^+ entrando a la célula, que incrementa el voltaje de membrana, seguido de un flujo saliente de K^+ que restablece el potencial del estado de reposo. Para mantener las concentraciones iniciales, un sistema llamado bomba de sodio-potasio, continuamente saca sodio del interior de la célula y mete potasio a la célula, recuperándose el gradiente de concentración después de que el potencial de acción se ha producido. Cada bomba mueve alrededor de 200 átomos de sodio y 130 iones de potasio por segundo. Una neurona puede tener millones de tales bombas moviendo cientos de millones de iones hacia dentro y fuera de la célula cada segundo. Estas moléculas orgánicas son mucho muy complejas pero literalmente bombean Na^+ hacia fuera y K^+ hacia dentro, en contra de sus gradientes de concentración por medio de secuencias de reacciones metabólicas y químicas que consumen obviamente energía.

Una célula que está disparando (produciendo un potencial de acción), reduce su nivel de potencial, comenzando desde un valor aproximadamente de -75 milivolts, hasta que súbitamente durante el disparo, el sodio y el potasio fluyen en sentidos contrarios, para que después de unos milisegundos, el interior de la célula llegue a ser aproximadamente unos 50 milivolts, positivo con respecto al exterior. Esta polaridad contraria, es responsable del impulso nervioso que corre a lo largo del axón hasta alcanzar las conexiones presinápticas (el axón es considerado como una línea de transmisión que puede tener ramificaciones que transportan mensajes de una zona a otra). Cuando el impulso llega a la terminal del axón, se encuentra en una zona neuronal que contiene grupos moleculares altamente especializados y que reaccionan a determinadas sustancias químicas.

Las **sinapsis** (puntos de comunicación) que normalmente ocurren entre las terminales de los axones y las dendritas, son básicamente de dos tipos: la eléctrica y la electroquímica. En las primeras, la señal se transfiere por medios eléctricos, ya que en este caso existe un contacto físico-eléctrico entre ellas. En la sinapsis electroquímica se encuentra un espacio entre las células involucradas, al que se le denomina espacio intersináptico. Éste impide el paso directo del potencial eléctrico, por lo que la transmisión de la información en esta sinapsis, se realiza por medio de sustancias químicas, los llamados neurotransmisores, que son liberados por medios eléctricos y son responsables de provocar variaciones eléctricas, en la neurona receptora.

Después de que un potencial de acción es generado, la célula entra en un periodo de relajamiento (de varios milisegundos en células cardíacas), durante el cual ésta regresa a su potencial de reposo y se prepara para generar otro pulso o disparo. Resumiendo este proceso, diremos que, la recepción de las primeras moléculas neurotransmisoras inician una disminución de aproximadamente entre 10 a 15 milivolts en la diferencia de voltaje de la célula en el reposo. En este momento los canales de sodio (Na^+) controlados por el voltaje son abiertos, permitiéndole al sodio ingresar al interior de la célula, provocándose con esto una nueva reducción del potencial. Al incrementar el flujo de sodio se crea un proceso que va creciendo y reforzándose por él mismo. Este proceso lleva a la célula a alcanzar un potencial positivo contrario al inicial que era negativo con respecto al exterior. Un corto tiempo después, los canales de sodio se cierran y los canales de potasio se abren, obteniéndose un flujo de potasio hacia afuera de la célula y los -75 milivolts del potencial interno son recobrados. La célula ahora está lista para la generación del próximo potencial de acción, el que se propaga rápidamente a lo largo de todo el axón (Fig. 2.1).

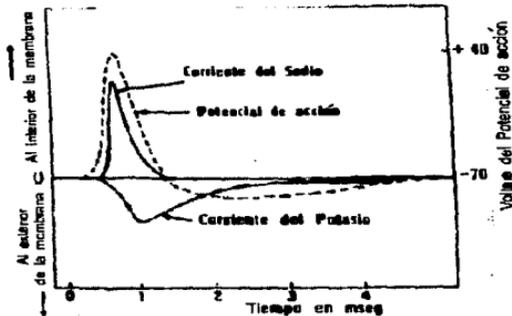


Fig. 2.1 Corrientes iónicas asociadas con los cambios de voltaje del potencial de acción. Las líneas sólidas muestran la corriente del sodio hacia el interior de la célula y la corriente del potasio hacia el exterior, las cuales resultan de cambios en la conducción de la membrana, a estos iones (ordenada izquierda). La línea punteada es el cambio de voltaje del potencial de acción, resultado de tales corrientes.

Los canales de sodio y potasio, responden a los cambios de potencial de la membrana, por ello se dice que pueden ser vistos como compuertas de voltaje. Otro tipo de canales son compuertas químicas, que se abren sólo cuando una molécula de un neurotransmisor específico está presente y éstos son casi insensibles al voltaje, tales canales son encontrados en las conexiones postsinápticas, sobre las dendritas. El compuesto acetilcolina al juntarse con un receptor, es uno de tales canales de compuerta química, que por incrementar positivamente el potencial de membrana, es considerado, como una excitación a que la célula dispare. Otros canales con compuerta química, sólo dejan pasar iones potasio hacia afuera de la célula, produciendo un decremento del voltaje. Por incrementar el potencial de membrana negativamente estos pulsos son vistos como **inhibitorios**, ya que ellos contribuyen a que la célula no dispare. El ácido gamma-aminobutírico es uno de los más comunes neurotransmisores inhibitorios. Estas moléculas activan mecanismos para incrementar el potencial negativo del interior de la célula y por lo tanto inhiben el disparo.

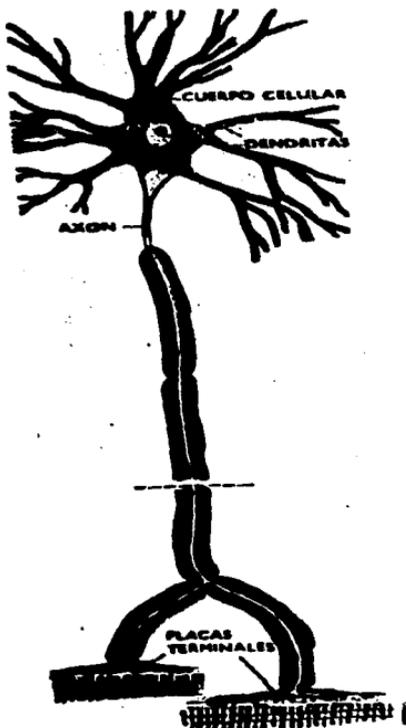


Fig. 2.2 Esquema de la célula nerviosa. Sobre las terminaciones de las dendritas se hacen las sinapsis con otras neuronas.

2.1.3 DESARROLLO DE AUTO-ORGANIZACIÓN NEURONAL

En los organismos biológicos existen sistemas que reconocen y responden a la variabilidad de atributos del medio ambiente que les rodea. La información recolectada entra por los sensores receptores, que están especializados para responder a diferentes formas de energía: magnética, térmica, calórica, química, luminica, mecánica, acústica, etc. Las estructuras específicas que ejecutan esta función son los órganos sensoriales que cuentan con prolongaciones nerviosas especializadas o adaptadas para transformar la energía que reciben, en señales eléctricas, provocando trenes de potenciales de acción que llegan a los centros nerviosos (conjunto de neuronas) donde se ejecutan procesos de recepción y evaluación de la información mucho muy complejos. Estos mecanismos se han tratado de entender a través de la historia del conocimiento humano, pero sin embargo, no es sino en años recientes, cuando en 1949 a través del libro titulado "The Organization of Behavior" escrito por Donald Hebb [10], en que se logra establecer un fundamento neuronal, surgiendo la teoría conexionista que ha permitido desarrollar la dinámica de aprendizaje en redes neuronales artificiales (R.N.A.).

El aprender y el reconocer formas y figuras, a partir de memorizar rasgos característicos para su identificación, es realmente una función de excepcional importancia de nuestro sistema nervioso, no sólo por nuestra habilidad de adaptación a variaciones ambientales imprevistas, sino también, por las tareas permanentes que tenemos que ejecutar, como son los procesos de la visión, del habla, etc., pues ellas son demasiado complicadas para ser genéticamente codificadas en su totalidad y tienen que ser adquiridas por aprendizaje en la temprana infancia.

La regla de aprendizaje de Hebb, como algunas veces es conocida, es muy usada hoy en día y es en esencia un proceso de aprendizaje por repetición. Se trata de un mecanismo repetitivo para inducir una respuesta, utilizando diferentes clases de patrones presentados persistentemente y en secuencias, tantas veces como sea necesario hasta que el aprendizaje, llamado hebbiano, se vea robustecido y la respuesta sea satisfactoria. A través de él, se va provocando una auto-organización neuronal, para condicionar la información y conseguir un reflejo a los cambios producidos en el medio ambiente, basados en la experiencia.

En 1949, Hebb en la pág. 62 de su libro [10], escribió el ahora famoso postulado neurofisiológico:

Cuando un axón de una célula A está lo suficientemente cerca para excitar a una célula B repetitivamente o persistentemente, entonces induce o toma parte en el disparo de ésta, activando algún proceso de crecimiento de respuesta metabólica que toma lugar en una o en ambas células tal que, la eficiencia de A como la eficiencia del disparo de la célula B, es incrementada.

¿dónde puede tener lugar el proceso de crecimiento de respuesta metabólica? Hebb sugirió un "tirador" sináptico de desarrollo y continuó diciendo: Si un patrón de estimulación provoca en ambas neuronas cercanas "i" y "j" disparos a una razón máxima y ellas están conectadas a través de una sinapsis excitatoria, tenemos entonces que esta conexión tiene más oportunidad de llegar a establecer la fuerza sináptica que caracterice al patrón de excitación, que otras señales entrando a través de otras sinapsis, las que tienen menos posibilidad de afectar el desarrollo de una respuesta metabólica asociada con este patrón.

Hebb a través de su libro nos lleva a pensar que la localización de la memoria, o la codificación en el sistema nervioso, debe hacerse por algún mecanismo biológico que facilite el desarrollo metabólico con lo cual las sinapsis excitatorias o inhibitorias, entonces expuestas a disparos simultáneos en ambas direcciones, ya sea transmisores o receptores se vayan tornando fuertes (eficientes), hasta adquirir una cierta calidad sináptica global. Esto nos permite implementar un aprendizaje, a través de ponderar el cambio sináptico; construyendo y mejorando esta calidad sináptica dentro de un gran número de conexiones, hacia una respuesta integral.

Aunque esto no es del todo cierto, sino que el problema es mucho más complicado de lo que se está diciendo, efectivamente como sabemos ahora, es en las sinapsis donde tenemos la mayor oportunidad en codificar y decodificar de alguna manera y en forma coherente, la información. Por tanto, el sistema neuronal debe medir y registrar los diferentes estímulos presentados por los distintos patrones, de tal forma que puedan ser representados principalmente a través de los cambios de la fuerza sináptica J_{ij} . Es decir las numerosas conexiones formadas entre las neuronas adquieren, desde este punto de vista, un papel activo sobre algún patrón externo a ser guardado en la red, a través de activar o provocar ciertos disparos (excitatorios o inhibitorios), que codifiquen la información utilizando para ello, un parámetro al que llamaremos la fuerza sináptica. Así se puede condicionar un auto-reflejo instantáneo adecuado o la

recuperación fiel del patrón, una vez que se logre almacenar o archivar de manera coherente, la información dentro de la red.

Hebb formuló sus ideas en claras palabras coloquiales, así que uno de los principales problemas que dejó, es la realización de la implementación matemática de sus ideas.

Normalmente se manejan las siguientes expresiones diciendo:

Una neurona, digamos "i", instantáneamente notifica lo que le sucede a una neurona "j", así que la acción de "j" vía la eficacia sináptica J_{ij} , como una experiencia de "i" en el tiempo t, es descrita como:

$$\text{Acción de } j = J_{ij} S_i(t) \quad (2.1)$$

donde $S_i(t)$ es la actividad de la neurona "i".

Ahora, si damos a J_{ij} un valor de eficacia sináptica antes de la sesión de aprendizaje cuya duración es T, tenemos que después de la sesión de aprendizaje, J_{ij} debe ser cambiada a $J_{ij} + \Delta J_{ij}$, con lo cual tenemos la siguiente expresión:

$$J_{ij} = J_{ij} + \Delta J_{ij} \quad (2.2)$$

donde ΔJ_{ij} es la experiencia adquirida durante la sesión de entrenamiento, la cual debe ser dirigida a través de una dinámica de aprendizaje, entre las que se encuentran, como las más populares por ser más eficientes, 1.) las que utilizan aproximaciones por retropropagación del error, para provocar las reacciones o respuestas deseadas y 2.) la de atractores o memoria asociativa a la Hopfield.

2.2 CÁLCULOS COLECTIVOS Y REDES NEURONALES .

Cuando se dice que el corazón es mucho muy parecido a una bomba, metiendo y sacando sangre a través de sus válvulas, o al comparar el estómago con un reactor químico, o al ojo con una cámara de televisión, uno queda más o menos satisfecho, pero cuando se habla de la mente y ésta es comparada con una computadora, vemos que nada tienen que ver una con la otra; son por demás ajenas, diferentes, tanto los conceptos como las propiedades de funcionalidad, de ahí que su capacidad y eficiencia sean incomparables. Estoy convencido sin lugar a dudas, que para las próximas décadas, uno de los grandes cambios intelectuales en el ámbito científico, atenderá a las cuestiones de la organización mental y es aquí donde el punto de vista físico juega un papel esencial. Probablemente lo más significativo entre el paralelismo de algunas propiedades genéricas, existente entre muchos sistemas físicos y biológicos, es que ambos tratan con muchos componentes interactuando entre sí y es precisamente en ello, donde estriba o se encierra, el misterio del procesamiento neuronal, es decir en su naturaleza colectiva y en su distribución azarosa. Teniendo en cuenta que la física más que otras disciplinas ha desarrollado métodos y herramientas matemáticas para estudiar sistemáticamente el multicomportamiento de tales sistemas, ha disparando en recientes años un rápido y fascinante desarrollo, dentro de esta disciplina, provocando intereses convergentes dentro de muchas otras ramas de la investigación y favoreciendo con ello que la llamada neurociencia sea una actividad multidisciplinaria.

2.2.1 RELACIÓN ENTRE DINÁMICA DE ATRACTORES Y ARQUITECTURA DE REDES NEURONALES ARTIFICIALES .

Un área donde las ideas de la teoría expuesta en el capítulo uno han jugado un importante papel, es dentro de las redes neuronales. Aquí la teoría tiene dos propósitos: primero, entender la naturaleza computacional colectiva en redes artificiales compuestas por muchos procesadores simples y segundo, el modelar algunas operaciones del sistema nervioso. Aunque estas áreas son demasiado grandes y complejas, se tiene algunas cuestiones generales donde, modelos relativamente simples, pueden ser capaces de proveernos de cierta experiencia.

Por ejemplo las redes atractoras. Estas consisten de una o dos capas de neuronas, interactuando entre sí, hasta alcanzar un estado estacionario. La salida de cada neurona en una capa dada, puede ser una entrada hacia cada una de las neuronas que se encuentran en otra capa y viceversa. Las investigaciones que se han planteado con este tipo de arquitecturas, comprenden trabajos tan populares como la red de Hopfield y el mapeo de auto-organización de Kohonen (memoria asociativa).

Una tarea esencial que se le pide a los algoritmos de entrenamiento de este tipo de redes neuronales, es el de optimizar el almacenamiento de los datos, los cuales se caracterizan por poderse manejar a través de una matriz de valores de pesos establecida, en relación al interconexiónado. Diferentes criterios para optimizar el almacenamiento son concebibles y corresponden a los diferentes algoritmos que han sido derivados para este tipo de arquitecturas.

El almacenamiento de la información en las redes atractoras está localizado de una forma distribuida, por lo que no podemos decir en cual estructura particular se encuentra tal o cual información, pues ella ni siquiera existe, ya que la información está por toda la red y se recupera por un proceso, donde el sistema se auto-conduce hacia algún estado estacionario, alcanzado en un espacio fase de alta dimensionalidad. La manera distribuida de almacenar la información, hace que estos sistemas sean muy tolerantes al daño parcial en su conectividad (por ejemplo la degradación de una unidad, afectará casi en nada el comportamiento previo que la red tenía), pero también introduce una tendencia a inferir errores de entrecruzamiento entre patrones similares. Este tipo de almacenamiento busca asociar la información contenida en los patrones almacenados con situaciones del mundo real. De hecho, almacena patrones prototipo, por ejemplo, si seleccionamos un número suficiente de imágenes (que pueden estar formadas por unidades en pixel) y a ellas las consideramos como pertenecientes a un conjunto de patrones prototipo, entonces en este conjunto podemos hacer una partición y utilizarla para lograr una clasificación adecuada que generalice todos los patrones existentes en el medio ambiente (es decir todo el conjunto de acontecimientos en el universo de estudio), de acuerdo a los prototipos almacenados.

El caso prototipo de esta habilidad, es la memoria asociativa simple, que es el proceso por el cual, la co-ocurrencia de dos o más rasgos particulares, resulta en lo que se llama mapeo heteroasociativo. Es decir, la presentación subsecuente de un patrón de entrada, (o un vector de atributos) conduce al sistema a establecer trayectorias o rutas hacia un estado de correspondencia con dicho patrón; esto se alcanza a través de lograr "almacenar los rasgos más sobresalientes de él". Por ejemplo, la entrada de un patrón sonoro de la palabra "manzana" evoca una actividad, en algún nivel, sobre aquellos elementos

neuronales que representen las características roja y redonda (entre algunas otras), las que son activadas cuando nosotros vemos una manzana real. La esencia de la autoasociación es la formación de cadenas entre los elementos que son activados, para obtener un estado del sistema particular. El resultado de este encadenamiento, es la consecuencia de haber entrado a lo que es conocido en el lenguaje de los sistemas dinámicos, como una cuenca de atracción perteneciente a un espacio de dimensión igual al número de neuronas en la red, dependiente de la interacción colectiva (del interconexionado), así como de los correspondientes estados de salida de cada neurona. Esto puede ser manejado, ya sea en forma binaria o continua, dependiendo de cómo el modelo es formulado. Si el sistema es persistentemente puesto en un estado dentro de la cuenca o base de atracción, la trayectoria del estado final estará en la dirección del atractor original. Esta propiedad es conocida como la consumación del patrón; de aquí se deja ver que no es necesario tener todo el detalle de la información, pues basta con la mínima suficiente que caiga en la cuenca, para así reconstruir todo el patrón original.

Realmente, estas ideas conexionistas aparecen en algunos lugares del libro de Hebb, quien pensó que los eventos sensoriales desencadenan reacciones que obedecen a experiencias pasadas de patrones semejantes, que corresponden a una activación neuronal específica. Él forjó el término de fase secuencial para referirse a las trayectorias en un intervalo temporal dentro del estado neuronal fundamental, causado por un encadenamiento de secuencias de eventos sucesivos, a partir de un motor censor. Desde el punto de vista de Hebb, las interconexiones en el cerebro habilitan la ocurrencia de lo que él llamó "actividad reverberatoria" (iniciada por un acontecimiento), lo cual se ha relacionado en forma equivalente a la existencia de atractores quasiperiódicos en el espacio neuronal. Él imaginó tal actividad, debido a la existencia de trampas responsables de un tipo de memoria de corto alcance, relacionada con aquellos elementos neuronales particulares que fueron activados mientras el sistema estuvo en un cierto ciclo límite de excitación particular, el cual tendría que estar comprometido con el fortalecimiento de sus interconexiones. De acuerdo con esto, el sistema podría ser conducido a la forjación de un recuerdo, dentro de lo que Hebb nombró "un ensamble celular". Esto es conceptualmente equivalente a una calda de potencial en una cuenca de atracción. Él pensó en tal ensamble celular como una característica intrínseca fundamental que emerge de la construcción del conexas neuronal.

A continuación seguiremos una de estas aparentemente sencillas historias, con un tratamiento más formal desde el punto de vista matemático, más no así desde el lado fisiológico, desde donde no se puede decir lo mismo.

2.2.2 UN ENREJADO ENFOCADO COMO UNA CAPA NEURONAL ARTIFICIAL EN DONDE SE PROMUEVE LA AUTO-ORGANIZACIÓN.

El sistema nervioso es un enorme ensamble de células ($\sim 10^{11}$ células) altamente interconectadas y una buena aproximación es que, cada disparo de una célula dada, es idéntico con el de cualquier otro disparo. Esto es, una célula tiene justo dos estados significativos generales: el de disparo y el de no disparo, por tanto, puede ser descrito con una variable binaria, que acorde a una analogía con el espin, podrá adquirir los valores $S_i = (\pm 1)$, ($i = 1, \dots, N$). En términos más abstractos considere un enrejado con N sitios, ocupado cada uno de ellos por valores de S_i ya sea que estén activados ($S_i = +1$) o desactivados ($S_i = -1$). También en este enrejado se deberá poder hablar de que las células interactúan entre sí ya sea para inhibir o excitar a sus vecinas, por tanto debemos pensar en el papel del voltaje celular asociado a la celda i , el cual será, en términos físicos, análogo al campo local h_i . Así el sitio S_i experimentará un campo de interacción con todos los demás S_j , dado por:

$$h_i = \sum_j J_{ij} S_j \quad (2.3)$$

donde J_{ij} es un número real que puede ser interpretado como la eficiencia de la interacción entre la actividad del sitio j y la actividad del sitio i . Esta ecuación en términos biológicos muy generales, podría decirse que es la entrada a la neurona "i". Las celdas del enrejado por lo general permiten una actualización con una secuencia azarosa, ya sea en serie o en paralelo (todas a un tiempo o una en seguida de la otra sin importar el orden). Así, pensando en una manera por la cual se cambie o se actualice la celda S_i en el tiempo, escribiremos:

$$S_i(t+1) = \text{sgn } h_i(t) \quad (2.4)$$

donde $(t+1)$ es un paso temporal y $\text{Sgn } h_i(t)$ es una función signo afectando la variable S_i dentro de la celda i , en cada paso temporal. La ecuación (2.4) describe la dinámica que experimenta la celda S_i dentro de un arreglo matricial de valores J_{ij} , interpretados como la ponderación de la eficiencia en la interacción sináptica. Es decir se le está diciendo a la celda i , calcula la ecuación (2.3) y del resultado sólo toma el signo, ésta es la dinámica Gardner de la mecánica estadística de los espines de vidrio. Pero por

supuesto, esto es toda una falsedad en una neurona real, por lo que algunos investigadores han propuesto una regla de umbral buscando una mejor aproximación para actualizar la celda i , diciendo: evalúe primero la ecuación (2.3) y compare si el valor está abajo o arriba de un cierto umbral pre-fijado; algunos otros utilizan una regla de evolución estocástica como la Monte Carlo, pero cualquiera de ellas son historias de casi lo mismo y en esencia no dan grandes diferencias substanciales.

Ahora bien, la estabilidad correspondiente a una situación especial del sistema se logra en una escala de tiempo mucho más larga, que la dada para los cambios de los pesos J_{ij} de cada paso temporal. Esto significa que si nosotros consideramos un conjunto de p arreglos $\xi^v, \in \{+1, -1\}$ ($i = 1, \dots, N; v = 1, \dots, p$) - estados de arribo posibles -, los debemos hacer corresponder poco a poco con el estado final que debería adquirir el sitio S_i , siempre y cuando el sistema inicie lo suficientemente cerca para establecer arreglos de la forma $\{S_i\} = \{\xi^v\}$, para lo cual debemos ir construyendo las adecuadas J_{ij} con cada paso temporal que dirigirá a cualquier perturbación acontecida en el medio ambiente, desde su arranque, hasta su estabilidad con el estado esperado correspondiente. Así, el acoplamiento interactivo J_{ij} de cambio, se construye de acuerdo a alguna función f :

$$J_{ij}(t+1) - J_{ij}(t) = f(\xi^v_i, \xi^v_j, h^v_i(t)) \quad (2.5)$$

con

$$h^v_i(t) = \sum_j J_{ij}(t) \xi^v_j \quad (2.6)$$

donde los parámetros J_{ij} describen la influencia o interacción de la célula j con la célula i , que hemos interpretado como la calidad o eficacia sináptica y lo que buscamos es tener un valor sobre la conexión i y j dentro de la red, que atienda o satisfaga eficientemente, a todas las perturbaciones posibles del medio ambiente, generalizando los patrones de entrenamiento $\{\xi^v\}$. Vea pues como la ecuación (2.5) debe ser una regla de aprendizaje, que nos conduce a un algoritmo para el manejo de la información local, pues el cambio de la eficacia J_{ij} depende sólo de la interacción de los dos sitios adyacentes (primeros vecinos, aunque en principio uno puede añadir alguna otra dependencia sobre h^v_i , pero no se espera tener ninguna ventaja).

La totalidad de los patrones $\{\xi_{v_i}\}$ pueden ser presentados azarosamente, secuencialmente o en paralelo; en el último caso f depende del conjunto completo $\{h_{v_i}(t)\}$.

El problema que ahora queda, es encontrar o describir la función f , es decir, el algoritmo con el cual, se le faculte al enrejado a auto-organizarse y reaccionar de manera coherente a la hora de procesar la información de su medio ambiente. Las perturbaciones del medio ambiente deberán ser auto-conducidas hacia una clasificación determinada (memoria asociativa). Esto significa que la ecuación (2.5) debe ir construyendo una matriz $[J_{ij}]$ con la cual, primero se asegure que todos los patrones o arreglos $\{\xi_{v_i}\}$ sean estados estables bajo la dinámica de la ecuación (2.4) y segundo, que cada patrón ξ_{v_i} sea un atractor específico dentro de esta dinámica, con una justa o adecuada base de atracción.

La primera condición es obviamente dada por: $\xi_{v_i} h_{v_i} > 0$ (2.7)

que no es más que la condición de estabilidad Liapunov [11, 12]. Es decir $\{\xi_{v_i}\}$ es un punto de estabilidad, si para toda $\epsilon = \xi_{v_i} h_{v_i} > 0 \exists \delta > 0$, tal que, $\| \{S_i\}_n - \xi_{v_i} \| < \delta$, para todo tiempo $t \geq 0$, siempre y cuando $\| \{S_i\}_n - \xi_{v_i} \| < \epsilon$ (donde $\{S_i\}_n$ denota el estado del enrejado después de n pasos temporales, obedeciendo la dinámica de la ecuación 2.4). Así las trayectorias que empiecen dentro de una distancia δ cerca de ξ_{v_i} permanecerán a una distancia ϵ de ξ_{v_i} después de n pasos temporales $\{S_i\}_n$.

La tarea que se busca realizar, es organizar el espacio del enrejado, en estados dentro de atractores, localizados alrededor de la cercanía del conocimiento a priori, de los estados de memoria deseados. Note que este problema es el inverso a los normalmente encontrados en la física de fenómenos colectivos, donde por lo general se busca encontrar los estados de equilibrio, cuando las interacciones son conocidas.

El desejar que al final tengamos $\{S_i\} \approx \{\xi_{v_i}\}$, significa que la red, al ser manejada por una fuente de estímulos externos, las cuales pensamos que inciden directamente sobre los acoples sinápticos J_{ij} , podemos llegar a considerar modificarlos con base en ciertas fuerzas dependientes del campo local, que garanticen que ξ_{v_i} llegue a ser un punto fijo bajo la dinámica (2.4). Así, aunque no es todavía bien conocido que propiedad de J_{ij} daría una buena base de atracción, parece plausible que una buena base considere el parámetro [12, 13]:

$$\Delta_{v,i} = \frac{\xi_{v,i} h_{v,i}}{\left(\sum_j J_{ij}^2 \right)^{1/2}} \quad (2.8)$$

para $i=1, \dots, N$ y $v=1, \dots, p$, sujeto a la condición $\sum_j J_{ij} = N$ (2.9)
por tanto, sustituyendo (2.6) en (2.8)

$$\Delta_{v,i} = \xi_v \sum_j \frac{J_{ij}}{N} \xi_j^v \quad (2.10)$$

Para patrones azarosos $\xi_{v,i}$, el espacio de todas las posibles matrices $\{J_{ij}\}$ ha sido investigado por Gardner, utilizando métodos de mecánica estadística del espín de vidrio [13, 14].

El parámetro de estabilidad máxima posible $\Delta = \max_{\{J_{ij}\}} \{ \min_{v,i} \Delta_{v,i} \}$ dependiente del campo local, ha sido calculado por Elizabeth Gardner, como una función del número p de patrones almacenados, obteniendo el número máximo $p=2N$ de patrones posibles para almacenar en una red de N neuronas.

De acuerdo a (2.7) y estando dentro de una cuenca de atracción (2.10), la regla de aprendizaje (2.5) deberá incrementar el campo, conforme nos aproximamos al punto fijo de atracción; pensando en el caso más simple, lo podemos obtener realizando el cambio de las sinapsis de manera proporcional al gradiente de estos campos:

$$J_{ij}(t+1) - J_{ij}(t) = \frac{\gamma}{N} \frac{\partial}{\partial J_{ij}} \xi_{v,i} h_{v,i} = \frac{\gamma}{N} \xi_v \xi_j^v \quad (2.12)$$

donde γ es una constante que puede ser interpretada como la rapidez del aprendizaje.

Si la matriz inicial es cero y si todos los patrones v son presentados una sola vez, es decir, practicando un barrido generacional (de duración igual al paso temporal), uno obtiene los acoplamientos:

$$J_{ij} = \frac{\gamma}{N} \sum_v \xi_v \xi_j^v \quad (2.13)$$

Este es el modelo de Hopfield, cuyos atractores han sido calculados analíticamente utilizando la teoría de los espines de vidrio [8].

Note como el enrejado puede ser considerado como una capa neuronal, homogénea y uniforme, entendiendo por esto que todos sus elementos son fundamentalmente similares y además, en este caso, las cuencas o bases de atracción son equivalentes y todas del mismo tamaño, cosa que no necesariamente tiene que ser así.

2.2.3 REDES MULTICAPAS DE ALIMENTACIÓN PROGRESIVA.

El marco de trabajo o más precisamente el espacio, queda especificado por la arquitectura de la red, la cual considera el número de capas y el número de nodos por capa, así como por la manera en que se van a relacionar las capas (el conexonado).

Por lo que se refiere al tipo de capas tenemos:

Capa de entrada.- Las unidades que conforman esta capa son llamadas unidades de entrada; ellas codifican el problema que ha sido seleccionado para que la red lo procese. Esta capa se encarga de desparramar la información o distribuirla, es decir, esta capa recibe un patrón de entrada y entonces, por virtud de sus interconexiones (canales de comunicación), lleva el patrón, usualmente sin cambio, a cada una de las neuronas en la capa subsecuente.

Capa escondida.- Los elementos procesadores en esta capa son llamados unidades escondidas porque no son directamente observables. Ellos proveen de no linealidad a la red.

Capa de salida.- Están formadas por unidades llamadas de salida, las cuales codifican los posibles conceptos que la red ha asignado al ejemplo que está bajo consideración, dando valores que pueden ser interpretados por el medio ambiente externo, como respuesta al problema. Por ejemplo, cada unidad de salida puede representar una clase de objetos.

Por lo que se refiere a la interacción (interconexonado), una red puede comprender una capa de entrada, una de salida y una o varias escondidas o también la capa de entrada puede funcionar a la vez como capa de salida o bien, un nodo puede estar conformado por todo un enrejado completo y ser tratado como una sola y simple unidad; en fin varias capas pueden permitir una riqueza de combinaciones y posibilidades en variabilidad de acoplamiento intercapas

enorme, sin embargo, en particular, atenderemos en los siguientes capítulos aquellas donde la salida de una de las capas puede ser la entrada a otra.

Estas redes pueden ser vistas como dispositivos que conducen una onda de actividad pasándola de capa en capa, en respuesta a la presencia de algún estímulo recogido o registrado por o sobre, una primera capa sensorial. Así la trayectoria de la señal va sólo en una dirección, de la capa de entrada a la capa de salida y ellas son llamadas redes de alimentación progresiva.

CAPITULO TRES

3 CARACTERIZACION CONEXIONISTA

3.1 Configurando una red neuronal multicapas de alimentación progresiva.

- 3.1.1. La neurona artificial. EL PERCEPTRÓN.
- 3.1.2. Estructurando la red artificial.

3.2 Propagación. Descripción general.

- 3.2.1. El manejo y propagación de la información.
- 3.2.2. Entrenamiento.

3.3 Retropropagación. Descripción general a bloques.

- 3.3.1 La regla delta.
- 3.3.2 El gradiente descendente del error.
- 3.3.3 Enseñanza - aprendizaje. Entrenamiento hebbiano.
(Basado en ejemplos y ejercicios).

CAPITULO TRES

EL PERCEPTRÓN

3 Caracterización Conexionista del Aprendizaje.

En las siguientes líneas se detallan los modelos neuronales de alimentación progresiva, contruidos a partir de una unidad neuronal artificial que se ha llamado *perceptrón*. Se verá cómo el aprendizaje que nos da el algoritmo de la retropropagación toma lugar en este tipo de arquitecturas. Estas unidades artificiales son utilizadas para manipular las señales de entrada vía la conexionalidad sináptica, donde la dinámica de funcionamiento de la red es modelada a través de las ponderaciones $[W_0]$ que pueden ser interpretadas como alguna propiedad característica sináptica. Los pesos son colocados en arreglos matriciales para que con ellos podamos calcular el vector de salida de la red, a partir de ir multiplicando un vector de entrada (excitabilidad) por una matriz de pesos (eficacia sináptica). De esta forma los estímulos son caracterizados como puntos en un espacio sensor, dados por la capa de entrada de la red; es decir, cada punto está determinado por una correspondiente activación simultánea de todas sus unidades de entrada y por virtud de la dinámica de la red, es transportado hacia un punto imagen en el espacio de respuestas o estados de reacciones sobre la capa de salida. Así la representación o la imagen neuronal de los acontecimientos del medio ambiente, llega a ser caracterizada gracias al rango de variabilidad de los pesos, lo que nos permite buscar y alcanzar la percepción adecuada.

El aprendizaje se entenderá, en esta caracterización conexionista, como la experiencia adquirida cada vez que se logra reducir el error de respuesta; es decir, se verá como el proceso de aprendizaje, que no es otra cosa que el ajuste de muchos pesos para así obtener el vector de salida adecuado. Mostraremos como el algoritmo de la retropropagación del error, nos lleva a lo que podría llamarse experiencia autoconducida por repetición.

3.1 CONFIGURANDO UNA RED NEURONAL MULTICAPAS DE ALIMENTACIÓN PROGRESIVA.

La unidad procesadora que se utiliza, tiene sus orígenes en la propuesta del perceptrón hecha en los años 50's por Frank Rosenblatt, un neurofisiólogo de la universidad de Cornell. Lo presenta en su libro titulado "Principles of Neurodynamics" donde con estas unidades, intenta modelar la retina y así imitar la habilidad del sistema visual para reconocer y clasificar patrones [15, 16].

3.1.1. LA NEURONA ARTIFICIAL. EL PERCEPTRON

Este es el nombre de una unidad que, en analogía con la neurofisiología, es concebida como una célula nerviosa muy primitiva, que puede ser construida artificialmente; ésta en esencia imita toscamente las cualidades más básicas de una neurona. Tal unidad neuronal, está constituida por cuatro componentes importantes, las cuales son:

1. **Conexiones de entrada:** a través de las cuales la unidad recibe la activación de otras unidades dándole una importancia individual a cada una de ellas; para ello se pondera cada señal de entrada por medio de un peso.
2. **Una función sumadora:** la que combina la acción de un conjunto de entradas en un solo valor.
3. **Una función de activación:** que hace la conversión del valor resultante de la suma, en una activación final, produciendo así una señal de salida.
4. **Conexiones de salida:** por las cuales la salida de activación de una unidad llega como señal de entrada a otras unidades del sistema.

Ver fig. 3.1

3.1.2. ESTRUCTURANDO LA RED ARTIFICIAL.

Se dice, con base en la teoría conexionista, que el comportamiento inteligente parece emerger del trabajo conjunto de un enorme número de neuronas interconectadas unas con otras; es decir se piensa que a través de las características sinápticas y la manera en que se va entretejiendo la red se logran obtener propiedades que hemos identificado como memoria asociativa, adaptabilidad a nuevos ambientes (aprendizaje), capacidad de cálculo, etc., por lo que la pregunta inmediata sería: ¿Cómo interconectamos nuestras unidades para obtener un sistema construido artificialmente que exhiba alguna de esta funcionalidad emergente? es decir ¿de qué forma debemos hacer el interconexiónado entre ellas, para que la información del medio exterior se propague sistemáticamente y cada unidad tenga la oportunidad de aportar una pequeña "opinión" sobre la solución de una tarea asignada? Los elementos podrían estar conectados de muchos modos, por ejemplo, éstos podrían ser conectados unos contra otros y todos contra todos, o tal vez formar grupos o capas que estén relacionadas entre si, etc. De hecho la respuesta a esta pregunta se ha buscado a través de la neurofisiología mediante el estudio de redes neuronales naturales [17, 18]. Nosotros seguiremos los trabajos previos de Frank Rosenblatt sobre el perceptrón [15], con el objeto de ilustrar la forma en que el aprendizaje toma lugar a través de la retropropagación y tratando de simplificar el problema, modelaremos, en forma gruesa, una clase de red neuronal que, a pesar de su tosca aproximación biológica, exhibe atributos que están fuertemente caracterizados en los seres vivos. Así presentaremos la siguiente forma de conexionado que adoptaremos en adelante, una arquitectura conocida como **red neuronal multicapas de alimentación progresiva**, la cual es susceptible de ser entrenada con la técnica de la retropropagación (ver fig.3.2). El término "alimentación progresiva" implica que la señal de salida de un elemento procesador, no puede ser una entrada para alguna unidad que se encuentre sobre la misma capa o capas anteriores y, por lo que se refiere a multicapas, es por que está formado por un número "n" de capas. La capa que recibe la información que está entrando a la red, procedente del medio exterior, es llamada capa de entrada; los elementos de esta capa, no ejecutan ninguna modificación a la señal; son sólo como puntos de distribución hacia la siguiente capa. La última capa, llamada capa de salida, proporciona la respuesta que la red ha generado como reacción al estímulo presentado. Cualquier otra capa en la arquitectura es llamada capa escondida, porque es interna a la red y por tanto, no tiene contacto directo con el medio ambiente externo.

El primer aspecto de las características de las componentes del perceptrón, mencionó la ponderación de las señales de entrada a una unidad procesadora cualquiera. Pues bien, esto puede verse como un valor de peso variable que puede ser modificado a voluntad y al encontrarse en todas y cada una de las conexiones, representará una forma muy eficiente de modular la activación de la red (Ver fig. 3.3).

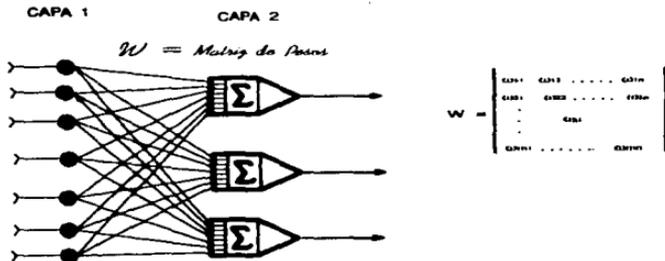


Fig 3.3 Ilustración de la equivalencia entre la conectividad y la matriz de pesos entre las capas 1 y 2, donde el renglón de la matriz, nos indica los pesos de las conexiones que arriban a una unidad en la capa 2 y las columnas, los pesos de las conexiones que parten de un nodo sobre la capa 1.

Tomando el término de la biología, diremos que una interacción excitatoria entre las unidades $U_{[i]} \rightarrow \omega_{ij} \rightarrow U_{[j]}$, estará representada con un valor de peso positivo ($+\omega$); de esta forma un peso negativo ($-\omega$) será tomado como una interacción inhibitoria y un valor cero se interpretará como una ausencia de conexión. En la fig. 3.3 se ilustra como los números ω_{ij} se pueden agrupar en una matriz de conectividad entre dos capas sucesivas, un arreglo matricial W , en el cual el elemento de entrada ω_{ij} representa la eficacia de la conexión procedente de la unidad i de la capa 1, a la unidad j de la capa 2 (Ver figura 3.3).

Es de notarse que si q es el número de capas que constituyen nuestra red, entonces tendremos $q - 1$ matrices de pesos, con las cuales podemos representar todo el conexionado en la red.

$$[W_1, W_2, W_3, \dots, W_i, \dots, W_{q-1}] = \text{Conjunto de matrices entre capas.}$$

Hasta aquí ha quedado constituida nuestra red y hemos establecido el conexionado de tal forma, que permite al sistema manejar cualquier entrada arbitraria, hacia una respuesta y al mismo tiempo, a través de la variabilidad de la ponderación en el interconexionado, se le da la flexibilidad o capacidad, para tener la posibilidad de autoajustar su respuesta. Ahora a continuación veremos como se llevan a cabo estos mecanismos.

3.2 PROPAGACIÓN : DESCRIPCIÓN GENERAL.

3.2.1. EL MANEJO Y PROPAGACIÓN DE LA INFORMACIÓN.

Se ha dicho que las entradas a un nodo j sobre una capa cualquiera, digamos $[k+1]$, son independientes, esto es, están pesadas separadamente (con excepción de las capas de entrada donde la calidad del sensor, es decir el vector de entrada, es dependiente sólo de la información obtenida del medio ambiente a partir de su sensibilidad) y ya que tenemos tantas llegadas como unidades haya en una capa anterior $[k]$, decimos que a este conjunto de ponderaciones, que se forman entre las capas $[k]$ y $[k+1]$, se le puede ver como un arreglo matricial de pesos (fig. 3.3). Ahora bien la importancia de cada señal sobre la conexión, es simplemente el producto de multiplicar la salida de la unidad correspondiente i , cuyo valor es $X_{i(k)}$, por el peso ω_{ij} asignado a la conexión entre las neuronas etiquetadas como i y j . De esta forma se construyen tantos productos como entradas tenga la neurona en cuestión, para ser sumados algebraicamente y poder posteriormente pasar el

valor resultante del operador suma, a la función de activación. Esto se expresa matemáticamente, en equivalencia con la ecuación 2.3, como:

$$\sum_i w_{ij} x_i[k] = g_i[k+1] \quad (3.1)$$

con $f[g_i] = x_i[k+1]$

donde $f[g_i]$ es la función de activación que da el estado de excitabilidad a la neurona i en la capa $[k+1]$, que en notación vectorial y en términos generales, la ecuación 3.1 podría escribirse como el producto entre la matriz de pesos $W[k-1]$ y el vector columna de activaciones $X[k]$ (excitabilidad). Para manejar en forma completa el conjunto de valores g se procederá a evaluar la función de activación $F(z)$, donde $z = (g_1, g_2, \dots, g_m)$. De esta manera se puede manejar una activación que pasa de capa en capa a través de la red, desde la presencia del estímulo, en la capa de entrada, hasta la reacción que la capa de salida ha de dar como respuesta.

Hablando ahora un poco de la función de activación, se debe decir que ésta necesita estar definida para aceptar cualquier valor de la operación suma y poder producir, con este valor, una señal de salida que será distribuida a todas las unidades de la siguiente capa. Uno de los propósitos de introducir una función de activación, es permitir a la unidad, variaciones en pasos temporales correspondientes a un cierto estado de activación de la neurona, las que deben mantener durante un intervalo lo suficientemente amplio, como para que se efectúe la dinámica completa del sistema. Después de este período, pasará a un nuevo estado de excitabilidad, dejando el viejo, en función del operador suma. La función de activación puede tomar un sinnúmero de diferentes formas, de hecho, el perfil gráfico que debe tener, es aún un área de investigación (donde las funciones de Lyapunov juegan un papel importante), pero frecuentemente se trata de establecer una dependencia no lineal, para que, de esta forma, superar la debilidad que los sistemas lineales manifiestan al transportar patrones de entrada (estímulos), a patrones de salida (reacciones) de manera correcta (un ejemplo clásico de esta debilidad lo constituye el problema de la X-OR tratado más adelante).

En la retropropagación [19, 20, 21], se implementa frecuentemente una función no lineal, con un perfil tipo S, que es una función continua, monótona, creciente y diferenciable, conocida como sigmoide; con ella se puede ampliar la variedad de aplicaciones y el número de tareas que la red puede ejecutar.

Matemáticamente, una de estas funciones de activación (fig 3.4), la cual es muy utilizada, queda descrita mediante la expresión 3.2:

$$f[\beta_i] = \frac{1}{1 + e^{-\beta_i[k]}} \quad (3.2)$$

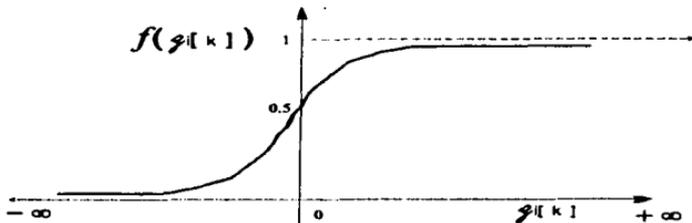


Fig. 3.4 Gráfica de la función de activación seleccionada.

Si siguiendo con nuestro modelo, nos permitiremos decir una vez más, que los únicos contactos que tiene nuestra red con el medio ambiente o mundo exterior son las capas de entrada / salida. Por tanto, enfocaremos nuestra atención a estas dos capas.

Sobre la capa de entrada se presentan los estímulos que supondremos pertenecen a un conjunto A (medio ambiente), el cual puede ser visto como un espacio vectorial de dimensión d , número que corresponde a la cantidad de nodos que existen en esta capa de entrada. Luego entonces, el conjunto de estímulos o patrones de entrada, son considerados como vectores procedentes o pertenecientes al conjunto A y están formados por d elementos, cada uno de los cuales está asociado con cada nodo perceptrón de la mencionada capa uno, o de entrada. Así la siguiente expresión representa o puede ser considerada, como el patrón de entrada.

$$\mathbf{X}_{[ent]} = (x_1, x_2, \dots, x_d)$$

donde $\mathbf{X}_{[ent]} \in A$

La capa de salida es por donde la red da su respuesta al mundo exterior, del estímulo recibido y también se puede representar por un vector, donde cada componente corresponde al valor o estado de activación, en que se encuentra la unidad correspondiente, dentro de esta última capa nombrada de salida. Así podemos construir un conjunto **B** de vectores respuesta, que formen también un espacio vectorial de dimensión, digamos m . De esta forma, la capa de salida debe tener m nodos y puede entonces ser representada por el siguiente vector:

$$\mathbf{X}_{[FIN]} = (x_1, x_2, \dots, x_m) \quad \mathbf{X}_{[FIN]} \in \mathbf{B}$$

El sistema que hemos creado, recibe un patrón procedente del conjunto de estímulos **A** sobre su capa de entrada, propaga la información a través de toda la red, de tal forma que pueda ser transportada directamente hacia el conjunto **B** de patrones de salida; éstos son exhibidos sobre la última capa y corresponden a las respuestas de los estímulos recibidos, para algún estado de conectividad dado.

Ahora bien, aplicando la regla de propagación (ecuación 3.1) a través de la red, a partir de la primera capa hasta el final y usando notación vectorial podemos escribir:

Para la capa 1

$$\begin{array}{l} \mathbf{VECTOR} \\ \mathbf{DE} \\ \mathbf{ENTRADA} \end{array} = \mathbf{X}_{[ent]} = \mathbf{X}_{[1]}$$

ya que son simples nodos distribuidores de la señal.

Para capas intermedias.

$$\begin{array}{l} \mathbf{X}_{[2]} = \mathbf{F}(Z_{[2]}) = \mathbf{F}((z_{21}), (z_{22}), (z_{23}), \dots) [z_1] \\ \vdots \\ \vdots \\ \vdots \end{array}$$

$$\mathbf{X}_{[q]} = \mathbf{F}(Z_{[q]})$$

con $q = 1, 2, 3, \dots, FIN - 1$ ya que todas las unidades se consideran idénticas y tienen y ejecutan la misma función de activación.

Para la última capa

$$\mathbf{VECTOR} \\ \mathbf{DE} \\ \mathbf{SALIDA} = \mathbf{X}_{\{\text{FIN}\}} = \mathbf{F}(\mathbf{z}_{\{\text{fin}\}})$$

Al conjunto de valores $\mathbf{Z}_{\{k+1\}}$ de una misma capa, se le puede ver como un vector y este vector está dado por la operación matricial:

$$\mathbf{Z}_{\{k+1\}} = \mathbf{W} \mathbf{X}_{\{k\}}$$

donde $\mathbf{X}_{\{k\}}$ es el vector columna de activaciones de la capa k y su respuesta nos la da la función:

$$\mathbf{F}(\mathbf{z}_{\{k\}}) = \mathbf{X}_{\{k\}}$$

con $k = 2, \dots, \text{FIN}$

Estas expresiones marcan la dinámica de la propagación hacia adelante, a través de ejecutar la serie de operaciones que experimenta la red al transportar el patrón procedente del conjunto **A** desde su entrada, hasta alcanzar la capa de salida, para formar parte, finalmente, de los patrones respuesta del conjunto **B**.



3.2.2. ENTRENAMIENTO.

Para darle coherencia a esta dinámica, es decir, alguna utilidad a la red en una tarea lógica o práctica, se le debe poder entrenar para que aprenda a reconocer el conjunto de patrones mostrados a su entrada. Para ello se le estimula con patrones que son presentados uno a la vez y para cada uno de ellos, la red será capaz de responder con una reacción conveniente, después de que el entrenamiento haya finalizado. El objetivo del entrenamiento es enseñar a la red a reconocer miembros de una misma clase, separando en subgrupos los elementos del conjunto A que tengan algo en común. Es decir, el sistema aprenderá a realizar un mapeo que transporte los estímulos de entrada, pertenecientes a un conjunto A, hacia un conjunto B, de tal forma, que se construya una partición, formando clases similares con alguna relación de equivalencia en el espacio de respuestas del conjunto B. Para esto, contamos o tenemos, antes de que el entrenamiento comience, la configuración de la red bien establecida, es decir, su arquitectura se encuentra totalmente definida: el número de nodos, el número de capas, así como las condiciones iniciales de las conexiones. Todo está listo para que dé comienzo el aprendizaje, como un alumno en el inicio de su primera clase, pero ¿cómo se va a llevar a cabo la enseñanza? El procedimiento de aprendizaje que proponemos involucra la modificación de los pesos en el interconexionado. Y ya que la información procedente de un patrón de entrada viene viajando a través de la red desde las primeras capas hasta su salida, a lo largo de todo el interconexionado, bastaría con modificar los pesos sinápticos, para establecer o implementar, un manejo de la información que obedezca a la experiencia que el sistema va adquiriendo cada vez que se enfrenta con una misma tarea, provocando poco a poco, de esta manera, la generación de la respuesta correcta (es decir construir una función de experiencia que se vaya ajustando a la función blanco demandada, a través de variar los pesos del interconexionado [21, 22, 23]).

3.3 RETROPROPAGACIÓN: DESCRIPCIÓN GENERAL A BLOQUES.

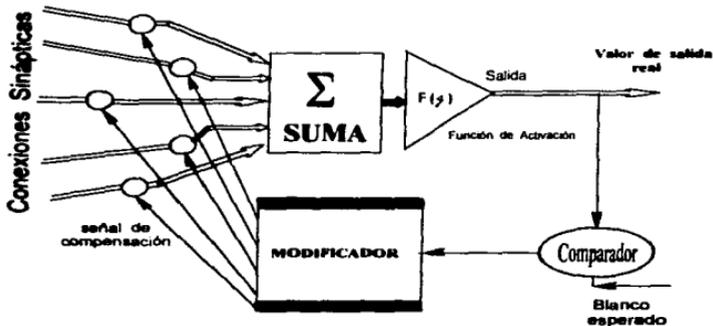


Fig. 3.5 Ilustración diagramática de como se lleva a cabo la variación de la fuerza sináptica.

En seguida se explica su funcionalidad matemática.

3.3.1. LA REGLA DELTA.

El interconexiónado de la red se encuentra, al principio del entrenamiento, inicializado con ciertos valores de pesos tomados al azar, que durante la enseñanza, el algoritmo irá modificando a medida que va aprendiendo, utilizando para ello una simple y elegante manera conocida como la Regla Delta, la cual consiste en lo siguiente: Primeramente el sistema usa el vector de entrada para generar su propio vector de salida, después compara este valor con lo que debería de ser la salida correcta, o como usualmente es llamado, con el vector blanco. Si no hay diferencia no toma lugar el aprendizaje, pero si no es así, los pesos son modificados para reducir la diferencia. Esta secuencia se sigue con cada barrido, tomando uno a uno el conjunto completo de vectores de

entrada, tantas veces como se considere necesario. El método de entrenamiento puede ser ilustrado con los siguientes pasos:

- 1.- Aplicar un vector de entrada al sistema y calcular o esperar su respuesta de salida (propagación).
 - a) Si la salida es correcta, ir al paso 2.
 - b) Si no, modificar los pesos para hacer más pequeña la diferencia (Retropropagación).

- 2.- Verificar con un criterio de evaluación si el entrenamiento es satisfactorio
 - a) No → Tomar otro vector de entrada e ir al paso 1.
 - b) Si → Termina

La regla para el cambio de pesos correspondientes a los datos de un par de patrones entrada / salida, está expresada por la ecuación 3.3. Ella establece que el aprendizaje o la construcción de la reacción deseada en respuesta a un patrón o estímulo presentado, es alcanzada a través de los cambios de pesos. Note como el ajuste de los pesos en cada paso temporal es proporcional a la discrepancia o delta formada entre la activación real $X_{[FIN]}$ producida por la propia red como respuesta al patrón de entrada y la activación deseada D , proporcionada por el conocimiento de algún maestro o instructor. Esta delta puede ser vista en una forma más general, como una diferencial, la cual es calculada a partir de una función de error. La discrepancia o delta del error, es transportada a la capa anterior, brincando así de capa en capa a través de la red, hasta alcanzar las unidades de la primera capa intermedia (la ejecución de este mecanismo se nombra retropropagación) y con este criterio de discrepancia, se realiza la modificación de los pesos en cada una de las unidades, bajo la ecuación 3.3, según corresponda a una capa de salida o a una capa intermedia:

Para un patrón P_i tenemos

$$\begin{aligned} \Delta W_{P_i}^{[FIN-1] \rightarrow [FIN]} &= \mu (D - X_{[FIN]}) F'(z) X_{[FIN-1]} \\ \Delta W_{P_i}^{[k-1] \rightarrow [k]} &= \mu \delta_{[k]} X_{[k-1]} \end{aligned} \quad (3.3)$$

donde $X_{[FIN]}$ es el vector de salida real producido por la presencia del vector de entrada asociado al patrón P_i ; D es el vector blanco al que el

vector de entrada tiene que llegar a igualar. En unidades escondidas la diferencia delta, denotada como $\delta[k]$ se estima y es la evaluación que el sistema realiza al retropropagar el error generado por la comparación entre el vector de salida propio de la red $X_{[PIN]}$ y la reacción que desearíamos que diera; μ es una constante de proporcionalidad (puede ser vista como un factor de la rapidez con que la red puede aprender); $\Delta W_{P_i}[(PIN-1) \rightarrow (PIN)]$ es el cambio que debe hacerse a la matriz de pesos que conecta la penúltima capa con la última capa, para el patrón P_i ; $\Delta W_{P_i}[(k-1) \rightarrow (k)]$ es el cambio sobre la matriz de pesos, formada por el interconexiónado entre dos capas sucesivas donde no esté involucrada la capa de salida. $X_{[PIN-1]}$ es el vector columna de excitaciones, existente en la penúltima capa. El vector blanco (D) es nuestra referencia para los valores de activación deseados, es decir los que nos gustaría que tuviera cada una de las neuronas de la capa de salida. La capa de salida es también el contacto que tiene la red con el medio exterior y muestra las reacciones que nosotros somos capaces de interpretar, por eso, es a través de ella que nos damos una idea de cómo andan las capas interiores, pues al estar "escondidas" no tienen un valor blanco bien establecido con respecto a nuestros sentidos, que nos pudiera dar idea inmediata de cómo va desarrollándose la actividad de la red y lo que es más, podrían pasar inadvertidos o transparentes para el mundo externo, en este caso, el instructor o evaluador. Debido a ello y con el fin de que quede más clara la forma en que la regla delta trabaja en la enseñanza de la red y con el objetivo de hacer que ella responda de una cierta manera, la cual deseamos, empezaremos el análisis por esta última capa, calcularemos el error originado en ella y lo retropropagaremos hacia la entrada a través de toda la red, para poder hacer con ello, posteriormente, una modificación de pesos en el interconexiónado. Este cambio en la matriz ΔW , según la ecuación 3.3, debe seguir o cumplir con la proporcionalidad:

$$\Delta W \propto \delta[k]$$

veamos:

Pensemos en un sistema que sólo tiene tres capas, del que derivaremos la ecuación 3.3. Para poder llegar a ella introduciremos primero, un pequeño formalismo con una definición de lo que entenderemos por error.

Proponemos que el procedimiento cambiante o estructura del conocimiento en el sistema, lo determina la variación inteligente de los pesos a partir de la delta, por tanto habrá que dar pasos en la dirección en que se minimice una función de error, definida o formada, por la suma de las discrepancias o

deltas, producidas en cada unidad; éstas δ 's se originan o principian, en la capa de salida, al hacer la comparación entre los valores reales de salida $X_{[FIN]}$ y los deseados (D), es decir sobre todas las unidades de la última capa y sobre todas las parejas de patrones entrada/salida. Por tal motivo, definiremos la medida del error en una unidad cualquiera debida a un patrón P como:

$$E_p = \frac{1}{2} \sum_{i=1}^m [d - X_{[FIN]}]^2$$

Número de unidades
en la última capa

$$E_p = \frac{1}{2} \sum_{i=1}^m [d - f_i(\varphi)]^2$$

Función de activación
de la unidad i .

(3.4)

donde $f_i(\varphi) = X_{i[k]}$ con $\varphi_{[k]} = \sum_{j=1}^n W_{ij} X_j^{[k-1]}$

Por tanto sacando la parcial de la función de error en una capa dada con respecto a lo que entra, tenemos en notación vectorial que:

$$\begin{aligned} \frac{\partial E}{\partial z} &= (D - F(z)) \left[\frac{\partial}{\partial z} (D - F(z)) \right] \\ &= (D - F(z)) \left(- \frac{\partial}{\partial z} F(z) \right) \\ &= - (D - F(z)) F'(z) \quad \text{donde} \quad F(z) = (f_1(z), f_2(z), \dots, f_m(z)) \end{aligned}$$

Es decir

$$\frac{\partial E}{\partial z[k]} = - \delta[k] \quad \longrightarrow \quad (3.5)$$

y demostraremos en el próximo inciso que: $-\frac{\partial E}{\partial w} = \delta X$

3.3.2 EL GRADIENTE DESCENDENTE DEL ERROR .

La dinámica para ir minimizando el error en todas las unidades de la red, utiliza un procedimiento basado en lo que se conoce como gradiente descendente del error, en el espacio de los parámetros W , que tiene la tarea de ir ajustando las "reacciones", de tal forma que se vayan acercando a las respuestas que nos gustaría que diera o adoptara la red.

Comenzaremos mostrando como la variación del error con respecto al cambio de pesos entre la última y penúltima capa es:

$$-\frac{\partial E}{\partial W} = \delta X_{[FIN-1]} \quad \begin{array}{l} \text{Activaciones de la penúltima capa} \\ \text{Debajo del error de las unidades} \\ \text{de la última capa.} \end{array}$$

((FN-1) → (FN))

Esta relación será como dar pasos en dirección descendente sobre una superficie (o línea) dentro de un espacio de pesos, cuyas alturas a cualquier punto de la superficie E , sería la medida del error.

Aplicando la regla de la cadena a la función de error, encontraremos la derivada de la medida del error con respecto a los pesos como sigue:

$$\frac{\partial E}{\partial W} = \frac{\partial E}{\partial Z_{[FN]}} \frac{\partial Z_{[FN]}}{\partial W} \quad \text{con } Z_{[FN]} = W X_{[FIN-1]} \quad (3.6)$$

Esta parte refleja la variación del error, como una función del cambio neto de lo que le llega a cada una de las unidades.

((FN-1) → (FN))

donde el término $X_{[FIN-1]}$ es el vector columna de las activaciones de la penúltima capa.

Además recordando de (3.5):

$$\frac{\partial E}{\partial Z_{[FN]}} = -\delta_{[FN]}$$

Ahora, por otro lado realizaremos el cálculo de: $\frac{\partial Z_{[FIN]}}{\partial W_{[FIN-1]}}$

$$\frac{\partial Z}{\partial W} = \frac{\partial}{\partial W} [W X_{[FIN-1]}] - X_{[FIN-1]} \quad (3.7)$$

luego entonces sustituyendo en la ecuación 3.6, las ecuaciones 3.5 y 3.7:

$$-\frac{\partial E}{\partial W} = \delta_{[FIN]} X_{[FIN-1]}$$

Esta ecuación nos dice como debe ir cambiando el error E, con respecto a la variación de pesos entre la última y penúltima capa después de que un ciclo completo ha finalizado, desde que entra en escena el patrón P_i, hasta que sale, proporcionando así, un criterio para ejecutar los cambios de pesos en la matriz $W_{[FIN-1] \rightarrow FIN}$. Es decir si hacemos la variación de ΔW proporcional a este gradiente descendente, aseguramos una disminución del error en cada ciclo, (esto referido al patrón con el que se está trabajando). Por tanto debemos hacer el cambio de pesos proporcional a la primera derivada parcial de la función de error con respecto a los pesos, es decir;

$$(\Delta w) \propto \frac{\partial E}{\partial W} = -\delta_{[FIN]} X_{[FIN-1]}$$

si ahora introducimos una constante de proporcionalidad μ podemos establecer la igualdad en la variación de las ponderaciones dentro de la matriz del conexionado entre la última y penúltima capa como:

$$\Delta W [(FIN-1) \rightarrow FIN] = \mu \delta_{[FIN]} X_{[FIN-1]} \quad (3.3.a)$$

donde diferenciando la función de error encontramos que la delta del error generada en la última capa por nuestra red es $\delta_{[FIN]} = (D - X_{[FIN]}) F'(Z)_{[FIN]}$, (adoptando la definición 3.4 de error). Ahora bien lo interesante y lo que le da el nombre a esta técnica, es la habilidad de propagar hacia atrás la delta o discrepancia, entre la respuesta real y la deseada y formar una señal delta de

error en retropropagación que, originándose en la última capa y estimada por medio de este pequeño cálculo, vaya saltando de capa en capa hasta alcanzar las unidades de la primera capa intermedia.

Pues bien, sigamos pensando en nuestra red de tres capas y que nos encontramos en la última capa, pero ahora conocemos los valores de las $\delta_{[FIN]}$ calculados previamente. En estas condiciones, daremos uno de esos pequeños saltos hacia atrás y de manera semejante a la ecuación 3.6, escribiremos:

$$\frac{\partial E}{\partial w_{\{[FIN-2] \rightarrow [FIN-1]\}}} = \frac{\partial E}{\partial z_{[FIN-1]}} \frac{\partial z_{[FIN-1]}}{\partial w} \quad (3.8)$$

En seguida realizaremos el cálculo de la variación de la señalización de lo que entra a la capa $[FIN-1]$; si provocamos un pequeño cambio de los pesos entre las capas $[FIN-2]$ y $[FIN-1]$:

$$\begin{aligned} \frac{\partial z_{[FIN-1]}}{\partial w} &= \frac{\partial}{\partial w} [W X_{[FIN-2]}] \\ &= X_{[FIN-2]}. \end{aligned} \quad (3.9)$$

y de la ecuación 3.5 tenemos además.

$$\frac{\partial E}{\partial z_{[FIN-1]}} = -\delta_{[FIN-1]}$$

esta es, la delta que es generada por el patrón P_i sobre las unidades de la capa $[FIN-1]$.

Usando nuevamente la regla de la cadena tenemos que $\delta_{[FIN-1]}$ es:

$$\delta_{[FIN-1]} = -\frac{\partial E}{\partial z_{[FIN-1]}} = -\frac{\partial E}{\partial x_{[FIN-1]}} \frac{\partial x_{[FIN-1]}}{\partial z_{[FIN-1]}} \quad (3.10)$$

donde podemos establecer la siguiente igualdad:

$$\frac{\partial X_{(PIN-1)}}{\partial Z_{(PIN-1)}} = \frac{\partial F_{(PIN-1)}}{\partial Z_{(PIN-1)}} = F'_{(PIN-1)} \quad (3.11)$$

recordando que $X_{(PIN-1)} = F_{(PIN-1)}$.

Así en esta penúltima capa, usaremos una vez más la regla de la cadena, en el cálculo de la derivada de $\partial E / \partial X_{(PIN-1)}$ (uno de los términos de $\delta_{(PIN-1)}$, en la Ec. 3.10).

$$\frac{\partial E}{\partial X_{(PIN-1)}} = \frac{\partial E}{\partial Z_{(PIN-1)}} \frac{\partial Z_{(PIN-1)}}{\partial X_{(PIN-1)}} \quad (3.12)$$

donde identificamos inmediatamente que $\frac{\partial E}{\partial Z_{(PIN-1)}} = -\delta_{(PIN-1)}$

ya calculada previamente (antes de dar el salto),

$$y \quad \frac{\partial Z_{(PIN-1)}}{\partial X_{(PIN-1)}} = \frac{\partial}{\partial X_{(PIN-1)}} [W_{X_{(PIN-1)}}] = W_{[(PIN-1) \rightarrow (PIN)]}$$

luego entonces, sustituyendo en la ecuación 3.12 estos dos resultados.

$$\frac{\partial E}{\partial X_{(PIN-1)}} = -\delta_{(PIN-1)} W_{[(PIN-1) \rightarrow (PIN)]} \quad (3.13)$$

así de las Ec. 3.11 y 3.13, la Ec. 3.10 resulta ser:

$$-\frac{\partial E}{\partial Z_{(PIN-1)}} = \delta_{(PIN-1)} = \delta_{(PIN-1)} W_{[(PIN-1) \rightarrow (PIN)]} F'_{(PIN-1)} \quad (3.14)$$

y ya que los pesos entre las capas se deben variar proporcionalmente a

$$\frac{\partial E}{\partial W_{[(PIN-2) \rightarrow (PIN-1)]}} \propto \Delta W_{[(PIN-2) \rightarrow (PIN-1)]}$$

encontramos al sustituir 3.9 y 3.10 en la ecuación 3.8 que:

$$(\Delta W) \propto \frac{\partial E}{\partial W} = -\delta_{[FIN-1]} X_{[FIN-2]}$$

Por lo que, entre la capa de entrada y la intermedia (pensando en la red de tres capas, una de salida, la intermedia y la entrada), tenemos introduciendo una constante de proporcionalidad μ que:

$$\Delta W_{[(FIN-2) \rightarrow (FIN-1)]} = \mu \delta_{[FIN-1]} X_{[FIN-2]} \quad (3.3.b)$$

donde según vimos $\delta_{[FIN-1]} = \delta_{[FIN]} W_{[(FIN-1) \rightarrow (FIN)]} F'_{[FIN-1]}$

Este primer brinco hacia atrás (o retropropagación), nos permite calcular las deltas en la penúltima capa, a partir de las δ 's disponibles en la capa de salida y con esto, poder realizar una variación en los pesos de manera inteligente, es decir, provocar un cambio en los elementos de la matriz de pesos, formada entre las capas de entrada e intermedia, $_{[FIN-2]}$ y $_{[FIN-1]}$, acorde a la ecuación 3.3.b, mientras que la ecuación 3.3.a nos dicta como debe ser la variación en la matriz de pesos, formada entre las capas intermedia y la de salida. De esta manera, vamos construyendo de brinco en brinco hacia atrás, un procedimiento recurrente, que nos permitirá estimar las δ 's de las unidades pertenecientes a una capa, a partir de las δ 's ya calculadas antes de dar el salto hacia atrás. Así podremos ir retropropagando la señal del error o delta hasta alcanzar las unidades de la primera capa intermedia.

La red podría tener k capas, no importa cuantas, tan sólo bastaría generalizar a la ecuación 3.3:

$$\begin{aligned} \Delta W_{P_i[(k-1) \rightarrow (k)]} &= \mu (D - X_{[k]}) F'_{[k]}(z) X_{[k-1]} \\ \Delta W_{P_i[(k-1) \rightarrow (k)]} &= \mu \delta_{[k]} X_{[k-1]} \end{aligned} \quad (3.3)$$

donde las δ 's serían generadas a partir de las anteriores. Tendríamos que realizar cambios de pesos (ΔW), en forma proporcional al gradiente descendente del error, generando así las deltas necesarias que aseguren una disminución del error en cada ciclo (patrón tratado y con el cual se esté

trabajando, designado por P_i). Con este procedimiento nos aseguramos que vamos acercándonos, cada vez más, a tener una similitud entre el vector blanco y el vector que realmente genera la red.

3.3.3 ENSEÑANZA - APRENDIZAJE. ENTRENAMIENTO HEBBIANO.

Como es sabido, el conexionado en un principio está inicializado con ciertos valores de pesos aleatorios y cada vez que variamos los pesos, en cada ciclo, sobre cada una de las conexiones, conforme a la ecuación 3.13, la red va disminuyendo la discrepancia entre lo que es y lo que debería de ser, acercándose cada vez más a la reacción requerida, para el estímulo asociado al patrón aplicado. A este ajuste del cambio de pesos, es a lo que se le da el nombre de aprendizaje. Los cambios de pesos se realizan muy lentamente, para que la discrepancia entre lo que era y lo que es, no sea muy grande y la red vaya conservando, en cada nuevo ciclo, las características más importantes del patrón anterior. Siguiendo la regla de Hebb, ecuación 2.2, podemos escribir en forma matricial, los valores de eficiencia sináptica o la variación de los pesos asignados, como:

$$W_{\text{ciclo}+1}[k] = W_{\text{ciclo}}[k] + \Delta W_{\text{Variación}} \quad (3.15)$$

donde ΔW es dada por la ecuación 3.3. Es decir durante el entrenamiento se irá adquiriendo una experiencia acumulable, haciendo predominar algunas características o detalles comunes, de un cierto grupo de patrones. Así, el procedimiento de aprendizaje consiste en ir modificando los pesos para producir, como deseamos, un mapeo adecuado de la entrada a la salida. Con esto podemos lograr que patrones de entrada similares sean mapeados hacia un grupo de reacciones de salida similares, provocándose generalizaciones razonables, que le permitan manejar casos de patrones futuros, que nunca antes habían sido presentados a la red, simplemente detectando alguna de esas características comunes en ellos. Así la productividad, en el campo de trabajo, consiste en que cada vez que se enfrenta a una nueva tarea (patrón que se presenta por vez primera), el sistema neuronal sepa resolver, respondiendo correctamente tan sólo por reconocer alguna de esas similitudes aprendidas.

CAPÍTULO CUATRO

4 AUTOMATIZACIÓN DE LA RED AUTO AJUSTABLE

4.1 Proceso de aprendizaje de la red neuronal multicapas.

4.1.1 El programa entrenador de una red neuronal.

4.1.2 Ejemplos simples de tareas de reconocimiento y clasificación.

4.1.2.1 La X-OR y la separabilidad lineal.

4.1.2.2 La escuela de Caperucita.

4.2 La red ya entrenada (su eficiencia).

4.2.1 Generalización.

4.2.2 Evaluación del trabajo realizado por la red.

CAPITULO CUATRO

AUTOMATIZACIÓN DE

LA RED

AUTO AJUSTABLE

4 Automatización

En este capítulo hemos implementado en "software" el modelo neuronal descrito en el capítulo tres, al que se le entrena para la ejecución de algunas tareas simples, que revelen alguna habilidad del perceptrón multicapas. Con estas unidades se forman estructuras o arquitecturas neuronales sencillas que, a través de manejar el algoritmo de aprendizaje de la retropropagación del error, son capaces de aprender a resolver tareas autónomamente, según indican las simulaciones computacionales realizadas.

Con algunos ejemplos, estimamos el comportamiento de estos sistemas para auto-corrigirse y aprender a resolver tareas con eficiencia, principalmente las que se encuentran en el área de reconocimiento y clasificación de patrones. Se da una idea de como construir estas estructuras, para que puedan tratar con problemas de clasificación no lineal; esto, a través de separar el espacio en regiones de decisión compleja (lo cual se apreciará con el problema clásico de la X-OR). Asimismo se trata brevemente el significado de generalizar el conocimiento (lo aprendido); es decir, saber de antemano las reacciones o el trato que la red debe dar a un patrón que nunca antes había visto, simplemente por reconocer en él algunas características peculiares pertenecientes a un cierto grupo.

También los ejercicios de simulación muestran que, a pesar de tener la habilidad de auto-ajustarse hasta alcanzar el aprendizaje, muchas veces éste llega después de mucho tiempo y en algunas ocasiones se bloquea para jamás llegar a él. En tales casos, existen métodos para acelerar el algoritmo de entrenamiento, utilizando la retropropagación del error. Estos métodos son llamados de aprendizaje rápido y en algunas otras situaciones se considera introducir términos adicionales al cambio de pesos, denominándolos momentos, con los cuales se pueden superar, en algunas ocasiones, los estancamientos en la ruta hacia el aprendizaje.

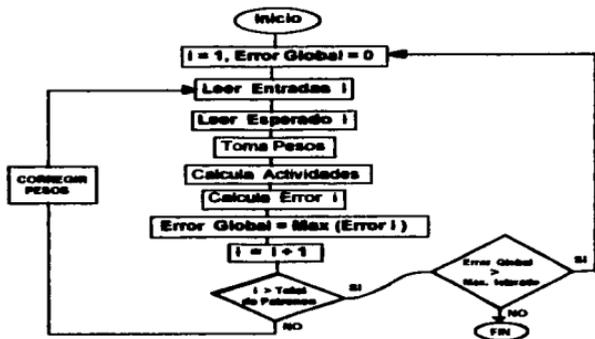


Fig. 4.1 Diagrama de flujo del programa utilizado para entrenar la red neuronal artificial.

4.1 PROCESO DE APRENDIZAJE DE LA RED NEURONAL MULTICAPAS

El proceso de aprendizaje radica en la adecuada actualización de los pesos de la fuerza del conexionado, dentro de la red. Esto se va realizando en forma repetitiva, basándonos en la comparación entre la respuesta real que da la red y la que desearíamos que diera.

El presentar a la red un cierto número de ejemplos de entrenamiento, tantas veces como sea necesario (hasta que se considere que el aprendizaje de todos ellos se ha llevado a cabo), implica la ejecución de un enorme número de iteraciones, que involucran operaciones aritméticas tediosas; por tal razón, es conveniente adaptar un proceso computacional automático que, basado en la teoría expuesta en los capítulos anteriores, nos lleve a determinar los valores idóneos de los elementos de las matrices de pesos W , dentro de la red. En este sentido se desarrollará un programa que supervise el aprendizaje de la red, utilizando para ello una PC, que pueda conseguir los cambios adaptativos que van dando la fuerza del interconexionado entre las neuronas, en esta etapa tan importante de aprendizaje.

4.1.1 EL PROGRAMA ENTRENADOR DE LA RED NEURONAL

A continuación se muestra un programa para entrenar redes neuronales multicapas de alimentación progresiva, utilizando la recursibilidad del algoritmo de retropropagación. Este procedimiento va minimizando la suma global del error sobre todo el conjunto de entrenamiento, basándose en las diferencias entre el valor real y el valor de salida deseado. En forma repetitiva, se va haciendo que la red se aproxime o converja poco a poco al aprendizaje. Se ha construido de manera general, para que pueda ser adaptado a casi cualquier arquitectura compuesta de tres capas, que disponga de un número arbitrario de neuronas y con cualquier conjunto de patrones ejemplos utilizados en el entrenamiento, simplemente cambiando las condiciones iniciales. Esta versatilidad hace que pueda ser adaptado para resolver un gran número de problemas. En la figura 4.1 se muestra el diagrama de flujo lógico; su implementación fue desarrollada en turbo pascal y el listado se encuentra en el apéndice D.

REAJUSTE EN LOS PESOS.

En una simulación programada de entrenamiento como ésta, se tiene la ventaja de poder sondear todo el conjunto de ejemplos, antes de realizar cualquier cambio en el interconexionado de la red.

La forma u orden en que se presentan los patrones y se realizan los ajustes en los pesos, nos marca una de las diferentes trayectorias que puede tomar la red, para conformar su conexionado y alcanzar el aprendizaje (entendiendo aprendizaje como un arreglo matricial particular de pesos en el conjunto de posibles soluciones equivalentes). Así el poder manejar o conducir al sistema por alguna de las rutas, representa una considerable ayuda para el instructor, por lo que se considera conveniente, anexar esta ventaja y en principio tendremos la capacidad para ejercitar a la red, en las dos maneras siguientes:

a.)- Trabajar con un patrón a la vez.

Se toma un ejemplo del conjunto de entrenamiento de forma azarosa y se procede a realizar el ajuste del conexionado, para disminuir el error de respuesta causado por este patrón en particular. Es decir, se barren uno a uno el conjunto de patrones de entrenamiento, sin importar el orden, hasta que todos ellos hayan pasado al menos una vez y tantas veces como se considere necesario.

b.)- Trabajar con todos a un tiempo.

Una vez que se le hayan mostrado a la red todos los patrones ejemplos, se sabe cuál de ellos originó el máximo error de respuesta (MaxError) y cuales son los valores de activación y las deltas necesarias, para hacer la corrección de pesos que nos permitan disminuir el MaxError a la salida, provocado por el patrón en cuestión, por así decirlo, "el más difícil de aprender", dándole a este patrón ejemplo, una atención especial, como si estuviéramos repasando con el ejercicio que más le cuesta a la red entender.

Al conducir a la red a la obtención de los elementos de las matrices de peso que resuelven el problema de los patrones ejercitados, se puede llegar a una gran eficiencia sobre estos ejemplos (un aprendizaje por memorización), pero no obstante, puede ser muy poco eficaz con patrones nuevos o también se puede caer en alguna ruta extremadamente larga o en estados estacionarios donde no podamos salir (barreras extremadamente altas). Para evitar o reducir estas posibles dificultades, tenemos durante la etapa de entrenamiento, la libertad de combinar ambas maneras de entrenamiento y de pasar de una a otra (de a \rightarrow b y viceversa), pero cualquiera que sea nuestra elección, tal como se dijo en el

capítulo 3, la ecuación 3.3 nos da el procedimiento cambiante de los pesos, para que la red alcance el conocimiento, a través de los incrementos calculados (ΔW) y sustituidos dentro de la ecuación 3.13 (regla de Hebb), que es la expresión escogida por nosotros en el algoritmo, para hacer la modificación de los pesos en las conexiones sinápticas y lograr con ello, una uniformidad en el conocimiento, en lo que se refiere a todos los patrones ejemplos, completando los ciclos o épocas que sean necesarios.

Los cambios entre las capas 2 y 3 se realizan haciendo uso de la ecuación 3.15 la cual es:

$$\underset{\substack{\text{Nuevos} \\ \text{Pesos}}}{W_{[2 \rightarrow 3]}} = \underset{\substack{\text{Viejos} \\ \text{Pesos}}}{W_{[2 \rightarrow 3]}} + \underset{\text{Variación}}{\Delta W_{[2 \rightarrow 3]}}$$

donde $\Delta W_{[2 \rightarrow 3]} = \Delta W_{[(FIN-1) \rightarrow FIN]} = \mu \delta_{[FIN]} X_{[FIN-1]}$ Ec. 3.3a.

mientras que los cambios de los pesos entre las capas 1 y 2 se harán utilizando:

$$\underset{\substack{\text{Nuevos} \\ \text{Pesos}}}{W_{[1 \rightarrow 2]}} = \underset{\substack{\text{Viejos} \\ \text{Pesos}}}{W_{[1 \rightarrow 2]}} + \underset{\text{Variación}}{\Delta W_{[1 \rightarrow 2]}}$$

donde $\Delta W_{[1 \rightarrow 2]} = \Delta W_{[FIN-2 \rightarrow FIN-1]} = \mu \delta_{[FIN-1]} X_{[FIN-2]}$ Ec. 3.3b.

donde la delta para una neurona particular, perteneciente a la capa número 2 según la ecuación 3.3b del capítulo 3 es:

$$\delta_{[FIN-1]} = \delta_{[FIN]} W F'_{[FIN-1]}$$

donde con nuestra función de activación $f(\varphi) = \frac{1}{1 - \text{Exp}(-\varphi)}$ encontramos

$$F' = F(\varphi)(1 - F(\varphi)) \Rightarrow$$

$$\delta_{[2]} = F(\varphi)(1 - F(\varphi)) * \delta_{[FIN]} * W_{[\text{Pesos que van de capa 1 a capa 2}]}$$

4.1.2 EJEMPLOS SIMPLES DE TAREAS DE RECONOCIMIENTO Y CLASIFICACIÓN

La mejor manera de probar el método de aprendizaje es asignándole alguna tarea específica, de relativa simplicidad, pero que a la vez muestre que la red tiene las propiedades básicas o la capacidad necesaria, para adquirir el conocimiento.

Uno de los primeros requisitos que se le pide a una red, es que sea capaz de resolver problemas no lineales, pues casi siempre los problemas interesantes en el reconocimiento y clasificación de patrones, presentan una no linealidad. El ejercicio clásico para evaluar esta cualidad, es el de la OR-exclusiva (X-OR); con este problema usualmente se revisa la no-linealidad de las unidades, por ser ésta una de las funciones no lineales más simples y por tanto fácil de manejar. Otra propiedad importante que se le pide a la red es la habilidad de generalizar, es decir, que pueda resolver tareas en base a la experiencia de lo aprendido. Lo que se pretende, es que a pesar de que las tareas asignadas sean nuevas, en el sentido que no fueron presentadas en la etapa de entrenamiento y aunque nunca hayan sido enseñadas a la red, ésta al presentar características semejantes a las que aprendió, sea capaz de resolverlas. También se ha tomado de la literatura infantil un cuento clásico con el fin de mostrar a manera de ilustración, con un ejemplo concreto, lo que la red neuronal puede hacer.

4.1.2.1 LA X-OR Y LA SEPARABILIDAD NO LINEAL

Desde que Frank Rosenblatt presentó su trabajo del Perceptrón y los mecanismos mentales, se demostró que, aunque todo parecía ir muy bien, su modelo no era capaz de aprender a resolver una función tan sencilla como era la X-OR (uno de los trabajos más completos fue el de Minsky en 1969 del Perceptrón [24]). Desde entonces se ha tomado el problema de la OR-exclusiva para verificar la no linealidad de una red neuronal [25, 26]. La idea es tener un sistema, el cual responda con un "uno" si éste recibe en sus entradas los valores $\{0, 1\}$ ó $\{1, 0\}$ y si no es así, responderá con un "cero"; esto es, tener una función que obedezca a la tabla de verdad 4.1. El sistema consta de dos entradas como muestra la fig. 4.2, llamadas entrada X y Y. El conjunto de patrones de entrada, consiste de cuatro combinaciones posibles que se pueden representar como puntos geométricos en el plano X-Y como se muestra en la fig. 4.3. Las combinaciones de entrada, que producirían una salida "cero" son dibujadas y nombradas A_0 y A_1 , mientras que las que producirían un "uno" son B_0 y B_1 .

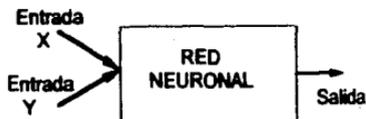


Fig. 4.2 Sistema neuronal, formado por dos unidades de entrada y una unidad de salida, la cual puede tomar los estados 1 = activado 0 = desactivado.

Puntos patrones	Valor X	Valor Y	Salida deseada
A ₀	0	0	0
B ₀	1	0	1
B ₁	0	1	1
A ₁	1	1	0

Tabla 4.1 Tabla de verdad de la OR-Exclusiva.

El trabajo de la red es combinar los valores que se presentan en sus entradas y dar una respuesta única. Si tuviéramos un sistema lineal el resultado neto de la combinación sería realizado bajo el siguiente cálculo:

$$\text{NET} = X W_1 + Y W_2 \quad (4.1)$$

luego este valor pasaría a una función de activación, que también sería lineal, la cual estaría encargada de dar la reacción de la red (activada = 1, desactivada = 0); la aseveración es entonces que no existen valores para los escalares de pesos W_1 y W_2 en la ecuación 4.1 que puedan producir la relación entrada salida de la tabla 4.1. Para entender esta limitación considere que NET puede ser mantenida constante en un valor digamos 0.5 y que la función de activación es la identidad, por tanto:

$$X W_1 + Y W_2 = 0.5 = \text{NET}$$

Así la combinación lineal de los valores de entrada que satisfacen esta ecuación deben caer sobre algún punto en la recta dentro del plano X-Y (fig. 4.3). Cualquier valor de entrada para X y Y sobre esta recta

podrá producir un valor de salida igual a 0.5. ya que nuestra función de activación, es la identidad $f(NET) = NET$.

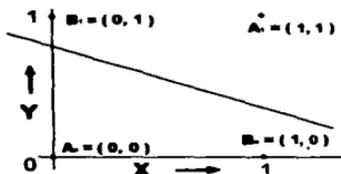


Fig. 4.3 Imposible separar con una línea recta los estados de actividad B's e inactividad A's en el problema de la X-OR. La ecuación de la recta trazada es

$$NET = \lambda W_1 + \frac{1}{\lambda} W_2$$

Ahora bien, combinaciones de los valores de W_1 , W_2 y de NET, podrían cambiar la pendiente y la posición de la recta, pero sin embargo para que la red produzca la función OR - exclusiva de la tabla 4.1 es necesario que el lugar de la línea sea tal, que todas las A's estén de un sólo lado y todas las B's estén del otro lado de la recta. Tratar de dibujar tal recta en la fig. 4.3 resulta imposible. Por ejemplo, si mantenemos constante la pendiente, los valores de salida sobre el lado superior de la recta producirían NET más grandes de 0.5 (los que se consideran como un 1 lógico) mientras que valores en el otro lado de la recta producen NET menores y por tanto una salida o respuesta menor que 0.5 (lo que es considerado como un cero lógico).

Lucgo entonces tratar de ajustar una recta a través de variar los pesos W_1 , W_2 de tal forma que los puntos A's estén separados de los B's, es geoméricamente imposible. Este resultado se extiende a redes con más de dos unidades de entrada y diremos, generalizando, que una clasificación de los patrones de entrada por una red perceptrón de este tipo está restringida a resolver problemas de separabilidad lineal.

Sin embargo este problema puede ser superado a través de adherir una no linealidad a la red, esto es agregando una capa más y reforzándola con funciones de activación no lineal. Con esto le damos a la red la posibilidad de que vaya componiendo la función no lineal que necesita para llevar en forma correcta un vector patrón del conjunto de entrada, a la salida correspondiente. La no linealidad de las unidades en nuestro programa, la verificaremos entonces construyendo la arquitectura mostrada en la fig. 4.4, a la que se le pedirá que aprenda la X-OR (la función definida por la tabla 4.1). Nuestro programa, al ir corriendo, nos va mostrando por medio de unas gráficas de error, cómo va disminuyendo éste, partiendo de

una configuración de pesos inicial, hasta conseguir finalmente las respuestas mostradas en la tabla 4.2, a través de ir actualizando los pesos en su interconexión, siguiendo la trayectoria de aprendizaje de las gráficas de la figura 4.5

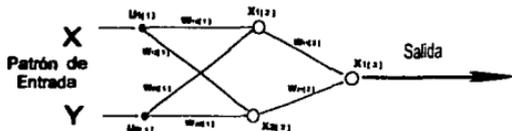


Fig. 4.4 En la entrada tenemos dos unidades que distribuyen los datos del patrón de excitación hacia la siguiente capa. Las unidades $X1[2]$ y $X2[2]$ forman la capa número dos. $X1[3]$ es otra unidad, representante de la capa 3, cuya actividad es la respuesta que da la red al patrón presente en la entrada.

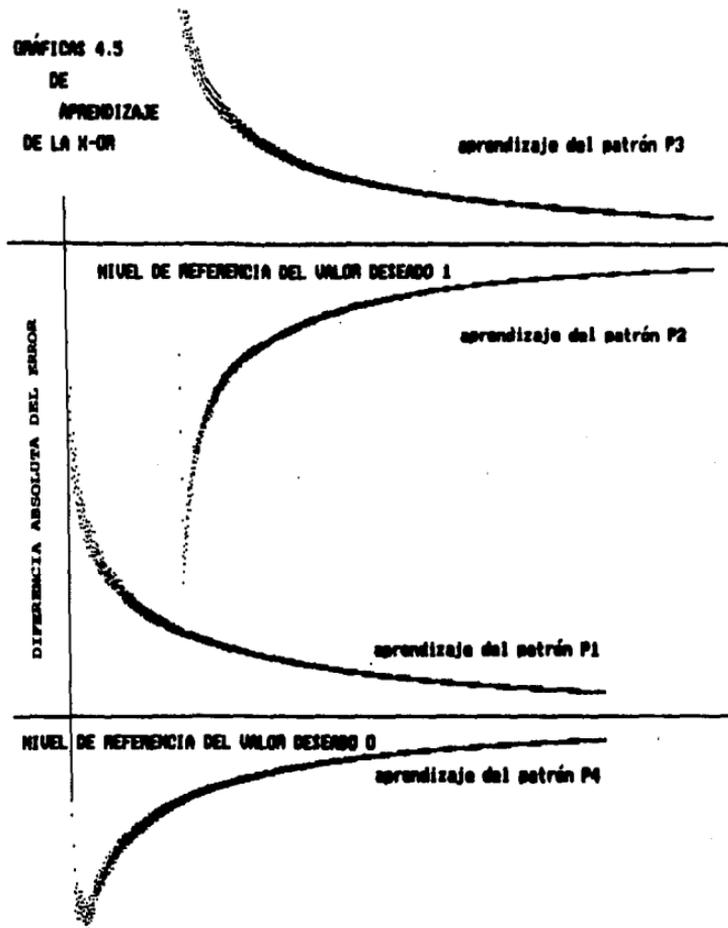
RESPUESTA O ACTIVACION QUE DA LA CAPA TRES	RESPUESTA DESEADA O BLANCO ESPERADO	ERROR DE DIFERENCIA ERROR3
$P_1 = (0, 0)$	0	-0.04913
$P_2 = (0, 1)$	1	0.06207
$P_3 = (1, 0)$	1	0.09207
$P_4 = (1, 1)$	0	-0.12286

MATRIZ DE PESOS ENTRE CAPAS 1 Y 2 \rightarrow $W11[1] = 0.9197$ $W21[1] = 0.9197$
 $W12[1] = 7.4884$ $W22[1] = 7.4950$

MATRIZ DE PESOS ENTRE CAPAS 2 Y 3 \rightarrow $W11[2] = -28.8664$ $W21[2] = 22.9404$

Tabla 4.2. Las actividades mostradas corresponden a los distintos patrones y son el resultado de efectuar 2000 ciclos (ajustes iterativos de los pesos en la arquitectura, donde un ciclo incluye una propagación y una retropropagación) y conseguir así los valores de peso necesarios en el interconexionado para las salidas deseadas.

GRÁFICOS 4.5
DE
APRENDIZAJE
DE LA H-OR



4.1.2.2. LA ESCUELA DE CAPERUCITA

Este ejemplo se relaciona con la fábula infantil de la pequeña Caperucita Roja, quien es mandada a la escuela a aprender como detectar lobos en el bosque. A la pequeña Caperucita le muestran una serie de imágenes de lobos hasta que ella caracteriza al patrón del lobo, como el ser que tiene grandes orejas, grandes ojos y grandes dientes. Estas características son detectadas a través de unidades especializadas de entrada (capa retinal) traduciéndolas después a una representación vectorial digital, cuyas componentes son ceros y unos; esta información es distribuida posteriormente a la capa intermedia donde se hace una representación interna del lobo, para que se manden las activaciones necesarias a la capa de salida, para así dar la reacción correspondiente final.

La Caperucita Roja podrá encontrar tres distintos seres en su mundo, que nosotros conocemos como lobo, abuelita y leñador: Ella debe aprender a ir a buscar al leñador, cuando tiene miedo, es decir, cuando ella detecte un ser con grandes orejas, grandes ojos y grandes dientes (el lobo). Así mismo debe aprender a acercarse, besar en la mejilla y alimentar a seres que son cariñosos, con arrugas y que tienen grandes ojos (la abuelita) y debe aprender a acercarse comida y coquetearle, a seres que son guapos, cariñosos y que tienen grandes ojos (el leñador).

Un modelo digital para el patrón de entrada, se presenta como un vector de ceros y unos; así el patrón de entrada para el lobo:

LOBO: grandes orejas
grandes ojos
grandes dientes

(1,1,1,0,0,0) = Lobo entrada

El vector blanco de salida que nosotros deseamos que la pequeña caperucita produzca, en respuesta al patrón de entrada del lobo, puede ser representado en forma similar, por tanto, el vector blanco de salida para el lobo:

CAPERUCITA: correr lejos
gritos de espanto
buscar al leñador

(1,1,1,0,0,0) = Respuesta para lobo

de esta forma podemos ir construyendo la tabla 4.3

PATRONES DE ENTRADA						
GRANDES OREJAS	GRANDES OJOS	GRANDES DIENTES	DULCE	ARRUGADO	GUAP0	
1	1	1	0	0	0	LOBO
0	1	0	1	1	0	ABUEJITA
0	1	0	1	0	1	LEÑADOR

REACCIONES DE SALIDA							
CORRER LEJOS	GRITOS DE ESPANTO	BUSCAR AL LEÑADOR	BESAR EN LA MEJILLA	APROXIMARSE	OFRECER COMIDA	COQUETEAR CON ...	
1	1	1	0	0	0	0	PARA EL LOBO
0	0	0	1	1	1	0	PARA LA ABUEJITA
0	0	0	1	1	1	1	PARA EL LEÑADOR

Tabla 4.3. Representación digital de los patrones entrada/salida. Los patrones de entrada contienen las características requeridas a reconocer. Las reacciones de salida son justo las respuestas deseadas que la red adoptará con cada correspondiente patrón de entrada.

Pensemos en la red de la figura 4.6, la cual está siendo entrenada para ir aprendiendo los patrones de entrada de la tabla 4.3, hasta alcanzar las reacciones de salida descritas en esa misma tabla. Así, antes de que el entrenamiento comience, todos los pesos del interconexiónado son inicializados con algún valor, el cual tiene una probabilidad muy baja de acertar a los pesos idóneos que darían las reacciones de salida deseadas. Rápidamente al ejecutar nuestro programa, sobre la red de la fig. 4.6, éste va modificando sus pesos en el interconexiónado, ajustándolos hasta alcanzar una construcción óptima que dará las reacciones de salida exigidas en la tabla 4.3. Así, la simulación de la red auto programable alcanzó la solución mostrada en la tabla 4.4, después de 1200 presentaciones de cada uno de los 3 patrones.

Para el PATRÓN 1 = (1, 1, 1, 0, 0, 0)

Activación	Capa 3	Nodo 1	=	0 97094	Blanco Esperado	=	1
Activación	Capa 3	Nodo 2	=	0 96947	Blanco Esperado	=	1
Activación	Capa 3	Nodo 3	=	0 96952	Blanco Esperado	=	1
Activación	Capa 3	Nodo 4	=	0 03325	Blanco Esperado	=	0
Activación	Capa 3	Nodo 5	=	0 02943	Blanco Esperado	=	0
Activación	Capa 3	Nodo 6	=	0 03090	Blanco Esperado	=	0
Activación	Capa 3	Nodo 7	=	0 00760	Blanco Esperado	=	0

Para el PATRÓN 2 = (0, 1, 0, 1, 1, 0)

Activación	Capa 3	Nodo 1	=	0 02993	Blanco Esperado	=	0
Activación	Capa 3	Nodo 2	=	0 02890	Blanco Esperado	=	0
Activación	Capa 3	Nodo 3	=	0 02040	Blanco Esperado	=	0
Activación	Capa 3	Nodo 4	=	0 98240	Blanco Esperado	=	1
Activación	Capa 3	Nodo 5	=	0 97143	Blanco Esperado	=	1
Activación	Capa 3	Nodo 6	=	0 98780	Blanco Esperado	=	1
Activación	Capa 3	Nodo 7	=	0 05363	Blanco Esperado	=	0

Para el PATRÓN 3 = (0, 1, 0, 1, 0, 1)

Activación	Capa 3	Nodo 1	=	0 00701	Blanco Esperado	=	0
Activación	Capa 3	Nodo 2	=	0 01156	Blanco Esperado	=	0
Activación	Capa 3	Nodo 3	=	0 02499	Blanco Esperado	=	0
Activación	Capa 3	Nodo 4	=	0 96974	Blanco Esperado	=	1
Activación	Capa 3	Nodo 5	=	0 98888	Blanco Esperado	=	1
Activación	Capa 3	Nodo 6	=	0 96951	Blanco Esperado	=	1
Activación	Capa 3	Nodo 7	=	0 92418	Blanco Esperado	=	1

PESOS ENTRE CAPA 1 Y 2

Pesos Entre	Entrada	1 y	Nodo	1	en	Capa	2	=	17152
Pesos Entre	Entrada	2 y	Nodo	1	en	Capa	2	=	-0 0075
Pesos Entre	Entrada	3 y	Nodo	1	en	Capa	2	=	-2 7152
Pesos Entre	Entrada	4 y	Nodo	1	en	Capa	2	=	-3 7137
Pesos Entre	Entrada	5 y	Nodo	1	en	Capa	2	=	0 6819
Pesos Entre	Entrada	6 y	Nodo	1	en	Capa	2	=	-1 3955
Pesos Entre	Entrada	1 y	Nodo	2	en	Capa	2	=	-0 6535
Pesos Entre	Entrada	2 y	Nodo	2	en	Capa	2	=	-0 7730
Pesos Entre	Entrada	3 y	Nodo	2	en	Capa	2	=	-0 3465
Pesos Entre	Entrada	4 y	Nodo	2	en	Capa	2	=	0 5735
Pesos Entre	Entrada	5 y	Nodo	2	en	Capa	2	=	3 0615
Pesos Entre	Entrada	6 y	Nodo	2	en	Capa	2	=	-3 4879
Pesos Entre	Entrada	1 y	Nodo	3	en	Capa	2	=	-0 5489
Pesos Entre	Entrada	2 y	Nodo	3	en	Capa	2	=	1 6436
Pesos Entre	Entrada	3 y	Nodo	3	en	Capa	2	=	-5 5489
Pesos Entre	Entrada	4 y	Nodo	3	en	Capa	2	=	2 1925
Pesos Entre	Entrada	5 y	Nodo	3	en	Capa	2	=	0 1861
Pesos Entre	Entrada	6 y	Nodo	3	en	Capa	2	=	-2 0063

PESOS ENTRE CAPA 2 Y 3

Peso Entre	Nodo 1	de la	Capa 2	y	Nodo de	Salida 1	=	3 0663
Peso Entre	Nodo 2	de la	Capa 2	y	Nodo de	Salida 1	=	3 2377
Peso Entre	Nodo 3	de la	Capa 2	y	Nodo de	Salida 1	=	3 7161
Peso Entre	Nodo 1	de la	Capa 2	y	Nodo de	Salida 2	=	1 3881
Peso Entre	Nodo 2	de la	Capa 2	y	Nodo de	Salida 2	=	0 8009
Peso Entre	Nodo 3	de la	Capa 2	y	Nodo de	Salida 2	=	-0 4448
Peso Entre	Nodo 1	de la	Capa 2	y	Nodo de	Salida 3	=	-5 0196
Peso Entre	Nodo 2	de la	Capa 2	y	Nodo de	Salida 3	=	-4 4993
Peso Entre	Nodo 3	de la	Capa 2	y	Nodo de	Salida 3	=	-3 6861
Peso Entre	Nodo 1	de la	Capa 2	y	Nodo de	Salida 4	=	3 2377
Peso Entre	Nodo 1	de la	Capa 2	y	Nodo de	Salida 4	=	3 7161
Peso Entre	Nodo 1	de la	Capa 2	y	Nodo de	Salida 4	=	-3 7695
Peso Entre	Nodo 1	de la	Capa 2	y	Nodo de	Salida 5	=	0 8009
Peso Entre	Nodo 1	de la	Capa 2	y	Nodo de	Salida 5	=	-0 4448
Peso Entre	Nodo 1	de la	Capa 2	y	Nodo de	Salida 5	=	0 8215
Peso Entre	Nodo 1	de la	Capa 2	y	Nodo de	Salida 6	=	-4 4993
Peso Entre	Nodo 1	de la	Capa 2	y	Nodo de	Salida 6	=	-3 6861
Peso Entre	Nodo 1	de la	Capa 2	y	Nodo de	Salida 6	=	3 4787
Peso Entre	Nodo 1	de la	Capa 2	y	Nodo de	Salida 7	=	3 7161
Peso Entre	Nodo 1	de la	Capa 2	y	Nodo de	Salida 7	=	-3 7695
Peso Entre	Nodo 1	de la	Capa 2	y	Nodo de	Salida 7	=	-3 2655

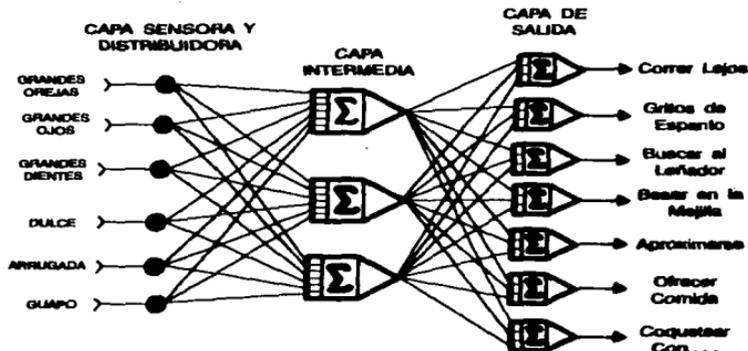


Fig. 4.6 Red neuronal multicapa de alimentación progresiva, entrenada automáticamente para aprender a representar el papel de la pequeña caperucita roja.

Las 6 unidades en la capa de entrada facilitan la adquisición de los 6 rasgos que caracterizan a cada habitante del bosque. Las 3 unidades en la capa intermedia interpretan la información y dan una caracterización interna de los seres del bosque, para que así la capa de salida active alguna de las 7 unidades según se requiera y se ejecute la acción respectiva.

4.2 LA RED YA ENTRENADA

En las soluciones demasiado rígidas, en las cuales la única forma de mejorar un programa es replantearlo y buscar una solución más adecuada, nos encontramos que siempre, después de su implementación, la eficiencia con la que se resuelve el problema queda determinada y no es posible variarla a menos que se reconstruya todo el sistema con una nueva solución. Aquí, donde normalmente se conoce, o se tiene a disposición, el conjunto universo de entradas y sus respectivas respuestas, lo que le conviene a la red neuronal es aprenderse

prácticamente, como de memoria, todos estos patrones o tareas a realizar y una vez que haya logrado memorizar todos ellos, la red nunca se equivocará en su reacción de salida, al ser utilizada en una vida activa en el campo de trabajo. Sin embargo, para estos problemas, casi siempre existen métodos computacionales convencionales, demasiado pulidos en su optimización, cuyas eficiencias son muy difíciles de mejorar a través de una simulación neuronal secuencial (no así en una implementación neuronal en hardware) por lo cual estos problemas no presentan, en este momento, un gran atractivo. Sin embargo existen problemas más interesantes, tareas donde las redes neuronales muestran ventajas sorprendentes y donde los algoritmos tradicionales no pueden resolver o que por ser demasiado largos e imprácticos, no son utilizables. En estos problemas, frecuentemente, es imposible que se aprenda todo el conjunto universo ya que éste suele ser demasiado grande, incluso algunas veces infinito y en otras ocasiones ni siquiera podemos disponer de él. Por tanto, en esta situación, la tarea más importante en la etapa de entrenamiento, es lograr que la red tenga una flexibilidad tal, que posea la habilidad para categorizar objetos que jamás ha visto, extrayendo, de los que sí ha visto, la estructura conceptual del mundo externo (generalización).

4.2.1 LA GENERALIZACIÓN.

Para lograr un entrenamiento óptimo en la red se necesita un gran número de detalles que atender, uno de ellos es el que se refiere a la capacidad de la red para generalizar, concepto conectado con que tan bien la red responda a los patrones, tanto a los aprendidos como a los nuevos. Así se dirá que la red aprendió, si al evaluar el trabajo ejecutado sobre patrones que nunca ha visto, sus predicciones son correctas, es decir, la ocurrencia de un error en tareas futuras es remota, con lo cual el conocimiento que adquirió en la etapa de entrenamiento es lo suficientemente general como para tratar con datos desconocidos.

La generalización depende por un lado de la arquitectura y del tamaño de la red y por otro, de la complejidad del problema y de la calidad y cantidad de los patrones de entrenamiento. Así, primeramente, la capacidad de generalización de una red neuronal puede ser mejorada al reducir la oportunidad de sobre adecuar los datos, a través de restringir el grado de conectividad en la red, a un valor que impida o reduzca la posibilidad de

sobre entrenarla, simplemente utilizando algunas conexiones desocupadas o irrelevantes, llegando rápidamente tan cerca de lo deseado que se pierda la generalidad sobre los datos no vistos, en vez de forzarla a buscar algún arreglo que le permita resolver el problema, consiguiendo, en el interconexiónado, un compromiso global sobre todos los patrones. Normalmente, no se conoce que tamaño de la red producirá el mejor resultado para un problema dado. Pero se ha verificado (mediante ensayos efectuados con varios tamaños de redes de alimentación progresiva), que configuraciones muy grandes, no son capaces de generalizar al mundo exterior, sino más bien, estos arreglos se aprenden como de memoria todo el conjunto de patrones de entrenamiento, sin sacar de él ningún conocimiento (por su gran número de conexiones les es más fácil hacerlo así, que buscar alguna generalidad útil), ya que al ser evaluada su eficiencia, se nota que la red sólo es capaz de resolver tareas idénticas a las mostradas en la etapa de entrenamiento y no así tareas nuevas o un poco distintas. Por otro lado configuraciones con pocas unidades no son capaces de aprender, ni siquiera el conjunto de entrenamiento, es decir no se les puede entrenar.

La red deberá tener un tamaño tal que capture la estructura de los datos (la capa de entrada), que modele o caracterice el problema y que arroje un resultado (las capas intermedias y de salida respectivamente). Con algún conocimiento específico de la estructura del problema, uno puede algunas veces, formar una buena estimación del tamaño apropiado de la red. Pero el tamaño idóneo de la red debe ser determinado por ensayo y error. Una técnica de aproximación es comenzar aumentando el tamaño de la red desde la más pequeña posible hasta que la ejecución llegue a un nivel de declinación. El tamaño de la red se incrementa al añadir más nodos en las capas fijas (intermedias) o por añadir nuevas capas. Otra alternativa, es seguir el camino contrario, comenzar con una red muy grande, aplicando una técnica de reducción para remover aquellas conexiones y nodos, los cuales tengan poca relevancia en la solución; en esta aproximación uno necesita conocer como seleccionar una red lo suficientemente grande para empezar. En lo que respecta al número de capas, en la mayoría de los casos, una red con tres capas (entrada - intermedia - salida), es suficiente para resolver casi cualquier problema de identificación y clasificación .

Ahora bien desde el punto de vista de la siguiente definición de generalidad: El conocimiento A es más general que el conocimiento B, si y sólo si, un ejemplo cubierto por B, es también cubierto por A pero no al revés. Es decir, A cubre más ejemplos que B, con tal definición se llega a la conclusión de lo que, le conviene a la red, es que el conjunto de entrenamiento sea lo más grande posible. De hecho, para entrenar a la red

tenemos que alimentarla con el mayor número de ejemplos y en cada caso, los valores de activación de las unidades de salida serán medidos y comparados con los valores deseados, de acuerdo a la entrada correspondiente para su previa asimilación.

4.2.2 EVALUACIÓN DEL TRABAJO REALIZADO POR LA RED.

¿ Como calificar el aprendizaje de un sistema ? Suponga que en la etapa de entrenamiento (aprendizaje), se mejora considerablemente la ejecución que desempeña la red. Entonces lo aprendido es válido y deberá estar bien calificado pues ha probado mejorar la ejecución significativamente. Si nosotros examinamos el conocimiento sólo de los datos de entrenamiento, el nivel de ejecución será alto y refleja muy bien el conocimiento apropiado para estos datos, pero ¿ qué tan confiable es esta estimación para los datos no vistos ? Esto guía al criterio, de que el aprendizaje debe conseguir que el sistema neuronal encuentre soluciones adecuadas de situaciones no antes vistas, mejorando su conocimiento, al extraerlo de los datos de entrenamiento. La evaluación de la ejecución debe medirse o calificarse con respecto a reacciones satisfactorias de datos no vistos en condiciones normales, en condiciones de ruido y con datos faltantes o con datos irrelevantes. Una calificación será así un indicador que estime indirectamente que tan bueno ha sido el aprendizaje.

El método de extensión es una simple prueba experimental de ensayo y error donde un conjunto de datos es particionado en dos subconjuntos disyuntos, uno usado para entrenamiento y el otro para prueba; si el medio ambiente o el mundo real es muy grande y disponemos de él, entonces podemos tener siempre un conjunto fuera del entrenamiento con el cual calificar la ejecución. Así, nosotros podemos llamar a esta técnica " incluir unos " y " omitir otros ", pues cada vez se va omitiendo un grupo para examinarlo y el resto para entrenamiento. Esta técnica emplea un gran consumo de tiempo para grandes ejemplos.

La idea básica detrás de la estimación del error o validación, es que el conjunto de prueba deberá ser independiente del conjunto de entrenamiento (patrones no antes vistos). La elección del subconjunto prueba deberá ser neutral, es decir al azar. Para evaluar una red que ha pasado por una etapa de entrenamiento, el criterio que emplearemos será el más natural de todos, el cual nos dice: de N patrones que jamás ha visto (aparentemente nuevos, en el sentido que no fueron parte del conjunto de entrenamiento), en cuántos acertó y en cuántos se equivocó, con lo cual daremos un porcentaje relativo en una cierta jornada de trabajo.

$$\text{Calificación de Efectividad} = \frac{\text{Número de Aciertos}}{N} \times 100$$

CAPITULO CINCO

SAYO DEL METODO

5.1 Utilización de una red neuronal artificial como ayuda en la interpretación del electrocardiograma.

5.1.1. El sistema cardiovascular.

5.1.2. El electrocardiograma. Huella de la dinámica del corazón.

5.2 Planteamiento general del sistema electrocardiográfico.

5.2.1. El amplificador.

5.2.2. El convertidor analógico - digital.

5.2.3. El detector de QRS.

5.3 Ensayo del método.

5.3.1 El programa neuronal reconecedor del complejo QRS.

5.3.2 Selección y almacenamiento de datos.

5.3.3 Etapa de enseñanza y entrenamiento.

5.3.4 Pruebas de generalización y evaluación.

CAPITULO CINCO

ENSAYO DEL MÉTODO

5 APLICACIÓN

En el área de la física médica, se implementará un sistema autoprogramable, basado en redes neuronales artificiales, que ayude a los cardiólogos en el monitoreo continuo del electrocardiograma de sus pacientes. Primero se presenta una serie de conceptos teóricos de la génesis del electrocardiograma, donde se irán dando las bases mínimas para un entendimiento de las interrelaciones físico - biológicas, con los aspectos electrofisiológicos del corazón.

5.1 UTILIZACIÓN DE UNA RED NEURONAL ARTIFICIAL COMO AYUDA EN LA INTERPRETACIÓN DEL ELECTROCARDIOGRAMA.

El electrocardiograma ha sido, muy probablemente, la primera señal estudiada, con fines diagnósticos, con la meta de automatizar su interpretación, utilizando la ayuda de una computadora digital. Los sistemas de electrocardiografía tienen como necesidad primordial el uso de un detector de QRS (una forma muy marcada en el trazo electrocardiográfico), pues a partir de su identificación se desarrollan cálculos que llevan al conocimiento de ciertos parámetros de interés, arrojando información que da, a los médicos, elementos de juicio para identificar algunas enfermedades o cubrir la necesidad de disparar algún tipo de alarma automáticamente, por lo que, tener un buen detector de QRS es fundamental. De aquí la preocupación por desarrollar detectores de QRS cada vez mejores y aunque existen algunos que han probado ser eficientes, sólo lo son en algunos casos y con ruido moderado, es decir, no existe uno que no presente algún tipo de problema. Tratando de mejorar tal deficiencia, se experimentará, en esta ocasión, con un programa para una computadora personal tipo PC, a fin de modelar una red neuronal artificial (RNA) de tres capas (fig 5.8), cuya arquitectura permita utilizar el método de aprendizaje por retropropagación. Esta red se ha estado entrenando en el reconocimiento del complejo QRS del electrocardiograma, utilizando segmentos de ECG normales y patológicos mientras que otros segmentos de ECG, se han estado empleando a fin de evaluar la eficiencia de la red. Así, se adaptará el simulador neuronal construido en el capítulo anterior, a un sistema de electrocardiografía, donde se explota la habilidad de las RNA para reconocer formas, específicamente, los rasgos característicos del complejo QRS.

5.1.1 SISTEMA CARDIOVASCULAR.

Dentro del sistema cardiovascular, le corresponde al corazón, la responsabilidad de mantener la circulación de la sangre en todo el cuerpo. El corazón es un órgano musculoso que bombea sangre hacia todos los tejidos del cuerpo, para nutrirlos con oxígeno y, a la vez, recoger con esta corriente de sangre, los productos de desecho metabólico celular. Para lograr esto, se requiere que las fibras musculares accionen en forma sincronizada, lo cual provoca que se generen señales eléctricas muy grandes sobre cualquier punto del volumen corporal. Esta actividad eléctrica que ocurre en el corazón cada vez que se contrae, se puede registrar y se conoce como electrocardiograma (ECG) [29, 30].

En el corazón, las dos cámaras superiores, más pequeñas, son las que reciben la sangre que retorna al corazón; se denominan aurícula izquierda y aurícula derecha y estos compartimientos están separados por una pared llamada tabique interauricular. Las dos cámaras inferiores, o ventrículos, separadas por el tabique interventricular, tienen la tarea de bombear la sangre fuera del corazón. El ventrículo derecho bombea sangre, no oxigenada, hacia los pulmones y el ventrículo izquierdo tiene un trabajo más exigente, al bombear sangre oxigenada hacia todos los tejidos del cuerpo y la cabeza (fig. 5.1) [27, 28].

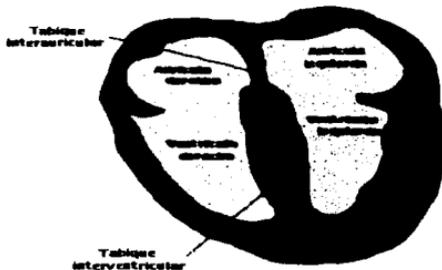


Fig. 5.1 Ilustración del corazón donde se muestran sus compartimientos.

El mecanismo que hace actuar al corazón como una bomba, es determinado por los eventos eléctricos de los cuales depende el patrón secuencial de excitación, de los subsecuentes procesos de contracción, que van de aurícula a ventrículo. El recorrido que lleva la sangre se ha ilustrado con los seis pasos que se muestran en la fig. 5.2. Aquí se ve, que la sangre no oxigenada retorna desde el cuerpo, hacia la aurícula derecha, fluyendo después hacia el ventrículo derecho, desde donde es bombeada hacia los pulmones, a través de la arteria pulmonar; ahí se oxigena y así queda lista para recircular por el cuerpo. Comienza primero su viaje, entrando a la aurícula izquierda a través de las venas pulmonares; luego, fluye hacia el ventrículo izquierdo y es bombeada hacia el cuerpo a través de la aorta, portando oxígeno a los tejidos.

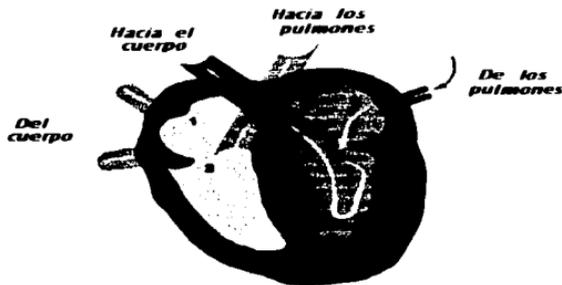


Fig.5.2 Diagrama Anatómico del Corazón, numerando la dirección del flujo sanguíneo por pasos.

- 1.- La sangre no oxigenada retorna a la aurícula derecha desde las venas cavas superior e inferior.
- 2.- La sangre sigue hacia el ventrículo derecho.
- 3.- La sangre es bombeada hacia los pulmones a través de la arteria pulmonar.
- 4.- La sangre oxigenada retorna a la aurícula izquierda a través de las venas pulmonares.
- 5.- La sangre fluye hacia el ventrículo izquierdo.
- 6.- La sangre es bombeada hacia el cuerpo, saliendo del corazón, a través de la aorta.

El mal funcionamiento de la acción mecánica de bombeo, puede llevar a anomalías graves y a la muerte misma. La contractibilidad o capacidad mecánica del corazón, de contraerse y relajarse, está controlada por los eventos bioeléctricos y su mal funcionamiento frecuentemente tiene fatales consecuencias, aún si el músculo cardíaco es capaz de una contracción normal y vigorosa. Los problemas en el sistema eléctrico, resultan en contracciones inapropiadas; teniendo una descoordinada actividad eléctrica, el bombeo puede incluso perderse, pues aunque los músculos estén vigorosos y viables, carecen de un apropiado disparador. Ésta es una de las principales causas por la cual, innecesariamente, mueren muchas personas en unos cuantos minutos u horas si no son atendidos con rapidez (ataque cardíaco por oclusión coronaria).

5.1.2 EL ELECTROCARDIOGRAMA .

El Electrocardiograma (E.C.G.) es el registro gráfico de los potenciales eléctricos producidos por el tejido cardíaco. La formación y conducción de estos impulsos eléctricos, producen pequeñas corrientes que fluctúan con el tiempo. Así, en pacientes en reposo, el E.C.G. se puede capturar sobre la superficie corporal, mediante un juego de electrodos conectados en puntos estratégicos sobre la piel, unidos a un aparato de registro.

La forma gráfica de la diferencia de potencial entre dos puntos cualesquiera del cuerpo, es dependiente, no solamente de la actividad eléctrica del corazón (distribución de carga) sino también, de la posición de él, en relación a los puntos escogidos para poner los electrodos; es decir la configuración espacial y otras particularidades, tales como la resistibilidad de los tejidos del cuerpo, son también factores influyentes. A pesar de esto, lo que se persigue es que las mediciones hechas, sólo dependan del corazón y no del resto del cuerpo, o de la posición de los electrodos. Esto ha sido un reto para la ciencia desde hace más de dos siglos, pasando por diversas etapas de investigación que han desembocado en perfeccionamientos tecnológicos y a la utilización de la computación digital [31, 32, 33]. Hasta el momento, la mejor aproximación a este ideal, requiere de la estimación del vector del campo eléctrico del corazón, calculado a partir de conocer las proyecciones ortogonales (sus componentes), sobre un sistema de coordenadas cartesianas, cuyo origen se coloca en el centro de gravedad del corazón; el eje y , apunta a los pies, el eje x hacia el brazo izquierdo y el eje z apunta a la espalda [27, 28].

Un arreglo básico particular de electrodos para medir estas componentes, se muestra en la figura 5.3, donde también se ilustra un trazo secuencial de la diferencia de potencial entre un par de electrodos cuyo registro puede ser obtenido con la ayuda de un galvanómetro.

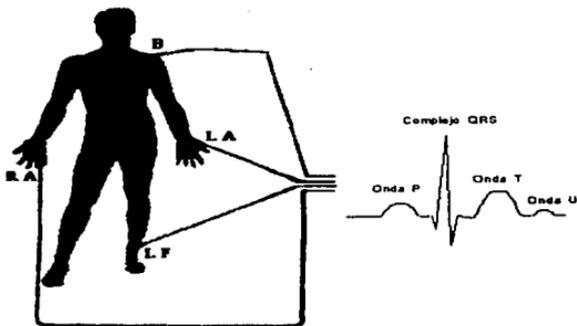


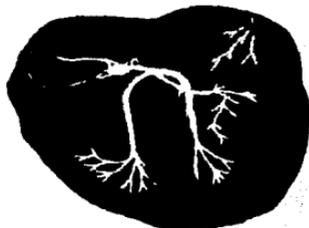
Fig. 5.3 Configuración estándar indicando los lugares donde son colocados los electrodos en este arreglo. La gráfica corresponde a un trazo estándar típico de la diferencia de potencial entre los electrodos RA vs. LF denominado conductión II.

Generalmente en el electrocardiograma hay cuatro desviaciones o deflexiones durante la duración de un ciclo cardíaco. La primera, designada la onda P, resulta de la despolarización auricular, excitación celular que contrae tales cavidades. La segunda, designada complejo QRS, resulta de la excitación de expansión a través de los ventrículos. La onda T, resulta de la recuperación de las células de los ventrículos y, una cuarta desviación, la onda U, es algunas veces vista siguiendo a la onda T. Su origen está aún en controversia [29, 34].

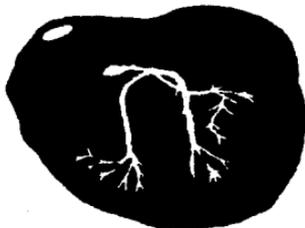
El seguimiento de la actividad eléctrica del corazón a través del pulso cardíaco marca la secuencia muscular. La siguiente serie de ilustraciones, nos ayudarán a correlacionar los sucesos eléctricos y mecánicos, con el trazo en los registros del ECG. (Atender a la región más oscura de las ilustraciones [34]).

NODO SINUSAL

NODO SINUSAL. El impulso cardíaco se origina en el nodo sinusal, denominado el marcapaso primario del corazón, ubicado en la parte superior de la aurícula derecha. Aquí se empiezan a obtener los primeros puntos del registro E.C.G. correspondientes a la onda P.



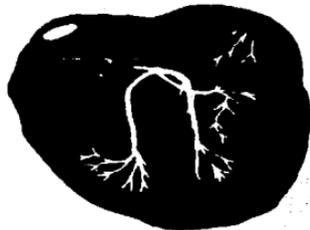
VAS INTERNODALES.



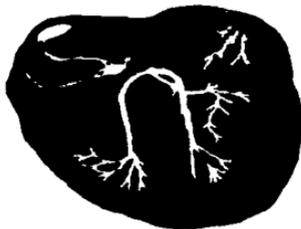
VAS INTERNODALES. El impulso es conducido a través de las vías internodales para la despolarización auricular, diseminándose a ambas aurículas y haciendo que se despolaricen y se contraigan, completándose finalmente la onda P. La frecuencia cardíaca usualmente se mide localizando dos ondas P consecutivas.

NODO AURICULOVENTRICULAR (AV)

La onda de despolarización llega al nodo AV que se ubica en el lado derecho del tabique interauricular, el cual se considera como un segundo marcapaso. El nodo AV requiere de más tiempo para activarse que el nodo sinusal, por tanto está sobramanejado por esto, es decir, si no recibe un impulso dentro del tiempo correspondiente a su periodo, iniciará su propio latido de escape. La onda se demora aquí aproximadamente 0.10 segundos, antes de pasar al haz de His, lo que es nombrado en el E.C.G. como segmento PR, intervalo entre el fin de la onda P y el inicio del QRS.



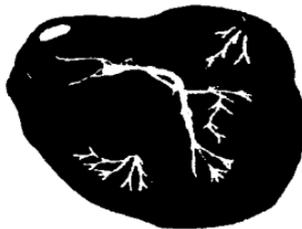
HAZ DE HIS



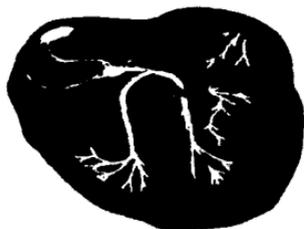
El impulso cardíaco se difunde desde el nodo AV hacia el delgado manojo de fibras denominadas haz de His, que conecta el nodo AV con las ramas que llevan la actividad hacia los ventrículos, para iniciar la actividad asociada con el QRS.

RAMA DERECHA E IZQUIERDA.

La rama derecha es un delgado fascículo que corre a lo largo del lado derecho del tabique interventricular y lleva el impulso eléctrico hacia el ventrículo derecho. Igualmente, la rama izquierda es la otra rama del haz de His y lleva el pulso eléctrico hacia el ventrículo izquierdo; corre a lo largo del lado izquierdo del tabique interventricular.



RAMA IZQUIERDA ANTERIOR



Lleva el pulso a las porciones anteriores y superiores del ventrículo izquierdo.

5.2 PLANTEAMIENTO GENERAL DEL SISTEMA ELECTROCARDIOGRÁFICO.

Para adquirir la señal eléctrica del corazón de un paciente (ECG) (ver arreglo de la fig 5.4), se le instala un juego de electrodos de acuerdo a las normas médicas de cardiología y a través de un cable de conexiones, se transportan las diferencias de potencial producidas entre los electrodos, a un sistema de adquisición de E.C.G., en cuya entrada encontramos un amplificador que tiene la función de llevar la señal eléctrica del corazón, del orden de milivoltios, a volts, para que posteriormente pase a un convertidor analógico digital, que digitalice y por tanto, se puedan obtener valores numéricos de la señal del ECG. Una vez que se tienen los valores numéricos del E.C.G., se pasan a una interfase de una PC, para que, por medio de un programa de control, se realice el análisis de la señal de E.C.G. [32].

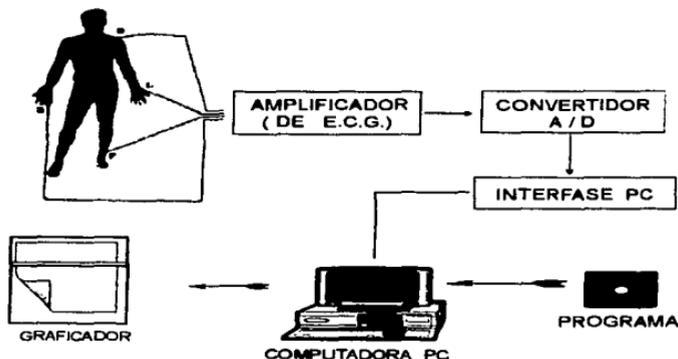


Fig 5.4 Descripción general a bloques del sistema electrocardiográfico

5.2.1 EL AMPLIFICADOR.

Las señales eléctricas del corazón, que se recogen a través de los electrodos y cables conectados al paciente, alimentan una interfase. Para que funcione, es necesario que la señal de E.C.G. tenga un nivel que corresponda al rango de operación de la interfase, por tanto, se hace un tratamiento previo a través de un amplificador que posee una muy alta impedancia de entrada (12^{12} Ohms), seguido de un convertidor A/D. Un grupo de circuitos amplificadores, magnifican la señal hasta 45 veces; a estos amplificadores se les han agregado filtros para dejar pasar señales sólo entre 0.05 y 100 Hz. En esta tablilla también se incorpora un sistema de protección al paciente, para no dejar pasar voltajes mayores de 700 milivoltios (del aparato al paciente). Así, de acuerdo a los estándares de electrocardiografía, se hacen pasar las señales procedentes de las derivaciones del brazo izquierdo (LA), brazo derecho (RA) y pie izquierdo (LF), hacia un arreglo resistivo denominado "red de Wilson" (ver fig 5.5), que pondera y entrega las señales derivadas, aVR+, aVR-, (voltajes aumentados del brazo derecho, positivo y negativo), aVL+, aVL- (voltajes aumentados del brazo izquierdo, positivo y negativo), aVF+, aVF- (voltaje aumentado del pie izquierdo, positivo y negativo) y la derivación común de referencia TCW (terminal común de Wilson) que simula el centro del corazón. Se incluye también un circuito generador de pulsos de 1 mV, que genera las señales de calibración Cal+ y Cal-, que se utilizan para ajustar los amplificadores a 1 cm/mV.

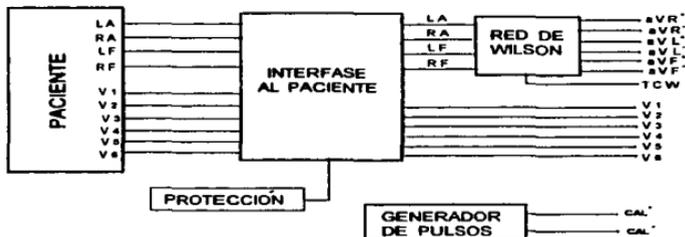


Fig.5.5 Adquisición de los datos del E.C.G. por el procesador. Los valores de los potenciales aumentados, están referidos al valor de TCW, que es el promedio instantáneo formado por los voltajes de las extremidades. El electrodo RF del pie derecho, normalmente es aterrizado, tanto para proteger al paciente de alguna descarga, como para fijar el valor común de tierra.

5.2.2 EL CONVERTIDOR ANALÓGICO DIGITAL.

Este módulo permite convertir hasta 12 señales analógicas, en señales digitales, utilizando técnicas de multiplexado en tiempo y conversión analógica a digital, (CAD) de aproximaciones sucesivas.

Como se muestra en el diagrama de la figura 5.6, el módulo o tabilla está formado por un *multiplexor*, que selecciona la señal de entrada, creando caminos o canales que son manejados por medio de un *controlador selector de canal*. Un *convertidor digital analógico* es utilizado como generador programable, que inicia una señal de realimentación de corriente directa, que sirve para ajustar el nivel de offset requerido en la señal de entrada a ser digitalizada (el nivel se maneja a través del microprocesador que manda una palabra digital a unos circuitos retenedores contenidos en el mismo circuito). Un *sumador* que sube o baja el nivel de offset por medio de una suma entre el canal seleccionado y la retroalimentación de corriente directa generada por el convertidor D/A. Un *amplificador programable* que ajusta la amplificación de la señal seleccionada. Un *convertidor analógico digital* que transforma la señal analógica seleccionada en digital, de acuerdo al registro selector de canal y que manda la señal de control necesaria para la amplificación que deba hacerse, poniendo en las líneas correspondientes los valores convertidos para que el procesador los lea. Un *registro de salidas digitales de control* que se programa para controlar los amplificadores de ECG. Un *registro de entradas digitales de control* para censar los estados de los amplificadores de ECG. Y un *decodificador* que controla el acceso a estos bloques.

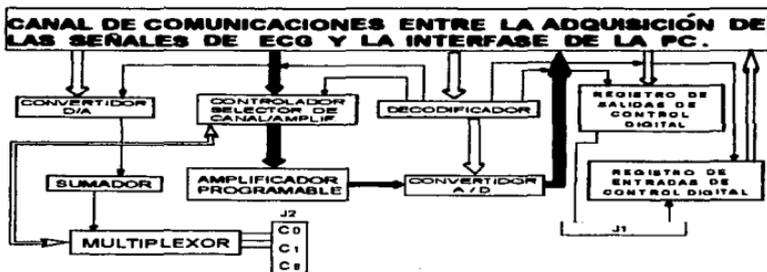


Fig. 5.6 Diagrama a bloques del convertidor A/D.

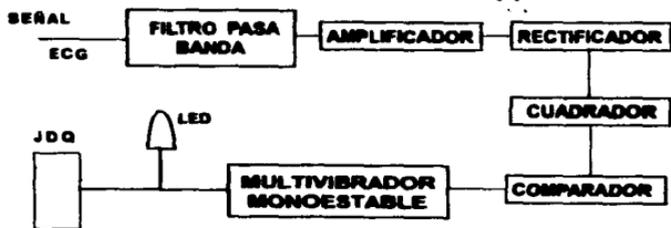


Fig. 5.7 Diagrama del detector de QRS.

5.2.3 EL DETECTOR DE QRS .

Se ha atacado el problema de la detección del complejo QRS de la señal electrocardiográfica, de dos maneras distintas: por programa ó software y por hardware ó instrumentación. De ellas el proceso más socorrido para la identificación de un posible QRS, es el que contempla pasar la señal de E.C.G. a través de un filtro de frecuencia de corte entre 15 y 20 Hz y con un ancho de banda de entre 5 y 10 Hz, que es donde normalmente se localizan la mayoría de los complejos QRS. Después de esto, ya sea que la señal se eleve al cuadrado o se rectifique, para que después, por alguna regla de decisión, se determine la autenticidad del posible QRS.

En el Departamento de Instrumentación del Instituto Nacional de Cardiología se ha estado trabajando en ambas líneas, contándose, en la parte de software, con un programa que, utilizando un algoritmo que calcula la segunda derivada del electrocardiograma, logra identificar el complejo QRS en tiempo real. Pruebas de evaluación sobre este programa, con diversos trazos de E.C.G.; han demostrado ser uno de los mejores que existen actualmente. Por otro lado, se cuenta con un equipo, diseñado en el mismo Departamento de Instrumentación, cuyo diagrama se presenta la Fig. 5.7, para detectar el complejo QRS. Empieza llevando la señal de E.C.G. a través de un filtro pasa banda, que deja libre el camino a señales cuyas frecuencias comprenden entre 10 y 20 Hz, donde se encuentran los complejos QRS; después se ajusta el nivel de amplitud con un amplificador, para que a continuación, la señal pase a un rectificador de media onda completa, que vuelve los pulsos QRS siempre positivos. Una vez que se tienen los QRS ajustados, se pasan por un elevador al cuadrado (cuadrador), que permite que las señales de ruido pequeñas, que puedan haber pasado, se vuelvan más pequeñas (pues se presuponen señales de ruido menores de 1 mV). Así las señales de QRS de interés se elevan al cuadrado y se logran separar ampliamente del ruido. Para seleccionar entonces los QRS y eliminar totalmente el ruido, se usa un comparador con una referencia que se encuentra arriba del ruido común; su salida da sólo dos niveles de voltaje (el +Vcc ó el 0), el nivel positivo dispara el multivibrador monoestable, que finalmente genera un pulso de 200 mS, que prende un LED indicador y avisa al procesador que ocurrió una onda R del ECG, para el cálculo de la frecuencia cardiaca del paciente.

5.3 ENSAYO DEL MÉTODO.

En los últimos años un gran número de detectores de QRS han sido desarrollados, los cuales trabajan bien en presencia de ruido moderado y sobre complejos más o menos normales. Sin embargo, usualmente, las señales que son sujetas a estudio son las que pertenecen a individuos que padecen algún problema cardiaco, por lo que no son precisamente normales; además, si tomamos en cuenta las diferencias que presentan los complejos QRS entre sujeto y sujeto, e incluso entre un mismo individuo y añadiéndole los casos patológicos que se presentan, hacen que los métodos tradicionales que existen para detectar el QRS, traigan consigo alguna dificultad asociada.

Por lo común, estos métodos sufren de la debilidad de las técnicas lineales, cuando son aplicadas a sistemas no lineales, es decir, es difícil, a una señal no lineal, adaptarle un modelo lineal. Y ya que el ECG es una señal no lineal generada por un sistema no lineal (el cuerpo humano) y dado que las redes neuronales artificiales son inherentemente modelos no lineales, éstas resultan ser potencialmente útiles en reconocer señales muy complicadas, como puede ser el complejo QRS. Por tanto se ha pensado en enfrentarlas a este problema, nada trivial de la física médica.

5.3.1 EL PROGRAMA NEURONAL RECONOCEDOR DEL COMPLEJO QRS.

Con base en redes neuronales, se ha desarrollado un programa para reconocer los complejos QRS. Este programa es como el presentado y desarrollado en los capítulos 3 y 4, que simula una red multicapas, como la que se muestra en la fig. 5.8, la cual está lista para aprender a reconocer el complejo QRS. Como se dijo en el capítulo 4, la red neuronal pasa por dos etapas: Una de entrenamiento, que le va dando, poco a poco, la habilidad para responder a la presencia de una entrada electrocardiográfica, reconociendo las características más sobresalientes del complejo QRS; en ella adquiere la experiencia necesaria, a través de mostrarle patrones ejemplos similares a los que podría encontrarse en una ejecución real (en el apéndice D se encuentra el listado del programa que simula esta red en lenguaje Pascal). La segunda etapa, aplica lo aprendido en la tarea de detección, haciendo que la red responda satisfactoriamente a la presencia del complejo QRS, cuando es estimulada con algún trazo de ECG cualquiera, para que, si se trata de un QRS, responda con una señal de aviso. Esta etapa en sí, es la prueba tangible a la que se ve sometida la red neuronal, cuando se evalúa su eficiencia. Estos puntos se discutirán con más detalle en los incisos que siguen.

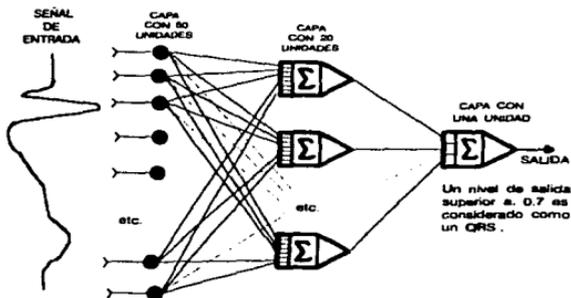


Fig. 5.8 Arquitectura o tamaño de la red conseguida mediante algunas ensayas previas de prueba y error, entrenada por retropropagación para que aprenda a reconocer automáticamente el QRS del electrocardiograma.

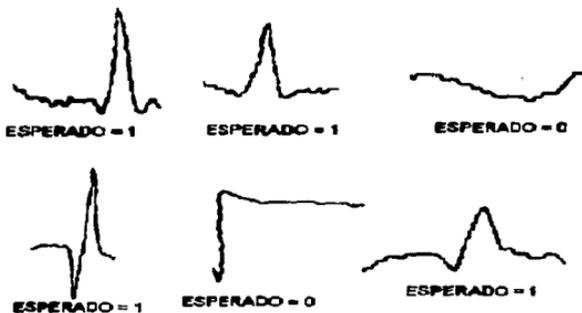


Fig. 5.9 Algunos patrones de entrenamiento formados por tramos de 50 puntos consecutivos.

Segmentos de ECG formados por

5.3.2 SELECCIÓN Y ALMACENAMIENTO DE DATOS .

La señal de E.C.G. que usamos para seleccionar los patrones de entrenamiento, es obtenida de la salida del convertidor analógico a digital. Este módulo proporciona los datos electrocardiográficos digitalizados que, a través de la interfase, se hacen llegar a la computadora (ver fig 5.4). Una vez en la computadora, se pueden manejar como a uno más le convenga, en este caso se dispuso grabarlos en un disco convencional, con una etiqueta de identificación. Mediante una breve inspección de todos los E.C.G. capturados, se advirtió que no todos los complejos QRS tienen el mismo tiempo de duración, por lo que se acordó tomar como patrones de entrenamiento, porciones de E.C.G. de longitud igual a 50 puntos (cada punto equivale a 4 milisegundos, por tanto, los 50 puntos tienen una duración de 200 milisegundos), medida que es lo suficientemente amplia como para contener cualquier complejo QRS (ver la fig. 5.9 donde se muestran algunos de estos patrones). Tomando en cuenta esta situación, se formaron unas especies de ventanas, donde se encontraba un trazo patrón formado por 50 puntos consecutivos del E.C.G. Estos valores, al ser almacenados o mantenidos en alguna localidad de memoria, junto con su respectivo blanco esperado, se ponen a disposición del programa simulador de la red neuronal (Neuro3QRS.PAS - red de la fig. 5.8), que los va tomando como imágenes proyectadas sobre su capa de entrada, donde cada unidad de entrada, tiene la localización de un punto imagen de la gráfica. En esta ventana se encuentra una porción del trazo electrocardiográfico. El blanco esperado se refiere a si la ventana no contiene un QRS o éste está incompleto; entonces su correspondiente esperado es igual a cero, en caso contrario sería igual a uno. De esta manera se cargan de forma azarosa tantos patrones de entrenamiento como sea posible para que, a través del programa de simulación del apéndice D adaptado a la red de fig. 5.8 (Neuro3QRS.PAS), se fueran repasando o asimilando hasta que la red aprenda a ver y distinguir en ellos, cualquier QRS, de entre ruido y otras ondas.

5.3.3 ETAPA DE ENSEÑANZA Y ENTRENAMIENTO.

El enseñar con ejemplos consiste en dirigir un patrón muestra, que es colocado en la entrada de la red, hacia la salida correspondiente o respuesta deseada; esto es, separar el conjunto de los que son QRS y el de los que no lo son. Para lograr esto, lo más simple sería introducir dentro del separador o clasificador de clase, una descripción exacta de los límites de tales clases. Sin embargo, en el caso del QRS, tenemos un conjunto universo infinito y las reglas que se han utilizado en los programas detectores de QRS, siempre han tenido algún tipo de problema.

Una persona frecuentemente es capaz de distinguir y reconocer objetos pertenecientes a alguna clase, debido a su experiencia acumulada, pero no es capaz de dar las instrucciones exactas de cómo hacer esto, a un programa o a una persona con escasa experiencia; sin embargo, es capaz de enseñarle a un inexperto, mostrándole ejemplos sin necesidad de instrucciones. Por ejemplo, los médicos enseñan a los jóvenes doctores a escuchar el corazón o los sonidos de la respiración en la práctica directa. Así, un matemático puede demostrar nuevos teoremas, no por que él tenga un algoritmo universal de cómo hacer esto, sino por el entrenamiento que él haya llevado. En realidad esta técnica de aprender con ejemplos es muy socorrida en la teoría de la enseñanza.

Escoger, de un sólo golpe, los patrones adecuados que traigan consigo los rasgos característicos que más le sirva a la red para identificar el QRS, es difícil; por tal motivo se cargaron tantos ejemplos como fue posible para que la red los fuera repasando. Todos los patrones fueron seleccionados al azar desde una base de datos construida con el sistema de electrocardiografía digitalizada descrito anteriormente; de aquí fueron tomados, tanto los ejemplos de entrenamiento, como los patrones de evaluación.

Para tener una noción de cómo se va dando el aprendizaje, se fueron proporcionando ó mostrando los ejemplos patrones a la red, en forma gradual. Es decir, se formaron grupos, con 49 patrones y antes de que los aprendiera, en un inicio, se verificó a cuales de ellos reconocía y a cuales no. Con esto el conjunto de entrenamiento fue creciendo en pasos de 49 en 49 (primera columna de la tabla 5.1). Así, antes de incorporarlos al conjunto de los patrones de entrenamiento, fueron usados para evaluar el incremento del aprendizaje experimentado por la red (segunda columna de la tabla 5.1), esto con la ayuda de un programa reconocedor (el mismo Neuro3QRS.PAS que sólo corre en esta ocasión la parte de propagación, en donde se detiene y da una

Comportamiento de la red al ir progresando su entrenamiento

Número de patrones de Entrenamiento	Patrones de Evaluación	Dictámenes FALSOS	Dictámenes VERDADEROS	Error en inferencia (%)
1 - 49	99 - 99	21	28	42.9
1 - 98	99 - 147	20	29	40.8
1 - 147	148 - 196	21	28	42.9
1 - 196	197 - 245	22	27	44.9
1 - 245	246 - 294	19	30	38.5
1 - 294	295 - 343	12	37	24.5
1 - 343	344 - 392	15	34	30.6
1 - 392	393 - 441	13	36	26.8
1 - 441	442 - 490	11	38	23.4
1 - 490	491 - 539	9	40	18.4
1 - 539	540 - 588	8	41	16.3

Tabla 5.1 Al ir evaluando a la red con patrones nuevos, se observa como al aprendizaje se va dando, conforme incrementamos los ejemplos o patrones de entrenamiento.

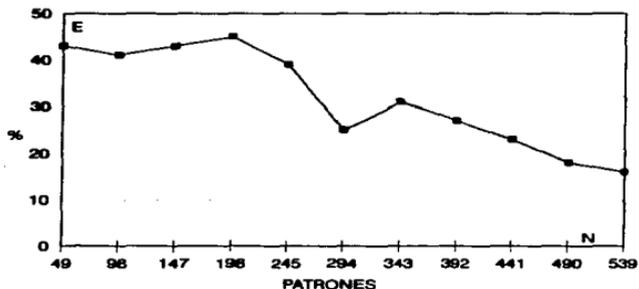


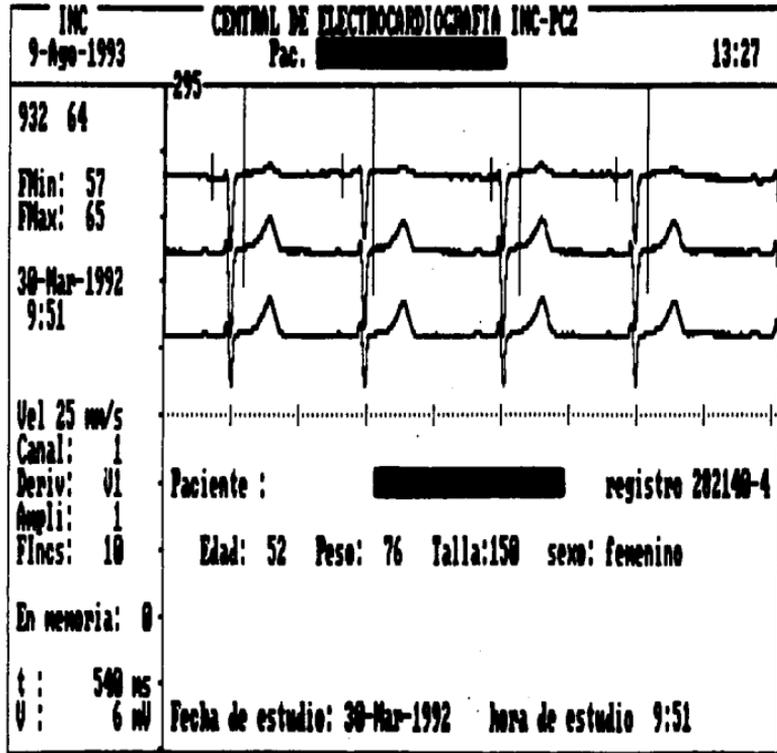
Fig. 5.10 Evolución del error total en los dictámenes de la red a medida que aumenta el conjunto de patrones exhibidos. Entre más ejemplos de entrenamiento se le muestran a la red, la equivocación en la respuesta se va reduciendo, adquiriendo un error en la inferencia cada vez menor.

respuesta del patrón presentado, sin intentar cambiar o reajustar el interconexiónado). Así sólo se ejecuta hacia delante la propagación del estímulo de entrada recibido para obtener una respuesta de salida en base a lo aprendido (columnas 3, 4 y 5 de la tabla 5.1). Una vez construida la información de la tabla 5.1., se utilizó para realizar una gráfica que estima el aprendizaje a través de las fallas y aciertos en función del número de patrones aprendidos. Así se metieron tantos grupos hasta que se consideró que el aprendizaje había sido fructuoso, es decir cuando el número de fallas se vio reducido, como puede observarse al traficar los valores de la tabla 5.1. (fig.5.10).

5.3.4 PRUEBAS DE GENERALIZACIÓN Y EVALUACIÓN.

La capacidad de generalización, se refiere al éxito con que la red enfrente nuevas situaciones, que no han aparecido en la entrada durante el proceso de aprendizaje. Así, para el reconocimiento de los patrones visuales del QRS, la red, o el joven médico, deberá aprender a distinguir el QRS entre un número finito o relativamente pequeño de patrones de enseñanza (muestras) y después, al finalizar la enseñanza, deberá conocer las reglas necesarias para clasificar el infinito o un número relativamente grande de situaciones, que puedan aparecer durante la vida activa o durante el proceso de prueba. La primera prueba satisfactoria de la capacidad o habilidad de la red para reconocer un QRS, la obtuvimos con el último grupo de 49 patrones, a los que reconoció casi sin equivocación. Sin embargo, se le sometió a una siguiente prueba que fuera más acorde a la realidad. Para esto, primero se grabaron ECG de pacientes durante un cierto tiempo, de tal forma, que la traza electrocardiográfica no se interrumpió, es decir, continuó durante un cierto intervalo de tiempo. En esta pequeña grabación se encontraban contenidos de 6 a 8 QRS, todo esto en forma continua (ver fig. 5.11). Después, se aplicó el programa de reconocimiento con una pequeña subrutina de barrido exploratorio, (un contador de longitud de ventana igual a 50 puntos) el cual iba mostrando a la red neuronal de 50 en 50 puntos

consecutivos o valores del trazo, para que ella determinara la autenticidad del QRS cuando se encontrara con él. Algunos de estos registros se ilustran en las figuras 11 a, b, c, en ellas se puede ver que en el inicio de una ventana que contiene un QRS, es decir el primer punto del trazo de longitud igual a 50 puntos, se marca con una pequeña raya vertical mientras que el final de los 50 puntos se indican con una raya más grande, esta raya únicamente se despliega en pantalla cuando se ha detectado un QRS junto con un sonido corto o biper. Aunque esta prueba de detección fue buena, en el sentido de que pudo localizar los QRS correctamente (como puede verse en las figuras 11), recordaremos que nuestro programa neuronal es una simulación secuencial de un modelo inherentemente paralelo y podría ocurrir que durante el tiempo de procesamiento utilizado por la computadora para analizar los primeros 50 puntos, se perdieran algunos datos importantes del siguiente segmento del trazo electrocardiográfico a procesar. Sin embargo esto no es necesariamente así, pues se puede superar con una máquina que tenga un procesador arriba de los 100 Mzh. Considerando que la adquisición de los 50 puntos se lleve a cabo en un tiempo de 200 milisegundos. Y si pensamos que este detector de QRS tiene la gran ventaja de que la red neuronal puede mejorar su eficiencia volviendo a la etapa de entrenamiento, si fuese requerida una exigencia mayor, resulta este método muy atractivo.



Legendo dato : 2944

Fig. 5.11 a.

INC
20-Jun-1996

CENTRAL DE ELECTROCARDIOGRAFIA INC-PC2
Pac. [REDACTED]

16:59

935Esc termina

FMin: 78
FMax: 78

[REDACTED]
17:45

Vel 25 mm/s
Canal: 1
Deriv: aVR
Ampli: 1
FIncs: 10

Paciente : [REDACTED] registro
Edad: 87 Peso: 42 Talla: 147 sexo: femenino

En memoria: 0

t : 876 ms
U : 0 mV

Fecha de estudio: [REDACTED] hora de estudio [REDACTED]

Leyendo dato : 2944

Fig. 5.11 b.

INC
20-Jun-1996

CENTRAL DE ELECTROCARDIOGRAFIA INC-PC2
Pac. [REDACTED]

17: 1

937Esc termina

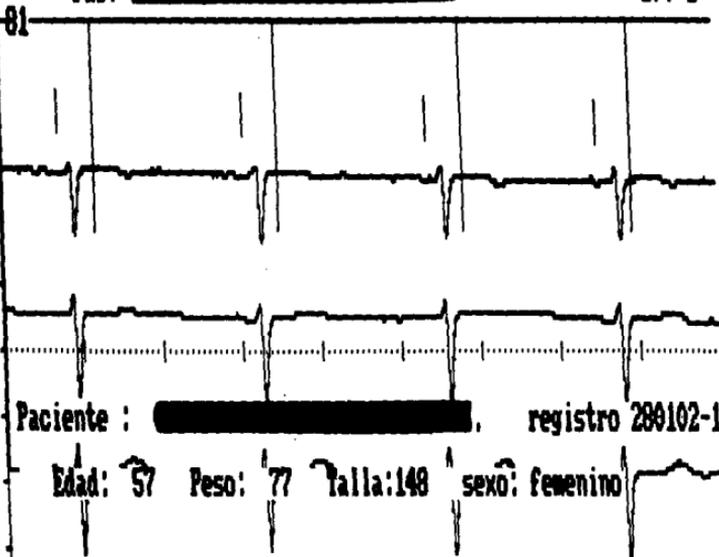
FMin: 54
FMax: 57

[REDACTED]
16:14

Vel 25 mm/s
Canal: 1
Deriv: VI
Ampli: 1
Fincs: 10

En memoria: 0

t : 838 ms
U : 12 mV



Leyendo dato : 2944

Fig. 5.11 c.

APÉNDICE A

La interacción aditiva en una retícula cuadrada de dos dimensiones cuyas celdas están ocupadas por dos tipos de átomos A y B es:

$$\sum_{\text{vecinos}} \frac{\epsilon_i + \epsilon_j}{2} = 2(N_A - N_B)$$

donde N_A , N_B son el número de átomos del tipo A, B respectivamente.

Demostración

Dado $\epsilon_i + \epsilon_j$

A	B	=	0
A	A	=	2
B	B	=	-2
B	A	=	0

Así que
$$\sum \frac{\epsilon_i + \epsilon_j}{2} = \sum \pm 1;$$

Con un arreglo alternado como el de la figura 1a, tenemos sumas de interacciones $(\epsilon_i + \epsilon_j) = 0$ entre pares de átomos que se van cancelando. Es decir parejas sólo de la forma (A, B) ó (B, A), por tanto.

$$\sum_{\text{vecinos}} \frac{\epsilon_i + \epsilon_j}{2} = 0$$

donde el número de átomos de tipo A es igual al número de átomos del tipo B.

Arreglo Alternado

A	B	A	B
B	A	B	A
A	B	A	B
B	A	B	A

De la figura
Vemos

$$2(N_A - N_B) = 0$$

Arreglo Aglomerativo

B	B	B	B
B	B	B	A
A	A	A	A
A	A	A	A

$$N_A > N_B$$

$$2(N_A - N_B) > 0$$

Fig. 1a *Enrejado con
disposición alternada*

Fig. 1b *Enrejado con aglomeraciones
o agrupación poblacional.*

En el segundo enrejado (fig. 1b.), si el número de átomos en las aglomeraciones, fuese $N_A = N_B$, tendríamos que las interacciones con valores +1 se cancelan con las de -1, y ocasionarían un resultado neto nulo.

$$\sum_{k \neq n} \frac{E_i + E_j}{2} = 2(N_A - N_B) = 0;$$

por lo que lo interesante es cuando hay un desequilibrio en el número de átomos que interactúan y en esta situación, ¿ Cuántos pares de interacciones ($E_i + E_j$) hay? Contémoslas para el caso de dos dimensiones:

Si L es igual al número de lazos totales en todo el arreglo, entonces este número también es igual al número de pares de interacciones ($E_i + E_j$). Es decir, tenemos tantas parejas ($E_i + E_j$) como nos indica el número L .

En el caso de dos dimensiones, el número de primeros vecinos que tiene un átomo cualquiera es 4. Ahora bien, debemos pensar en forma ideal, que no existan fronteras u orillas, es decir cualquier átomo está en una posición interna. Así el número total de lazos y en consecuencia de interacciones es $4(N_A + N_B)$; pero de todas estas parejas, hay que encontrar cuales nos dan -1, cuales +1 y cuántas 0. Pues bien, esto depende del número de átomos de la clase A así como del número de átomos de la clase B que tengamos. Por ejemplo, si todos los átomos del enrejado son del tipo A, entonces todas las parejas de ($E_i + E_j$) son +2;

así tenemos $4N_A = \text{Lazos} = \text{parejas} = \text{Interacciones}$

por tanto dividiendo por 2

$$\sum_{k \neq n} \frac{E_i + E_j}{2} = 2N_A$$

encontrando así, que en este caso el número de +1 es igual a $2N_A$

Luego si introducimos un solo átomo del tipo B, tenemos que ocho parejas $E_i + E_j$ son (0) y todas las demás serían (+2):

Es decir con $N_B = 1$ tenemos $4(N_A + N_B) = \text{Número total de lazos}$;

De los cuales en este caso 8 son cero y todos los demás +2. Se debe considerar, al contar las interacciones de primeros vecinos, tanto a ($E_i + E_j$) como a ($E_j + E_i$).

Así, por un lado, el número total de lazos es: $4(N_A + 1) = 4N_A + 4$
 (de ellos 8 son cero) y por tanto el número de parejas ($E_i + E_j$) diferentes de cero sería igual a $4N_A + 4 - 8 = 4N_A - 4$
 $= 4(N_A - 1) = 4(N_A - N_B)$

ahora dividiendo entre 2 encontramos

$$\sum_{\text{lazos}} \frac{E_i + E_j}{2} = 2(N_A - N_B)$$

Ahora, si introducimos dos átomos del tipo B, $N_B = 2$, donde estos dos átomos del tipo B son vecinos, tendríamos 2 parejas ($E_i + E_j$) con valor (-2) y 12 parejas (0). Todas las demás (+2).

Es decir si $N_B = 2$ y éstos son primeros vecinos*, tendremos:

$4(N_A + N_B) =$ Número total de lazos (de los cuales en este caso 12 son (0), 2 son (-2), todos los demás (+2)) [De los dos que son (-2), irán a cancelar a dos correspondientes (+2), teniendo 4 lazos más, netamente cero]

Así, por un lado el número total de lazos es $= 4(N_A + 2) = 4N_A + 8$

y por consecuencia el número total de parejas ($E_i + E_j$) que participan en la suma sería $= 4N_A + 8 - 12 - 4$
 $= 4N_A - 8 = 4(N_A - 2)$
 $= 4(N_A - N_B)$

∴

$$\sum_{\text{lazos}} \frac{E_i + E_j}{2} = 2(N_A - N_B)$$

Si introducimos 3 átomos del tipo B, $N_B = 3$ (donde ellos estén formando un aglomerado de 3), tendríamos 4 parejas ($E_i + E_j$) con valor (-2) y 16 parejas (0). Todos los demás (+2). Es decir si $N_B = 3$, tenemos

$4(N_A + N_B) =$ Número total de lazos (de los cuales, en este caso, 16 son cero y 4 son (-2). Todos los demás (+2). Poniendo atención a estas 20 interacciones de las cuales 4 son (-2), observamos que éstas van a cancelar a cuatro correspondientes interacciones con valor +2

* Si no fueran vecinos entonces tendríamos 16 interacciones en cero.

En el segundo enrejado (fig. 1b.), si el número de átomos en las aglomeraciones, fuese $N_A = N_B$, tendríamos que las interacciones con valores +1 se cancelan con las de -1, y ocasionarían un resultado neto nulo.

$$\sum_{\text{lazos}} \frac{E_i + E_j}{2} = 2(N_A - N_B) = 0;$$

por lo que lo interesante es cuando hay un desequilibrio en el número de átomos que interactúan y en esta situación, ¿ Cuántos pares de interacciones $(E_i + E_j)$ hay? Contémoslas para el caso de dos dimensiones:

Si L es igual al número de lazos totales en todo el arreglo, entonces este número también es igual al número de pares de interacciones $(E_i + E_j)$. Es decir, tenemos tantas parejas $(E_i + E_j)$ como nos indica el número L .

En el caso de dos dimensiones, el número de primeros vecinos que tiene un átomo cualquiera es 4. Ahora bien, debemos pensar en forma ideal, que no existen fronteras u orillas, es decir cualquier átomo está en una posición interna. Así el número total de lazos y en consecuencia de interacciones es $4(N_A + N_B)$; pero de todas estas parejas, hay que encontrar cuales nos dan -1, cuales +1 y cuántas 0. Pues bien, esto depende del número de átomos de la clase A así como del número de átomos de la clase B que tengamos. Por ejemplo, si todos los átomos del enrejado son del tipo A, entonces todas las parejas de $(E_i + E_j)$ son +2;

así tenemos $4N_A = \text{Lazos} = \text{parejas} = \text{Interacciones}$

por tanto dividiendo por 2

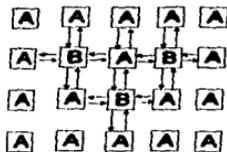
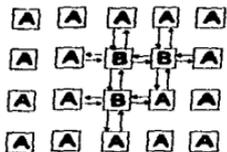
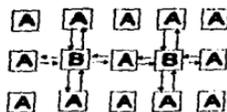
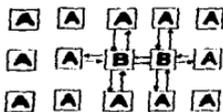
$$\sum_{\text{lazos}} \frac{E_i + E_j}{2} = 2N_A$$

encontrando así, que en este caso el número de +1 es igual a $2N_A$

Luego si introducimos un solo átomo del tipo B, tenemos que ocho parejas $E_i + E_j$ son (0) y todas las demás serían (+2):

Es decir con $N_B = 1$ tenemos $4(N_A + N_B) =$ Número total de lazos;

De los cuales en este caso 8 son cero y todos los demás +2. Se debe considerar, al contar las interacciones de primeros vecinos, tanto a $(E_i + E_j)$ como a $(E_j + E_i)$.



Así, por un lado, el número total de lazos es: $4(N_A + 1) = 4N_A + 4$
 (de ellos 8 son cero) y por tanto el número de parejas $(E_i + E_j)$ diferentes de cero
 sería igual a $4N_A + 4 - 8 = 4N_A - 4$
 $= 4(N_A - 1) = 4(N_A - N_B)$

ahora dividiendo entre 2 encontramos

$$\sum_{\text{lazos}} \frac{E_i + E_j}{2} = 2(N_A - N_B)$$

Ahora, si introducimos dos átomos del tipo B, $N_B = 2$, donde estos dos átomos del tipo B son vecinos, tendríamos 2 parejas $(E_i + E_j)$ con valor (-2) y 12 parejas (0) . Todas las demás $(+2)$.

Es decir si $N_B = 2$ y éstos son primeros vecinos*, tendremos:

$4(N_A + N_B) =$ Número total de lazos (de los cuales en este caso 12 son (0) ,
 2 son (-2) , todos los demás $(+2)$) [De los dos que son
 (-2) , irán a cancelar a dos correspondientes $(+2)$, teniendo
 4 lazos más, netamente cero]

Así, por un lado el número total de lazos es $= 4(N_A + 2) = 4N_A + 8$

y por consecuencia el número total de parejas $(E_i + E_j)$ que participan en la suma
 sería $= 4N_A + 8 - 12 - 4$
 $= 4N_A - 8 = 4(N_A - 2)$
 $= 4(N_A - N_B)$

∴

$$\sum_{\text{lazos}} \frac{E_i + E_j}{2} = 2(N_A - N_B)$$

Si introducimos 3 átomos del tipo B, $N_B = 3$ (donde ellos estén formando un
 aglomerado de 3), tendríamos 4 parejas $(E_i + E_j)$ con valor (-2) y 16 parejas (0) .
 Todos los demás $(+2)$. Es decir si $N_B = 3$, tenemos

$4(N_A + N_B) =$ Número total de lazos (de los cuales, en este caso, 16 son
 cero y 4 son (-2)). Todos los demás $(+2)$. Poniendo
 atención a estas 20 interacciones de las cuales 4 son (-2) ,
 observamos que éstas van a cancelar a cuatro correspondientes
 interacciones con valor $+2$

* Si los átomos vecinos estuvieran formando un aglomerado de 3.

Así, por un lado, el número total de lazos es $= 4(N_A+3) = 4N_A + 12$
 y por consecuencia el número total de parejas ($E_i + E_j$) que participan en la suma
 sería $= 4N_A + 12 - 16 - 8$
 $= 4N_A - 12 = 4(N_A - 3)$
 $= 4(N_A - N_B)$

∴

$$\sum_{\text{lazos}} \frac{E_i + E_j}{2} = 2(N_A - N_B)$$

Si seguimos con el caso en que, $N_A > N_B$ hasta que $N_B = N_A - 2$, donde el número total de (+2) en esta situación es:

$N_B = N_A - 2$ tenemos por un lado que el número total de lazos es $= 4[N_A + (N_A - 2)] = 4(2N_A - 2) = 8N_A - 8$

De todos estos lazos, los que nos interesan son los correspondientes a los dos átomos más que tenemos del tipo A. Y por tanto el número de parejas $E_i + E_j$ que nos darían (+2), sería 8, ya que todos los demás lazos se anulan.

$$\sum_{\text{lazos}} \frac{E_i + E_j}{2} = 4$$

haciendo uso de la fórmula $2(N_A - N_B)$ encontramos que nos da exactamente este mismo número.

$$\begin{aligned} \sum_{\text{lazos}} \frac{E_i + E_j}{2} &= 2(N_A - N_B) \\ &= 2[N_A - (N_A - 2)] \\ &= 4 \end{aligned}$$

APÉNDICE B

Para una retícula cuadrada de N celdas, hay $2N$ interacciones o lazos de primeros vecinos, pensando en que $E_i E_j = E_j E_i$, es decir que se trata de la misma interacción.

Átomos	Lazos
$N = 9 = 3^2 = n^2$	$2N = 2 [\# \text{ anterior de átomos} + 2n - 1]$
	$18 = 2 [4 + (2 \times 3 - 1)] = 18$
1° 2° 3°	<i>primera fila</i> (E1E2) (E2E3) (E3E4)
	<i>segunda fila</i> (E4E5) (E5E6) (E6E4)
	<i>tercera fila</i> (E7E8) (E8E9) (E9E7)
4° 5° 6°	<i>primera columna</i> (E1E4) (E4E7) (E7E1)
	<i>segunda columna</i> (E2E5) (E5E8) (E8E2)
	<i>tercera columna</i> (E3E6) (E6E9) (E9E3)

Átomos	Lazos
$N = 16 = 4^2 = n^2$	$2N = 2 [\# \text{ anterior de átomos} + 2n - 1]$
	$32 = 2 [9 + (2 \times 4 - 1)] = 32$
1° 2° 3° 4°	<i>primera fila</i> (E1E2) (E2E3) (E3E4) (E4E1)
	<i>segunda fila</i> (E5E6) (E6E7) (E7E8) (E8E5)
	<i>tercera fila</i> (E9E10) (E10E11) (E11E12) (E12E9)
	<i>cuarta fila</i> (E13E14) (E14E15) (E15E16) (E16E13)
9° 10° 11° 12°	<i>primera columna</i> (E1E5) (E5E9) (E9E13) (E13E1)
	<i>segunda columna</i> (E2E6) (E6E10) (E10E14) (E14E2)
	<i>tercera columna</i> (E3E7) (E7E11) (E11E15) (E15E3)
	<i>cuarta columna</i> (E4E8) (E8E12) (E12E16) (E16E4)

Átomos supuestos	Lazos
$N = n^2$ entonces	$2N = 2 [\# \text{ anterior de átomos} + 2n - 1]$
	$= 2 [(n-1)^2 + 2n - 1]$
	$= 2n^2$

Ahora por demostrar que si nuestros átomos son $N = (n+1)^2$

$$\begin{aligned}
 \text{Lazos} \\
 2N &= 2(n+1)^2 = 2 [\# \text{ anterior de átomos} + (2n+1) - 1] \\
 &= 2 [n^2 + 2n + 2 - 1] = 2(n+1)^2
 \end{aligned}$$

por tanto para una retícula de N átomos hay $2N$ lazos

APÉNDICE C

Se muestra que el mapa logístico de la ecuación $X_{i+1} = \lambda X_i(1 - X_i)$, tiene un doble ciclo para toda λ en $\lambda \in [3, 1 + \sqrt{6}]$.

Un doble ciclo existe si y sólo si, existen dos puntos p y q tales que

$$f(p) = q \text{ y } f(q) = p$$

equivalentemente, tal p debe satisfacer que $f(f(p)) = p$ donde $f(x) = \lambda X(1 - X)$.

Así p es un punto fijo en el mapa de segunda interacción $f^2(x) = f(f(x))$ y $f^2(x)$ es un polinomio de cuarto grado. Su gráfica para $\lambda > 3$ es mostrada en la fig. C.1

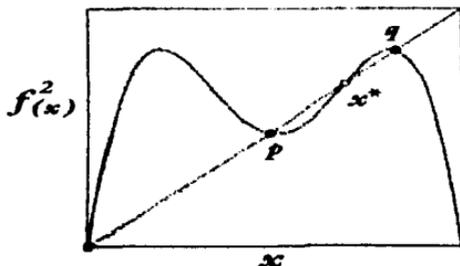


fig. C.1

Para encontrar p y q se necesita resolver la ecuación para los puntos donde la gráfica intersecta la diagonal, es decir, es necesario resolver la ecuación de grado cuatro $f^2(x) = x$. Esto suena muy difícil de realizar, no obstante puede darse cuenta, de la sección 1.2.1 que los puntos fijos $X^* = 0$ y $X^* = 1 - 1/\lambda$ son soluciones triviales de esta ecuación, ya que ellos satisfacen $f(x^*) = X^*$ y en forma automática $f^2(x^*) = X^*$.

Después de sacar los factores correspondientes a los puntos fijos, el problema se reduce a resolver una ecuación cuadrática:

$$\begin{aligned}
 P(x) &= f[\lambda X(1-X)] - \lambda[\lambda X(1-X)][1-\lambda X(1-X)] \\
 &= \lambda^2 X(1-X)[1-\lambda X(1-X)]
 \end{aligned}$$

Así que la expresión $f(x) - X = 0$ nos da $\lambda^2 X(1-X)[1-\lambda X(1-X)] - X = 0$
 $X[\lambda^2(1-X)[1-\lambda X(1-X)] - 1] = 0$

$$\therefore X^0 = 0 \quad \text{y} \quad \lambda^2(1-X)[1-\lambda X(1-X)] - 1 = 0$$

$$\begin{aligned}
 &-\lambda^3 X^3 + 2\lambda^3 X^2 - \lambda^2(1+\lambda)X + \lambda^2 - 1 = 0 \\
 &[X - (1 - 1/\lambda)][-\lambda^3 X^2 + \lambda^2(1+\lambda)X - \lambda(1+\lambda)] = 0
 \end{aligned}$$

$$\therefore X^0 = 1 - 1/\lambda \quad \text{y} \quad -\lambda^3 X^2 + \lambda^2(1+\lambda)X - \lambda(1+\lambda) = 0$$

por tanto después de resolviendo la ecuación cuadrática se obtiene el par de raíces:

$$p, q = \frac{\lambda + 1 \pm \sqrt{(\lambda - 3)(\lambda + 1)}}{2\lambda} \quad \text{las cuales son reales para } \lambda > 3$$

Así, el doble ciclo existe para todo $\lambda > 3$ como se exige. En $\lambda = 3$ las raíces coinciden, siendo iguales a $X^0 = 1 - 1/\lambda = 2/3$. Para $\lambda < 3$ las raíces son complejas lo cual significa que el doble ciclo no existe.

A continuación haremos el cálculo de la suma y el producto de estos valores.

$$\begin{aligned}
 P &= \frac{\lambda + 1 + \sqrt{(\lambda - 3)(\lambda + 1)}}{2\lambda} & q &= \frac{\lambda + 1 - \sqrt{(\lambda - 3)(\lambda + 1)}}{2\lambda}
 \end{aligned}$$

$$p+q = \frac{\lambda+1 + \sqrt{(\lambda-3)(\lambda+1)}}{2\lambda} + \frac{\lambda+1 - \sqrt{(\lambda-3)(\lambda+1)}}{2\lambda}$$

$$p+q = \frac{2(\lambda+1)}{2\lambda} = \frac{\lambda+1}{\lambda}$$

$$pq = \frac{\lambda+1 + \sqrt{(\lambda-3)(\lambda+1)}}{2\lambda} \cdot \frac{\lambda+1 - \sqrt{(\lambda-3)(\lambda+1)}}{2\lambda} \quad |$$

$$= \frac{(\lambda+1)^2 - (\lambda-3)(\lambda+1)}{4\lambda^2} = \frac{4(\lambda+1)}{4\lambda^2}$$

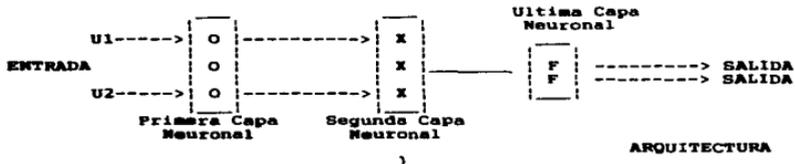
$$pq = \frac{(\lambda+1)}{\lambda^2}$$

APÉNDICE D

Listado del programa neuronal (en pascal). construido con la colaboración y ayuda del investigador Oscar Infante del Departamento de Diseño e Instrumentación del Instituto Nacional de Cardiología "Ignacio Chávez".

Program Neuronal;

{ Simulación de aprendizaje mediante RETROPROPAGACION de una red Neuronal de 3 Capas. Entrenamiento automatico para aprender a realizar una tarea }



USES {*****}

 crt ;

CONST

 NodosDeEntrada = 1 ;
 TotalDePatrones = p ;
 NodosEnCapa2 = 3 ;
 NodosEnCapa3 = k ;
 ErrorMin=0.01 ;

TYPE

 tEntrada = Record
 Entrada:Array [1..NodosDeEntrada] of Integer;
 Esperado:Integer;
 END;
 tPasosC1aC2 = Array [1..NodosDeEntrada,1..NodosEnCapa2] of Real;
 tPasosC2aC3 = Array [1..NodosEnCapa2,1..NodosEnCapa3] of Real;
 tActivaCapa2 = Array [1..NodosEnCapa2] of Real;
 tActivaCapa3 = Array [1..NodosEnCapa3] of Real;
 tErrorSalida = Array [1..NodosEnCapa3] of Real;
 tErrorC3 = Array[1..NodosEnCapa3] of Real;
 tDeltaErrorC2 = Array [1..NodosEnCapa] of Real;

VAR

PesosC1aC2 : tPesosC1aC2;
PesosC2aC3 : tPesosC2aC3;
Entradas: Array [1..TotalDePatrones] of tEntrada;
ActivaCapa2:tActivaCapa2;
ActivaCapa3:tActivaCapa3;
ErrorSalida:tErrorSalida;
ErrorC3:tErrorC3;
DeltaErrorC2:tDeltaErrorC2;

AuxEntradas: tEntrada;
AuxActivaCapa2:tActivaCapa2;
AuxActivaCapa3:tActivaCapa3;
AuxErrorC3:tErrorC3;

ArchPesos1 : File of tPesosC1aC2;
ArchPesos2 : File of tPesosC2aC3;
ArchEntradas : File of tEntrada;

NomArchEntradas: String;
NomArchPesos1: String;
NomArchPesos2: String;

AuxErrorC2 : Real;
MaxError:Real;
PatronConMaError: Integer;
ErrorGral:Real;

Ciclos:integer;
tecla:char;

{*****}

PROCEDURE PesosIniciales;

{ Construya la Matriz de Pesos entre las Capas }
{ uno-dos fijando el valor de sus Elementos de }
{ Manera Azarosa }

Var

i,j:integer; { i x j = Numero de Conexiones entre la capa 1 --> 2 }

BEGIN

CICLOS :=0;

Randomize;

For i:=1 to NodosDeEntrada DO
For j:=1 to NodosEnCapa2 DO BEGIN
PesosC1aC2[i,j]:=Random-Random;

END;

END;

```

PROCEDURE Pesos2Iniciales; { Construye la Matriz de Pesos entre }
                           { las Capas dos-tres Fijando el valor }
                           { de sus Elementos de manera Azarosa }

```

```

Var

```

```

    i,j:Integer;

```

```

BEGIN      ( i X j = Numero de Conexiones entre la capa 2 --> 3 )

```

```

    Randomize;

```

```

    For i:=1 To NodosEnCapa2 DO
    For j:=1 to NodosEnCapa3 DO BEGIN
        PesosC2aC3[i,j]:=Random-Random;
    END;

```

```

                                { PesosC2aC3 [ Nodo Capa 3, Nodo Capa 2 ] }

```

```

(*****

```

```

PROCEDURE LeerEntradas;

```

```

    { Abre el Archivo para leer un Patron y su Correspondiente }
Var      { Esperado Posesionandose en el lugar adecuado }

```

```

    i, ip:Integer;
    NuPat:string[3];

```

```

BEGIN      { La extension del archivo en disco }
           { Corresponde al numero de patron actual }

```

```

    For ip:=1 to TotalDePatrones DO BEGIN      { Escoje el campo de entrada ;
                                                { que tiene un numero de }
    Str(ip,NuPat);                             { extension igual al valor del }
    NomArchEntradas := 'Entradas.'+ NuPat ;    { patron actual, selecciona }
    Assign (ArchEntradas,NomArchEntradas);    { PatronActual = NuPat }
    Reset (ArchEntradas);
    Read(ArchEntradas,Entradas[ip]);
    Close(ArchEntradas);

```

```

WriteLn;
    for i:=1 to NodosDeEntrada Do
    Write(Entradas[ip].Entrada[i]:8);
    WriteLn;

    for i:=1 to NodosEnCapa3 Do
    Write(Entradas[ip].Esperado[i]:8);
    Writeln;

    end;
readln;
END;

```

```

(*****

```

PROCEDURE LeePesos;

Var { Captura los pesos optimos emanados }
{ del Programa Correspondientes a las }
Res:Char; { conexiones entre las capas dos - tres }

BEGIN

(SI-
NomArchPesos1 := 'Pesos1';
Assign (ArchPesos1, NomArchPesos1); { para vaciarlos o }
Reset (ArchPesos1); { leerlos del disco }
If IOResult=0 Then BEGIN { S=Pesos ya establecidos en disco }
Write('¿Se toman pesos previos?'); { N=Pesos Asignados al Azar }
res:=ReadKey;
If res in ['s','S'] Then BEGIN

WriteLn('Cual fue LA ULTIMA EPOCA ? ');
ReadLn(Ciclos);

Read(ArchPesos1, PesosC1aC2);
NomArchPesos2 := 'Pesos2'; { Conexiones entre las Capas 2 a 3 }
Assign (ArchPesos2, NomArchPesos2); { para vaciarlos o }
Reset (ArchPesos2); { leerlos del disco }
Read(ArchPesos2, PesosC2aC3);
Close(ArchPesos2);
END;

Close(ArchPesos1);
END;
END;

{*****}

PROCEDURE EscPesos;

BEGIN

NomArchPesos1 := 'Pesos1'; { Captura los Pesos Optimos Emanados }
Assign (ArchPesos1, NomArchPesos1); { del Programa para Grabarlos en Disco }
Rewrite (ArchPesos1); { o Vacía los Pesos Contenidos en Disco }
Write(ArchPesos1, PesosC1aC2); { al Programa }
NomArchPesos2 := 'Pesos2';
Assign (ArchPesos2, NomArchPesos2);
Rewrite (ArchPesos2);
Write(ArchPesos2, PesosC2aC3);

END;

{*****}

PROCEDURE MuestraPesos;

Var

res:Char;
N1,N2:Integer;
i,j:Integer;

BEGIN

WriteLn;
WriteLn;
WriteLn(' PESOS ENTRE CAPA 1 Y 2 ');
WriteLn;

For i:=1 to NodosEnCapa2 Do
For j:=1 to NodosDeEntrada Do
WriteLn(' Pesos Entre Entrada ',j:6,
' y Nodo en Capa2',i:4 ,PesosC1aC2[j,i]:10:4);

WriteLn;
ReadLn;
ReadLn;
WriteLn(' PESO ENTRE CAPA 2 Y 3 ');
WriteLn;

For i:=1 to NodosEnCapa3 Do
For j:=1 to NodosEnCapa2 Do
WriteLn(' Peso Entre Nodo ',j:6,
' de la Capa 2 y Nodo de Salida' ,i:4,PesosC2aC3[j,i]:10:4);

END;

{*****}

PROCEDURE ValoresIniciales;

BEGIN

PesosIniciales; { Manda una Semilla de Pesos, Entre las Capas 1-2 }
Pesos2Iniciales; { Siembra una Semilla de Pesos, Entre las Capas 2-3 }
LeePesos;

END;

{*****}

{*****}

{*****}

```

PROCEDURE CalculaActivaCapa2(ip:Integer);
  Var
    i,j:Integer;                               [ contadores ]
    SumaXConexion:real;                         ( Suma de estímulos por nodo)
  BEGIN
    For j:=1 to NodosEnCapa2 DO BEGIN          ( Barre nodos en capa 2 )
      SumaXConexion:=0;
      For i:=1 to NodosDeEntrada DO BEGIN     ( Barre nodos de entrada )
        SumaXConexion:=(SumaXConexion+PesosC1aC2[i,j]*
          Entradas[ip].Entrada[i]);
      END;
      ActivaCapa2[j] := 1/(1+exp(-SumaXConexion));
    END;
  END;
  (*****);

PROCEDURE CalculaActivaCapa3;
  Var
    i,j:Integer;                               [ Contadores ]
    SumaXConexion:real;                         ( Suma de estímulos por nodo )
  BEGIN
    For i:=1 to NodosEnCapa3 DO BEGIN        ( Barre nodos en capa 3 )
      SumaXConexion:=0;
      For j:=1 to NodosEnCapa2 DO BEGIN      ( Barre nodos de entrada )
        SumaXConexion:= (SumaXConexion + PesosC2aC3[i,j]
          *ActivaCapa2[j]);
      END;
      ActivaCapa3[i]:=1/(1+exp(-SumaXConexion));
    END;
  END;
  (*****);

PROCEDURE SalvaActivaciones;
  BEGIN
    AuxActivaCapa2:=ActivaCapa2; [ Excitacion del Patron Que Genero el Maximo ]
    AuxActivaCapa3:=ActivaCapa3; [ Error. ]
  END;

PROCEDURE SalvaEntradas(ip:Integer);
  BEGIN
    AuxEntradas:=Entradas[ip];
  END;

PROCEDURE SalvaErrores;
  BEGIN
    AuxErrorC3:=ErrorC3
  END;

```

```
PROCEDURE ErrorCapa3(ip:Integer);
```

```
    { Da el error en capa 3 .      ip es el patron a evaluar }  
    { se calcula el error de cada unidad en capa 3 y se sumaran }
```

```
Var
```

```
  i:Integer;      { Calcula el Error en Capa 3 generada por la }  
  AbsErr:Real;    { Entrada Actual, para ser comparado con los }  
  AbsMax:Real;    { Errores provocados por los otros patrones. }
```

```
BEGIN
```

```
  For i:1 to NodosEnCapa3 DO BEGIN  
    ErrorSalida[i] := Entradas[ip].Esperado [i] - ActivaCapa3[i];  
    ErrorC3[i] := ActivaCapa3[i] * (1-ActivaCapa3[i])  
                * ErrorSalida[i] ;
```

```
  Write('   Activacion Capa 3 Nodo',i, ' ', ActivaCapa3[i]:7:5);  
  WriteLn('         Blanco Esperado', ' ', Entradas[ip].Esperado[i]:7);  
  WriteLn(' Discrepancia de Salida Nodo',i, ' ', ErrorSalida[i]:7:5);
```

```
END;
```

```
  AbsErr := 0;
```

```
  For i : to NodosEnCapa3 DO BEGIN  
    AbsErr := AbsErr + Abs(ErrorSalida[i]);
```

```
  END;
```

```
    If AbsErr > MaxError Then BEGIN  
      MaxError:= AbsErr;
```

```
  SalvaActivaciones;  
  SalvaEntradas(ip);  
  SalvaErrores;
```

```
END;
```

```
END;
```

```
(.....)
```

```
PROCEDURE DeltaCapa3(ip:Integer);
```

```
Var
```

```
  ErrorC3:Real;  
  ErrorSalida:Real;
```

```
BEGIN
```

```
  ErrorSalida:=Entradas[ip].Esperado[NC3]-ActivaCapa3[NC3];  
  ErrorC3 := ActivaCapa3[NC3]*(1-ActivaCapa3[NC3])*ErrorSalida;
```

```
{ Write ('Activacion Capa 3', ' ', ' ', ActivaCapa3[NC3]:7:5);  
{ Write (' Blanco Esperado', ' ', ' ', Entradas[ip].Esperado[NC3]:7);  
{ WriteLn('Error Capa 3 ' , ' ', ' ', ErrorC3:7:5); }  
{ WriteLn; }
```

```
  ErrorMaxOcurridoC3:=ErrorC3;  
  MaxError:=Abs(ErrorSalida);  
  SalvaActivaciones;  
  SalvaEntradas(ip);
```

```
END;
```

```

PROCEDURE ErrorCapa2(uni:Integer);
Var j:Integer { Calcula el Error en la capa 2 generado }
      { por la Entrada que provoca el Maximp Error.}
      ErrorC2:Real
      { En esta capa tenemos n neuronas por tanto el }
      { patron de entrada genera n Errores. }
BEGIN
  ErrorC2:=0;
  For j:=1 to NodosEnCapa3 DO BEGIN
    ErrorC2 := ErrorC2+(AuxActivaCapa2[uni])*(1-AuxActivaCapa2[uni])
      *(PesosC2aC3[j,uni]*AuxErrorC3[j]);
    { WriteLn('AuxActivaCapa2 NODO ', i, ', ', AuxActivaCapa2[i]); }
    { WriteLn('PesosC2aC3[i,j]',PesosC2aC3[i,j]); }
    { WriteLn('Delta de Error C3 NODO ',j,', ', AuxErrorC3[j]); }
    { WriteLn('Delta de Error C2 NODO ', uni, ', ', DeltaErrorC2[j]); }
    { WriteLn('Delta de Error Suma Capa2 NODO ', uni, ', ', DeltaErrorCapa2[uni]); }
    { Readln; }
  end;
  DeltaErrorC2[uni]:=ErrorC2;
END;

(*****
PROCEDURE CorrigePesos;
Const
  mu=0.07;
Var
  { incl = Numera las Conexiones que llegan }
  { a un NODO i de la Capa 2 }
  { inc2 = Numera Ya Sea; la Conexion que Sale }
  { del NODO j de la Capa 1, 0 el Numero de }
  { Conexion que llega al NODO k de la Capa 3 }
  incl,inc2,inc3:Integer;
      { CORRECCION ENTRE LAS CAPAS 2 a 3 }
BEGIN
  For inc2:=1 to NodosEnCapa3 DO BEGIN
    For inc3:=1 to NodosEnCapa2 DO BEGIN
      PesosC2aC3[incl,inc2] := PesosC2aC3[inc2,inc3] +
        (mu*AuxErrorC3[inc2]*AuxActivaCapa2[inc3]);
    End;
    { CORRECCION ENTRE LAS CAPAS 1 -> 2 }
  For incl:=1 to NodosDeEntrada DO
    For inc2:=1 to NodosEnCapa2 DO BEGIN
      PesosC1aC2[incl,inc2] := PesosC1aC2[incl,inc2] +
        mu*DeltaErrorC2[inc2]*
          AuxEntradas.Entrada[incl];
    End;
  END;

```

PROCEDURE Entrenamiento1;

```
Var
  ip,uni:Integer;      { Acumula Experiencia, del Patron ip }
  AuxErrorC2:Real;
BEGIN
  MaxError:=0;
  For ip:=1 to TotalDePatrones DO BEGIN
    CalculaActivaCapa2(ip);
    CalculaActivaCapa3; {***** PROPAGACION *****}
    ErrorCapa3(ip):={ Supervision Del Patron Mas
                      Dificil de APRENDER}
  End;
  For uni:=1 to NodosEnCapa2 DO BEGIN
    ErrorCapa2(uni);
  End;
  Writeln;
  CorrigePesos;
END;
```

PROCEDURE Entrenamiento2;

```
Var
  ip:Integer;
BEGIN
  FOR ip:=1 to TotalDePatrones DO BEGIN
    CalculaActivaCapa2(ip);
    CalculaActivaCapa3;
    DeltaCapa3(ip,1);
    CorrigePesos;
  END;
END;
```

```
{*****}
{*****}
{*****}
{*****}
```

{*****PROGRAMA PRINCIPAL*****}

BEGIN

tecla:=#0;

ClrScr;

ValoresIniciales: { llama a la Sub-Rutina que Asigna los Pesos }
{ o el valor de Conexion, en toda la Red. }

{ ** Hay Dos Posibilidades Se Toman De Disco o se Asignan al Azar ** }

LeerEntradas;

REPEAT

Entrenamiento1 ;

If KeyPressed Then BEGIN

tecla:=ReadKey; { CICLO DE ENTRENAMIENTO }

tecla:=UpCase(tecla);
{ if tecla='P' then ModificaPesos;}

END;

Ciclos :=Ciclos + 1 ;

Sound(500);

WriteLn('-----Ciclos','Ciclos:8;)

WriteLn;

NoSound;

UNTIL (MaxError < ErrorMin) or (tecla = 'T');

MuestraPesos;

WriteLn;

ReadLn;

EscPesos; {MuestraPesos;}

{ Siempre Para Aprender el Patron Mas Dificil, es Decir }
{ el que Causa el Mayor Error }

END.

{ Para Salir de este Ciclo de Entrenamiento se Debe Cumplir }
{ la Condicion o Apretar la Tecla "T" }

BIBLIOGRAFÍA

1. STATISTICAL MECHANICS A SET OF LECTURES; Frontiers in physics, R. P. Feynman, Addison-Wesley Publishing Company, 17th. printing, 1994.
2. COURSE OF THEORETICAL PHYSICS, Vol. 5, Statistical Physics, Part I; L. D. Landau and M. Lifshitz, 3rd Edition, Pergamon Press Ltd, 1980.
3. QUANTUM THEORY OF MAGNETISM, White, R.M; McGraw-Hill, New York, 1983.
4. SPIN GLASSES Cambridge studies in magnetism: I K H. Fischer & J.A. Hertz, Ed. By D. Edwards and D. Melville, 1993
5. CALOR Y TERMODINÁMICA. Mark W. Zemansky y Richard H. Dittman, Ed. By McGraw-Hill, 1981.
6. CRITICAL POINT PHENOMENA : Universal physics at large length scales. Alastair Bruce and David Wallace Cambridge University Press 1989
7. CYBERNETICS , N. Wiener. Wiley, New York, and Hermann, Paris, 1948.
8. NEURAL NETWORKS AND PHYSICAL SYSTEMS WITH EMERGENT COLLECTIVE COMPUTATIONAL ABILITIES, J.J. Hopfield, Proc Natl. Acad. Sci USA, Vol. 79, pp 2554-2558, April 1982.
9. SPIN GLASSES AND BIOLOGY, Directions in Condensed Matter Physics, Vol. 6, Edited by Daniel L. Stein, World Scientific Publishing Co., 1992.
10. THE ORGANIZATION OF BEHAVIOR, Hebb,D.O., Ed. Wiley, N.Y., 1949.
11. NONLINEAR DYNAMICS AND CHAOS. With Applications to Physics, Biology, Chemistry, and Engineering Steven H. Strogatz, pp 324, Atractores, Doc.
12. PHYSICS OF NEURAL NETWORKS . MODELS OF NEURAL NETWORKS Domany E., Van Hemmen J.L., Schulten K., Eds., , Springer - Verlag edit., Germany, 1992
13. E. Gardner : J. Phys A: Math. Gen. 21,257 1988
14. SPIN GLASS THEORY AND BEYOND. M. Mezard, G. Parisi, and M.A. Virasoro (World Scientific, Singapore 1987).
15. PRINCIPLES OF NEURODYNAMICS, Perceptrons and the Theory of Brain Mechanisms, Rosenblatt, Frank, Spartan Books, Washington, D.C. 1962.
16. NEURAL MECHANISM OF VISUAL PERCEPTON, Proceeding of the Retina Research Foundation Symposia Vol. I and II, Man-kit Lam D., Gilbert, Charles,D. Portfolio Publishing Co., Houston, Texas, 1989.
17. NEUROSCIENCE AND CONNECTIONIST THEORY, Edited by Mark A. Gluck, David E Rumelhart , Lawrence erlbaum associates, publishers, 1992.
18. NEURAL COMPUTING ARCHITECTURES : THE DESIGN OF BRAIN-LIKE MACHINES, Edited by Aleksander Igor, North Oxford Academic Publishers Ltd., 1989.
19. MODELING BRAIN FUNCTION, Amit D J., Cambridge Univ. Press, 1988.

20. NEURAL NETWORKS, CONCEPTS, APPLICATIONS, AND IMPLEMENTATIONS. Vol 1, Antognetti Paolo, Milutinovic Veljko, Editors, Prentice Hall Advanced Reference Series Engineering, 1991
21. LEARNING INTERNAL REPRESENTATIONS BY ERROR PROPAGATION, in Parallel Distributed Processing, Exploration in the Microstructure of Cognition, Rumelhart, D.E., Hinton, G.E. and Williams, R. J. Mit Press, Cambridge, MA, 1986
22. LEARNING-LOGIC, Invention Report, Parker DB., S81-64, File 1, Office of Technology Licensing, Stanford University, October, 1982
23. BEYOND REGRESSION- NEW TOOLS FOR PREDICTION AND ANALYSIS IN THE BEHAVIORAL SCIENCES, Werbos, P. J. thesis in applied mathematics, Harvard University, Aug 1974
24. PERCEPTRONS, Minsky, M Papert, MIT Press, Cambridge, MA 1969
25. STATISTICAL PATTERN RECOGNITION WITH NEURAL NETWORKS, Kohonen, T., Barna, G. and Christley, R., Benchmarking Studies, " Proceedings of International Conference on Neural Network " (ICNN-88), Vol 1, 61-68, 1988
26. AN INTRODUCTION TO COMPUTATION GEOMETRY, Minsky, M Papert, 2nd ed, MIT Press, 1988
27. MEDICAL PHYSICS ELECTROCARDIOGRAPH, Johnston F.D. In Glajjer O (Ed), The year book publisher, 1944.
28. THEORETICAL FOUNDATIONS OF MEDICAL PHYSICS, Vol II Klip W. University of Alabama Press, 1969
29. ON THE THEORY OF THE ELECTROCARDIOGRAM, David B Geselowitz, Proceedings of the IEEE, Vol. 77, No. 6, 1989.
30. AN APPROACH TO QRS COMPLEX DETECTION USING MATHEMATICAL MORPHOLOGY, IEEE Transactions on Biomedical Engineering Trahanias P.E., Vol 40 N° 2, 1993.
31. A HISTORY OF ELETROCARDIOGRAPHY, Burch, G.E., and DePaquale, N.P. Year Book Medical Publishers, Inc. Chicago, 1964.
32. ELECTROCARDIOGRAFO MULTICANAL COMPUTARIZADO Rojas M., Rodríguez G., Garcia Moreira C., Rev. Mex. Ing. Biomédica 5 (1) 1984.
33. NEURAL - NETWORK - BASED ADAPTIVE MATCHED FILTERING FOR QRS DETECTION., Qiuzhen Xue, Yu Hen Hu, IEEE Transactions on Biomedical Engineering, Vol. 39, No. 4, April, 1992.
34. INTERPRETACIÓN DEL ECG, Davis D., Editorial Médica Panamericana, Argentina, 1991.