

8  
2g.



Universidad Nacional Autónoma de México  
Facultad de Ciencias

**Estudio del caos en sistemas  
dinámicos discretos**

Tesis para la obtención del título de  
Matemático

presenta

Jose Luis Herrera Aguilar

Director de tesis  
Dr. Federico Zertuche Mones



México, D.F. a Diciembre de 1999

2-13-2-16

TESIS CON  
FALLA DE ORIGEN



Universidad Nacional  
Autónoma de México

Dirección General de Bibliotecas de la UNAM

**Biblioteca Central**



**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  
AVENIDA LE  
MEXICO

**MAT. MARGARITA ELVIRA CHÁVEZ CANO**  
**Jefa de la División de Estudios Profesionales de la**  
**Facultad de Ciencias**  
**Presente**

Comunicamos a usted que hemos revisado el trabajo de Tesis:

Estudio del caos en sistemas dinámicos discretos

realizado por José Luis Herrera Aguilar

con número de cuenta 9038907-0 , pasante de la carrera de matemáticas

Dicho trabajo cuenta con nuestro voto aprobatorio.

Atentamente

Director de tesis  
Propietario Dr Federico Zertuche Mones

Propietario Dr Henri Waelbroeck

Propietario Dr Ramón López-Peña

Suplente Dr Humberto Carrillo Galvet

Suplente Dr Christopher Stephens

*Julio Cesar Guevara Bravo*  
*Ramón López Peña*  
*Humberto Carrillo Galvet*  
*Christopher Stephens*

Consejo Departamental de matemáticas  
Matemático Julio Cesar Guevara Bravo

FALTAN PAGINAS

De la: **i**

a la: **viii**

## AGRADECIMIENTOS

Esta tesis ha sido escrita como requerimiento para obtener el título de matemático de la facultad de ciencias de la UNAM. El objetivo principal de escribir esta tesis es el de llevar a cabo una introducción al estudio de los sistemas dinámicos discretos, en especial a las redes neuronales las cuales en los últimos años han tenido un importante desarrollo. En especial el primer capítulo es una introducción básica de las redes neuronales. En los capítulos subsecuentes se realiza investigación respecto a un tema muy reciente de las redes neuronales y la teoría del caos.

En esta tesis se encuentran plasmados casi dos años de trabajo, además es el final del trabajo realizado durante poco más de cuatro años de licenciatura. En especial deseo agradecer a mis padres y a mi hermano el apoyo que me brindaron para la realización de esta tesis, ya que sin el hubiera sido imposible realizarla. Un agradecimiento especial para mi director de tesis Dr. Federico Zertuche Mones, por su paciencia. deseo hacer una mención especial a Pinto diseño por su colaboración en la impresión de esta tesis. A Todos aquellos que de alguna manera me apoyaron o ayudaron, muchas gracias.

# CONTENIDO

	Página
Prefacio . . . . .	ix
 <b>CAPÍTULO 1: Introducción.</b>	
Introducción . . . . .	1
El cerebro y la neurociencia . . . . .	3
Historia . . . . .	5
 <b>CAPÍTULO 2: Modelo de Hopfield.</b>	
Memoria asociativa . . . . .	7
El modelo de Hopfield . . . . .	8
Caso 1: un patrón . . . . .	9
Caso 2: $P$ patrones . . . . .	10
Capacidad de almacenamiento de la red . . . . .	12
La función de energía . . . . .	15
Mecánica Estadística de los sistemas magnéticos . . . . .	17
Dinámica a baja temperatura . . . . .	19
Teoría del campo medio . . . . .	20
Redes neuronales con dinámicas estocásticas . . . . .	22
Teoría del campo medio . . . . .	23
Capacidad de la red neuronal . . . . .	25

### **CAPÍTULO 3: Modelo Asimétrico**

El modelo Asimétrico . . . . .	31
Estabilidad inicial de la red . . . . .	33
Capacidad de Almacenamiento . . . . .	34
Teoría del campo medio . . . . .	35
Resultados de las simulaciones . . . . .	35

### **CAPÍTULO 4: Modelo de Crisanti**

El modelo de Crisanti . . . . .	43
Análisis del modelo . . . . .	44
Resultados de las simulaciones . . . . .	45

### **CAPÍTULO 5: Redes neuronales con memoria**

Teoría del caos discreto . . . . .	53
Función de Correlación . . . . .	55
Modelo de caos discreto . . . . .	56
Resultados de las simulaciones . . . . .	57
Cálculo de la función de correlación y distancias de Hamming . . . . .	57
Cálculo de la función de correlación con distancia beta . . . . .	65
Análisis de la sensibilidad de las condiciones iniciales . . . . .	70
Modelo con memoria estilo Crisanti . . . . .	77
El modelo . . . . .	77
Resultados de las simulaciones . . . . .	78
Análisis de la sensibilidad de las condiciones iniciales . . . . .	83

**CAPÍTULO 6: Conclusiones.**

Conclusiones . . . . . 91

**BIBLIOGRAFÍA . . . . . 93**

## INTRODUCCIÓN

En un principio el interés en el estudio de las redes neuronales se enfocaba en la neurociencia, es decir en como funciona nuestro cerebro. A partir de la neurociencia se desarrollaron diversos modelos de redes neuronales. Uno de los problemas de los modelos matemáticos de las redes neuronales es la memoria asociativa, dentro de este problema se encuentra el problema de como asignar los valores correctos a la matriz de pesos para que la red tenga una buena memoria asociativa. En los primeros años de estudio se propusieron diversos algoritmos, estos algoritmos no solucionaban por completo los problemas existentes de la correcta asignación de los pesos de la red.

Fue hasta 1982 que Hopfield<sup>[1,2]</sup> introdujo el concepto de función de energía en las redes neuronales que se consiguieron importantes avances, poco antes Little<sup>[1,2,3]</sup> había mostrado la analogía de las redes neuronales y los sistemas desordenados de *spin*, mejor conocidos como vidrios de spin. ▸

A partir de ese momento el estudio de las propiedades térmicas de estos sistemas\* se ha incrementado considerablemente. El estudio térmico de estos sistemas permite el uso de la mecánica estadística y la teoría del campo medio, este interés trajo consigo importantes avances. La gran variedad de modelos generados a partir de las ideas de Hopfield y el gran interés en el estudio del caos para las funciones en  $\mathbb{R}^n$  ha llevado a los investigadores a estudiar redes neuronales asimétricas, estas reciben su nombre a partir de su matriz de pesos, ya que la matriz de pesos de la red es no simétrica. En algunos casos la matriz puede ser antisimétrica o puede ser formada por dos matrices distintas, dando como resultado una matriz asimétrica. Las redes neuronales asimétricas pueden presentar ciclos límite como atractores, lo que no se puede presentar en caso de redes neuronales con matrices de pesos simétricas, en las cuales se logra un equilibrio como es el caso del modelo de Hopfield, el cual estudiaremos más adelante. A pesar de la presencia de ciclos límite mucha gente considera que algunas redes neuronales pueden ser vistas como sistemas dinámicos caóticos, en especial si los ciclos límite son de período muy grande.

---

\* En realidad en las redes neuronales no existe la temperatura física del sistema. la temperatura es un parámetro usado para determinar el grado de aleatoriedad del sistema.

## 2 Estudio del caos en sistemas dinámicos discretos

El decir que una red neuronal es capaz de presentar una dinámica caótica es confuso ya que el concepto de caos para los sistemas dinámicos discretos\*, como las redes neuronales, no se encuentra totalmente definido. En la actualidad se han llevado a cabo algunos esfuerzos, por ejemplo: Algunos autores consideran que una red neuronal es caótica si los ciclos límite de la red tienen períodos de orden  $e^{\gamma N}$ , con  $N$  igual al número de neuronas en la red y  $\gamma$  una constante. Otros autores dicen que las redes neuronales no son capaces de tener una dinámica caótica, ya que la dinámica de la red es totalmente determinística y por tal motivo la evolución de la red siempre será a un atractor determinado (un punto fijo o a un ciclo límite). Todas estas controversias llevaron a presentar una nueva teoría de lo que se podría llamar un "caos discreto"<sup>[4]</sup>. La teoría del caos discreto está basada en los conceptos de caos para funciones en  $\mathbb{R}^n$ . La idea es la de generalizar los conceptos de caos para funciones de  $\mathbb{R}^n$  a los sistemas discretos representados por el espacio de los estados binarios ( $\Omega$ ).

En esta tesis utilizaremos los conceptos del caos discreto<sup>[4]</sup>, para el estudio de las redes neuronales asimétricas. Enfocaremos el estudio en ciertos ejemplos de redes neuronales con y sin memoria, veremos que las redes neuronales sin memoria no pueden proveer de ejemplos de caos, ya que la métrica del espacio está dada por la distancia de Hamming. Al estudiar redes neuronales con memoria tenemos que extender el espacio a  $\Omega$ , en el cual la métrica es distinta. Lo que nos permite generalizar el concepto de caos para estos sistemas. En particular estudiaremos dos ejemplos de redes neuronales con memoria, que presentan evidencia de caos discreto.

El resto de este capítulo hablará de los inicios de las redes neuronales, como es que las neuronas reales trabajan y de la historia de las redes neuronales. El capítulo dos se enfocará en el estudio del modelo propuesto por Hopfield<sup>[1,2]</sup>, a manera de introducción formal al estudio de las redes neuronales. En el capítulo tres llevaremos a cabo la simulación de una red neuronal asimétrica tipo Hopfield. La simulación nos permitirá estudiar la dinámica de la red neuronal. En el capítulo cuatro estudiaremos una red propuesta por Crisanti, Falcioni y Vulpiani<sup>[5]</sup>, nuevamente realizaremos la simulación de la red neuronal para poder estudiar la dinámica de la red. En el capítulo cinco se estudiarán dos modelos de redes neuronales asimétricas con memoria. El primero de ellos es una red que en principio presenta un caos discreto para la cual ampliaremos los estudios para determinar si cumple con las definiciones de un caos discreto. El segundo modelo es una red neuronal que proponemos y la cual en los análisis que se han realizado es una red que presenta características de una dinámica caótica. En el capítulo seis presentaremos las conclusiones del estudio del caos en las redes neuronales.

---

\* Al hablar de los sistemas dinámicos discretos nos referimos a sistemas cuyo espacio fase es discreto.

## EL CEREBRO Y LA NEUROCIENCIA

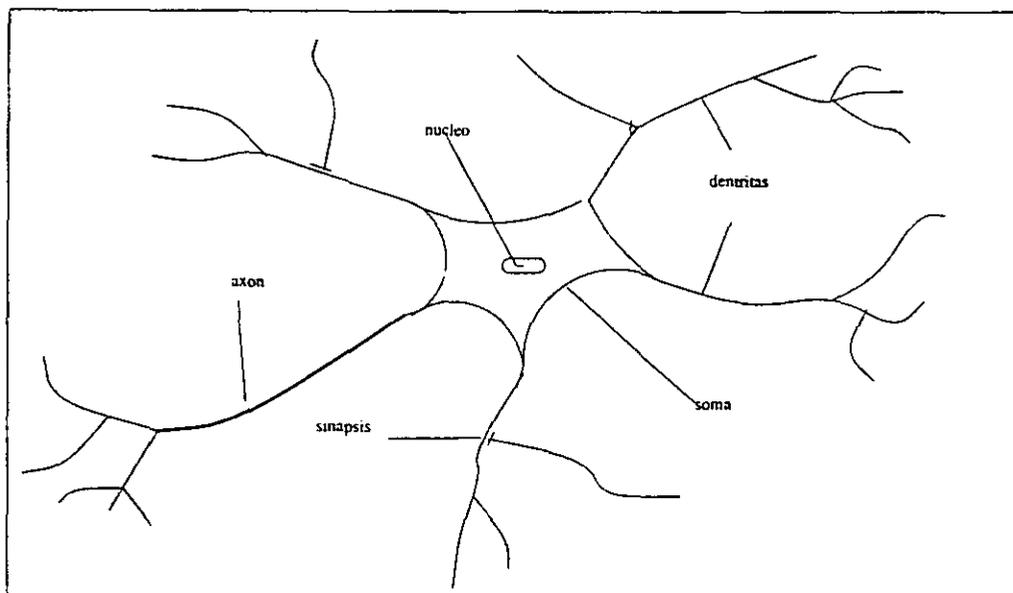
Las redes neuronales están inspiradas en el funcionamiento de nuestro cerebro<sup>[1,2,6]</sup>. El cerebro humano tiene una red neuronal de aproximadamente  $10^{11}$  neuronas y cada una de ellas se encuentra conectada con aproximadamente  $10^4$  neuronas. En la actualidad tenemos una idea general del funcionamiento de las neuronas, que son las unidades del cerebro y de la red neuronal. Básicamente una neurona tiene dos estados, el estado de *excitación* y el estado de *inhibición*. El estado en el que se encuentra la neurona está determinado por los impulsos provenientes de otras neuronas, los cuales son procesados de alguna forma dentro de la neurona y si estos impulsos son lo suficientemente fuertes para sobrepasar un valor determinado (*umbral*) entonces la neurona emite un impulso o se *dispara* (se *excita*). Si los impulsos no son lo suficientemente fuertes entonces la neurona no emitirá impulso alguno y se mantendrá *apagada* (se *inhibe*).

Las neuronas son células nerviosas y como todas las células tienen un núcleo, este se encuentra en el soma o cuerpo de la célula. Del cuerpo de la célula se desprenden unos delgados filamentos llamados *dentritas* estos filamentos son los conductos por los cuales la célula recibe los impulsos provenientes de las otras neuronas. Los impulsos salientes de las neuronas son transmitidos por el *axón*, que es un filamento que sale del cuerpo de la neurona, este filamento es ligeramente más grueso que las *dentritas*. La *sinapsis* es la unión de dos neuronas la cual se lleva a cabo cuando se ponen en contacto las prolongaciones del axón de una neurona con las *dentritas* de otra neurona (*fig1.1*).

En la realidad la transmisión de las señales de una a otra neurona es un proceso químico muy complejo. El cual involucra sustancias químicas transmisoras y también involucra un proceso eléctrico a diversos niveles, ambos procesos producen un incremento o un decremento en el potencial eléctrico de la célula y cuando este potencial eléctrico alcance un valor determinado (*umbral*) la neurona reacciona emitiendo un impulso a lo largo del axón. Esto quiere decir que la neurona ha disparado (se encuentra *excitada*). Una vez que la neurona ha realizado un disparo esta tiene que esperar un período de tiempo para poder disparar de nuevo, a este período de tiempo se le conoce como período de refracción y apenas dura unas millonésimas de segundo.

A pesar de todo lo que sabemos de las neuronas y de la red neuronal que forman. Los modelos matemáticos de las redes neuronales son mucho más sencillos, esto se debe a la complejidad que presentan las redes neuronales reales. En esta tesis trabajaremos con redes neuronales basadas en el modelo que propusieron McCulloch y Pitts<sup>[1,2,6]</sup> en 1943. en el propusieron que las neuronas fueran consideradas como unidades binarias con un cierto umbral. El modelo

#### 4 Estudio del caos en sistemas dinámicos discretos



*fig 1.1, Forma básica de una neurona.*

inicial sólo expresaba los pesos de entrada de las demás neuronas como una combinación lineal y tenía como salida un uno o un cero, este valor dependía de la combinación lineal, si ésta era mayor o menor que el valor del umbral. El valor de la neurona estaba dado por:

$$n_i(t+1) = \Theta \left( \sum_j w_{ij} n_j(t) - \mu_i \right) \quad (1.1)$$

En esta ecuación la  $n_i$  representa a la  $i$ -ésima neurona de la red, las neuronas sólo pueden tomar los valores de cero o uno. En la ecuación anterior  $t$  es el tiempo y es una variable discreta. Los pesos de las entradas están dados por  $w_{ij}$ . Los pesos representan la “fuerza” de la conexión entre la neurona  $i$  y la neurona  $j$ , en este caso los pesos pueden ser positivos o negativos, los pesos toman el valor de cero en el caso de que no exista conexión entre la neurona  $i$  y la neurona  $j$ . El umbral de la  $i$ -ésima neurona está dado por  $\mu_i$ . La función  $\Theta(x)$  se define como:

$$\Theta(x) = \begin{cases} 1, & x \geq 0; \\ 0, & x < 0. \end{cases} \quad (1.2)$$

Desde un punto de vista computacional podemos pensar que las neuronas son como pequeños chips que se encuentran conectados entre sí, por lo cual el cerebro es como una computadora formada por alrededor de  $10^{11}$  chips, cada uno de estos procesadores se encuentra conectado con alrededor de  $10^4$  chips más formando un procesador muy complejo.

En la actualidad los modelos de redes neuronales han intercambiado la combinación lineal de los pesos de las neuronas por una función general  $g(x)$ , la cual puede ser no lineal. Dando lugar a diversos tipos de redes neuronales.

## HISTORIA

A continuación hablaremos un poco de la historia de las redes neuronales. En realidad el interés en el estudio y desarrollo de las redes neuronales tiene dos orígenes, uno es el estudio sobre el funcionamiento del cerebro humano y el otro es el interés de poder desarrollar computadoras que sean capaces de tener un proceso de aprendizaje\*.

El inicio real del estudio de las redes neuronales se dio en el año de 1943 con el modelo propuesto por McCulloch y Pitts<sup>[1,2,6]</sup>, el modelo de McCulloch y Pitts difiere de una computadora esencialmente en que lleva a cabo un procesamiento en paralelo de los datos a diferencia de una computadora que lleva a cabo un procesamiento de forma serial, a pesar de este gran avance realizado por McCulloch y Pitts, todavía faltaba mucho por hacer, ya que se tenía el problema de como seleccionar los pesos apropiados para que la red funcionará correctamente, fue hasta los primeros años de la década de los sesenta que hubo importantes avances en el problema de la asignación correcta de los pesos de las conexiones entre las neuronas.

En 1961 Caianiello<sup>[1,2]</sup> propuso un algoritmo para la correcta asignación de los pesos. este algoritmo se basaba en la regla de aprendizaje propuesta por Hebb<sup>[1,2]</sup> en 1949. muchas de las ideas de este algoritmo fueron tomadas a partir de la teoría de la mecánica estadística. Sin embargo ya desde un año antes [1960] existía un grupo de científicos dedicados entre otras cosas al estudio de las redes

---

\* El proceso de aprendizaje en una computadora es mejor conocido como Inteligencia Artificial.

## 6 Estudio del caos en sistemas dinámicos discretos

neuronales, a la cabeza de este grupo se encontraba Frank Rosenblatt, el grupo de Rosenblatt estaba enfocado a resolver el problema de la asignación correcta de los pesos, para lo cual se concentraron en un sólo tipo de redes neuronales llamadas *perceptrones*<sup>[1,2,6]</sup>. En 1962 Rosenblatt y su grupo proponen un algoritmo de aprendizaje para los perceptrones más simples\* , resolviendo el problema de la asignación de los pesos.

Los perceptrones son redes neuronales formadas por capas, la primera capa es la capa de entrada de los datos, posteriormente los datos pasan a las capas subsecuentes hasta llegar a la capa de salida. la particularidad de los perceptrones es que todas las neuronas de la capa  $i$  están conectadas con todas las neuronas de la capa  $i + 1$ , además las neuronas que se encuentran en una misma capa no están conectadas entre sí.

Con el algoritmo propuesto por Rosenblatt mucha gente creyó que el problema de la correcta asignación de los pesos estaba resuelto, pero para el año de 1969 Marvin Minsky y Seymour Papert<sup>[1,2,6]</sup> demostraron que los perceptrones de una sola capa no eran capaces de resolver problemas tan sencillos como la regla XOR, sin embargo algunos científicos creían que los perceptrones de capas múltiples serían capaces de resolver estos y otros problemas más complejos, pero al no existir un algoritmo para la correcta asignación de los pesos para los perceptrones de capas múltiples estos fueron abandonados durante algún tiempo, dando lugar a que los científicos se enfocarán en otro tipo de redes neuronales. Fue hasta 1974 que William Little<sup>[1,2]</sup> mostró la similitud entre las redes neuronales y los llamados *vidrios de spin*. para 1978 Little y Shaw<sup>[1,2]</sup> amplían la información sobre como estos sistemas almacenan información y para 1982 John Hopfield<sup>[1,2]</sup> introduce en las redes neuronales el concepto de una función de energía y a partir de ese momento el estudio de las redes neuronales se enfoca en el estudio de las propiedades termodinámicas, análogamente a los sistemas desordenados de *spin* mejor conocidos como *vidrios de spin (spin glasses)*.

En los últimos años el estudio y desarrollo de las redes neuronales se ha llevado a cabo a partir de los modelos de Little y de Hopfield, a pesar de que ha habido un resurgimiento de los perceptrones. En general el estudio de las redes neuronales se ha incrementado debido al interés existente en el estudio de los sistemas dinámicos no lineales y su estrecha relación con el *caos*. además de las múltiples aplicaciones que se pueden desarrollar a partir de las redes neuronales.

---

\* Los perceptrones más simples son aquellos que sólo constan de la capa entrada y de la capa de salida.

## MODELO DE HOPFIELD

## MEMORIA ASOCIATIVA.

A continuación estudiaremos el modelo propuesto por Hopfield<sup>[1,2]</sup> en 1982. El cual nos servirá como introducción formal a las redes neuronales. El primer problema será el problema de la memoria asociativa en las redes neuronales, a partir de ese problema estudiaremos la estabilidad de la red y posteriormente estudiaremos la función de energía. Después hablaremos un poco de la relación entre los llamados vidrios de spin y las redes neuronales, lo cual nos abre las puertas a la teoría del campo medio y la mecánica estadística, con estas herramientas podremos estudiar la capacidad de almacenamiento de la red neuronal.

El problema de la memoria asociativa es uno de los muchos problemas en el estudio de las redes neuronales. En el caso de redes neuronales asimétricas el problema se conoce como el reconocimiento secuencial de patrones. Básicamente el problema de la memoria asociativa es el siguiente: Supongamos que tenemos una red neuronal la cual queremos que reconozca las letras del abecedario, el problema radica en que hay varios tipos de letras, supongamos que la red "aprende" a reconocer la letra "A" desde el punto de vista matemático sería el patrón  $\xi$ . Nosotros ingresamos a la red otro tipo de letra, por ejemplo  $A$ , esta letra es un nuevo patrón al cual denominaremos como  $\zeta$ . Deseamos que la red nos indique que al patrón  $\zeta$  se le debe asociar el patrón  $\xi$ , en otras palabras, queremos que la red nos diga que efectivamente la letra  $A$  es una A. a esto se le llama memoria asociativa. Lo que está haciendo la red es asociar el patrón  $\zeta$  con el patrón  $\xi$ , los asocia ya que estos dos patrones se encuentran muy cercanos.

Antes de explicar la noción de cercanos haremos una serie de definiciones que usaremos a lo largo de esta tesis. Supongamos que tenemos una red neuronal con  $i = 1, 2, 3, \dots, N$ , neuronas, y con  $\mu = 1, 2, 3, \dots, P$ , patrones. Definiremos a un estado de la red al tiempo  $t$  como  $S(t)$ , cada posible estado de la red es un vector que consta de  $N$  entradas  $S_i(t)$  cuyo valor es  $\pm 1$  y denotaremos por  $\xi_i^\mu$  como al estado del  $\mu$ -ésimo patrón en la  $i$ -ésima neurona, llamaremos  $\xi^\mu$  al  $\mu$ -ésimo patrón, el cual es un vector con  $N$  entradas binarias, por último denotaremos como  $\zeta$  al patrón de prueba (vector con  $N$  entradas binarias).

## 8 Estudio del caos en sistemas dinámicos discretos

Regresando al concepto de cercanía entre dos estados, este concepto involucra el concepto de distancia entre dos estados binarios de dimensión  $N$ , los cuales son generados con una dinámica discreta. Usaremos la distancia de Hamming definida como:

$$d_H(\xi^\mu, \zeta) = \sum_{i=0}^N [\xi_i^\mu (1 - \zeta_i) + (1 + \xi_i^\mu) \zeta_i] \quad (2.1)$$

La distancia de Hamming entre dos estados binarios es el número de bits distintos entre los dos estados, por lo que la máxima distancia entre  $\xi^\mu$  y  $\zeta$  es  $N$ .

Una red neuronal que tenga una buena memoria asociativa debe de ser capaz de poder recuperar o reconstruir la información a partir de información que presente errores o de información incompleta.

### EL MODELO DE HOPFIELD

Existen dos formas básicas de llevar a cabo la actualización de las neuronas. La actualización serial (asíncrona) en la cual se debe actualizar una neurona por cada unidad de tiempo, la neurona se elige de forma aleatoria. Para el caso del modelo de Hopfield actualizaremos la red de forma serial. Notemos que en el caso de la actualización serial la diferencia entre el estado  $S(t)$  y el  $S(t+1)$  se encuentra en la  $i$ -ésima neurona. La actualización en paralelo (síncrona) es actualizar todas las neuronas de la red por cada unidad de tiempo.

A continuación definiremos el modelo básico que propuso Hopfield [1,2] en 1982, con él comenzaremos formalmente el estudio de las redes neuronales. La regla de evolución de la red está dada por:

$$S_i(t+1) = \text{sign} \left( \sum_{j=0}^N (w_{ij} S_j(t) - \theta_i) \right) \quad (2.2)$$

En la ecuación anterior podemos observar que el último estado de la red esta dado sólo por el valor de la red en la  $i$ -ésima neurona. La función  $\text{sign}(x)$  es la función signo, la cual podemos definirla como:

$$\text{sign}(x) = \begin{cases} 1, & \text{if } x \geq 0; \\ -1, & \text{if } x < 0. \end{cases} \quad (2.3)$$

Los  $w_{ij}$  son los pesos de las conexiones entre la neurona  $i$  y la neurona  $j$ . Los pesos de la red forman una matriz de  $N \times N$ . El valor del umbral de la  $i$ -ésima neurona esta dado por  $\theta_i$ , en este capítulo fijaremos el valor del umbral como cero, por lo que el modelo se reduce a:

$$S_i(t+1) = \text{sign} \left( \sum_{j=0}^N w_{ij} S_j(t) \right) \quad (2.4)$$

Al hacer el valor del umbral igual a cero hemos simplificado el modelo, a partir de esta modificación el estado de la neurona sólo depende de la fuerza de las conexiones con las otras neuronas y si estas están o no excitadas. El tiempo ( $t$ ) es considerado como una variable discreta. Durante este capítulo supondremos que cada uno de los patrones iniciales ( $\xi_i^\mu$ ) está generado de forma aleatoria e independiente.

### CASO 1: UN PATRÓN

Empecemos con el estudio del modelo de Hopfield, primero estudiaremos el caso de un solo patrón, que es el caso más sencillo. Tenemos una red neuronal con  $N$  neuronas y con  $P = 1$  patrones, deseamos que nuestra red neuronal logre "memorizar" el patrón  $\xi$ , el cual es nuestro estado inicial  $S(0)$  al realizar una iteración la red se debe mantener en ese estado inicial, matemáticamente hablando queremos:

$$\text{sign} \left( \sum_{j=0}^N w_{ij} \xi_j \right) = S_i(1) = \xi_i \quad (2.5)$$

En particular si exigimos que se cumpla la igualdad :

$$\sum_{j=0}^N w_{ij} \xi_j = \xi_i \quad (2.6)$$

A partir de esta ecuación se puede observar que los pesos de la red son proporcionales a los patrones de la red, en especial si tomamos:

## 10 Estudio del caos en sistemas dinámicos discretos

$$w_{ij} = 1/N(\xi_i \xi_j) \quad (2.7)$$

En la ecuación anterior se puede ver que el patrón inicial es un estado estable, con esto hemos logrado encontrar un valor para los  $w_{ij}$  en función de los patrones de la red, pero ¿Cuántos bits distintos puede tener el estado inicial  $\xi$  con respecto al estado o patrón a memorizar  $\zeta$ ? Para que el patrón inicial sea un estado estable. Veamos esto con mayor cuidado. Supongamos que  $\xi$  tiene  $k$  bits distintos de  $\zeta$ . Entonces la suma de la ecuación (2.4) la dividiremos en dos sumas, la primera suma tendrá los  $N - k$  bits iguales entre  $\xi$  y  $\zeta$ , la segunda suma tendrá los restantes  $k$  bits (distintos).

$$\text{sign} \left( \sum_{\xi=\zeta} w_{ij} \xi_j + \sum_{\xi \neq \zeta} w_{ij} \xi_j \right) \quad (2.8)$$

Como  $\xi_i \neq \zeta_i$  quiere decir que  $\xi_i = -\zeta_i$ , los pesos tienen el valor dado por la ecuación (2.7) podemos deducir :

$$\begin{aligned} (N - k/N) &> (k/N) \\ (N/2) &> k \end{aligned} \quad (2.9)$$

La condición esencial para la estabilidad inicial de la red es, que el máximo número de bits distintos entre  $\xi$  y  $\zeta$  no sea mayor que la mitad del número total de neuronas en la red. Si  $N/2 > k$  quiere decir que la red es capaz de corregir errores, esto implica que la red tiene una buena memoria asociativa.

### CASO 2: P PATRONES

Este caso es más complicado debido a que ya no vamos a tener un solo patrón a memorizar, vamos a tener  $P$  distintos patrones a memorizar, antes de empezar definiremos  $h_i^\nu$  como:

$$h_i^\nu = \sum_j w_{ij} \xi_j^\nu \quad (2.10)$$

El modo más sencillo de estudiar la estabilidad de la red para este caso es analizar la estabilidad de un patrón, como en el caso anterior las condiciones de estabilidad que requerimos son:

$$\text{sign}(h_i^\nu) = \xi_i^\nu \quad (2.11)$$

Al igual que en el caso anterior, podemos proponer un valor para  $w_{ij}$  en función de los patrones de la red.

$$w_{ij} = 1/N \left( \sum_{\mu=1}^P \xi_i^\mu \xi_j^\mu \right) \quad (2.12)$$

Esta forma de definir la matriz de pesos es una forma muy natural, lo único que estamos haciendo es generalizar la ecuación (2.7), claro que en este caso debemos tomar en cuenta los  $P$  patrones y no sólo un patrón. Veamos que efectivamente con la ecuación (2.11) los patrones iniciales son estables. Para que cada patrón inicial sea estable queremos que la configuración inicial de la red corresponda a algún patrón inicial ( $\xi^\mu$ ) y la red se mantenga en ese patrón inicial.

Usando la ecuación (2.12) obtenemos.

$$h_i^\nu = \sum_j w_{ij} \xi_j^\nu = 1/N \left( \sum_j \sum_\mu \xi_i^\mu \xi_j^\mu \xi_j^\nu \right) \quad (2.13)$$

En donde separando la suma en dos partes, la primera con  $\mu = \nu$  y la segunda con todos los términos restantes.

$$h_i^\nu = \xi_i^\nu + 1/N \left( \sum_j \sum_{\mu \neq \nu} \xi_i^\mu \xi_j^\mu \xi_j^\nu \right) \quad (2.14)$$

Podemos garantizar la estabilidad inicial de la red si el segundo término (término cruzado) es cero o es lo suficientemente chico para no alterar el signo de  $h_i^\nu$ . Si la ecuación (2.11) se cumple para todo  $\nu = 1, 2, 3, \dots, P$ , entonces la estabilidad de la red será para todos los patrones iniciales de la red, coincidiendo con el caso uno. Aquí surge una nueva pregunta, ¿Cuándo podemos garantizar que el término cruzado sea lo suficientemente chico para que no altere el signo de  $h_i^\nu$ ? En la siguiente sección discutiremos la capacidad de almacenamiento de la red, lo cual dará una respuesta a la pregunta anterior.

## CAPACIDAD DE ALMACENAMIENTO EN LA RED.

A continuación analizaremos el caso más sencillo. fijaremos el valor de  $P$  como un valor fijo y finito, el término cruzado es una suma de  $N(P-1)$  términos independientes con media cero y varianza uno, por lo que la varianza del término cruzado es muy parecida a  $P/N$  de tal manera que el término cruzado es del orden de  $\sqrt{P/N}$ \*. Como  $P$  es finito el orden del término cruzado es aproximadamente  $\sqrt{1/N}$ , si hacemos crecer a  $N$  entonces el término cruzado tiende a cero y como consecuencia no influye en el signo de  $h_i^\nu$ , este análisis solo es válido para  $P$  finito y  $N$  muy grande.

Ahora analizaremos el caso en que  $P$  no está fijo propiciando que  $\sqrt{P/N}$  no tienda a cero. Partiremos de la idea de calcular la probabilidad que tienen cada una de las entradas  $\xi_i^\mu$  de tomar los valores  $\pm 1$  y a partir de esos cálculos extendernos a la teoría de probabilidad. Primeramente definiremos  $C_i^\nu$  como  $-\xi_i^\nu$  veces el término cruzado.

$$C_i^\nu = -\xi_i^\nu / N \left( \sum_j \sum_{\mu \neq \nu} \xi_i^\mu \xi_j^\mu \xi_j^\nu \right) \quad (2.15)$$

Si  $C_i^\nu$  es negativo entonces el término cruzado es del mismo signo que  $\xi_i^\nu$ , por tal motivo no afecta en el valor final de  $h_i^\nu$ . Si  $C_i^\nu$  es positivo, entonces el término cruzado tiene el signo contrario a  $\xi_i^\nu$ , en ese caso puede cambiar el signo de  $h_i^\nu$ , sólo si  $C_i^\nu$  es mayor que uno. Por lo que nosotros queremos calcular la probabilidad de que  $C_i^\nu$  sea mayor que uno, si logramos hacer que la probabilidad de  $C_i^\nu > 1$  sea muy pequeña garantizaremos que la red sea estable.

$$P_{error} = P(C_i^\nu > 1) \quad (2.16)$$

Como los  $\xi_i^\mu$  están generados de forma aleatoria e independiente, cada uno de los  $\xi_i^\mu$  tiene la misma probabilidad de tomar los valores  $\xi_i^\mu = +1$  y  $\xi_i^\mu = -1$ . Como estamos considerando el caso en el que tanto  $P$  como  $N$  son mucho mayores que uno† podemos aproximar  $P_{error}$  por medio de la distribución Normal<sup>[7,8]</sup>. La función normal tendrá media cero y varianza  $\sigma^2 = P/N$ .

\* Por el teorema del límite central se puede ver que el orden es  $\sqrt{p/n}$ .

† Estamos considerando que  $C_i^\nu$  es  $1/N$  veces la suma de  $N(P-1)$  números aleatorios ( $\pm 1$ ), independientes entonces la  $\sigma^2 = NP$ , por lo que  $C_i^\nu$  tiene una distribución binomial con media cero y varianza  $P/N$ .

$$C_i^y \sim N(0, P/N) \quad (2.17)$$

La probabilidad que queremos calcular es la cola derecha de una distribución Normal (valores mayores a uno), por tal motivo el valor de  $P_{error}$  se incrementa conforme se incrementa el número de patrones y se mantiene fijo el número de neuronas.

$$P_{error} = \frac{1}{\sqrt{2\pi}\sigma} \int_1^{\infty} e^{-x^2/2\sigma^2} dx \quad (2.18)$$

Para poder simplificar un poco los cálculos a continuación definiremos la función error.

$$erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x exp(-u^2) du \quad (2.19)$$

Si tomamos el límite cuando  $x \rightarrow 0$  de la función error, este vale cero. Propondremos un cambio de variable para poner la  $P_{error}$  en función de  $erf(x)$ . Si hacemos  $u = t/\sqrt{2\sigma^2}$ .

$$P_{error} = 1/2[1 - erf(1/\sqrt{2\sigma^2})] \quad (2.20)$$

Claramente el valor de  $P_{error}$  es una función que depende directamente de  $P$  y de  $N$ . Por lo que al sustituir el valor de  $\sigma^2$  obtenemos.

$$P_{error} = 1/2[1 - erf(\sqrt{N/2P})] \quad (2.21)$$

Para poder calcular la probabilidad del error (con una distribución normal o con la función error) tenemos que buscar en tablas<sup>[9]</sup> el valor de  $P_{error}$  para una probabilidad dada\* y a partir de esos valores podemos conocer los valores máximos de  $P$  con  $N$  fija.

Existe una forma alternativa para poder calcular la capacidad de la red<sup>1</sup>, usaremos la aproximación asintótica<sup>[10]</sup> a la función error siempre y cuando  $N \rightarrow \infty$ . Para este caso sólo usaremos el primer término de la aproximación asintótica†.

$$1 - erf(x) \rightarrow e^{-x^2} / \sqrt{\pi}x, \quad x \rightarrow \infty \quad (2.22)$$

\* En general se quiere que  $P_{error} < 0.001$ .

† Utilizaremos el primer término de la aproximación asintótica de  $1 - erf(x)$ .

## 14 Estudio del caos en sistemas dinámicos discretos

Sabemos que la función error tiende a uno en el límite cuando  $x$  tiende a infinito, la aproximación asintótica tiende a cero conforme  $x$  tiende a infinito.

$$e^{-x^2}/\sqrt{\pi x} \rightarrow 0, x \rightarrow \infty \quad (2.23)$$

Usando la ecuación (2.21) obtenemos.

$$P_{error} \approx \frac{1}{2} e^{-x^2}/\sqrt{\pi x} \quad (2.24)$$

A partir de la ecuación (2.24) podemos encontrar nuevamente los valores máximos para  $P$  y  $N$ , estos valores se encuentran a partir de sustituir el valor de  $x = \sqrt{N/2P}$  y con la condición de  $P_{error} < k/N$ . Donde  $k$  es la probabilidad de que un bit tenga el signo contrario\* (i.e. esté erróneo). Los valores que dan para este caso de  $N$  y  $P$  son:

$$P_{max} = \frac{N}{4 \ln N} \quad (2.25)$$

En esta última ecuación podemos observar claramente que el valor máximo de  $P$  depende del valor de  $N$ . Como esta ecuación es resultado de una aproximación asintótica de la función error,  $N$  tiene que tomar valores suficientemente grandes.

Hasta este momento sólo hemos calculado la probabilidad del error en el término  $C_i^y$ , si este es mayor o menor que +1 produciendo un cambio en el signo de  $h_i^y$ , esto nos garantiza que a lo más un pequeño porcentaje de los bits de cada patrón tengan el signo contrario produciendo que los patrones sean estables sólo inicialmente. En el caso que el patrón inicial de nuestra red neuronal sea en un patrón distinto de los patrones iniciales no podemos garantizar que la red evolucione a un patrón estable. Más adelante hablaremos de la capacidad máxima de la red considerando la estabilidad total de la red y no sólo la estabilidad inicial de los patrones como en este caso.

---

\* Si queremos un 99% de bits correctos en cada uno de los patrones el valor de  $k$  es  $k = 1 - 0.99$ , por lo que  $P_{error} < 0.01/N$ .

## LA FUNCIÓN DE ENERGÍA

Dentro del modelo que propuso Hopfield<sup>[1,2]</sup> en 1982 se presentó por vez primera una función de energía. En el caso de las redes neuronales la función de energía presenta características de mucha ayuda para su estudio. Una de las propiedades fundamentales de la función de energía radica en que es una función decreciente durante toda la evolución del sistema, esto quiere decir que la función jamás aumentará su valor, en el peor de los casos la función podría mantenerse constante.

El espacio fase de la función de energía es sobre todas las posibles configuraciones del sistema  $\{S_i\}$  cruz los valores de la función de energía, recordemos que el sistema tiene  $2^N$  posibles estados. Dentro del espacio fase se forman pequeños valles de atracción y como la función de energía es decreciente la red debe evolucionar hacia alguno de estos valles. cuyo punto mínimo es un mínimo local. en general queremos que los mínimos locales sean cada uno de los  $\xi^\mu$ .

Definición de la Función de Energía.

$$H = -1/2 \sum_{ij} w_{ij} S_i S_j \quad (2.26)$$

La suma es una doble suma sobre todas las  $i$  y todas las  $j$ . En este caso deseamos encontrar los puntos donde la función de energía alcance sus valores mínimos locales. Cabe mencionar que la función de energía generalmente existe en redes neuronales con pesos simétricos ( $w_{ij} = w_{ji}$ ), en el caso de las redes neuronales si la matriz de pesos de la red no es simétrica la función de energía no es decreciente esto pasa ya que la función de energía se encuentra estrechamente relacionada con la dinámica de la red. Si la red tiene una matriz de pesos asimétrica la red es capaz de exhibir un comportamiento periódico, esto provoca que la "función de energía" no sea decreciente. Este comportamiento periódico no se puede dar en el caso de una matriz de pesos simétrica ya que el sistema logra un equilibrio.

A continuación demostraremos que la función de energía para el modelo de Hopfield es una función decreciente. Sea  $S_k$  un estado de la red y sea  $S'_k$  el siguiente estado en la evolución de la red. Recordemos que para este modelo estamos actualizando la red de forma serial (asíncrona), por lo que existen dos casos. El caso en el que  $S_k = S'_k$  en donde la energía se mantiene constante y el caso contrario donde  $S'_k = -S_k$ , en donde la energía presenta una variación.

$$S'_k = \text{sign} \left( \sum_j w_{kj} S_j \right) \quad (2.27)$$

En este caso las funciones de energía tienen valores distintos, si logramos demostrar que  $H' - H < 0$ , la función de energía siempre será decreciente.

$$\begin{aligned} H' - H &\leq 0 \\ H' &\leq H \end{aligned} \quad (2.28)$$

Simplificaremos un poco la función de energía.

$$\begin{aligned} H &= -\frac{1}{2} \sum_{i < j} w_{ij} S_i S_j - \frac{1}{2} \sum_{i > j} w_{ji} S_j S_i \\ &= -\frac{1}{2} \sum_{i < j} w_{ij} S_i S_j - \frac{1}{2} \sum_{i < j} w_{ij} S_i S_j \\ &= -\sum_{i < j} w_{ij} S_i S_j \end{aligned} \quad (2.29)$$

Escribamos las diferencias de las funciones de energía:

$$\begin{aligned} H' - H &= -\sum_{k < j} w_{kj} S'_k S'_j + \sum_{k < j} w_{kj} S_k S_j \\ &= -\sum_{k < j} w_{kj} S'_k S_j + \sum_{k < j} w_{kj} S_k S_j \\ &= -\sum_j w_{kj} S_j (-S'_k + S_k) \\ &= 2S_k \sum_{k < j} w_{kj} S_j \leq 0 \end{aligned} \quad (2.30)$$

A partir de aquí se puede ver claramente que  $H' - H < 0$ , en la última ecuación (2.30) tenemos  $2S_k$  por la suma sobre  $j$  que tiene el mismo signo que  $S'_k$ , por lo que la función de energía siempre decrece\*. En particular si la red llega a

---

\* Notemos que estamos haciendo uso de la forma de actualización de la red en el caso del modelo de Hopfield la red es actualizada en forma asincrónica  $S_k = -S'_k$  y  $S_i = S'_i \forall i \neq k$ .

un punto como algún  $\xi_i^\mu$ , que son puntos estables dentro del espacio fase, entonces la función de energía se volverá constante, provocando que el  $\xi_i^\mu$  sea un mínimo local. Lo anterior se puede garantizar sólo para el caso de  $P$  suficientemente pequeño y  $N$  grande, que es el caso cuando  $\xi_i^\mu$  es un estado estable.

Acabamos de demostrar que cada uno de los  $\xi_i^\mu$  son mínimos locales, pero no son los únicos mínimos. Los puntos  $-\xi_i^\mu$  también son mínimos locales y además tienen la misma energía que el correspondiente  $\xi_i^\mu$ , esto se debe a la simetría de la red. Pero esos no son los únicos mínimos, existen estados de mezcla<sup>[1,2]</sup> que no son iguales a algún patrón, pero están formados por una combinación lineal de un número impar de patrones, estos puntos son mínimos locales.

$$\xi_i^{mixto} = \text{sgn}(\pm \xi_i^{\mu_1} \pm \xi_i^{\mu_2} \pm \xi_i^{\mu_3}) \quad (2.31)$$

En este caso existen ocho posibles combinaciones para el estado mixto, estos estados mixtos deben de cumplir con las condiciones de estabilidad inicial (ecuación (2.11)) para que la función de energía alcance un mínimo local en el estado mixto (ecuación (2.28)). El caso de una combinación lineal de un número par de patrones en algunas ocasiones el  $\xi_i^{mixto}$  toma el valor de cero, por lo que no se encuentra en el espacio fase del sistema.

Por último para el caso en que  $P$  sea muy grande aparecen mínimos locales que no tienen correlación alguna con un número finito de patrones  $\xi_i^\mu$ . Estos estados y los estados de mezcla forman los llamados estados espúreos.

Las redes neuronales tipo Hopfield que sean actualizadas de una forma asíncrona siempre deben de alcanzar un estado estable que es un mínimo local, siempre y cuando el estado inicial de la red sea algún estado dentro del espacio de las posibles configuraciones del sistema.

## MECÁNICA ESTADÍSTICA DE LOS SISTEMAS MAGNÉTICOS

El modelo de Hopfield es uno de los modelos de redes neuronales más estudiados debido a la gran similitud existente con los modelos de los materiales magnéticos<sup>[1,2]</sup>. La analogía entre estos sistemas permite que posteriormente se ingrese el concepto de temperatura en la red cuando usamos unidades estocásticas (neuronas que tienen una cierta probabilidad de tomar los valores  $\pm 1$ ).

Los materiales magnéticos consisten en pequeñas partículas que representan la estructura de los cristales. estas pequeñas partículas se les da el nombre de *spin*. En general los *spin* pueden tomar varias direcciones las cuales dependen

## 18 Estudio del caos en sistemas dinámicos discretos

del material al que pertenecen, a continuación hablaremos del caso más sencillo en el cual cada *spin* sólo tiene dos direcciones posibles. En el caso de los *spins* del modelo de Ising los ferromagnetos pueden ser modelados como variables  $S_i$  que toman sólo dos valores posibles  $\pm 1$  para cada  $i$ . Si el *spin* está orientado para arriba  $S_i = 1$  y si el *spin* está orientado para abajo  $S_i = -1$ . por lo que la analogía con las redes neuronales tipo Hopfield es obvia, ya que en la red neuronal si una neurona está excitada el valor de  $S_i = 1$  y si está inhibida el valor de  $S_i = -1$ .

En el caso de los *spin* tenemos que cada uno de ellos interactua con *spins* cercanos. Esto provoca que algunos átomos cambien su *spin* hasta que todo el sistema alcance un equilibrio. Cada uno de los *spin* está influenciado por un campo magnético  $h$ , este campo magnético consta de un campo magnético externo  $h^{ext}$  y un campo magnético interno. El campo magnético interno es proporcional a cada uno de los *spin*. El campo magnético externo es inducido, por lo que el campo magnético total para cada *spin* está dado por:

$$h_i = \sum_{j=1}^n w_{ij} S_j + h^{ext} \quad (2.32)$$

Los  $w_{ij}$  representan la interacción entre el *spin*  $S_i$  y el *spin*  $S_j$ , para los sistemas magnéticos las interacciones son simétricas  $w_{ij} = w_{ji}$ , estas interacciones cambian dependiendo de la distancia entre los átomos. El campo magnético  $h_i$  es el que controla la evolución del sistema. Cuando el sistema se encuentra a baja temperatura los *spin* tienden a alinearse paralelamente al campo magnético  $S_i = \text{sign}(h_i)$ . Por último la función de energía correspondiente a este modelo es:

$$H = -1/2 \sum_{ij} w_{ij} S_i S_j - h^{ext} \sum_i S_i \quad (2.33)$$

En el caso en donde fijamos  $h^{ext} = 0$  el modelo es completamente idéntico al modelo de Hopfield. La ecuación (2.33) es idéntica a la función de energía del modelo de Hopfield (ecuación (2.26)), podemos decir que ambos sistemas son equivalentes, en el caso de que la red de *spins* sea unidimensional.

## DINÁMICA A BAJA TEMPERATURA.

Los sistemas magnéticos están expuestos a los efectos de la temperatura, conforme se eleva la temperatura del sistema surgen las llamadas fluctuaciones térmicas que tienden a desalinearse los *spin*. Provocando que los *spin* se encuentren influenciados por el campo magnético y las fluctuaciones térmicas. Las fluctuaciones térmicas siempre están presentes mientras que la temperatura no alcance el cero absoluto\*. La fuerza de las fluctuaciones varía según la temperatura a mayor temperatura las fluctuaciones térmicas tienen mayor fuerza. si la temperatura es muy alta entonces las fluctuaciones térmicas dominan y los *spin* tienden a alinearse en sentido contrario a su campo magnético.

Debido a estas fluctuaciones térmicas la dinámica original se debe modificar por una dinámica estocástica, propiciando que cada actualización de los *spin* se haga de forma aleatoria. La nueva regla de evolución es.

$$S_i := \begin{cases} +1, & \text{con probabilidad } g(h_i); \\ -1, & \text{con probabilidad } 1 - g(h_i). \end{cases} \quad (2.34)$$

En donde la función  $g(h)$  es una función que depende de la temperatura y la  $i$  varía sobre cada *spin*<sup>[1,2]</sup>.

Para este caso utilizaremos la forma:

$$g(h) \equiv \frac{1}{1 + \exp(-2\beta h)} \quad (2.35)$$

Donde  $\beta$  es inversamente proporcional a la temperatura†.

$$\beta = \frac{1}{K_B T} \quad (2.36)$$

$K_B$  es la constante de Boltzmann cuyo valor es de  $1.38 \times 10^{-16} \text{erg/K}$  y  $T$  es la temperatura, además la función  $g(-h) = 1 - g(h)$ , permitiendonos escribir la regla de evolución en forma simétrica.

$$\text{prob}(S_i = \pm 1) = g(\pm h_i) \equiv \frac{1}{1 + \exp(\mp 2\beta h_i)} \quad (2.37)$$

---

\* El cero absoluto corresponde a  $-273$  grados centígrados.

† La temperatura se mide en grados Kelvin.

## 20 Estudio del caos en sistemas dinámicos discretos

La probabilidad de  $S_i$  es una función que depende del valor de  $h_i$  que en este caso es el valor del campo magnético total. La función  $g(h)$  es una función de probabilidad continua y por tal motivo es una función creciente en el intervalo  $(0, 1)$  y en el límite cuando  $h \rightarrow -\infty$ ,  $g(h) \rightarrow 0$ , para el caso de  $h \rightarrow \infty$ .  $g(h) \rightarrow 1$ .

La temperatura controla el salto\* de la función  $g(h)$  en los valores cercanos a cero de  $h$ . En el límite para  $\beta \rightarrow \infty$  (i.e.  $T \rightarrow 0$ ) tenemos que la función  $g(h)$  se convierte en la función signo, en ese caso recuperamos el modelo original.

### TEORÍA DEL CAMPO MEDIO.

Primero analizaremos el caso para el cual el sistema tiene un solo *spin* en equilibrio<sup>[1]</sup>. El análisis lo realizaremos por medio de la mecánica estadística. En la sección anterior  $S$  es una variable aleatoria por tal motivo podemos calcular su promedio<sup>†</sup> que denotaremos como  $\langle S \rangle$ .

$$\begin{aligned}\langle S \rangle &= \text{prob}(+1) * (+1) + \text{prob}(-1) * (-1) \\ &= \frac{1}{1 + e^{-2\beta h}} - \frac{1}{1 + e^{2\beta h}} \\ &= \frac{e^{\beta h}}{e^{\beta h} + e^{-\beta h}} - \frac{e^{-\beta h}}{e^{-\beta h} + e^{\beta h}} \\ &= \tanh(\beta h)\end{aligned}\tag{2.38}$$

Como podemos observar el promedio de  $S$  está dado por la función tangente hiperbólica que es una función continua muy similar a la función  $g(h)$ <sup>‡</sup>. En promedio  $S$  varía de  $-1$  a  $1$ , pero  $S$  toma los valores de  $-1$  ó  $+1$  de acuerdo con  $g(h)$ . Cuando  $\beta \rightarrow \infty$  la función tangente hiperbólica se aproxima a la función signo, en ese caso recuperamos el modelo determinístico.

Ahora revisaremos el caso en que tenemos varios *spin* interactuando entre sí, este caso es mucho más complejo. El promedio de un *spin* está determinado por el valor de  $h_i$ , el cual depende de los valores que tomen cada uno de los  $S_j$

---

\* La palabra salto se refiere a que tan rápido varía la función  $g(h)$  de valores cercanos a 0 a valores cercanos a 1, para valores alrededor de  $h = 0$ .

† En realidad estamos calculando el promedio térmico.

‡ Podemos escribir la función tangente en términos de la función  $g(h)$  como:  $\tanh(x) = 2g(x) - 1$ .

y no depende de sus valores promedio. Además la función  $g(h)$  es una función no lineal. esto provoca que no podamos tomar el promedio en el argumento de la función  $g(h)$  de manera directa. En general no hay una forma directa de resolver este problema, pero existe una aproximación por medio de la teoría del campo medio. Básicamente la aproximación del campo medio es substituir  $h_i$  por su promedio  $\langle h_i \rangle$ .

$$\langle g(h_i) \rangle \rightarrow g(\langle h_i \rangle) = f \left( \sum_j w_{ij} \langle S_j \rangle \right) \quad (2.39)$$

La ecuación (2.39) nos permite calcular el promedio de cada  $S_i$  como:

$$\begin{aligned} \langle S_j \rangle &= \langle g(h_i) \rangle - \langle g(-h_i) \rangle \\ &\rightarrow g(\langle h_i \rangle) - g(\langle -h_i \rangle) \\ &= \tanh(\beta \langle h_i \rangle) \end{aligned} \quad (2.40)$$

La aproximación del campo medio queda:

$$\langle S_i \rangle = \tanh(\beta \langle h_i \rangle) = \tanh \left( \beta \sum_j w_{ij} \langle S_j \rangle + \beta h_i^{ext} \right) \quad (2.41)$$

Este es un sistema de  $N$  ecuaciones no lineales con  $N$  incógnitas, pero no contiene ningún tipo de variable aleatoria. En realidad lo que estamos haciendo es tomar el promedio para cada *spin*. Recordemos que esta aproximación es válida sólo para muestras grandes, es decir  $N \rightarrow \infty$ .

La aproximación del campo medio es una aproximación que se vuelve exacta para el límite  $N \rightarrow \infty$ , para ver que tan buena es la aproximación estudiaremos la diferencia entre  $h_i$  y  $\langle h_i \rangle$ .

$$\begin{aligned} \sigma^2 &= \langle (h_i - \langle h_i \rangle)^2 \rangle \\ &= \langle h_i^2 \rangle - \langle h_i \rangle^2 \end{aligned} \quad (2.42)$$

En realidad lo que estamos calculando es la varianza de  $h_i$ , la cual para el caso de  $N \rightarrow \infty$  es aproximadamente  $\sqrt{1/N}$ . Como resultado para el caso  $N \rightarrow \infty$  obtenemos:

$$\begin{aligned} \langle (h_i - \langle h_i \rangle)^2 \rangle &\rightarrow 0 \\ h_i &\rightarrow \langle h_i \rangle \end{aligned} \quad (2.43)$$

### REDES NEURONALES CON DINÁMICAS ESTOCÁSTICAS.

Ahora aplicaremos al modelo de Hopfield lo que acabamos de revisar en las secciones anteriores, aquí introduciremos la noción de temperatura para las redes neuronales<sup>[1,2]</sup>. La introducción del concepto de temperatura nos permitirá hacer uso de la teoría del campo medio, por último podremos calcular la capacidad de la red, recordando, siempre que tomemos la temperatura igual a cero regresaremos al modelo determinístico.

Para empezar haremos que las neuronas se comporten de forma aleatoria, esto se logra de la misma forma que en el modelo de Ising.

$$prob(S_i = \pm 1) = g(\pm h_i) = \frac{1}{1 + \exp(\mp 2h_i \beta)} \quad (2.44)$$

Recordemos que la actualización de las neuronas es llevada a cabo de forma serial. En las redes neuronales reales cada vez que las neuronas disparan los impulsos salientes son de distinta magnitud, además el tiempo entre cada disparo es variable, estas y otras situaciones parecidas producen lo que comunmente se conoce como ruido. Al hacer que las neuronas se comporten de forma aleatoria estamos reproduciendo parte de ese ruido. En este modelo el parámetro  $\beta$  no lo podemos relacionar con la temperatura física del sistema\*, podemos pensar que  $\beta$  es un parámetro que controla el nivel de ruido en el sistema o como un parámetro de que tanto se viola la regla de evolución determinística. Por último definiremos a la temperatura como la inversa del parámetro  $\beta$ .

$$\beta \equiv \frac{1}{T} \quad (2.45)$$

Notemos que la temperatura determina la suavidad de la función  $g(h)$  para valores de  $h$  cercanos a cero, y en el límite cuando la temperatura tiende a cero

---

\* Al no poder relacionar  $\beta$  con la temperatura física podemos excluir la constante de Boltzmann de la ecuación (2.36).

recuperamos el modelo determinístico de Hopfield. La utilización de unidades estocásticas nos permite que durante la evolución de la red el sistema no quede atrapado en estados mínimos locales como pueden ser los estados espúreos\* que tienen un mayor grado de energía que los patrones a memorizar. otra ventaja de manejar unidades estocásticas es que cada vez que pongamos a trabajar a la red esta siempre evolucionará de manera distinta por lo que a partir de ahora nos fijaremos en el comportamiento global de la red, al hablar de un equilibrio en la red debemos pensar que siempre que esta esté evolucionando debe alcanzar un punto estable y ese punto de equilibrio no necesariamente debe ser el mismo para cada evolución de la red.

### TEORÍA DEL CAMPO MEDIO.

Ahora aplicaremos la teoría del campo medio para el cálculo de la capacidad de almacenamiento<sup>[1]</sup>. En este caso  $P \ll N$ , además  $P$  es finito. el caso cuando  $P$  es mayor será revisado después. Usando la ecuación (2.41). las ecuaciones del campo medio en este caso quedan:

$$\begin{aligned} \langle S_i \rangle &= \tanh(\beta \langle h_j \rangle) \\ &= \tanh \left( \frac{\beta}{N} \sum_{j,\mu} \xi_i^\mu \xi_j^\mu \langle S_j \rangle \right) \end{aligned} \quad (2.46)$$

El valor de  $h_i$  es el valor dado en la ecuación (2.10). este sistema de ecuaciones consta de  $N$  ecuaciones no lineales. Para poder resolver el sistema realizaremos la siguiente hipótesis, supondremos que  $\langle S_i \rangle$  es directamente proporcional a alguno de los patrones a memorizar.

$$\langle S_i \rangle = m \xi_i^\nu \quad (2.47)$$

Recordemos que en el caso del modelo determinístico patrones como este son puntos estables, lo que equivale en este sistema es que en el límite de  $T = 0$  estos puntos son estables. pero obviamente lo que queremos es que sean estables también para  $T \neq 0$ .

---

\* Los estados espúreos tienen mayor energía que los patrones a memorizar, ver función de energía.

## 24 Estudio del caos en sistemas dinámicos discretos

Procediendo como en el modelo determinístico la estabilidad del patrón inicial se da si se cumple:

$$m\xi_i^\nu = \tanh \left( \frac{\beta}{N} \sum_{j\mu} \xi_i^\mu \xi_j^\mu m\xi_j^\nu \right) \quad (2.48)$$

En donde nuevamente separaremos la suma en dos partes. la primera con  $\nu = \mu$  y la segunda con todos los términos restantes.

$$m\xi_i^\nu = \tanh \left( \beta m\xi_i^\nu + m\beta \frac{1}{N} \sum_j \sum_{\mu \neq \nu} \xi_i^\mu \xi_j^\mu \xi_j^\nu \right) \quad (2.49)$$

En donde el segundo término es  $m\beta$  veces el término cruzado (ecuación 2.14) y como se observó en la sección de capacidad de almacenamiento de la red conforme  $N$  crece y  $P$  se mantiene fijo el término cruzado cada vez influye menos en el argumento de la  $\tanh(x)$ . Este análisis es válido para valores de  $T$  pequeños por lo que la ecuación (2.49) la podemos simplificar para valores grandes de  $N$ . En realidad como el término cruzado no influye en el valor final lo podemos quitar del argumento de la tangente hiperbólica. La función  $\tanh(x)$  es una función impar y los  $\xi_i^\nu$  toman valores de  $\pm 1$  podemos sacar del argumento de la tangente hiperbólica el  $\xi_i^\nu$ . como resultado obtenemos.

$$m = \tanh(\beta m) \quad (2.50)$$

Desde un punto de vista geométrico esta ecuación tiene tres posibles soluciones para  $\beta > 1$  de la forma  $m = \pm x, 0$  con  $x \in [-1, 1]$ . En el límite de  $T \rightarrow 0$  las soluciones de la ecuación son de la forma  $m = \pm 1.0$ . Los estados serán estables para temperaturas menores a 1. en el caso de temperaturas superiores a uno la única solución que existe es  $m = 0$ , que es la solución trivial. Lo que quiere decir que la temperatura crítica es  $T_c = 1$ .

Usando la ecuación (2.47) obtenemos:

$$\begin{aligned}
 m &= \langle S_i \rangle / \xi_i^\nu \\
 &= \text{Prob}(\text{bit}_i(\text{correcto})) - \text{Prob}(\text{bit}_i(\text{incorrecto})) \\
 \text{sea } P_i &= \text{Prob}(\text{bit}_i(\text{correcto})) \\
 m &= P_i - (1 - P_i) \\
 m &= 2P_i - 1 \\
 \frac{1}{2}(m + 1) &= P_i \\
 \sum_{i=1}^N \frac{1}{2}(m + 1) &= \sum_{i=1}^N P_i \\
 \frac{N}{2}(m + 1) &= \langle N_{\text{bits correctos}} \rangle
 \end{aligned} \tag{2.51}$$

En consecuencia el promedio de bits correctos que permite a los patrones iniciales ser puntos estables es:

$$\langle N_{\text{bits correctos}} \rangle = \frac{1}{2}N(m + 1) \tag{2.52}$$

Cuando la temperatura se incrementa\* a partir de la temperatura crítica ( $T_c$ ),  $m$  tiende a cero por lo que el promedio de bits correctos tiende a ser  $N/2$ , cuando la temperatura se aproxima a cero el valor de  $m$  se aproxima a 1 por lo que el promedio de bits correctos tiende a  $N$ . La variación en el promedio de bits correctos es indicativo de un cambio de fase conforme varía la temperatura. Este cambio está relacionado con el número de estados estables. Al igual que en el caso determinístico hasta ahora hemos demostrado que los patrones a memorizar son puntos estables, sin embargo no son los únicos estados estables. Aquí también existen los llamados estados espúreos, y los estados inversos a los patrones. todos estos también son puntos estables dentro del espacio fase.

## CAPACIDAD DE LA RED NEURONAL.

Por último revisaremos el cálculo de la capacidad de almacenamiento de la red<sup>[1]</sup>, como capacidad de almacenamiento de la red entendemos el número

---

\* En este caso  $T \leq 1$ .

## 26 Estudio del caos en sistemas dinámicos discretos

máximo de patrones iniciales que la red pueda memorizar correctamente entre el número de neuronas de la red. en la sección anterior revisamos el caso de  $P$  finito y  $N$  grande.

$$\alpha = P/N \quad (2.53)$$

En donde  $\alpha$  es la fracción de patrones a memorizar en función del número total de neuronas en la red, en este caso  $\alpha$  toma valores mayores a cero. a diferencia de la sección anterior, en la cual el valor de  $\alpha$  era cercano a cero.

En este caso el término cruzado no lo podemos despreciar. ya que el valor de  $P$  no esta fijo el término cruzado no tiende a cero, por tal motivo tenemos que fijarnos en la relación entre el estado  $\langle S_i \rangle$  y los patrones a memorizar. la cual la definimos como:

$$m_\nu = \frac{1}{N} \sum_i \xi_i^\nu \langle S_i \rangle \quad \forall \nu = 1, \dots, P. \quad (2.54)$$

Al igual que en la sección anterior empezaremos suponiendo que el promedio de  $S_i$  es proporcional a alguno de los patrones a memorizar. Supongamos que deseamos memorizar el primer patrón  $\langle S_i \rangle \propto \xi_i^1$ , por lo tanto  $m_1$  es del orden de la unidad, mientras que los términos  $m_\nu$  con  $\nu \neq 1$  son del orden de  $\frac{1}{\sqrt{N}}$ .

Definiremos  $r$  como la desviación estandar del sistema.

$$r = \frac{1}{\alpha} \sum_{\nu \neq 1} m_\nu^2 \quad (2.55)$$

El cálculo de la capacidad requiere del uso de la teoría del campo medio por lo que la ecuación (2.54) queda de la siguiente forma.

$$m_\nu = \frac{1}{N} \sum_i \xi_i^\nu \tanh \left( \beta \sum_\mu \xi_i^\mu m_\mu \right) \quad (2.56)$$

Separaremos la suma en tres partes  $\mu = 1$ ,  $\mu = \nu$  y para  $\mu \neq \nu$ :

$$m_\nu = \frac{1}{N} \sum_i \xi_i^\nu \xi_i^1 \tanh \left( \beta \left( m_1 + \xi_i^\nu \xi_i^1 m_\nu + \sum_{\mu \neq 1, \nu} \xi_i^\mu \xi_i^1 m_\mu \right) \right) \quad (2.57)$$

Recordemos que la función tangente hiperbólica es una función impar y  $\xi_i^\nu = \pm 1$ , el primer término dentro del argumento de la tanh es de orden de la unidad, mientras que el segundo término es de orden  $\frac{1}{\sqrt{N}}$  por último el tercer término tiene alrededor de  $P$  términos. A continuación realizaremos una expansión en serie de Taylor para la ecuación (2.57).

$$m_\nu = \frac{1}{N} \sum_i \xi_i^\nu \xi_i^1 \tanh \left[ \beta \left( m_1 + \sum_{\mu \neq 1, \nu} \xi_i^\mu \xi_i^1 m_\mu \right) \right] + \frac{\beta}{N} \sum_i \left[ 1 - \tanh^2 \left( \beta \left[ m_1 + \sum_{\mu \neq 1, \nu} \xi_i^\mu \xi_i^1 m_\mu \right] \right) \right] m_\nu \quad (2.58)$$

A partir de la ecuación (2.55) los  $m_\mu$  con  $\mu \neq 1$  son variables aleatorias independientes con media cero y varianza  $\alpha r/P$ . Al término que involucra la  $\tanh^2$  le aplicaremos el teorema del límite central ya que los  $\xi_i^\mu \xi_i^1$  son variables aleatorias independientes de  $m_\mu$  y  $N^{-1} \sum_i$  es un promedio\* sobre  $\sum_{\mu \neq 1, \nu} \xi_i^\mu \xi_i^1 m_\mu$  cuya media es cero y la varianza es  $\alpha r$ .

$$m_\nu = \frac{N^{-1} \sum_i \xi_i^\nu \xi_i^1 \tanh \left[ \beta \left( m_1 + \sum_{\mu \neq 1, \nu} \xi_i^\mu \xi_i^1 m_\mu \right) \right]}{[1 - \beta(1 - q)]} \quad (2.59)$$

En donde :

$$q = \int \frac{dz}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \tanh^2 (\beta [m_1 + \sqrt{\alpha r} z]) \quad (2.60)$$

A partir de la ecuación (2.59) podemos calcular  $r$ . tenemos que elevar al cuadrado para poder encontrar el cuadrado de cada  $m_\nu$ .

$$m_\nu^2 = \left[ \frac{1}{1 - \beta(1 - q)} \right]^2 \frac{1}{N^2} \sum_{ij} \xi_i^\nu \xi_i^1 \xi_j^\nu \xi_j^1 \tanh \left[ \beta \left( m_1 + \sum_{\mu \neq 1, \nu} \xi_i^\mu \xi_i^1 m_\mu \right) \right] \tanh \left[ \beta \left( m_1 + \sum_{\mu \neq 1, \nu} \xi_j^\mu \xi_j^1 m_\mu \right) \right] \quad (2.61)$$

---

\* Como los  $\xi_i^\mu$  son variables aleatorias tenemos que  $E(g(x)) = \int g(x)f(x)$  donde  $f(x)$  es función de distribución de  $x$ .

## 28 Estudio del caos en sistemas dinámicos discretos

Separaremos los términos  $i = j$ , son  $N$  términos iguales.

$$\begin{aligned}
 m_\nu^2 &= \left[ \frac{1}{1 - \beta(1 - q)} \right]^2 \frac{1}{N^2} \sum_{i \neq j} \xi_i^\nu \xi_i^1 \xi_j^\nu \xi_j^1 \tanh \left[ \beta \left( m_1 + \sum_{\mu \neq 1, \nu} \xi_i^\mu \xi_i^1 m_\mu \right) \right] \\
 &\quad \tanh \left[ \beta \left( m_1 + \sum_{\mu \neq 1, \nu} \xi_j^\mu \xi_j^1 m_\mu \right) \right] + \left[ \frac{1}{1 - \beta(1 - q)} \right]^2 \times \\
 &\quad \frac{1}{N^2} \sum_i \tanh^2 \left[ \beta \left( m_1 + \sum_{\mu \neq 1, \nu} \xi_i^\mu \xi_i^1 m_\mu \right) \right] \tag{2.62}
 \end{aligned}$$

En el caso de la suma de  $i \neq j$  podemos conocer el promedio de  $\xi_i^\nu \xi_i^1 \xi_j^\nu \xi_j^1$  independientemente del promedio de la  $\tanh(x)$ , ya que  $\nu$  es la variable y no se encuentra dentro de la  $\tanh(x)$  este promedio es cero, por lo que sólo sobrevive el término de la  $\tanh^2(x)$ .

$$\begin{aligned}
 m_\nu^2 &= \left[ \frac{1}{1 - \beta(1 - q)} \right]^2 \frac{1}{N^2} \sum_i \tanh^2 \left[ \beta \left( m_1 + \sum_{\mu \neq 1, \nu} \xi_i^\mu \xi_i^1 m_\mu \right) \right] \\
 &= \frac{1}{N} q \times \left[ \frac{1}{1 - \beta(1 - q)} \right]^2 \tag{2.63}
 \end{aligned}$$

En la ecuación anterior hemos encontrado el valor para  $m_\nu^2$ , por la ecuación (2.55) podemos obtener:

$$r = \frac{q}{[1 - \beta(1 - q)]^2} \tag{2.64}$$

Por último la ecuación para  $m_1$  la podemos encontrar de la misma forma que la ecuación (2.60).

$$m_1 = \int \frac{dz}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \tanh(\beta [m_1 + \sqrt{\alpha r} z]) \tag{2.65}$$

Ahora necesitamos resolver simultáneamente las ecuaciones (2.60), (2.64) y (2.65) para  $m_1$ ,  $q$  y  $r$ . se estudiará en particular el caso para  $T \rightarrow 0$ , en

este límite la cantidad  $C := \beta(1 - q)$  permanece finita. Primero realizaremos la simplificación de las ecuaciones (2.60) y (2.65). Calculemos la cantidad  $(1 - q)$ .

$$(1 - q) = \int \frac{dz}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} (1 - \tanh^2 \beta[m_1 + \sqrt{\alpha r} z]) \quad (2.66)$$

Observemos que en el límite  $T \rightarrow 0$  podemos aproximar la ecuación (2.66) como:

$$(1 - q) \approx \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \Big|_{-m_1/\sqrt{\alpha r}} \int (1 - \tanh^2 \beta[m_1 + \sqrt{\alpha r} z]) dz$$

A partir de la cual podemos obtener:

$$(1 - q) \approx \sqrt{\frac{2}{\pi \alpha r}} \frac{1}{\beta} e^{-\frac{m_1^2}{2\alpha r}} \quad (2.67)$$

Para la ecuación (2.65) obtenemos:

$$\begin{aligned} m_1 &= \int \frac{dz}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \tanh(\beta[m_1 + \sqrt{\alpha r} z]) \\ T \rightarrow 0 &\rightarrow \int \frac{dz}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \text{sign}[m_1 + \sqrt{\alpha r} z] \\ &= 2 \int_{\frac{-m_1}{\sqrt{\alpha r}}}^{\infty} \frac{dz}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} - 1 \\ &= \text{erf} \left( \frac{m_1}{\sqrt{2\alpha r}} \right) \end{aligned} \quad (2.68)$$

Nuestras tres ecuaciones se convierten en:

$$C = \sqrt{\frac{2}{\alpha r \pi}} \exp \left( -\frac{m_1^2}{2\alpha r} \right) \quad (2.69)$$

$$r = \frac{1}{(1 - C)^2} \quad (2.70)$$

$$m = \text{erf} \left( \frac{m_1}{\sqrt{2\alpha r}} \right) \quad (2.71)$$

### 30 Estudio del caos en sistemas dinámicos discretos

La capacidad de la red la podemos encontrar resolviendo estas tres ecuaciones numéricamente, gráficamente podemos hacer  $y = m_1/\sqrt{2\alpha r}$  y obtener la ecuación.

$$\operatorname{erf}(y) = y \left( \sqrt{2\alpha} + \frac{2}{\sqrt{\pi}} e^{-y^2} \right) \quad (2.72)$$

A partir de resolver esta última ecuación se puede observar que existe un punto crítico a partir del cual las soluciones no triviales ( $m_1 \neq 0$ ) desaparecen. La evaluación numérica da un  $\alpha$  crítico ( $\alpha_c$ ). Recordemos que este análisis es válido para cuando  $T \rightarrow 0$ .

$$\alpha_c \approx 0.138 \quad (2.73)$$

A partir del valor del  $\alpha_c$  la red no presenta una buena memoria asociativa, al no presentar una buena memoria asociativa podemos decir que la red se vuelve inestable.

## MODELO ASIMÉTRICO

Ahora estudiaremos una red neuronal asimétrica. Nuevamente estudiaremos la estabilidad inicial de la red y la capacidad de almacenamiento de la red para el caso determinístico. Posteriormente revisaremos para que valores de  $\alpha$  la red es inestable. Nuestro objetivo es el estudio del caos en sistemas dinámicos discretos, para lo cual realizaremos una simulación de la red neuronal. La simulación nos permitirá observar que para valores de  $\alpha$  mayores al valor crítico ( $\alpha_c$ ) de este modelo la red se vuelve inestable. La pérdida de la estabilidad de la red ocasiona la pérdida del reconocimiento secuencial de patrones, provocando que la red vaya de un estado a otro sin que esta alcance un estado estable o un ciclo.

El estudio de las redes neuronales asimétricas es de gran importancia ya que presentan una actividad periódica. La actividad periódica no puede ocurrir cuando la red logra un equilibrio como en el caso del modelo de Hopfield<sup>[1,2]</sup>. A pesar de la actividad periódica de la red, en algunos casos los ciclos son de período muy largo, esto produce que el comportamiento de la red antes de alcanzar el ciclo sea en ocasiones muy complejo, esta situación ha provocado que en algunos casos se diga que esta dinámica es caótica. En este capítulo presentamos una red neuronal tipo Hopfield con una dinámica asimétrica y una evolución en paralelo, la simulación es de una red neuronal con dinámica determinística (i.e. temperatura igual a cero). Podemos considerar que este modelo es una modificación del modelo de Hopfield<sup>[1,2]</sup>, lo que nos permite estudiar el modelo de forma semejante. La métrica del espacio estará dada por la distancia de Hamming (ecuación (2.1)).

## EL MODELO ASIMÉTRICO

A continuación definiremos el modelo con el cual trabajaremos a lo largo de este capítulo, y al cual denominaremos como modelo asimétrico. La regla de evolución para temperatura distinta de cero y finita es:

$$\text{prob}(S_i = \pm 1) = \frac{1}{1 + \exp(\mp 2h_i \beta)} \quad (3.1)$$

## 32 Estudio del caos en sistemas dinámicos discretos

La regla de evolución de la red para temperatura cero está dada por:

$$S_i(t+1) = \text{sign} \left( \sum_{j=0}^N w_{ij} S_j(t) \right) \quad (3.2)$$

La función signo esta definida en la ecuación (2.3) y al igual que en el modelo anterior los  $w_{ij}$  son los pesos de las conexiones entre la neurona  $i$  y la neurona  $j$ . Los  $w_{ij}$  forman una matriz de  $N \times N$ .  $S_i$  es el valor de la  $i$ -ésima neurona. El tiempo está representado por  $t$  y se toma como una variable discreta. Durante este capítulo supondremos que los patrones iniciales de la red ( $\xi_i^\mu$ ) están generados de forma aleatoria e independiente.

Como se observa en las ecuaciones (3.1) y (3.2) la forma general del modelo no presenta variación alguna con respecto al modelo de Hopfield (ecuaciones (2.4) y (2.41), respectivamente). Una de las modificaciones que realizaremos durante este capítulo se refiere a la actualización de la red, la red evolucionará en paralelo, esto quiere decir que actualizaremos todas las neuronas de la red en cada unidad de tiempo. Como consecuencia de la actualización en paralelo el estado  $S(t)$  es un vector con  $N$  entradas. La otra modificación es la forma de construir la matriz de pesos de la red ( $w_{ij}$ ), la cual dejará de ser una matriz simétrica para convertirse en una matriz asimétrica. La suma sobre  $\mu$  es una suma módulo  $P^*$ .

Los elementos de la matriz de pesos de la red los definimos como:

$$w_{ij} = \frac{1}{N} \left( \sum_{\mu=1}^P \xi_i^{\mu+1} \xi_j^\mu \right) \quad (3.3)$$

Los  $\xi_i^\mu$  son los patrones iniciales de la red.

---

\* La suma módulo  $P$  se refiere a:  $\xi^{x+P} = \xi^x$ .

ESTABILIDAD INICIAL DE LA RED.

Estudiaremos la estabilidad inicial de la red para el caso determinístico (temperatura igual a cero). Nuevamente deseamos que la red memorice inicialmente los  $P$  patrones, nuevamente usaremos:

$$h_i^\nu = \sum_j w_{ij} \xi_j^\nu \quad (3.4)$$

Para que el conjunto de patrones inicial sea estable queremos que si la configuración inicial de la red corresponde a algún  $\xi^\mu$  la red se mantenga en ese estado inicial.

$$\text{sign}(h_i^\nu) = \xi_i^\nu \quad \forall i = 1, \dots, N. \quad (3.5)$$

Veamos que pasa para una neurona, usando la ecuación (3.2) obtenemos:

$$h_i^\nu = \sum_j w_{ij} \xi_j^\nu = 1/N \left( \sum_j \sum_\mu \xi_i^{\mu+1} \xi_j^\mu \xi_j^\nu \right) \quad (3.6)$$

En donde separando la suma en dos partes, la primera con  $\mu = \nu$  y la segunda con todos los términos restantes.

$$h_i^\nu = \xi_i^{\nu+1} + 1/N \left( \sum_j \sum_{\mu \neq \nu} \xi_i^{\mu+1} \xi_j^\mu \xi_j^\nu \right) \quad (3.7)$$

Nuevamente si el término cruzado es lo suficientemente chico no influye en el signo de  $h_i^\nu$ . En ese caso la ecuación (3.7) nos indica que el patrón que la red está reconociendo es uno de los patrones iniciales. Podemos pensar que la red sí es estable inicialmente\*. Para que la red sea estable el término cruzado deber ser lo suficientemente chico para toda  $i = 1, \dots, N$ .

Esta diferencia en el concepto de estabilidad de la red se debe a la forma de construir la matriz de pesos de la red, ya que al construirla usamos una interacción entre los patrones a memorizar. Además estamos llevando a cabo una actualización en paralelo de las neuronas. Cabe recordar que los patrones son

---

\* Este tipo de estabilidad sólo se puede dar si la red es actualizada en paralelo.

generados de manera aleatoria e independiente. por lo que no existe correlación alguna entre ellos. En este capítulo al decir que la red es estable queremos decir que si iniciamos la red en algún punto de un conjunto previamente determinado y la evolución inicial de la red no sale de ese conjunto entonces la red es estable. Este tipo de estabilidad es conocida como reconocimiento secuencial de patrones. En la siguiente sección analizaremos para que casos el término cruzado no influye en el valor final de  $h_i^\nu$ .

### CAPACIDAD DE ALMACENAMIENTO EN LA RED

A continuación hablaremos un poco del cálculo de la capacidad de almacenamiento para este modelo, el cálculo se lleva a cabo de manera similar al del modelo de Hopfield. Al igual que en el caso del modelo de Hopfield. cuando  $P$  tiene un valor finito y  $N$  es muy grande el término cruzado tiende a cero. por lo que no influye en el valor de  $h_i^\nu$ . Esto se debe a que nuevamente tenemos una suma de  $N(P - 1)$  términos independientes.

Ahora analizaremos el caso en el que  $P$  no esta fijo, partiremos de la idea de calcular la probabilidad que tienen cada una de las entradas  $\xi_i^\mu$  de tomar los valores  $\pm 1$  y a partir de esos cálculos extendernos a la teoría de probabilidad. Veamos que pasa en el caso de una neurona, primeramente definiremos  $C_i^{\nu+1}$  como  $-\xi_i^{\nu+1}$  veces el segundo término de la ecuación (3.7).

$$C_i^{\nu+1} = -\xi_i^{\nu+1} 1/N \left( \sum_j \sum_{\mu \neq \nu} \xi_i^{\mu+1} \xi_j^\mu \xi_j^\nu \right) \quad (3.8)$$

$C_i^{\nu+1}$  influye en el valor final de  $h_i^\nu$  sólo si es mayor que uno, nuevamente la probabilidad que queremos calcular es la probabilidad  $C_i^{\nu+1}$  sea mayor que uno.

$$P_{error} = P(C_i^{\nu+1} > 1) \quad (3.9)$$

Como cada uno de los  $\xi_i^\mu$  está generado de forma aleatoria e independiente,  $C_i^{\nu+1}$  es una función de los  $\xi_i^\nu$  y además  $N, P$  son mucho mayores que uno\*, obtenemos:

$$C_i^{\nu+1} \sim N(0, P/N) \quad (3.10)$$

---

\* Para más detalles ver el capítulo 2. capacidad de almacenamiento de la red.

Como estamos realizando una actualización en paralelo la  $i$  corre sobre todas las neuronas  $i = 1, 2, \dots, N$ . por tal motivo para calcular la probabilidad del error total tenemos que calcular la probabilidad conjunta de todos los  $C_i^{\nu+1}$ .

$$P_{\text{error total}} = P(C_i^{\nu+1} > 1 \forall i = 1, 2, \dots, N) \quad (3.11)$$

Como  $C_i^{\nu+1}$  tiene una distribución normal para todas las  $i$ , la probabilidad que deseamos conocer es la probabilidad conjunta de  $N$  variables aleatorias con distribución normal. Observemos que la capacidad de almacenamiento de este modelo es distinta a la encontrada para el modelo de Hopfield, sin embargo se calcula de forma muy similar.

## TEORÍA DEL CAMPO MEDIO.

En este modelo también se puede aplicar la teoría del campo medio para el análisis de la capacidad de la red neuronal en el caso de temperatura distinta de cero. A partir de la teoría del campo medio<sup>[11]</sup> se puede calcular el valor crítico de  $\alpha$ , en este caso el valor crítico de  $\alpha$  es:

$$\alpha_c \approx 0.278 \quad (3.12)$$

A partir de este valor crítico de  $\alpha$  la red se vuelve inestable, presenta ciclos límite como atractores, en la siguiente sección podremos observar la inestabilidad de la red.

## RESULTADO DE LAS SIMULACIONES.

En esta sección mostraremos los resultados de la simulación de la red neuronal con la dinámica asimétrica. el objetivo de este estudio sobre el comportamiento de la red. es observar que efectivamente para valores mayores al  $\alpha_c$  la red se vuelve inestable. Primero tomaremos un valor pequeño de  $\alpha$  para posteriormente aumentarlo poco a poco. lo que nos permitirá visualizar claramente los efectos del incremento de  $\alpha$ .

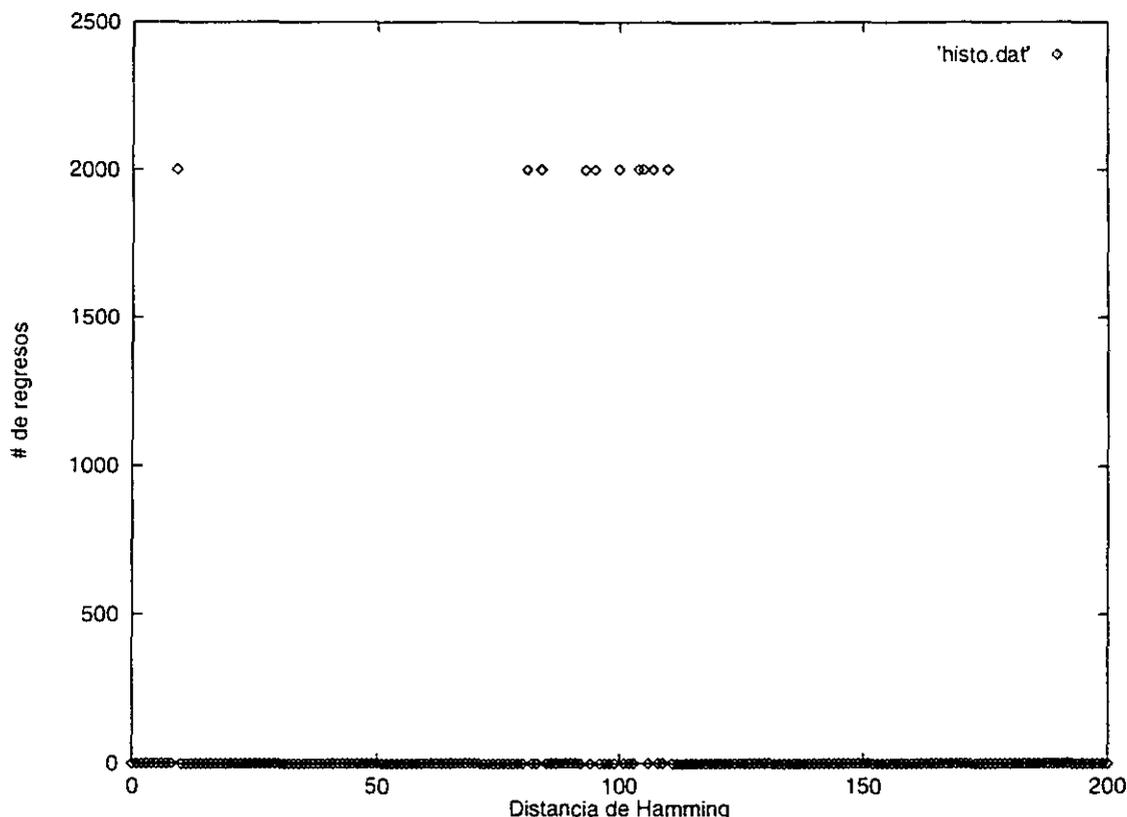


fig 3.1. Histograma de las distancias de Hamming de una red neuronal con  $N = 200$  neuronas y  $\alpha = 0.05$ . En esta gráfica podemos observar que la estabilidad de la red ya que presenta un ciclo, por lo cual podemos decir que la red es estable.

Recordemos que la distancia usada es la distancia de Hamming (ecuación (2.1)), por lo que la máxima distancia es  $N$ . Graficaremos las longitudes de los ciclos de la red. primero se fija el estado inicial  $S'$  y en cada unidad de tiempo se toma la distancia de Hamming del estado obtenido  $S$  al estado inicial  $S'$ . Después se lleva a cabo un análisis de regresión para estimar el ajuste que tiene la curva obtenida a una curva gaussiana\*. A pesar de que el análisis de regresión indica que para valores grandes de  $\alpha$  el comportamiento de la red se ajusta a una curva gaussiana la diferencia entre la varianza ( $\sigma^2$ ) de los datos obtenidos y el valor de la varianza gaussiana (Es el valor esperado de la varianza  $\sqrt{50}$ ) es relativamente grande, esto nos indica que la red no tiene un comportamiento ciento por ciento aleatorio.

\* El ajuste se realiza a una curva gaussiana para poder comparar este proceso con un proceso aleatorio.

A partir de este momento la red neuronal que simulamos consta de  $N = 200$  neuronas. Vamos a incrementar el número de patrones  $P$  en la red, con lo cual el valor de  $\alpha$  se incrementará poco a poco. Todas las simulaciones se llevaron a cabo con 20.000 iteraciones.

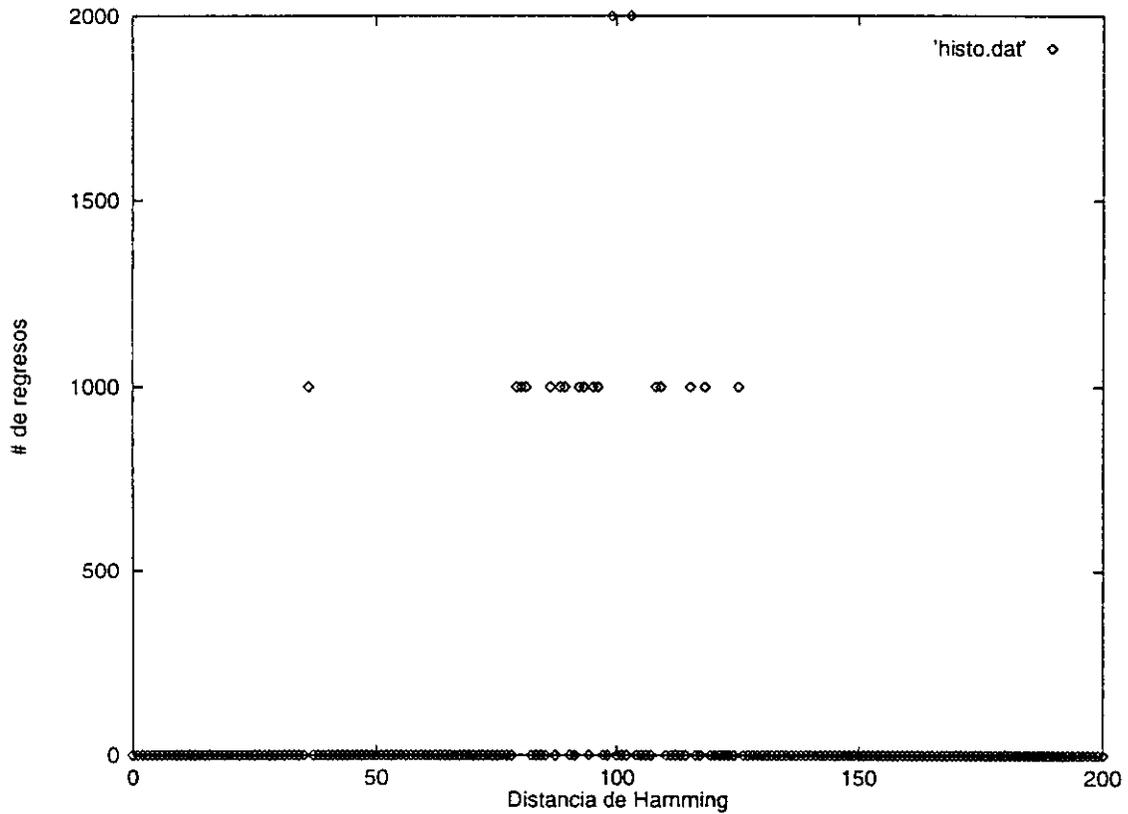


fig 3.2. Histograma de las distancias de Hamming de una red neuronal con  $N = 200$  neuronas y  $\alpha = 0.1$ , en este caso la red sigue presentando un comportamiento estable, a pesar de que existen estados predominantes.

En la fig 3.1 el número de patrones es  $P = 10$ .  $\alpha = 0.05$ . El valor de  $\alpha$  es muy pequeño, se encuentra por debajo del valor crítico, lo que observamos es que la red evoluciona a ciertos puntos que podemos considerar como puntos estables. También podemos observar que los diez puntos estables tienen el mismo número de regresos\*. cada uno tiene 2000 regresos.

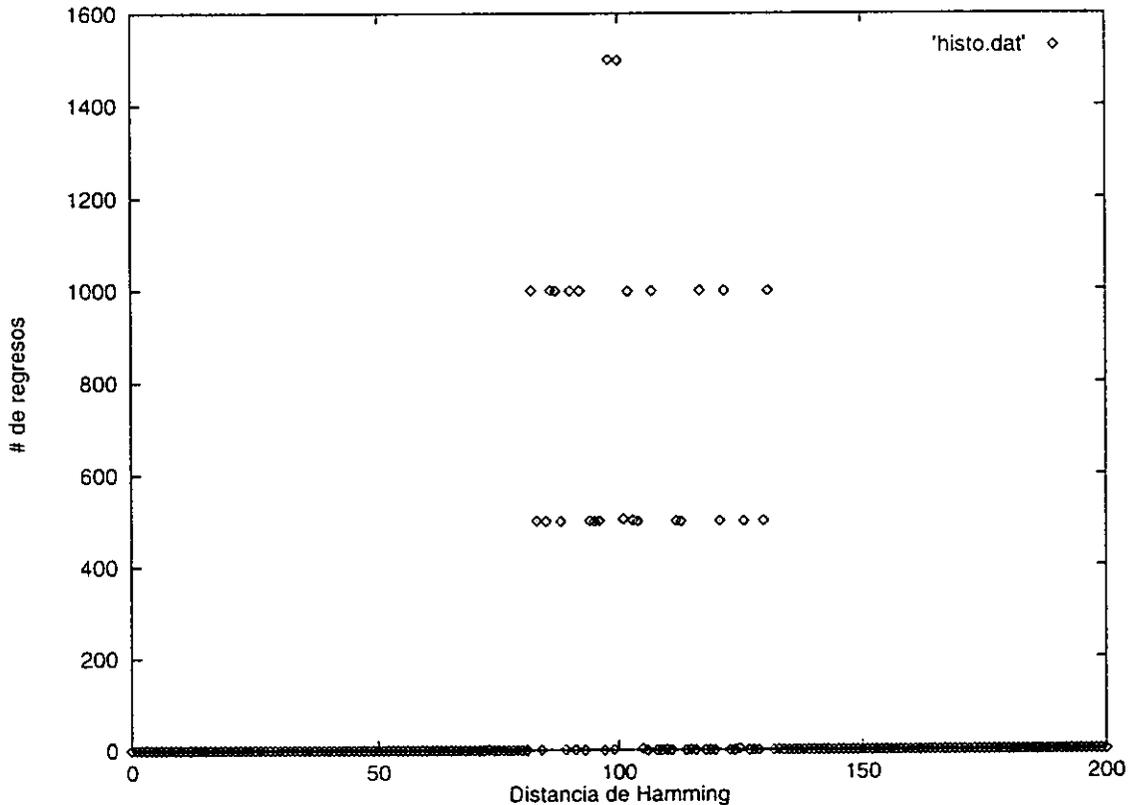
---

\* Al referirnos a regresos queremos decir el número de veces que se repite cierta distancia

### 38 Estudio del caos en sistemas dinámicos discretos

En conclusión para el valor de  $\alpha = 0.05$  la red tiene un buen reconocimiento secuencial de patrones.

Ahora aumentaremos el valor de  $P$ ,  $P = 20$  y  $\alpha = 0.1$ . Este valor de  $\alpha$  es un valor por debajo del valor crítico, por lo cual esperamos que la red mantenga cierta estabilidad (*fig3.2*).



*fig 3.3. Histograma de las distancias de Hamming de una red neuronal con  $N = 200$  neuronas y  $\alpha = 0.2$ . como este valor de  $\alpha$  es muy cercano al valor crítico de la red, esta presenta un comportamiento no muy estable.*

Como se puede observar la red comienza a ser inestable. existen distancias de Hamming a las cuales la red regresó sólo una vez. Esto nos indica que para algunos patrones iniciales la red no es estable, a pesar de la periodicidad la red no regresó a esos patrones iniciales.

En la siguiente gráfica el valor de  $\alpha$  será muy cercano al valor crítico.  $P = 40$ ,  $\alpha = 0.2$ . Esperamos que la red siga manteniendo cierta estabilidad (*fig 3.3*).

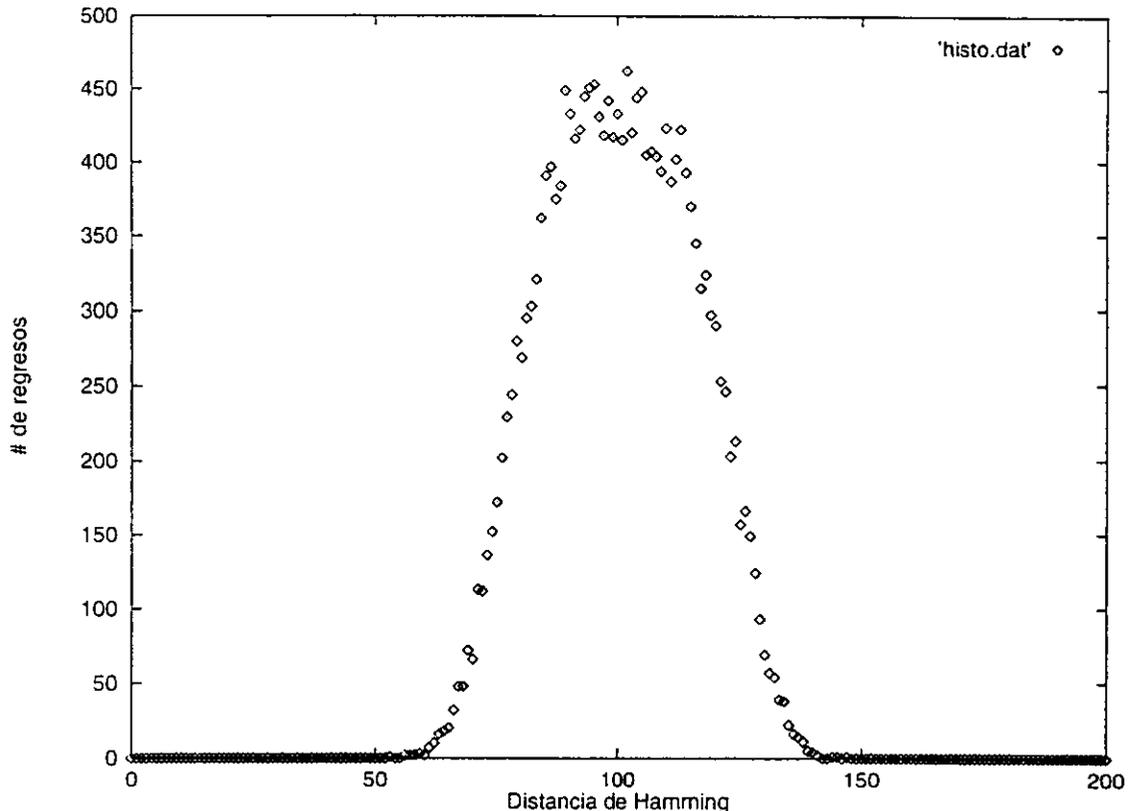


fig 3.4. Histograma de las distancias de Hamming de una red neuronal con  $N = 200$  neuronas y  $\alpha = 0.3$ , en este caso ya hemos rebasado el valor crítico, por lo que la red ya no tiene un comportamiento cíclico.

La gráfica muestra que la red cada vez es más inestable, se puede decir que la red se ha vuelto desordenada. En conclusión la red ya no presenta un buen reconocimiento secuencial de patrones.

A partir de este momento el número de patrones iniciales en la red seguirá incrementando, provocando que el valor de  $\alpha$  sea superior al valor del  $\alpha_c$ . En consecuencia la red se vuelve desordenada, como se podrá observar en la siguiente gráfica (fig 3.4).

En la gráfica (fig 3.4) tomaremos el valor de  $P = 60$  y  $\alpha = 0.3$ , en esta gráfica se ve claramente como la red ha perdido toda estabilidad y todo el reconocimiento secuencial de patrones.

Como se puede observar en la gráfica de la fig 3.4 conforme incrementamos el valor de  $\alpha$  la curva de la gráfica se parece a la curva de la distribución Normal (gaussiana). si incrementamos más el valor de  $\alpha$  la curva se ajustará cada vez más

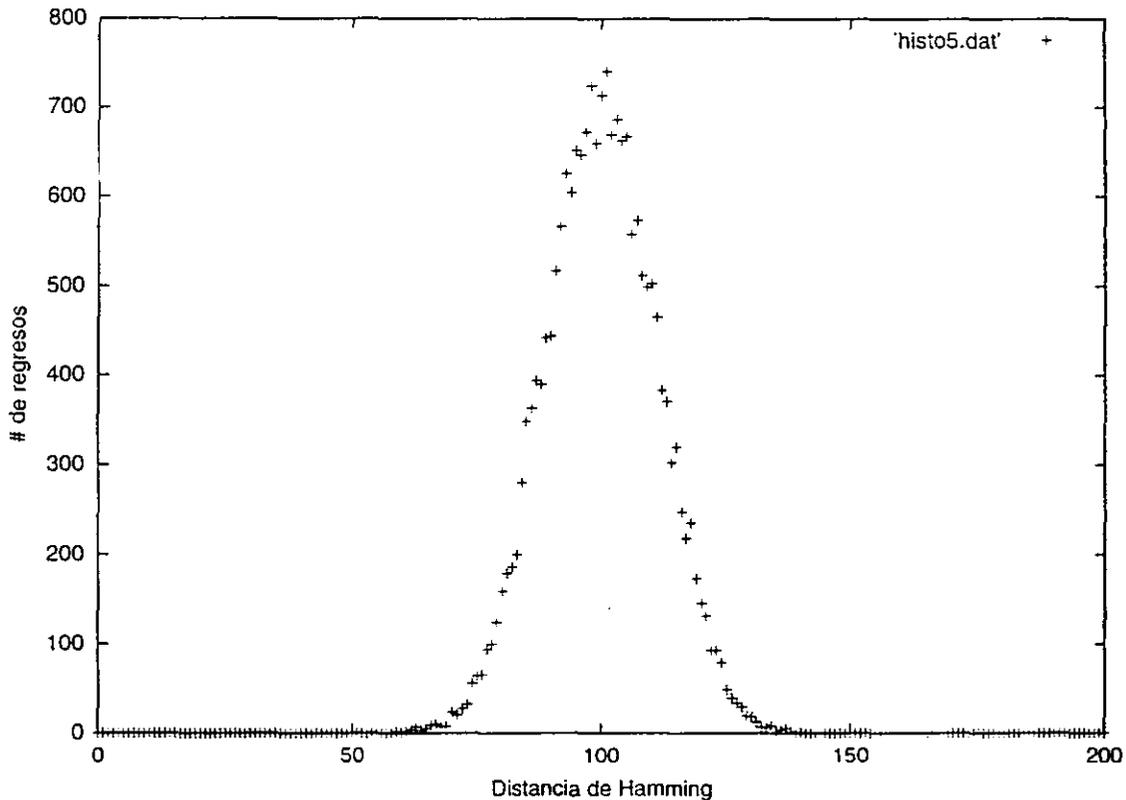


fig 3.5. Histograma de las distancias de Hamming de una red neuronal con  $N = 200$  neuronas y  $\alpha = 0.5$ . conforme se incrementa el valor de  $\alpha$  observamos que la gráfica de la función se asemeja a una gráfica de un proceso aleatorio.

a la curva de la Normal. Como ejemplo de esto tomaremos el valor de  $P = 100$  y el valor de  $\alpha = 0.5$ . este valor de  $\alpha$  ya es muy superior al valor crítico (fig 3.5).

En la gráfica fig3.5 la red presenta tantos puntos que no es posible identificar los patrones iniciales como puntos estables. Como se observa en la gráfica la mayoría de los estados tienen una distancia de Hamming alrededor de  $N/2$ , notemos que no existen regresos a distancias de Hamming menores que  $N/4$ .

Por último presentaremos los resultados del análisis de regresión<sup>[12]</sup> que hemos realizado a los datos, como se pudo observar en general la forma de la gráfica para  $\alpha > \alpha_c$  se asemeja mucho a una curva gaussiana (curva de una función Normal). el objetivo es identificar si existe una diferencia entre la dinámica de la red neuronal y una dinámica totalmente aleatoria. Para tener varios puntos de comparación se graficaron diez valores de  $\alpha$  y a cada uno de ellos se le realizó el análisis de regresión. En la siguiente tabla podemos observar

los resultados. En la primera columna se encuentra el valor de  $\alpha$ . En la segunda columna se encuentra el valor del coeficiente de correlación\* y por último en la tercera columna tenemos el promedio del incremento entre  $\sigma^2$  obtenida de la red y la  $\sigma^2$  de la curva gaussiana†.

El valor de  $\alpha = 10.0$  presenta un mejor coeficiente de correlación y una menor diferencia entre los valores de  $\sigma^2$ . Como se esperaba conforme aumentamos el valor de  $\alpha$  la curva de la red neuronal se ajustó mejor a la curva de la función Normal, pero el promedio del incremento de  $\sigma^2$  es relativamente grande aun para el valor de  $\alpha = 10.0$ , por tal motivo no se puede asegurar que la red tenga un comportamiento totalmente aleatorio. Observemos que para valores menores que  $\alpha < 0.3$  el coeficiente de correlación es muy bajo y en consecuencia el ajuste es malo. Para el valor de  $\alpha = 0.05$  el análisis de regresión no es posible llevarlo a cabo. La *fig3.6* muestra la gráfica de los coeficientes de regresión.

$\alpha$	$coef_{corr}$	$\Delta\sigma^2/\sigma^2$
0.1	0.122525	1.031135
0.2	0.200444	0.796133
0.3	0.973083	0.699881
0.4	0.977594	0.736258
0.5	0.978056	0.614557
1.0	0.978626	0.480434
2.0	0.972810	0.426392
5.0	0.970629	0.393868
10.0	0.981275	0.274354

*fig3.6, en este cuadro podemos observar los resultados del análisis de regresión realizado para ajustar el comportamiento de la red a un proceso aleatorio. En la primera columna se puede observar el valor de  $\alpha$ . En la segunda columna se observa el coeficiente de correlación obtenido (mientras el valor del coeficiente de correlación se aproxime más a uno el ajuste es mejor). En la última columna podemos observar el porcentaje de la variación entre el valor esperado de la varianza de un proceso aleatorio y el valor encontrado para la red neuronal (en este caso mientras más pequeña sea la diferencia el ajuste es mejor).*

\* mientras el valor del coeficiente de correlación se aproxime más a uno tendremos un mejor ajuste.

† El promedio está calculado como:  $\Delta\sigma^2/\sigma^2 = 1 - N/(4 * \sigma^2)$ .

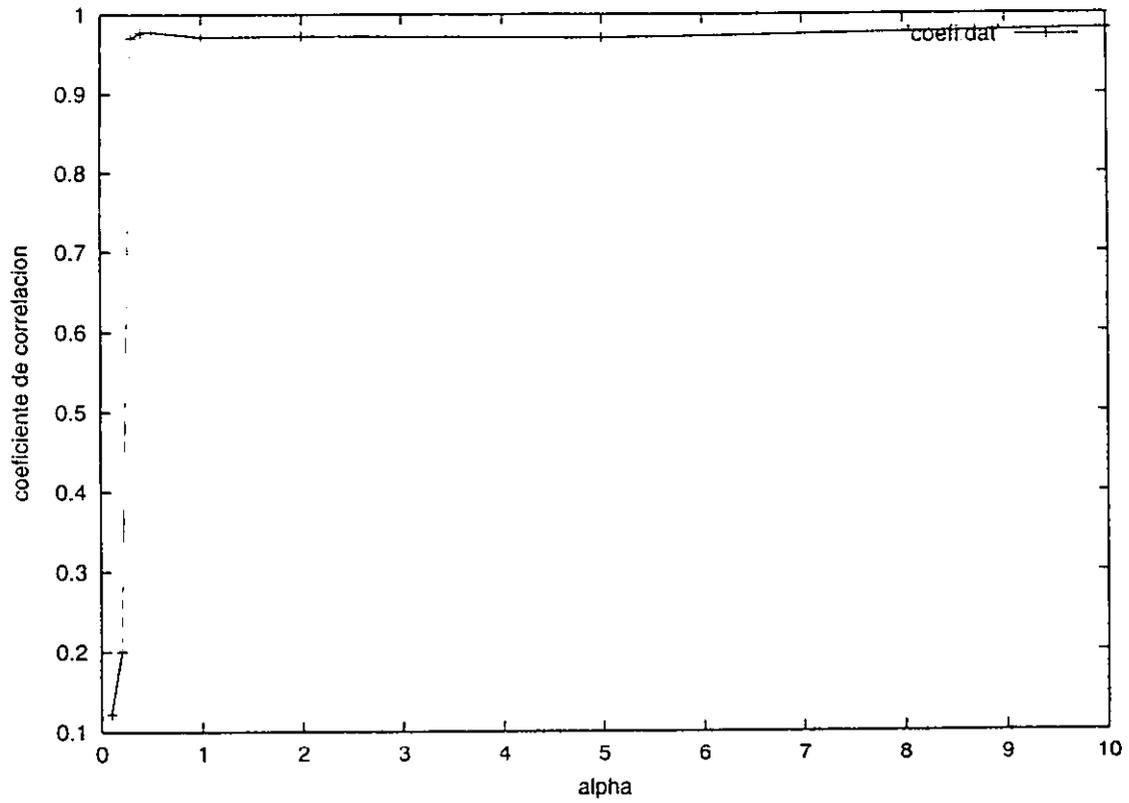


fig 3.7. Gráfica de los coeficientes de correlación con respecto a  $\alpha$ , en esta gráfica podemos observar que a partir de  $\alpha = 0.3$  se tiene un buen ajuste.

## MODELO DE CRISANTI.

A continuación estudiaremos una red neuronal asimétrica, con una dinámica determinística propuesta por Crisanti, Falcioni y Vulpiani<sup>[5]</sup>. El grado de asimetría de la red lo mediremos por medio del parámetro  $k$ . en esta red no habrá patrones a memorizar. Al igual que en el capítulo anterior la red volverá a ser inestable para valores de  $k$  superiores al  $k_c$ . La dinámica de la red es una dinámica determinística. en consecuencia la red sólo tiene un número finito de posibles estados dentro del espacio fase, esto provoca que después de un determinado tiempo la red se vuelva periódica, esto no quiere decir que el comportamiento de la red antes de volverse periódica sea trivial. Algunas personas consideran que el comportamiento de la red previo a la periodicidad puede ser visto como caótico<sup>[5]</sup>.

En este modelo la dinámica de la red sólo está definida para el caso determinístico, por lo que no podemos tener el concepto de temperatura. La matriz de pesos de la red es asimétrica para valores de  $k$  distintos de cero, como la red presenta ciclos límite como atractores, no tiene sentido hablar de una función de energía. La evolución de la red se llevará a cabo en paralelo.

## EL MODELO DE CRISANTI

Ahora definiremos el modelo con el cual trabajaremos a lo largo de este capítulo, y al cual denominaremos como modelo de Crisanti<sup>[5]</sup>. La regla de evolución es:

$$S_i(t+1) = \text{sign} \left( \sum_{j=1}^N w_{ij} S_j(t) \right) \quad (4.1)$$

Donde  $\text{sign}(x)$  es la función signo definida en la ecuación (2.3). De igual modo que en el modelo anterior los  $w_{ij}$  son los pesos de las conexiones entre la neurona  $i$  y la neurona  $j$ . los pesos forman una matriz de  $N \times N$ .  $S_i$  es el valor de la  $i$ -ésima neurona. El tiempo está representado por  $t$  y es una variable discreta.

## 44 Estudio del caos en sistemas dinámicos discretos

Una de las particularidades de este modelo es que no tiene patrones iniciales.

Nuevamente la red evolucionará en paralelo, lo que quiere decir que actualizaremos todas las neuronas de la red en cada unidad de tiempo. Los elementos de la matriz de pesos de la red los definiremos como:

$$w_{ij} = w_{ij}^s + kw_{ij}^a \quad k \geq 0 \quad (4.2)$$

En la ecuación anterior los  $w_{ij}^s$  son de la forma  $w_{ij}^s = w_{ji}^s$  y definen una matriz simétrica\* de  $N \times N$ . Los  $w_{ij}^a$  se definen como  $w_{ij}^a = -w_{ji}^a$ , forman una matriz antisimétrica de  $N \times N$ . Cada una de las entradas de las matrices es generada de forma aleatoria e independiente con una distribución normal con media cero y varianza  $\sigma^2 = 1/(N-1)(1+k^2)$ †. El parámetro  $k$  mide el grado de antisimetría de la red.

### ANÁLISIS DEL MODELO

En este modelo no hay patrones a memorizar ni patrones iniciales, tampoco existe  $\alpha$ . En este modelo tenemos el parámetro  $k$ , que es un parámetro semejante a  $\alpha$ , en este caso si  $k = 0$  la red es simétrica,  $k \neq 0$  la red es asimétrica y para  $k \rightarrow \infty$  la red es antisimétrica.

Como esta es una red neuronal con un número finito de posibles estados y la dinámica de la red es totalmente determinística, todas las configuraciones iniciales evolucionan a un atractor definido. El atractor puede ser de dos tipos un punto fijo o un ciclo, a pesar de esto no podemos calificar el comportamiento de la red como trivial, la razón es que el tiempo que el sistema tarda en alcanzar el atractor puede ser muy largo, como consecuencia Crisanti, et al...<sup>[5]</sup> consideran que el comportamiento de la red antes de alcanzar los ciclos límite puede ser descrito como caótico. Para valores por abajo del valor crítico de  $k$ , la longitud de los ciclos alrededor del atractor son potencias de  $N$ . El caso de valores superiores al  $k_c$ , la longitud de los ciclos crecen de forma exponencial y las trayectorias son

---

\* En el caso de la matriz simétrica pediremos que los elementos de la diagonal sean cero, es decir  $w_{ii}^s = 0 \forall i = 1, 2, \dots, N$ .

† Los renglones de la matriz de pesos tienen en promedio norma uno (los renglones de la matriz de pesos son variables aleatorias con una distribución Normal con media cero y varianza uno), independientemente del tamaño del sistema o de la simetría.

muy irregulares. Estas variaciones permiten que el comportamiento de la red durante el transitorio sea muy irregular. Se considera que este modelo presenta ciertas características de los sistemas caóticos. Por lo que en cierto sentido este modelo puede ser visto como caótico<sup>[5]</sup>.

En este modelo vamos a variar el parámetro  $k$ . en el caso con  $k = 0$  la matriz de pesos es una matriz simétrica provocando que la red tenga puntos fijos como atractores. Cuando  $k > 0$  la red puede presentar ciclos límite o puntos fijos como atractores. si  $k > k_c$  la red presenta ciclos límite de período<sup>[5]</sup>  $e^N$ . En este modelo el valor crítico fue calculado por Crisanti, et. al ...<sup>[5]</sup> y es de:

$$k_c = 0.5,$$

Crisanti, et.al... Encontraron que para  $k = 0.5$  la dinámica muestra un cambio de fase. característica de los sistemas dinámicos. además se calcula la entropía de la red. En esta red neuronal el estudio de la dinámica se realiza durante el período transitorio, esto es antes de que la red alcance el atractor. Esto se debe a que los atractores de la red sólo son de dos tipos un ciclo o un punto fijo, además el tiempo que la red tarda en alcanzar al atractor es del orden  $e^N$  por lo que es razonable estudiar el comportamiento de la red durante el transitorio.

## RESULTADO DE LAS SIMULACIONES.

En esta sección mostraremos como al hacer una simulación de una red neuronal con la dinámica de Crisanti. conforme incrementamos el valor del parámetro  $k$  la red poco a poco empieza a exhibir un comportamiento no trivial. En este modelo el valor crítico es  $k_c = 0.5$ . a partir del cual la red presenta ciclos límite como atractores. Pero si el valor de  $k$  aumenta demasiado entonces la red alcanza un estado en el que aparentemente los atractores nuevamente son puntos fijos\*.

---

\* En las simulaciones hemos realizado el valor aproximado es  $k \approx 1.3$ .

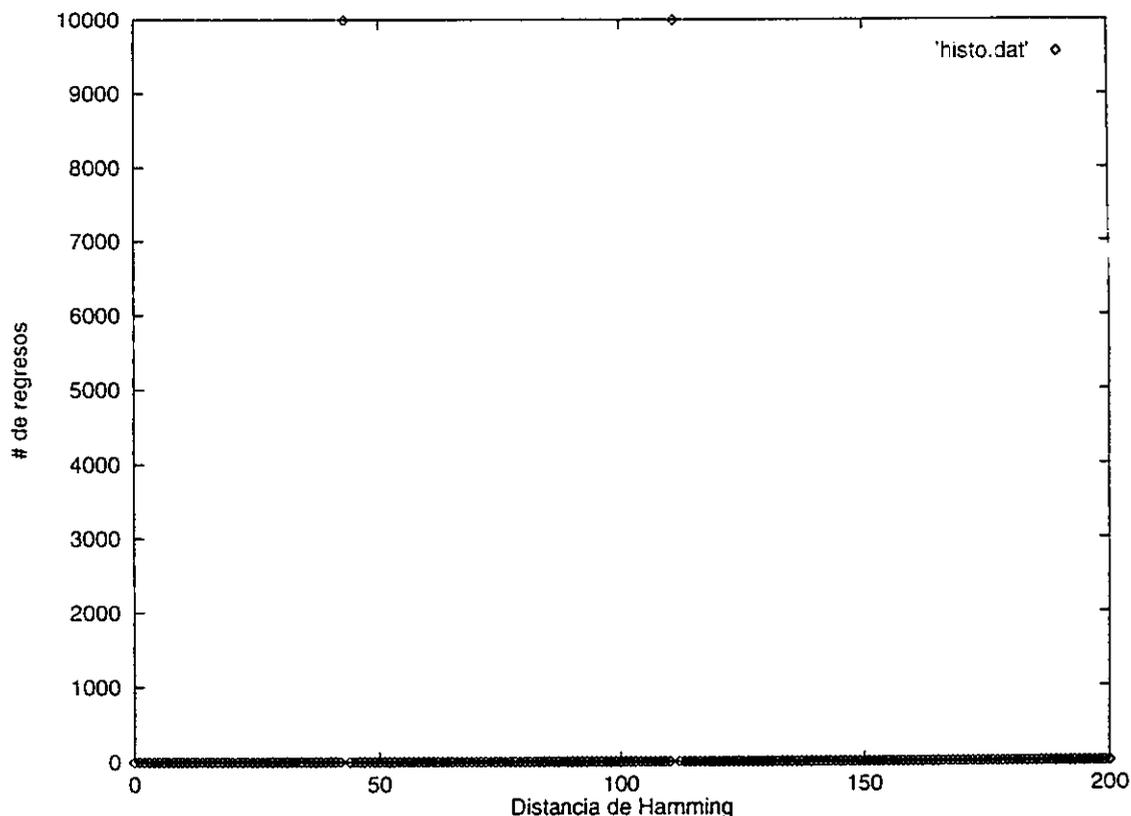


fig 4.1. Histograma de las distancias de Hamming de una red neuronal con  $N = 200$  neuronas y  $k = 0.0$ , en esta gráfica se puede observar que la gráfica presenta un ciclo, por lo tanto es estable.

Al igual que en el capítulo anterior graficaremos los ciclos de la red para la distancia de Hamming (2.1). Posteriormente realizaremos un análisis de regresión para estimar la curva de la función Normal, también calcularemos el promedio del incremento de  $\sigma^{2*}$ . La simulación se realizó con una red neuronal con  $N = 200$  neuronas y con 20,000 iteraciones. el parámetro  $k$  irá aumentando poco a poco. En la *fig4.1* tenemos la red neuronal con  $k = 0.0$ , como se puede observar esta cuenta con dos puntos fijos como atractores al igual que en la gráfica *fig4.2* en donde el valor es  $k = 0.2$ .

---

\* El promedio del incremento es calculado de la misma forma que en el capítulo anterior, ver *fig3.7*.

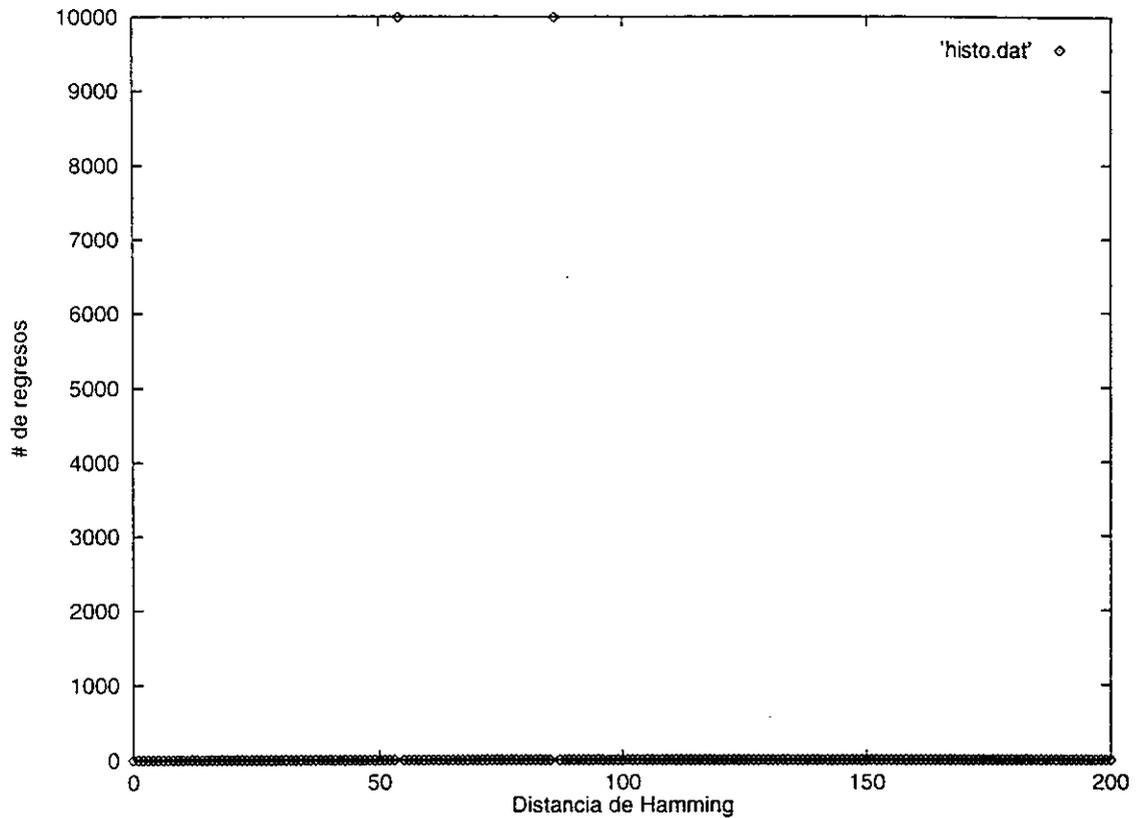
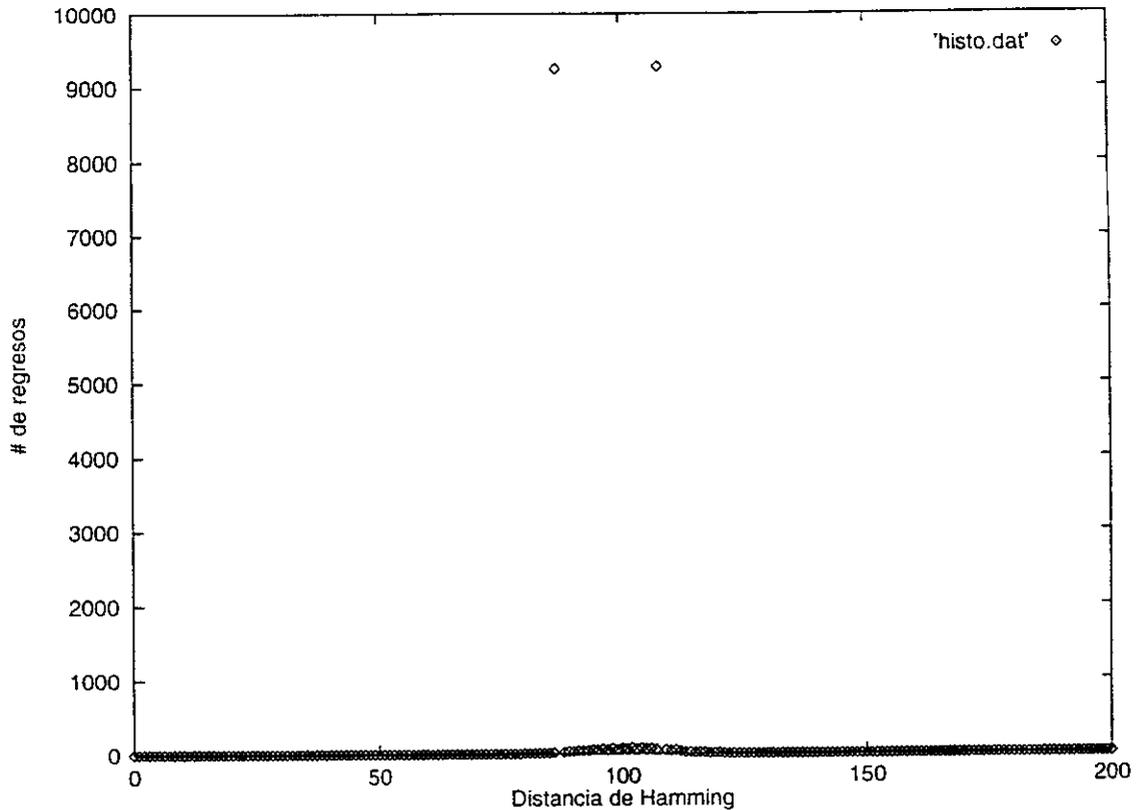


fig 4.2. Histograma de las distancias de Hamming de una red neuronal con  $N = 200$  neuronas y  $k = 0.2$ , nuevamente la red es estable ya que el valor de  $k$  es inferior al valor crítico.

Como podemos observar la red es estable para valores de  $k$  menores al  $k_c$ , tal y como se esperaba la red presenta puntos fijos como atractores. Recordemos que el valor crítico de  $k$  en el modelo es  $k = 0.5$ , a partir del cual la red empieza a presentar ciclos límite y puntos fijos como atractores como se puede observar en la gráfica fig4.3.



*fig 4.3. Histograma de las distancias de Hamming de una red neuronal con  $N = 200$  neuronas y  $k = 0.5$ . En este caso hemos alcanzado el valor crítico por lo que la red ya no presenta un comportamiento cíclico, se puede observar que para distancias de Hamming alrededor de cien existen regresos.*

En la gráfica *fig4.3* el valor de  $k = 0.5$ . Se observa que la red empieza a ser inestable, esto lo podemos observar en que ya no sólo existen dos puntos fijos como atractores. También podemos observar que para las distancias alrededor de  $N/2$  la red presenta regresos. Esto es indicativo de la inestabilidad de la red. Para el valor de  $k = 0.6$  la red ya perdió toda la estabilidad *fig4.4*. en esta gráfica ya se puede observar que la forma de la curva que describe es semejante a una curva gaussiana.

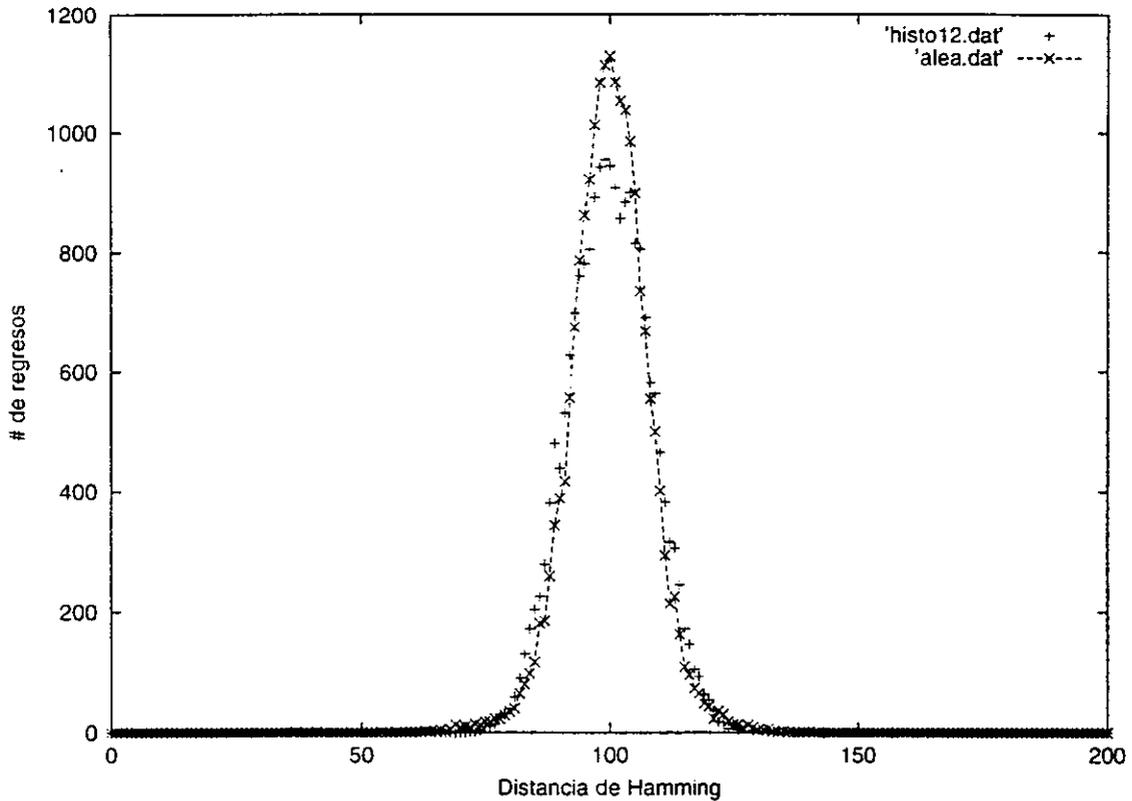
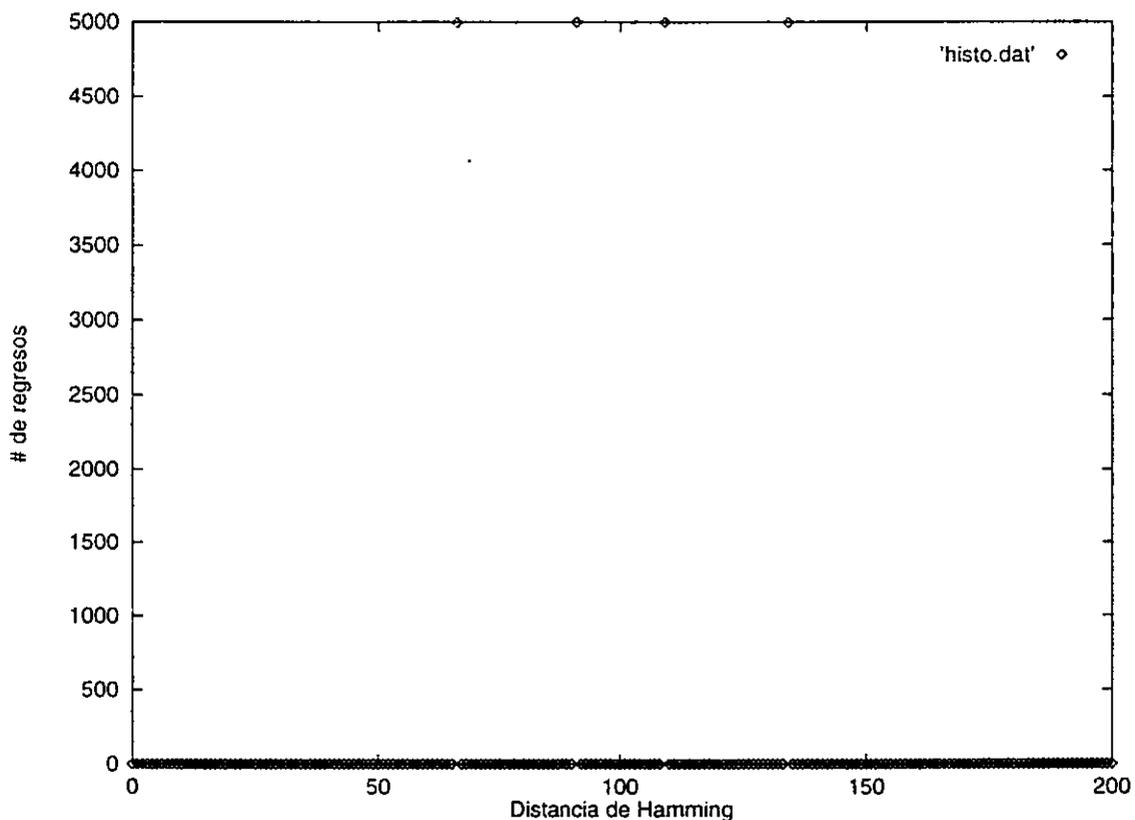


fig 4.4. Histograma de las distancias de Hamming de una red neuronal con  $N = 200$  neuronas y  $k = 0.6$ . El valor de  $k$  ya es superior al valor crítico y muestra un claro comportamiento desordenado, el cual se asemeja nuevamente a un proceso aleatorio. Como se observa los datos encontrados (+) estos difieren del proceso aleatorio (-) muy poco.

Mientras más se aumente el valor de  $k$  la gráfica de la curva se ajustará mejor a la función normal, siempre y cuando el valor  $k$  sea menor que  $k \approx 1.3$ . A partir del cual aparentemente la red comienza a relajarse nuevamente a un estado donde los atractores son puntos fijos. Esto lo visualizaremos mejor con el análisis de regresión. Para el valor de  $k = 3.0$  la gráfica de la red ya muestra cuatro puntos fijos como atractores (fig4.5).

## 50 Estudio del caos en sistemas dinámicos discretos

El análisis de regresión se realizó para 28 valores distintos de  $k$ . La siguiente tabla *fig4.7* se pueden observar los resultados del análisis de regresión. La primera columna muestra el valor de  $k$ . La segunda columna muestra el valor del coeficiente de regresión y en la tercera columna se observa el promedio del incremento de  $\sigma^2$ . El promedio del incremento de  $\sigma^2$  nos permite diferenciar el comportamiento de la red neuronal de un proceso totalmente aleatorio, si el comportamiento de la red fuera un proceso aleatorio el valor de  $\Delta\sigma^2/\sigma^2$  sería muy cercano a cero. sin embargo el valor más pequeño que obtuvimos es  $\Delta\sigma^2/\sigma^2 = 0.240301$  con  $k = 1.3$ . El cual es un valor relativamente grande.



*fig 4.5. Histograma de las distancias de Hamming de una red neuronal con  $N = 200$  neuronas y  $k = 3.0$ . En este caso el valor de  $k$  ya es muy superior al valor crítico provocando que la red regrese a un estado de estabilidad, nuevamente aparecen ciclos. esto se debe a que la red se vuelve antisimétrica.*

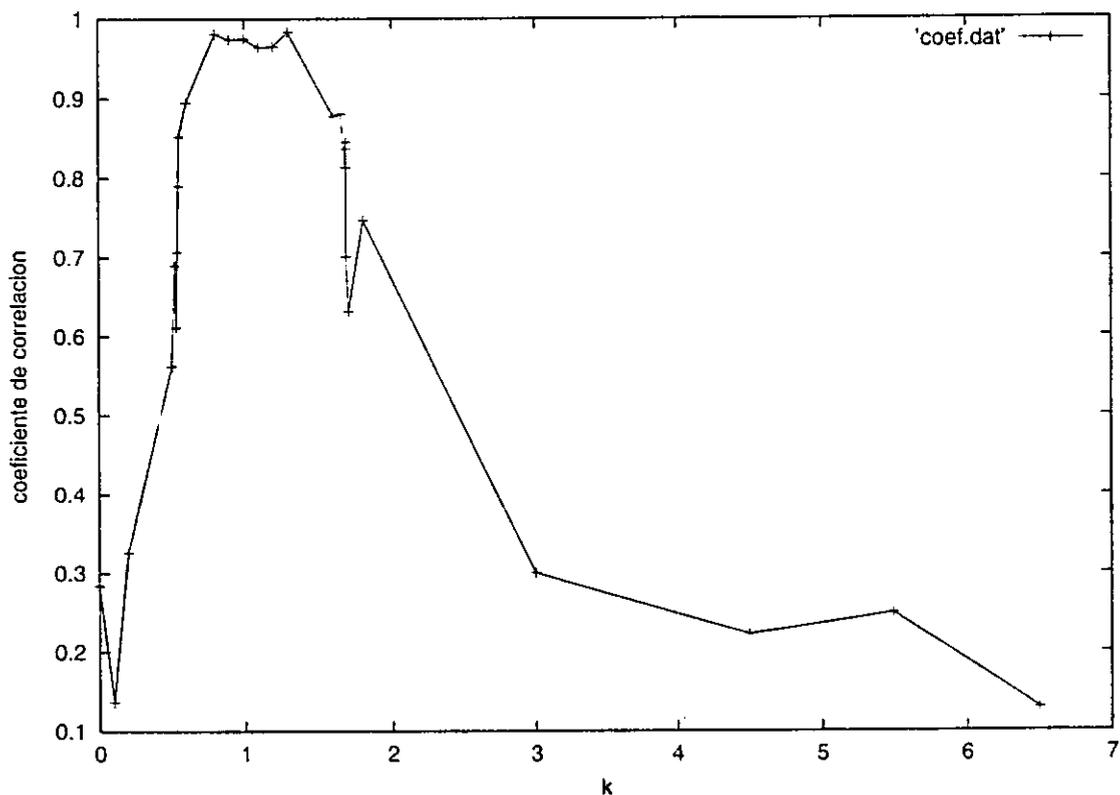
Nuevamente se calculó el incremento de  $\sigma^2$  y al igual que en el capítulo anterior es indicativo de la variación entre la  $\sigma^2$  de un proceso aleatorio y la  $\sigma^2$  de nuestra red neuronal. podemos observar que el comportamiento de la red neuronal no es un comportamiento aleatorio.

$k$	$coef_{corr}$	$\Delta\sigma^2/\sigma^2$
0.0000	0.283638	0.783282
0.1000	0.136976	0.917616
0.2000	0.325805	0.766399
0.5000	0.562238	0.764967
0.5200	0.690014	0.715795
0.5300	0.611088	0.689310
0.5400	0.706837	0.659765
0.5450	0.790421	0.584105
0.5500	0.852043	0.529718
0.6000	0.894972	0.449466
0.8000	0.981198	0.318029
0.9000	0.974327	0.275656
1.0000	0.974880	0.337595
1.1000	0.964688	0.352970
1.2000	0.965443	0.366841
1.3000	0.983763	0.240301
1.6000	0.878199	0.537646
1.6500	0.880899	0.537519
1.6820	0.813539	0.599834
1.6830	0.836782	0.599089
1.6845	0.845607	0.584875
1.6850	0.701030	0.537519
1.7000	0.631787	0.699620
1.8000	0.746749	0.620709
3.0000	0.300000	0.908725
4.5000	0.222168	0.952269
5.5000	0.249164	0.957342
6.5000	0.129060	0.979044

fig4.6. en este cuadro podemos observar los resultados del análisis de regresión realizado para ajustar el comportamiento de la red a un proceso aleatorio. En la primera columna se puede observar el valor de  $k$ . En la segunda columna se observa el coeficiente de correlación obtenido (mientras el valor del coeficiente de correlación se aproxime más a uno el ajuste es mejor). En la última columna podemos observar la variación entre el valor esperado de la varianza de un proceso aleatorio y el valor encontrado para la red neuronal (en este caso mientras más pequeña sea la diferencia nos indica que el proceso se aproxima mucho a un proceso aleatorio).

## 52 Estudio del caos en sistemas dinámicos discretos

Como podemos observar en la tabla (*fig4.6*) conforme aumentamos el valor de  $k$  el ajuste a la curva normal cada vez es mejor hasta el valor de  $k = 1.3$ , el cual es el valor que más se ajusta a la curva normal y a partir de ese momento el coeficiente de correlación empieza a disminuir hasta llegar nuevamente a valores cercanos a cero. la gráfica del coeficiente de correlación es la *fig4.7*.



*fig 4.7. Gráfica de los coeficientes de correlación con respecto a  $k$ . en esta gráfica se observa los coeficientes de correlación, podemos observar que existe una zona en la que los coeficientes de correlación son muy buenos para valores de  $k$  entre  $[0.5, 2.0]$  tenemos los mejores coeficientes de correlación.*

## REDES NEURONALES CON MEMORIA

## TEORÍA DEL CAOS DISCRETO.

En los capítulos anteriores estudiamos sistemas de redes neuronales sin memoria y como hemos visto el comportamiento que presentan es no trivial. En los modelos anteriores la métrica estaba dada por la distancia de Hamming. La teoría del caos discreto<sup>[4]</sup> ha extendido los conceptos de caos a los sistemas dinámicos discretos, en especial a los sistemas binarios, como lo son las redes neuronales. Las cuales se representan por variables binarias  $S_i$  con  $i = 1, \dots, N$  ó como un vector  $S$  con  $N$  entradas binarias. El espacio fase de las redes neuronales tiene  $2^N$  estados posibles y la métrica del espacio está dada por la distancia de Hamming (ecuación (2.1)).

En general el estudio de las redes neuronales se realiza para valores grandes de  $N$  (i.e.  $N \rightarrow \infty$ ), pero en ese caso la distancia de Hamming deja de ser una distancia\*. En el caso de que la distancia de Hamming esta dada por la ecuación (2.1), en el límite  $N \rightarrow \infty$  tenemos que  $d_h \rightarrow \infty$ , por tal motivo la distancia no cumple con la propiedad (a). Si definimos una distancia de Hamming normalizada (i.e.  $D_h = \frac{1}{N}d_h$ ) el problema de acotamiento queda resuelto ( $D_h = \frac{1}{N}d_h \leq 1$ ), sin embargo surge el problema de que existen estados  $S, S'$  distintos cuya distancia en el límite  $N \rightarrow \infty$  es cero provocando que no se cumpla la propiedad (b), como ejemplo tomemos  $S_1 \neq S'_1$  y  $S_i = S'_i \forall i = 2, 3, \dots, N$  en este caso la distancia normalizada es  $1/N$  y al tomar el límite  $N \rightarrow \infty$  la distancia tiende a cero. Esto provoca que la topología del espacio no esté bien definida, por estas razones la teoría del caos discreto<sup>[4]</sup> ha extendido el espacio de las posibles configuraciones de la red a  $(\Omega)$ , para tener un número infinito de estados, y tener una métrica bien definida. Los elementos de este nuevo espacio los denotaremos como  $S$ .

---

\* La función distancia definida en un espacio métrico  $(X)$  tiene las siguientes propiedades:

- (a)  $0 \leq d(x, y) < \infty \forall x, y.$
- (b)  $d(x, y) = 0 \Leftrightarrow x = y.$
- (c)  $d(x, y) = d(y, x) \forall x, y.$
- (d)  $d(x, y) \leq d(x, z) + d(z, y) \forall x, y, z.$

## 54 Estudio del caos en sistemas dinámicos discretos

Primero empezaremos por considerar a  $\Omega$  como el producto cartesiano de copias infinitas de  $\{S\}$ . los elementos de  $\Omega$  los denotaremos como:

$$\mathbf{S} = (S(0), S(1), \dots) \quad (5.1)$$

$S(t)$  es un vector con  $N$  entradas binarias. Convertiremos a  $\Omega$  en un espacio topológico introduciendo la siguiente base<sup>[4]</sup>.

$$N_n(\mathbf{S}) = \{S' \in \Omega \mid S'(m) = S(m), \forall m < n\} \quad (5.2)$$

En la ecuación anterior  $n = 1, 2, 3, \dots$ . El espacio  $\Omega$  esta muy relacionado al espacio de símbolos de la dinámica simbólica. En general la métrica usada es:

$$d(\mathbf{S}, \mathbf{S}') = \sum_{n=0}^{\infty} \frac{1}{2^n} d_H^n(\mathbf{S}, \mathbf{S}') \quad (5.3)$$

En la ecuación (5.3)  $d_H^n(\mathbf{S}, \mathbf{S}')$  es la distancia de Hamming (ecuación (2.1)) de la  $n$ -ésima copia. Al haber extendido el espacio a  $\Omega$  con la métrica (5.3), permite que el análisis del sistema sea completamente distinto.

A continuación daremos las definiciones de cuando una función  $F : \Omega \rightarrow \Omega$  es caótica, para la teoría del caos discreto<sup>[4]</sup>:

**Definición 5.1:**  $F$  tiene sensibilidad a las condiciones iniciales en  $A \subset \Omega$  si  $\exists n \in \mathbb{N} \ni \forall S \in A$  y  $\forall N_m(S) \exists S' \in N_m(S) \cap A$  y  $k \in \mathbb{N}$  tal que  $F^k(S') \notin N_n(F^k(S))$ .

**Definición 5.2:** Sea  $A \subset \Omega$  un conjunto cerrado e invariante  $F : \Omega \rightarrow \Omega$  es topológicamente transitiva en  $A \subset \Omega$  si para cualesquiera conjuntos abiertos  $U, V \subset A \exists n \in \mathbb{Z} \ni F^n(U) \cap V \neq \emptyset$ .

**Definición 5.3:** Sea  $A \subset \Omega$  un conjunto compacto,  $F : A \rightarrow A$  es caótica en  $A$  si  $F$  tiene sensibilidad a las condiciones iniciales y es topológicamente transitiva en  $A$ .

En los capítulos anteriores estudiamos sistemas en los cuales graficamos las distancias de Hamming del último estado generado por la dinámica ( $S_0$ ) de la red al estado inicial  $S'$ . Se observó en las gráficas que no existen regresos cercanos\*, al no existir regresos cercanos las redes neuronales no muestran indicios de la transitividad topológica. lo que nos indica que el comportamiento de la red no

---

\* Al referirnos a regresos cercanos nos referimos a estados  $S$  cuya distancia de Hamming al estado  $S'$  (estado inicial) sea mucho menor que  $N/4$ .  $d_H = 1, 2, 3$ .

puede ser visto como caótico. Desde un punto de vista del caos discreto. En este capítulo estudiaremos dos redes que cumplen con la definición de caos discreto. El primer modelo es estudiado por H Waelbroeck y F Zertuche<sup>[4]</sup>, el cual en principio presenta una dinámica caótica. llevaremos a cabo un estudio más detallado de este modelo. Por último se propone un modelo basado en la dinámica de Crisanti. El cual en principio presenta una dinámica caótica.

## FUNCIÓN DE CORRELACIÓN.

En los sistemas dinámicos de  $\mathbb{R}^N$  se requiere de calcular la dimensión fractal del atractor alrededor del cual evoluciona la red<sup>[4]</sup>. Se usará el método de Grassberger y Procaccia<sup>[13]</sup>, para el cual se define la función de correlación  $C(\rho)$  como el promedio del número de estados  $S_{t'}$  que tienen distancia menor que  $\rho$  a  $S_t$ ,  $S_t = F^t(S)$ . En  $\mathbb{R}^N$  el volumen de los hipercubos de radio  $\rho$  crece como  $\rho^N$ . Esto permite que para valores pequeños de  $\rho$  la dimensión de correlación  $D_a$  del atractor se asemeje al grado de crecimiento del promedio de puntos del atractor en el hipercubo de radio  $\rho$ , es decir  $C(\rho) \sim \rho^{D_a}$ . A partir de esta observación podemos calcular la dimensión del atractor como:

$$D_a = \lim_{\rho, \rho' \rightarrow 0} \left( \frac{\log(C(\rho)) - \log(C(\rho'))}{\log(\rho) - \log(\rho')} \right) \quad (5.4)$$

en donde  $C(\rho)$  es la función de correlación. A continuación calcularemos la dimensión de correlación en  $\Omega$ , para lo cual es necesario conocer el número de estados a  $S'$  cuya distancia sea menor que  $\rho$ . Como  $\Omega$  es homogéneo podemos tomar  $S' = 0$ , la función de correlación la podemos expresar como:

$$C(\rho) = \lim_{T \rightarrow \infty} \frac{1}{2^{NT}} \sum_{\{S\}} \Theta(\rho - d(S, 0)), \quad (5.5)$$

tomaremos  $T$  estados de  $\Omega$  y posteriormente tomaremos el límite cuando  $T \rightarrow \infty$ . La distancia estará dada por la ecuación (5.3).

La evaluación de esta suma y posteriormente del límite no es sencilla de resolver de manera directa\*. Lo primero que hay que hacer es expresar la función  $\theta(x)$  en términos de su transformada de Fourier una vez expresada se tiene que

---

\* En la referencia <sup>[4]</sup> se explica una forma de evaluar la ecuación (5.5).

## 56 Estudio del caos en sistemas dinámicos discretos

evaluar la expresión de la suma sobre  $\{S\}$ , y por último se hace uso de los métodos de variable compleja para poder expresar la función de correlación como:

$$C(\rho) = \frac{1}{2^N N!} \sum_{k=0}^{[\rho/2]} (-1)^k \binom{N}{k} (\rho - 2k)^N \quad (5.6)$$

La forma de la función de correlación no es una potencia de  $\rho$  como en  $\mathbb{R}^N$ , pero en realidad nosotros estamos interesados en calcular  $C(\rho)$  para  $\rho \rightarrow 0$ , podemos simplificar la ecuación (5.6) para  $\rho \leq 2$  [4].

$$C(\rho) = \frac{1}{2^N N!} \rho^N \quad (5.7)$$

Permitiéndonos usar las mismas técnicas en  $\Omega$  que en el caso de  $\mathbb{R}^N$ , la ecuación (5.4) nos servirá para calcular la dimensión del atractor.

### MODELO DE CAOS DISCRETO.

A continuación definiremos el modelo con el cual trabajaremos en esta parte del capítulo. Este modelo es una red neuronal con memoria la cual en principio presenta una dinámica caótica[4]. La evolución de la red se realiza en paralelo y la regla de evolución de la red está dada por:

$$S_i(t+1) = \text{sgn} \left( \sum_{n=0}^{T-1} \frac{1}{2^n} \sum_{j=1}^N w_{ij}^{(n)} S_j(t-n) \right) \quad (5.8)$$

En este caso la primera suma se refiere a la memoria de la red, en principio la red tiene un número infinito de estados en la memoria, pero en la práctica eso no es posible. La matriz de pesos está dada por:

$$w_{ij}^{(n)} = \frac{1}{N} \sum_{\mu=1}^P \xi_i^{\mu+n+1} \xi_j^{\mu} \Theta(S_{\mu}^2 - \eta^2) \quad (5.9)$$

En donde:

$$S_{\mu} = \frac{1}{N} \sum_{i=1}^N S_i \xi_i^{\mu} \quad (5.10)$$

es la correlación del estado de la red con el patrón  $\xi_i^{\mu}$ . Al igual que en los modelos anteriores  $\xi_i^{\mu} = \pm 1$ , con  $\mu = 1, \dots, P + T$  son generados de manera aleatoria y tiene la misma probabilidad de tomar los valores  $\pm 1$ . Como se puede observar se ha introducido un umbral ( $\eta$ ) en la dinámica de la red. Este umbral se introduce en las redes neuronales para incrementar la capacidad\* de la red neuronal<sup>[11,14]</sup>, esto se logra eliminando de la sinapsis aquellos patrones que tengan una correlación con el estado de la red menor que  $\eta$ . La función  $\Theta(x)$  está definida en la ecuación (1.2).

## RESULTADOS DE LAS SIMULACIONES.

En esta sección presentaremos los resultados de la simulación de una red neuronal con memoria. La dinámica de la red está dada por la ecuación (5.8). La actualización es en paralelo. La simulación se realizó con  $N = 49$ ,  $P = 19$  y  $\eta = 0$ . Como se mencionó anteriormente en la práctica la memoria de la red no es infinita, en este caso usaremos una red neuronal con treinta estados de memoria† ( $T = 30$ ).

Primero graficaremos los ciclos de la red.  $d_h$  vs # de regresos. Recordemos que si la red presenta regresos cercanos en la distancia de Hamming la red puede presentar un comportamiento caótico, ya que la existencia de los regresos cercanos nos indica la transitividad topológica de la red. Después estudiaremos la dimensión del atractor, para lo cual tendremos que realizar el cálculo de la función de correlación.

---

\* Se busca que el valor de  $\alpha_c$  se incremente.

† En la práctica el número de estados en la memoria de la red depende de las limitaciones de computo.

## CÁLCULO DE LA FUNCIÓN DE CORRELACIÓN Y DISTANCIAS DE HAMMING.

A continuación calcularemos los regresos cercanos de Hamming y la función de correlación. Es decir que calcularemos los regresos cercanos y la función de correlación de  $S'(t_0 + l)$  con todos los estados, la  $l = 1, 2, \dots, L$ , en donde  $L$  es el número máximo de iteraciones.

En esta sección la red tendrá como parámetros  $N = 49$ ,  $P = 19$ ,  $T = 30$  y  $\eta = 0$ , los estados que analizaremos son estados de la red que ya pasaron el transitorio.

En la primera gráfica la red tiene mil iteraciones. La red genero alrededor de mil estados a los cuales calculamos las distancias de Hamming totales (*fig.5.1*).

Como se puede observar los regresos a distancia de hamming cero existen, a continuación llevaremos a cabo el análisis de regresión de los datos de la función de correlación (*fig5.2*). Para el análisis de la función de correlación primero se llevo a cabo una transformación de los datos para que estos se ajustarán a una recta, lo que nos llevó a realizar un análisis de regresión lineal de los datos. El haber transformado los datos nos facilita el análisis de regresión y además nos permite llevar a cabo una comparación de los datos con otro tipo de ajustes (el ajuste exponencial y el ajuste potencia). Todos los ajustes que realizamos son con respecto a los datos transformados de la función de correlación.

El número de datos para la regresión es: 20.

La regresión se realizó para valores de  $\rho \in [0.013, 0.4]$ .

### Resultados de la regresión lineal.

El modelo es  $y(x) = m * x + b$ .

Pendiente  $m = 1.984495$ .

Ordenada  $b = 2.260799$ .

Coefficiente de correlación  $r = 0.987548$ .

Dimensión del atractor  $D_a = 1.984495$ .

Constante  $e^b = 9.590749$ .

### Resultados de la regresión exponencial.

El modelo es  $\ln(y) = m * x + \ln(b)$ .

Pendiente  $m = 0.236355$ .

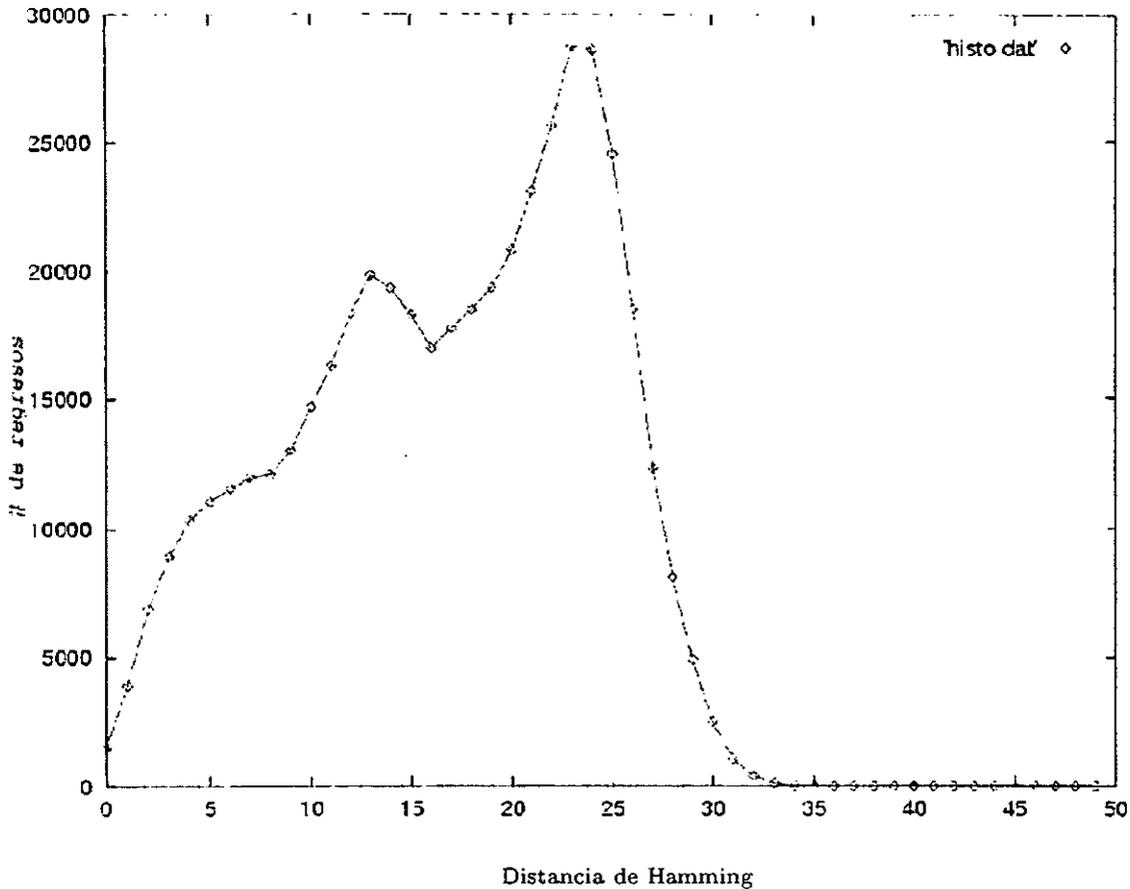


fig 5.1. Histograma de las distancias de Hamming totales de una red neuronal con  $N = 49$ ,  $P = 19$ ,  $T = 30$  y  $\eta = 0$ . En esta gráfica se puede observar que existen estados cuya distancia de Hamming es muy pequeña o es cero, esto es un indicativo de la presencia de transitividad topológica. En este caso a pesar de que algunos estados tienen distancia de Hamming cero no hemos alcanzado un ciclo ya que la distancia real esta dada por (5.3), la cual no es cero.

Ordenada  $\ln(b) = 1.366751$ .

Coefficiente de correlación  $r = 0.966982$ .

Resultados de la regresión potencia.

El modelo es  $\ln(y) = m * \ln(x) + \ln(b)$ .

pendiente  $m = 0.766312$ .

Ordenada  $\ln(b) = 1.268869$ .

Coefficiente de correlación  $r = 0.982709$ .

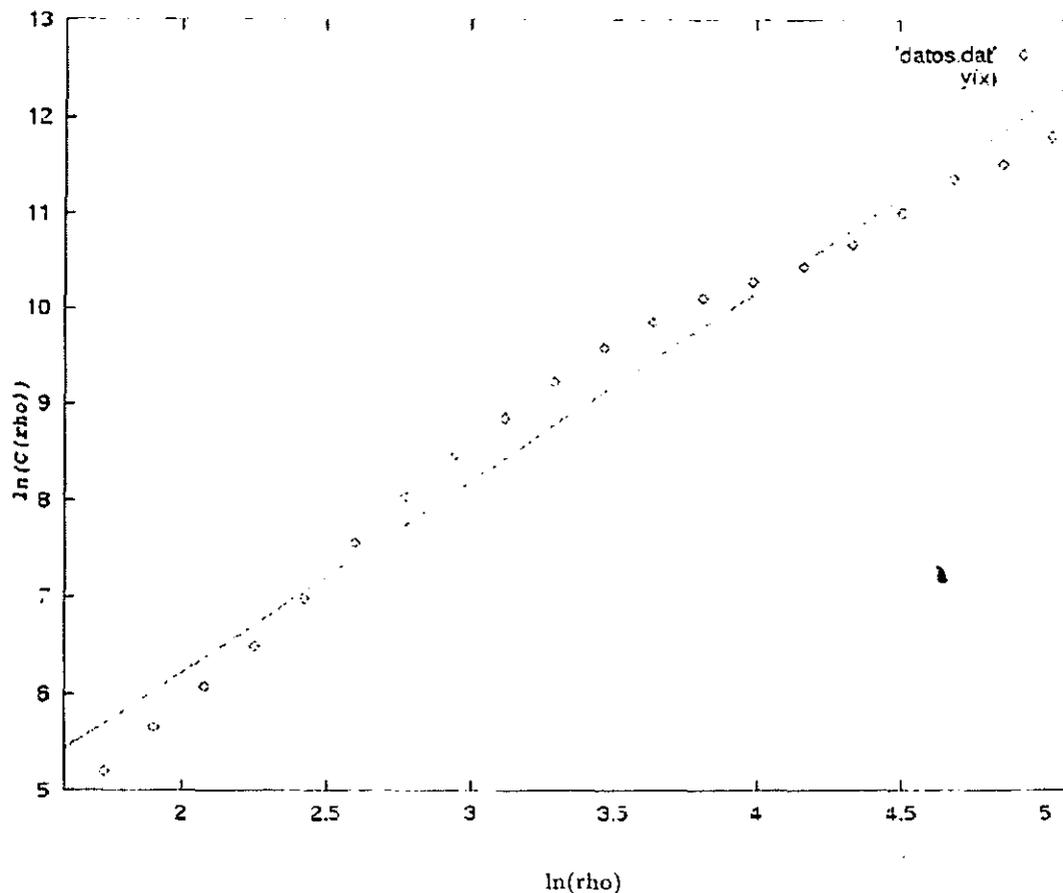


fig 5.2. Gráfica de la función de correlación  $\ln(C(\rho))$  y la curva ajustada  $y(x) = m * x + b$ . En esta gráfica se observa el análisis de regresión realizado a los datos de la función de correlación, este análisis se realiza para poder encontrar la dimensión del atractor alrededor del cual evoluciona la red, la pendiente de la recta ajustada nos indica la dimensión del atractor, en este caso el ajuste es regular, ya que el coeficiente de correlación de 0.98.

En el análisis de regresión lineal podemos observar que la dimensión del atractor es  $D_a = 1.98$ . con un buen coeficiente de correlación. En estos análisis de regresión los resultados son bastante confiables, ya que estos cuentan con los suficientes datos, propiciando que la estadística sea muy confiable.

En la siguiente gráfica la red tendrá  $10^4$  datos para realizar los análisis, la gráfica de las distancias de Hamming es fig5.3.

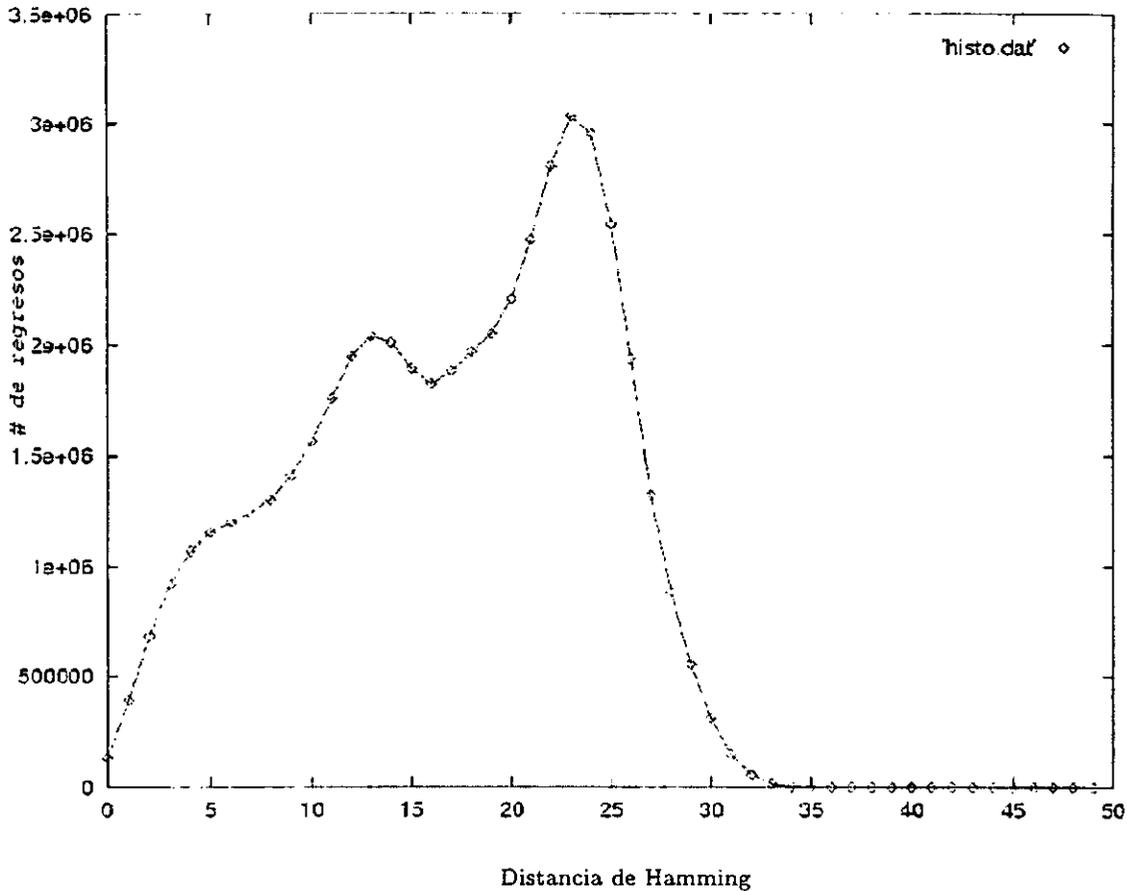


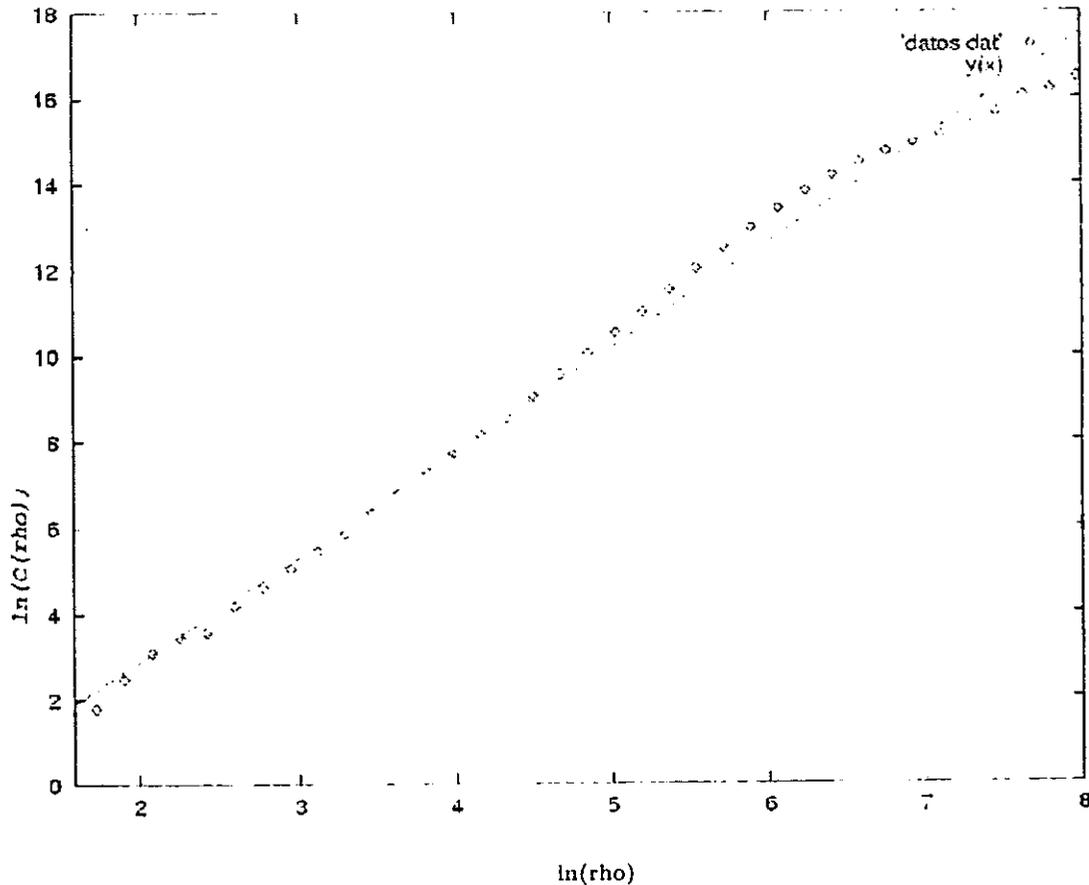
fig 5.3 Histograma de las distancias de Hamming totales de una red neuronal con  $N = 49$ ,  $P = 19$ ,  $T = 30$  y  $\eta = 0$ . Esta gráfica consta de  $10^4$  iteraciones. En esta gráfica se puede observar que existen estados cuya distancia de Hamming es muy pequeña o es cero, esto es un indicativo de la presencia de transitividad topológica. En este caso a pesar de que algunos estados tienen distancia de Hamming cero no hemos alcanzado un ciclo ya que la distancia real esta dada por (5.3), la cual no es cero, se puede observar que al incrementar el número de iteraciones el número de estados con distancia de Hamming cero aumentó. El comportamiento de la red es básicamente el mismo, lo que indica que la red mantiene un comportamiento muy complejo, el que la red se comporte cualitativamente igual es que la red se encuentra evolucionando alrededor de un atractor, lo cual impide que las trayectorias evolucionen en todo el espacio fase.

## 62 Estudio del caos en sistemas dinámicos discretos

El análisis de la función de correlación para la gráfica *fig5.3*, lo podemos observar en la gráfica *fig5.4*:

El número de datos para la regresión es: 37.

La regresión se realizó para valores de  $\rho \in [0.0007, 0.4]$ .



*fig 5.4. Gráfica de la función de correlación  $\ln(C(\rho))$  y la curva ajustada  $y(x) = m * x + b$ . Nuevamente podemos observar el ajuste de los datos de la función de correlación, en esta gráfica el ajuste es mejor y los datos se asemejan a una recta el coeficiente de correlación tiene un valor de 0.99 el cual es bueno, en esta gráfica podemos observar que la red sufre una reducción en la dimensión del atractor, la reducción del atractor es una característica de los sistemas dinámicos caóticos.*

## Resultados de la regresión lineal.

El modelo es  $y(x) = m * x + b$ .

Pendiente  $m = 2.441778$ .

Ordenada  $b = -2.003325$ .

Coefficiente de correlación  $r = 0.996293$ .

Dimensión del atractor  $D_a = 2.441778$ .

Constante  $e^b = 0.134886$ .

## Resultados de la regresión exponencial.

El modelo es  $\ln(y) = m * x + \ln(b)$ .

Pendiente  $m = 0.302743$ .

Ordenada  $\ln(b) = 0.673762$ .

Coefficiente de correlación  $r = 0.952658$ .

## Resultados de la regresión potencia.

El modelo es  $\ln(y) = m * \ln(x) + \ln(b)$ .

pendiente  $m = 1.354063$ .

Ordenada  $\ln(b) = 0.118999$ .

Coefficiente de correlación  $r = 0.992611$ .

El coeficiente de correlación ha mejorado y se nota una reducción en la dimensión del atractor, lo cual es un indicativo de los sistemas dinámicos caóticos. Por último presentaremos los resultados de los análisis para una red con 34,435 estados. que es equivalente a tener cien mil iteraciones en la red (*fig5.5*). Recordemos que de las cien mil iteraciones hay que eliminar aquellos estados que se encuentren en el transitorio, el comportamiento de la red en general es muy semejante a la gráfica anterior, el que la red se comporte caulitativamente igual nos permite realizar un análisis más confiable.

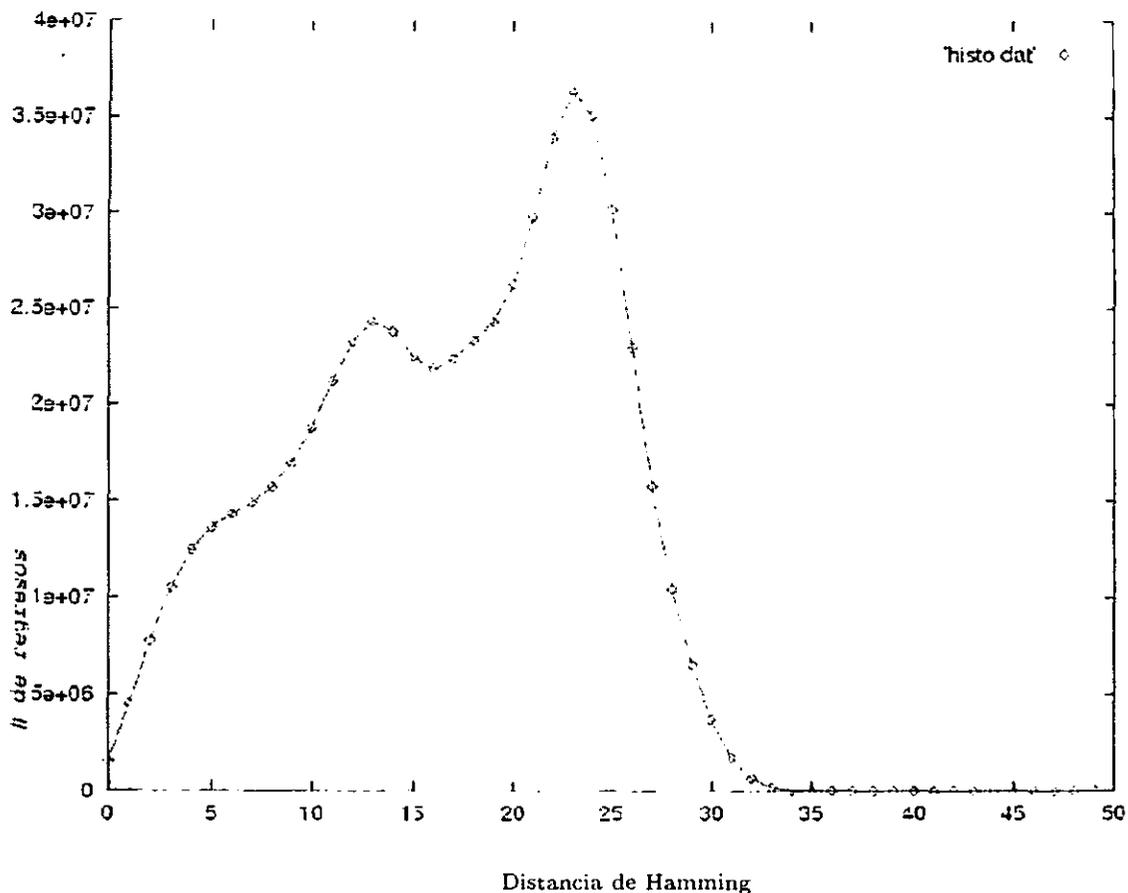


fig 5.5. Histograma de las distancias de Hamming totales de una red neuronal con  $N = 49$ ,  $P = 19$ ,  $T = 30$  y  $\eta = 0$ . Esta gráfica consta de  $10^5$  iteraciones. En esta gráfica se puede observar que existen estados cuya distancia de Hamming es muy pequeña o es cero, esto es un indicativo de la presencia de transitividad topológica. En este caso a pesar de que algunos estados tienen distancia de Hamming cero no hemos alcanzado un ciclo ya que la distancia real esta dada por (5.3), la cual no es cero, se puede observar que al incrementar el número de iteraciones el número de estados con distancia de Hamming cero aumentó y el comportamiento de la red es básicamente el mismo.

Por último presentaremos el análisis de regresión para el cálculo de la función de correlación de la gráfica fig5.5. (fig5.6):

El número de datos para la regresión es: 45.

La regresión se realizó para valores de  $\rho \in [0.0005, 0.4]$ .

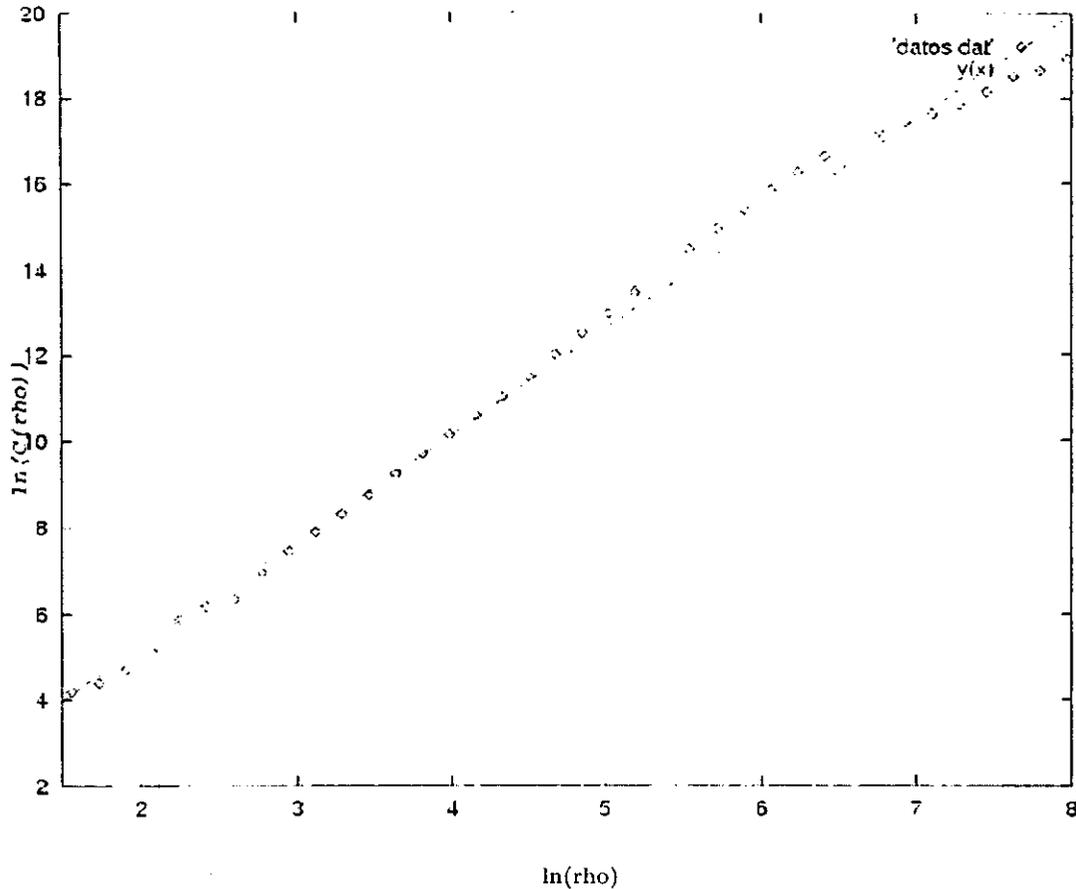


fig 5.6. Gráfica de la función de correlación  $\ln(C(\rho))$  y la curva ajustada  $y(x) = m * x + b$ . Nuevamente podemos observar el ajuste de los datos de la función de correlación, en esta gráfica el ajuste es mejor y los datos se asemejan a una recta el coeficiente de correlación tiene un valor de 0.99 el cual es bueno, en esta gráfica podemos observar que la red sufre una reducción en la dimensión del atractor. la reducción del atractor es una característica de los sistemas dinámicos caóticos.

#### Resultados de la regresión lineal.

El modelo es  $y(x) = m * x + b$ .

Pendiente  $m = 2.465910$ .

Ordenada  $b = 0.304496$ .

Coefficiente de correlación  $r = 0.996592$ .

Dimensión del atractor  $D_a = 2.465910$ .

## 66 Estudio del caos en sistemas dinámicos discretos

Constante  $e^b = 1.355941$ .

Resultados de la regresión exponencial.

El modelo es  $\ln(y) = m * x + \ln(b)$ .

Pendiente  $m = 0.232377$ .

Ordenada  $\ln(b) = 1.289567$ .

Coefficiente de correlación  $r = 0.971740$ .

Resultados de la regresión potencia.

El modelo es  $\ln(y) = m * \ln(x) + \ln(b)$ .

pendiente  $m = 0.991928$ .

Ordenada  $\ln(b) = 0.941889$ .

Coefficiente de correlación  $r = 0.996245$ .

Los resultados obtenidos en esta sección son muy similares a los encontrados en la teoría del caos discreto<sup>[4]</sup>. En los resultados del caos discreto se observa una reducción de la dimensión del atractor. La dimensión del atractor se encuentra alrededor de dos. Además la red muestra indicios de transitividad topológica. todos esos resultados son consistentes con los resultados que nosotros hemos obtenido. Obviamente hace falta un análisis a la sensibilidad a las condiciones iniciales (definición 5.1), para poder determinar si la red puede presentar caos. Antes de realizar el análisis a las condiciones iniciales repetiremos el cálculo de la dimensión del atractor pero con una nueva distancia.

### CÁLCULO DE LA FUNCIÓN DE CORRELACIÓN Y DISTANCIAS DE HAMMING CON DISTANCIA BETA.

A continuación definiremos una nueva distancia. con la cual repetiremos los cálculos de la dimensión del atractor. Al repetir estos cálculos nos permitirá establecer que tanto dependen nuestros resultados anteriores de la métrica del espacio.

$$d_{\beta}(S, S') = \beta^n, \quad 0 < \beta < 1 \quad (5.11)$$

Recordemos que  $S$  es una matriz de  $T \times N$  y en la ecuación (5.11) la  $n = 0, \dots, T - 1$ . Donde  $S(k) = S'(k) \forall k < n$  y  $S(k) \neq S'(k)$ . El valor de  $n$

estará determinado por el primer bit distinto entre  $S$  y  $S'$ . este bit se encontrará en el renglón  $k$ -ésimo (i.e. el bit distinto se encontrara en  $S(k)$ ).

La función de correlación nuevamente la calculamos como el promedio del número de estados cuya distancia entre  $S'$  y  $S_0$  es menor que  $\rho$ :

$$C(\rho) = \lim_{T \rightarrow \infty} \frac{1}{2^{NT}} \sum_{\{S\}} \Theta(\rho - d_{\beta}(S, S')) \quad (5.12)$$

En la ecuación (5.12) hemos tomado  $T$  estados del espacio, al aplicar el límite cuando  $T \rightarrow \infty$  recuperamos el espacio total, la suma es sobre todos los posibles estados. Para la evaluación de la función de correlación supondremos que  $\rho = \beta^n$ , entonces la suma sobre todas las configuraciones la podemos calcular directamente:

$$\begin{aligned} C(\rho) &= \lim_{T \rightarrow \infty} \frac{1}{2^{NT}} (2^{(T-n)N}) \\ C(\rho) &= 2^{-nN} \end{aligned} \quad (5.13)$$

Por otro lado tenemos que:

$$\begin{aligned} \rho &= \beta^n \\ \ln(\rho) &= n \ln(\beta) \\ \frac{\ln(\rho)}{\ln(\beta)} &= n \end{aligned} \quad (5.14)$$

Substituyendo el valor de  $n$  en la ecuación (5.13) obtenemos:

$$\begin{aligned} \ln(C(\rho)) &= -nN \ln(2) \\ \ln(C(\rho)) &= -\frac{\ln(\rho)}{\ln(\beta)} N \ln(2) \\ \ln(C(\rho)) &= -\frac{\ln(2)}{\ln(\beta)} N \ln(\rho) \\ C(\rho) &= \rho^{-\frac{\ln(2)}{\ln(\beta)} N} \end{aligned} \quad (5.15)$$

En el caso particular de  $\beta = 1/2$  tenemos:

$$\ln(\beta) = -\ln(2)$$

Sustituyendo el valor de  $\ln(\beta)$  en la ecuación (5.15) obtenemos:

$$C(\rho) = \rho^N \quad (5.16)$$

Como vemos la función de correlación la podemos expresar como una potencia esto nos permite usar las mismas técnicas que en la sección anterior.

En nuestras simulaciones el valor de  $\beta$  siempre será  $\beta = 1/2$ . El número de neuronas en la red será  $N = 49$ . El número de patrones será  $P = 19$  y tendremos  $T = 30$  estados en la memoria de la red. El valor del umbral  $\eta = 0$ .

Los resultados del cálculo de la dimensión de atractor se realizó para  $10^4$  estados y para 34,435 estados. Primero mostraremos los resultados de  $10^4$  estados, en la gráfica (fig5.7) se puede observar graficamente la curva de ajuste y los datos:

El número de datos para la regresión es: 9.

La regresión se realizó para valores de  $\rho \in [0.002, 1.0]$ .

Resultados de la regresión lineal.

El modelo es  $y(x) = m * x + b$ .

Pendiente  $m = 2.141845$ .

Ordenada  $b = -0.132287$ .

Coefficiente de correlación  $r = 0.993097$ .

Dimensión del atractor  $D_a = 2.141845$ .

Constante  $e^b = 0.876089$ .

Resultados de la regresión exponencial.

El modelo es  $\ln(y) = m * x + \ln(b)$ .

Pendiente  $m = 0.494285$ .

Ordenada  $\ln(b) = 0.062461$ .

Coefficiente de correlación  $r = 0.967547$ .

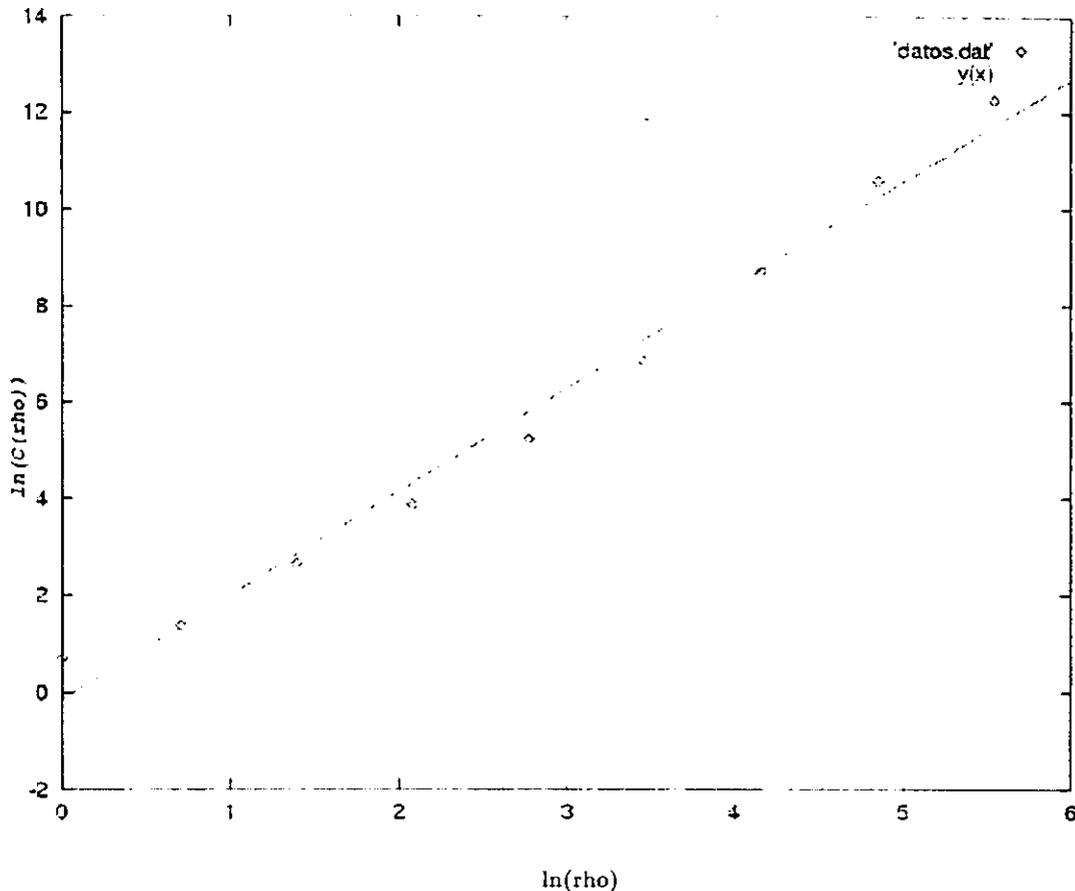


fig 5.7. Gráfica de la función de correlación  $\ln(C(\rho))$  y la curva ajustada  $y(x) = m * x + b$ . En esta gráfica se puede observar el análisis de regresión realizado a la función de correlación, a diferencia de las gráficas (fig 5.2, fig 5.4, fig 5.6) la función de correlación fue calculada con la distancia (5.11), en esta gráfica tenemos  $10^4$  iteraciones, a pesar de haber cambiado la distancia el análisis de regresión nos indica una reducción en la dimensión del atractor muy semejante a la encontrada para la distancia (5.3).

El análisis de regresión potencia no se puede llevar a cabo debido a que  $x$  toma el valor de cero, provocando que  $\ln(x)$  tienda a infinito. Los siguientes resultados son de una red con 34.435 estados (fig 5.8).

El número de datos para la regresión es: 9.

La regresión se realizó para valores de  $\rho \in [0.002, 1.0]$ .

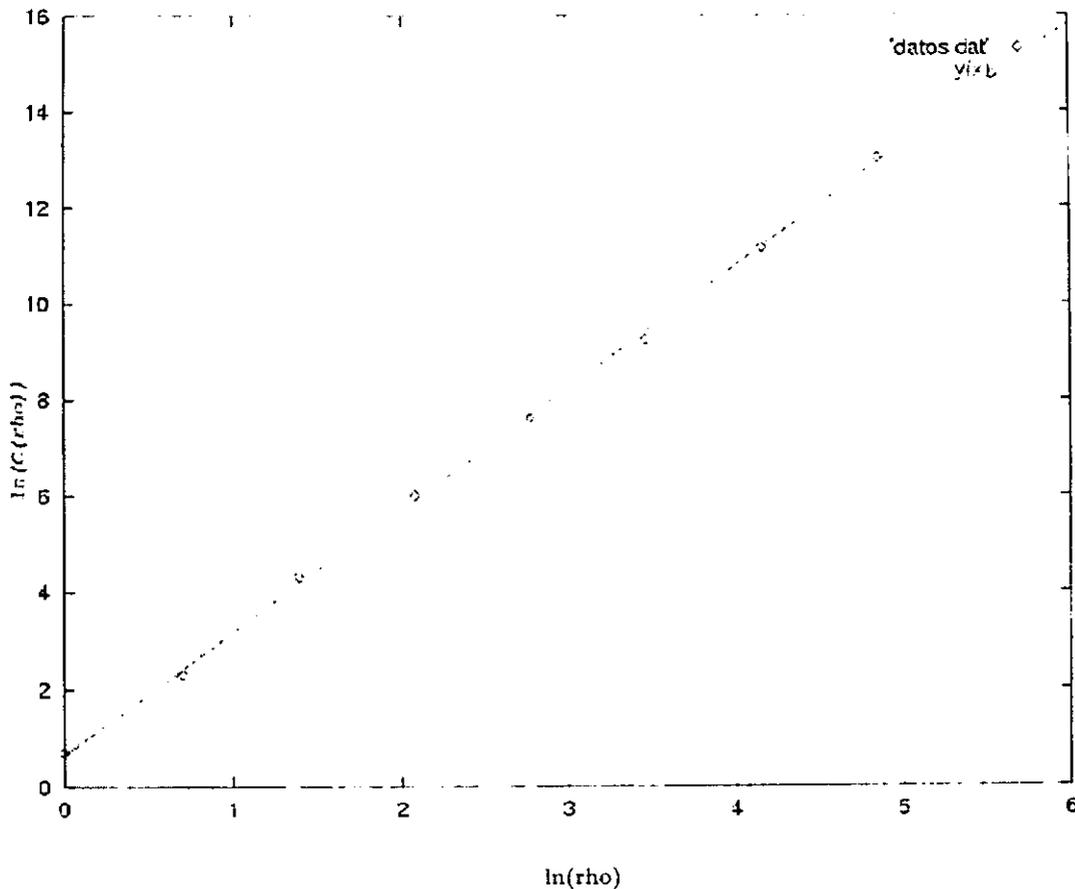


fig 5.8. Gráfica de la función de correlación  $\ln(C(\rho))$  y la curva ajustada  $y(x) = m * x + b$ . En esta gráfica se puede observar el análisis de regresión realizado a la función de correlación calculado con la distancia (5.11), en esta gráfica tenemos  $10^5$  iteraciones, nuevamente el análisis de regresión nos indica una reducción en la dimensión del atractor muy semejante a la encontrada para la distancia (5.3).

#### Resultados de la regresión lineal.

El modelo es  $y(x) = m * x + b$ .

Pendiente  $m = 2.530869$ .

Ordenada  $b = 0.664503$ .

Coefficiente de correlación  $r = 0.999746$ .

Dimensión del atractor  $D_a = 2.530869$ .

Constante  $e^b = 1.943524$ .

Resultados de la regresión exponencial.

El modelo es  $\ln(y) = m * x + \ln(b)$ .

Pendiente  $m = 0.474905$ .

Ordenada  $\ln(b) = 0.422436$ .

Coefficiente de correlación  $r = 0.919192$ .

ANÁLISIS DE LA SENSIBILIDAD DE LAS CONDICIONES INICIALES.

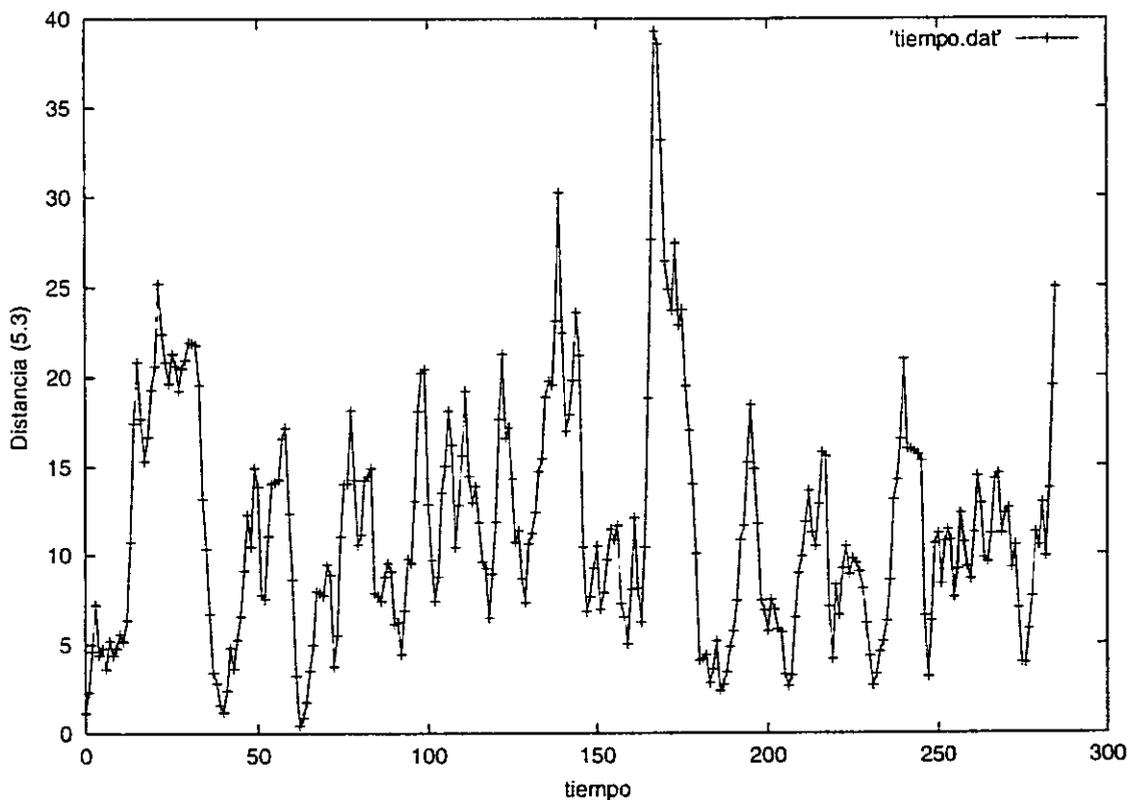
<i>estado</i>	$d_H$	$d$
0	0	0.000000
1689	0	6.584806
2801	0	3.117676
5989	0	4.216129
7005	0	4.788762
7426	0	3.360330
7718	0	5.367251
8616	0	4.369336
10794	0	3.140106
11086	0	2.262836
12725	0	4.725931
13601	0	5.712282
15364	0	1.721301
15684	0	1.149989
16453	0	1.834846
16554	0	2.354884
18423	0	4.370901
19995	0	2.950980
21348	0	2.205613
27225	0	1.325576

fig 5.9. En este cuadro podemos observar los estados que tienen una menor distancia al estado inicial. En la primera columna se observa el número de estado. En la segunda columna se observa la distancia de Hamming entre el renglón  $S(0)$  de  $S$  y le renglón  $S'(0)$  de  $S'$ . En la tercer columna se observa la distancia (5.3) al estado inicial. A partir de este cuadro podemos ver que estados se encuentran más cercanos al estado inicial, esto lo realizamos con el objetivo de realizar un análisis de la sensibilidad a las condiciones iniciales (definición 5.1).

## 72 Estudio del caos en sistemas dinámicos discretos

Los resultados del estudio con la distancia beta son consistentes con los resultados obtenidos a partir de la distancia definida por (5.3), por tal motivo llevaremos acabo el análisis a la sensibilidad de las condiciones iniciales. Primero tendremos que establecer los estados que tengan la menor distancia posible\* con  $S'$ , que es nuestro estado fijo.

Los cálculos se realizarán para una red con  $10^6$  iteraciones. La tabla nos indicará los estados que tienen distancias de Hamming cero con respecto a  $S'$  en el primer renglón. Además nos mostrará la distancia (5.3) que hay entre ellos.



*fig 5.10. Comparación de las trayectorias entre  $S_0$  y  $S'$ . En esta gráfica podemos observar la distancia entre las trayectorias de  $S_0$  y de  $S'$ . El estado  $S'$  es el estado con menor distancia a  $S_0$ , estos estados se encuentran cercanos y al cabo de unas cuantas iteraciones sus trayectorias se separan considerablemente, esto nos sugiere que la red puede presentar sensibilidad a las condiciones iniciales.*

---

\* La distancia usada es la ecuación (5.3).

El estado que tiene una menor distancia es el estado número 15,684, al cual llamaremos  $S_0$ . Tiene una distancia de Hamming igual a cero y además la distancia que hay entre  $S'$  y  $S_0$  es la menor del total de los estados\*.

Ahora llevaremos a cabo la comparación de trayectorias entre  $S_0$  y  $S'$ . Hemos elegido a  $S_0$  debido a que en nuestra simulación es el estado más cercano a  $S'$ , la gráfica (fig5.10) nos mostrará como varía la distancia en la trayectorias a partir de  $S_0$  y  $S'$ , se graficaron 285 iteraciones.

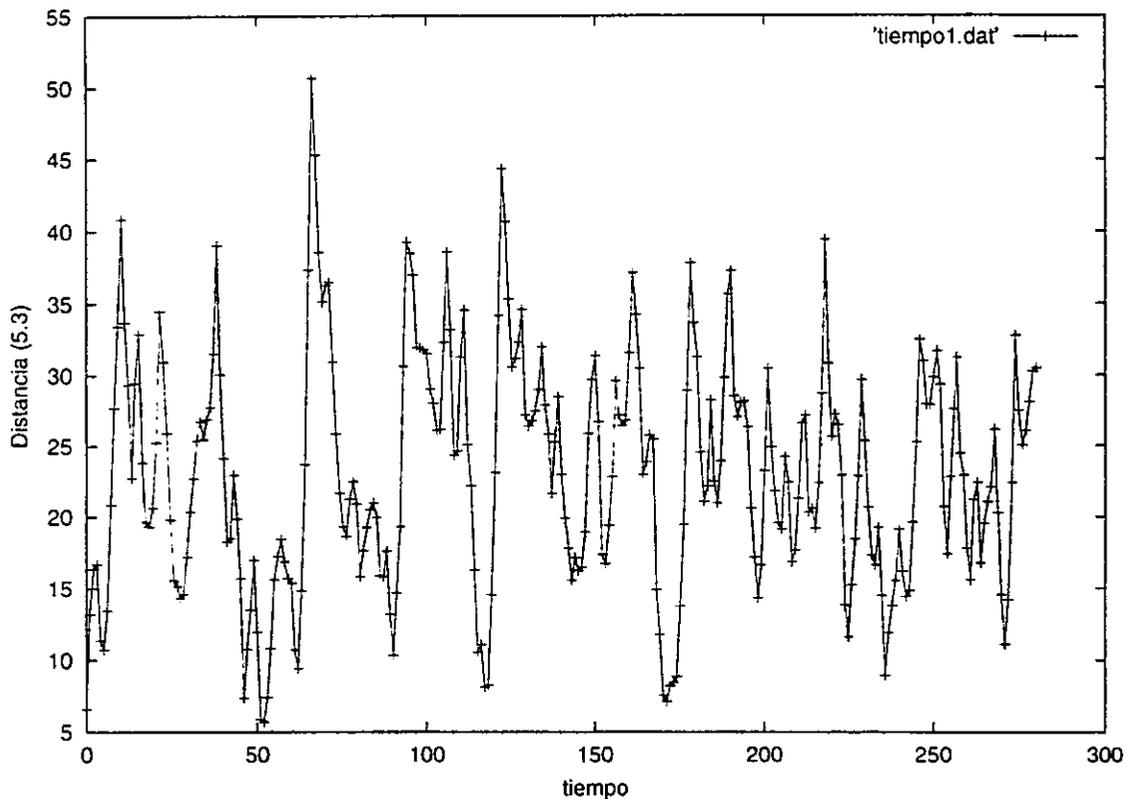


fig 5.11. Comparación de las trayectorias entre  $S_0$  y  $S'$ . En esta gráfica hemos podemos observar la distancia entre las trayectorias de  $S_0$  y de  $S'$ . En este caso el estado  $S'$  es un estado que es cercano a  $S_0$  y nuevamente al cabo de unas cuantas iteraciones sus trayectorias se separan considerablemente, esto nos sugiere que la red sí presenta sensibilidad a las condiciones iniciales. Comparamos dos estados para asegurarnos que los cálculos realizados no dependen únicamente del estado elegido.

\* Es la menor distancia entre  $S'$  y los estados de la evolución de la red.

En la gráfica (fig5.11) compararemos las trayectorias entre un nuevo  $S_0 = 1,689$  y  $S'$ . Como podemos observar la gráfica es muy similar a la anterior esto nos indica que para estados relativamente cercanos la red sí tiene sensibilidad a las condiciones iniciales.

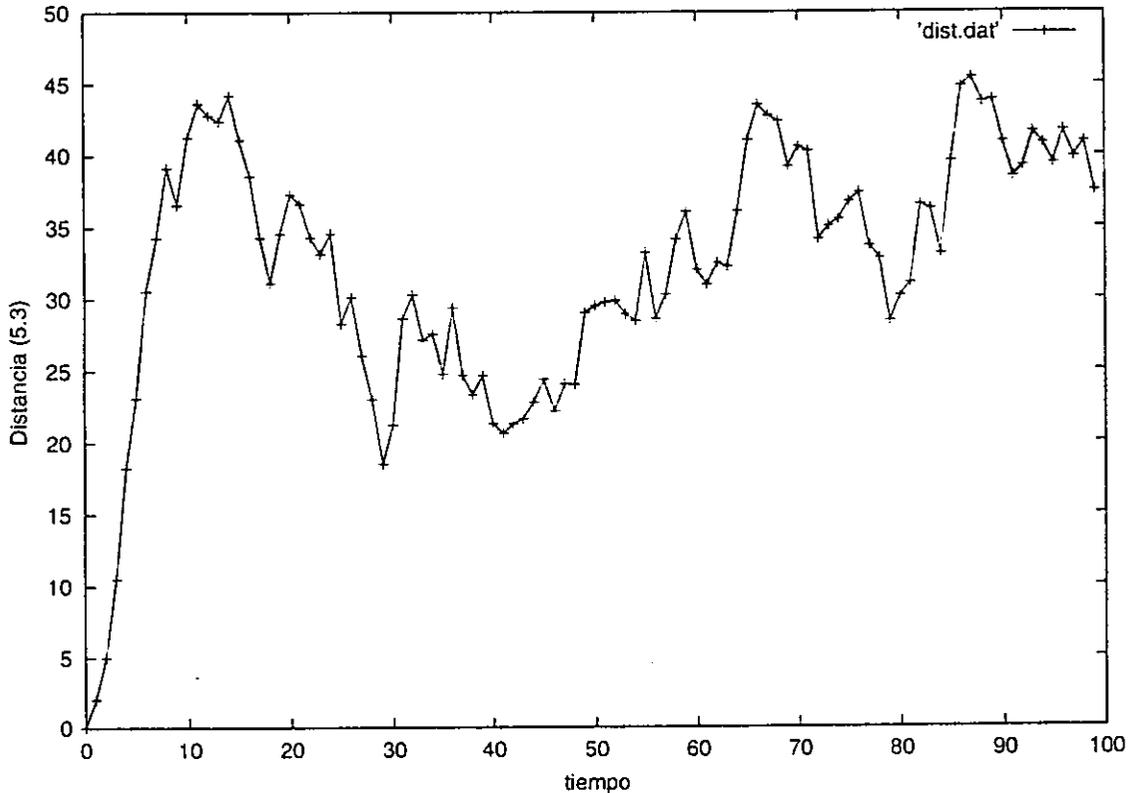


fig 5.12. Comparación de las trayectorias entre  $S_0$  y  $S'$ . En este caso hemos generado un estado muy cercano a nuestro estado inicial para poder estudiar si realmente estos estados tan cercanos muestran sensibilidad a las condiciones iniciales. En la gráfica observamos la distancia entre las trayectorias de los dos estados y se puede observar que la red sí presenta sensibilidad a las condiciones iniciales.

Las gráficas anteriores muestran indicios que existe sensibilidad a las condiciones iniciales, sin embargo podemos observar que en ambos casos la distancia entre  $S_0$  y  $S'$  no es tan pequeña como nosotros quisieramos. Hemos generado un estado  $S'$  y un estado  $S_0$  lo suficientemente cercanos, con una distancia menor que  $10^{-7}$ . En esta simulación graficaremos la distancia entre las trayectorias de  $S_0$  y de  $S'$ . La simulación se llevo a cabo con 100 iteraciones. Los nuevos estados

fueron generados de forma aleatoria y la diferencia se encuentra en el último renglón de los estados que también es generado de manera aleatoria. La gráfica (fig 5.12) nos mostrará la diferencia en las trayectorias.

En la gráfica (fig5.12) la distancia inicial entre  $S_0$  y  $S'$  era de :  $3.7252903 \times 10^{-8} < 10^{-7}$  . Como se observa en la gráfica la red es altamente sensible a las condiciones iniciales. Por último presentaremos una simulación con 300 iteraciones y con una distancia inicial muy semejante. Como se puede observar nuevamente la red presenta sensibilidad a las condiciones iniciales (fig5.13).

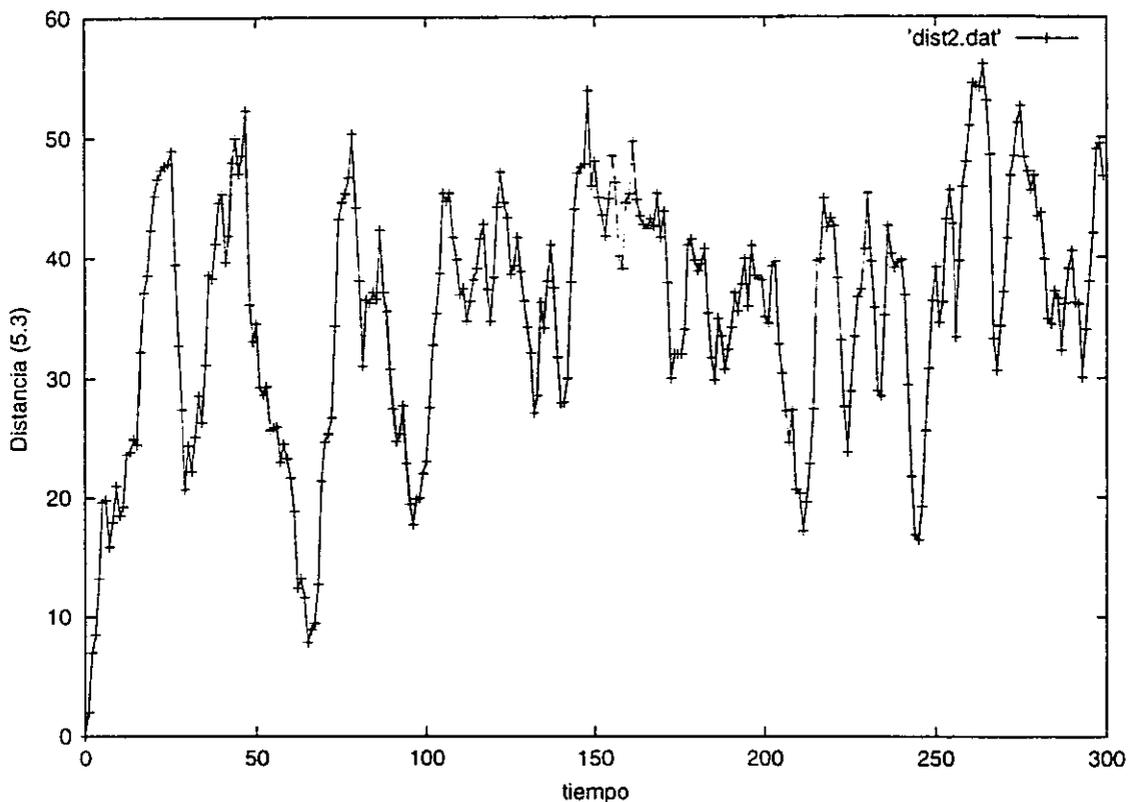


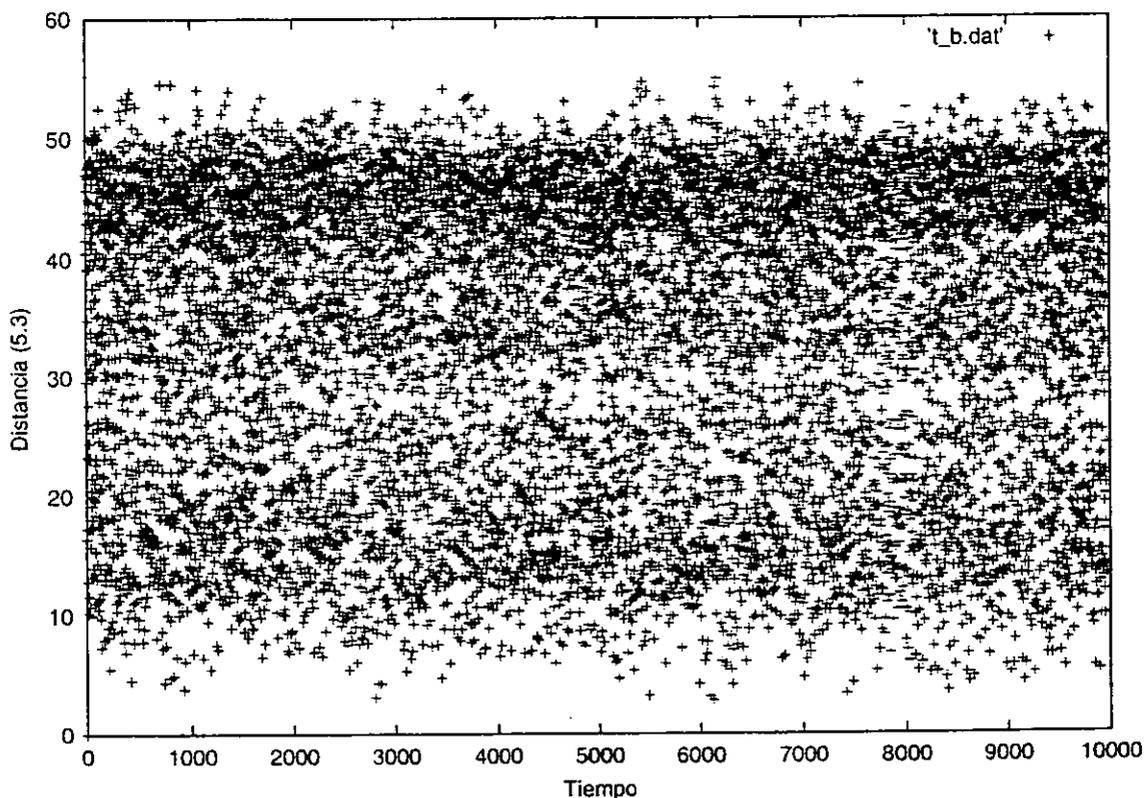
fig 5.13. Comparación de las trayectorias entre  $S_0$  y  $S'$ . Nuevamente tenemos dos estados muy cercanos y se observa que al cabo de algunas iteraciones las trayectorias se separan considerablemente, por lo que podemos afirmar que la red sí presenta sensibilidad a las condiciones iniciales.

En la gráfica observamos que la distancia entre los dos estados se incrementa bruscamente y posteriormente se mantiene en un rango de distancia mayor a diez, esto para las dos gráficas (fig5.12 y fig 5.13). Una de las razones para realizar dos

gráficas como las anteriores es para asegurarnos que los resultados no dependen del estado que elegimos. con lo que pueden ser generalizados a cualquier estado  $S$  dentro del espacio fase.

En este momento ya podemos decir que la red sí presenta una clara evidencia de sensibilidad a las condiciones iniciales, de hecho es altamente sensible. Tomamos dos puntos muy cercanos y después de unas cuantas iteraciones estos se encuentran muy lejanos, con esto hemos terminado nuestro análisis de la sensibilidad a las condiciones iniciales.

Para concluir el análisis de este modelo estudiaremos la trayectoria de  $S'$  para lo cual graficaremos la distribución de la distancia (5.3) y la distancia de Hamming. Graficaremos las distancias de Hamming (*fig5.15*) y la distancia (5.3) contra el tiempo (*fig5.14*), lo cual nos servirá para visualizar como están distribuidas las distancias.



*fig 5.14. Distribución de la distancia (5.3). En esta gráfica podemos observar la distribución de la distancia (5.3), se puede ver hay una zona comprendida entre distancia 40 y 50 que tiene una mayor incidencia.*

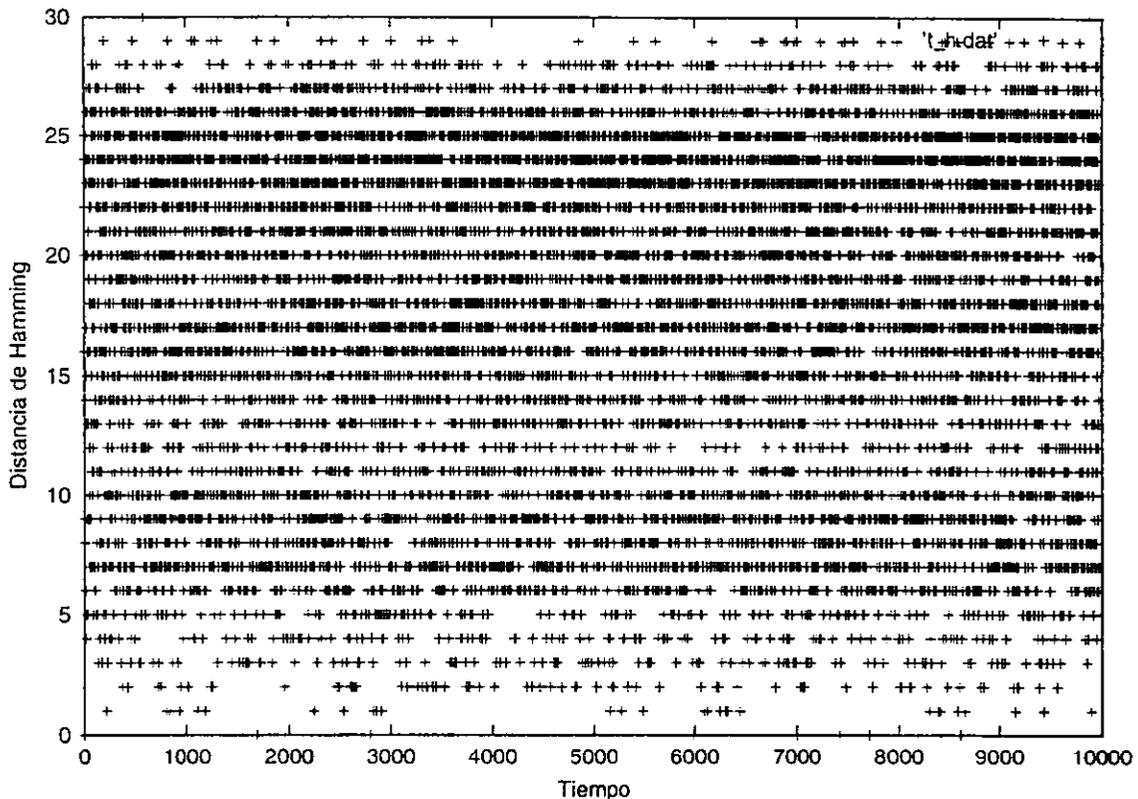


fig 5.15. Distribución de la distancia de Hamming, en esta gráfica podemos observar como se distribuye la distancia de Hamming. La cual se distribuye más homogéneamente que la distancia (5.3). En esta gráfica no se puede identificar una zona con un número mayor de regresos.

Comparando las dos gráficas vemos que la distancia binaria tiende a tener un mayor grado de incidencia para distancias entre 40 y 50, a diferencia de la distancia de Hamming que no tiene una distancia dominante.

Con esto hemos terminado el estudio de este modelo y en conclusión podemos decir que el modelo presenta evidencia de sensibilidad a las condiciones iniciales y además hay evidencia de que es topológicamente transitivo. Podemos considerar que el modelo como un modelo caótico. Además la dimensión del atractor experimenta una reducción la cual es característica de los sistemas caóticos en  $\mathbb{R}^n$ . En este caso podemos decir que la dimensión del atractor es  $D_a \approx 2$ .

## MODELO CON MEMORIA ESTILO CRISANTI.

A continuación estudiaremos un modelo basado en el modelo de Crisanti et al.<sup>[5]</sup>, el cual será una red neuronal con memoria. El estudio que llevaremos a cabo nos da indicios de una dinámica caótica de la red. Nosotros enfocaremos el estudio en la dimensión del atractor y si esta dimensión del atractor presenta una reducción tal vez la dinámica sea caótica, posteriormente estudiaremos la sensibilidad a las condiciones iniciales.

## EL MODELO.

El modelo que propondremos a continuación está basado en el modelo de Crisanti<sup>[5]</sup>. La evolución del sistema estará dada por:

$$S_n(t+1) = \operatorname{sgn} \left( \sum_{m=0}^T \sum_{j=0}^N w_{nm} S_m(t) \right) \quad (5.12)$$

En donde la matriz de pesos de la red ( $w_{nm}$ ) está dada por:

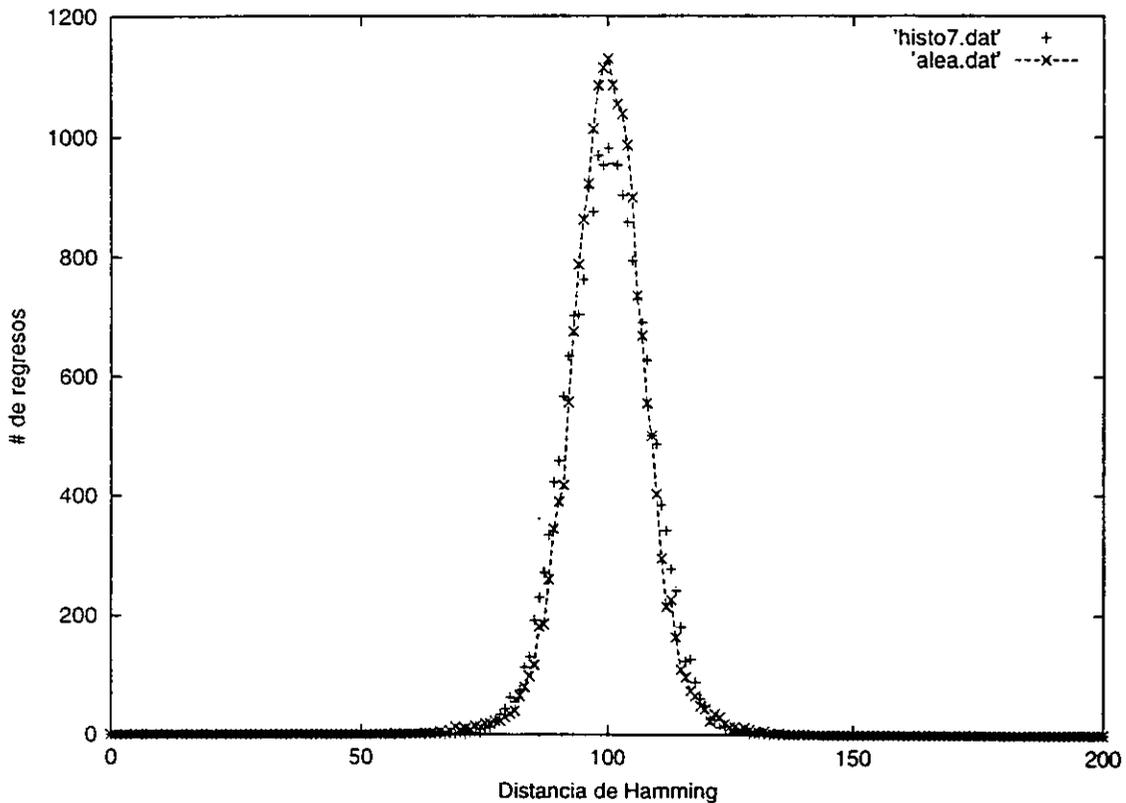
$$w_{nm} = \begin{cases} w_{nm}^c * R^{T+m-n-1}, & \text{si } m < n \\ w_{nm}^c * R^{m-n-1}, & \text{si } n < m \\ 0 & \text{si } n = m. \end{cases} \quad (5.13)$$

La  $i = 1, \dots, N$ ,  $N$  es el número de neuronas y  $n, m = 0, \dots, T$ ,  $T$  es el número de estados en la memoria de la red.  $w_{nm}^c$  es la matriz de pesos estilo Crisanti (ecuación (4.2)).  $R$  es un parámetro de decaimiento de la dinámica de Crisanti,  $R \in [0, 1]$ . Ha diferencia de los demás modelos la matriz de pesos de este modelo es de  $T \times T$ .

El modelo que estudiaremos a continuación es un modelo "vertical", debido a que constará de una neurona y  $T$  estados de memoria.

## RESULTADOS DE LAS SIMULACIONES.

El modelo se analizó con  $N = 1$ ,  $T = 100$  y un grado de antisimetría con  $k = 1.0$  para la matriz de Crisanti. Recordemos que en el modelo de Crisanti<sup>[5]</sup>  $k = 1.0$  la red ya presenta un comportamiento semejante al de un proceso gaussiano\*, como se puede observar en la siguiente gráfica *fig(5.16)*.



*fig 5.16. Histograma de las distancias de Hamming de una red neuronal con dinámica de Crisanti con  $N = 200$  neuronas y  $k = 1.0$ . En esta gráfica podemos observar que para valores de  $k = 1.0$  la red estilo Crisanti se comporta de manera muy semejante a un proceso aleatorio.*

A continuación presentaremos el análisis realizado para el cálculo de la dimensión del atractor. Si la dimensión presenta una reducción entonces es posible

\* Para más detalles ver capítulo 4. Resultados de la simulación.

que la dinámica de la red sea caótica, por desgracia como el modelo que estudiamos tiene sólo una neurona. Los regresos de Hamming no pueden ser tomados con un indicativo de la transitividad topológica, ya que la distancia de Hamming será sólo cero ó uno.

La distancia usada en este modelo para el cálculo de la función de correlación es la definida por la ecuación (5.3). La dimensión del atractor se define por la ecuación (5.4). En este modelo el parámetro que irá variando será el decaimiento ( $R$ ).

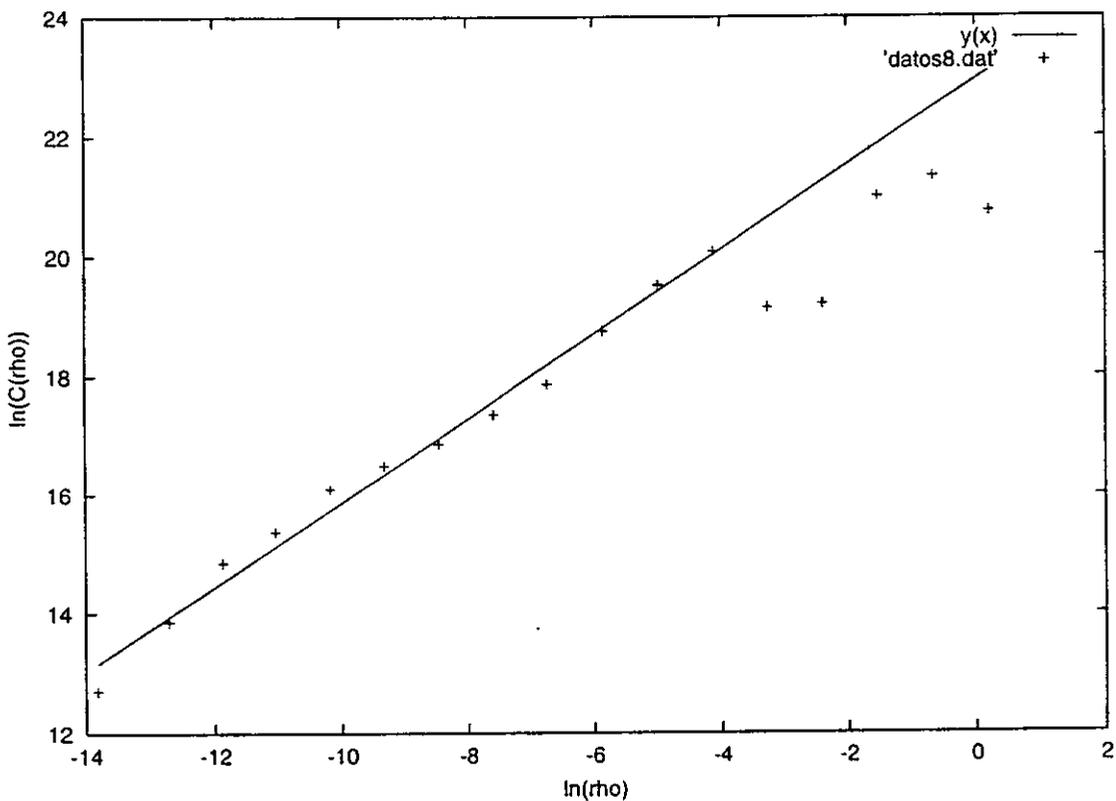


fig 5.17. Gráfica de la función de correlación  $\ln(C(\rho))$  y la curva ajustada  $y(x) = m \cdot x + b$ . En este modelo el análisis de regresión realizado a los datos de la función de correlación nuevamente nos servirá para establecer si existe una reducción en la dimensión del atractor, ya que la pendiente de la curva ajustada es una aproximación de la dimensión del atractor. En este modelo el parámetro que varía es  $R$  por lo que este primer ajuste corresponde a  $R = 0.8$ . En los resultados vemos que existe una zona en la cual la recta se ajusta bastante a los datos y presenta una reducción en la dimensión del atractor.

La simulación de la red se llevó a cabo con cien mil iteraciones y un decaimiento de  $R = 0.8$ . La gráfica resultante nos indica el logaritmo de  $\rho$  vs el logaritmo de  $C(\rho)$ . Este cálculo es de la función de correlación total. El resultado del análisis de regresión lo podemos observar a continuación (*fig5.17*):

El número de datos para la regresión es: 12.

Resultados de la regresión lineal.

El modelo es  $y(x) = m * x + b$ .

Pendiente  $m = 0.708367$ .

Ordenada  $b = 22.947754$ .

Coefficiente de correlación  $r = 0.993590$ .

Dimensión del atractor  $D_a = 0.708367$ .

Constante  $e^b = 9248745692.094458$ .

Resultados de la regresión exponencial.

El modelo es  $\ln(y) = m * x + \ln(b)$ .

Pendiente  $m = 0.043305$ .

Ordenada  $\ln(b) = 3.188908$ .

Coefficiente de correlación  $r = 0.987533$ .

Como se observa la dimensión del atractor es inferior a la dimensión del sistema la cual es una característica de los sistemas dinámicos caóticos. el coeficiente de correlación para el ajuste lineal es muy bueno.

Los resultados del análisis de regresión para  $R = 0.9$  y cien mil iteraciones son (*fig5.18*):

El número de datos para la regresión es: 23.

Resultados de la regresión lineal.

El modelo es  $y(x) = m * x + b$ .

Pendiente  $m = 0.686012$ .

Ordenada  $b = 14.518275$ .

Coefficiente de correlación  $r = 0.997494$ .

Dimensión del atractor  $D_a = 0.686012$ .

Constante  $e^b = 2019327.916553$ .

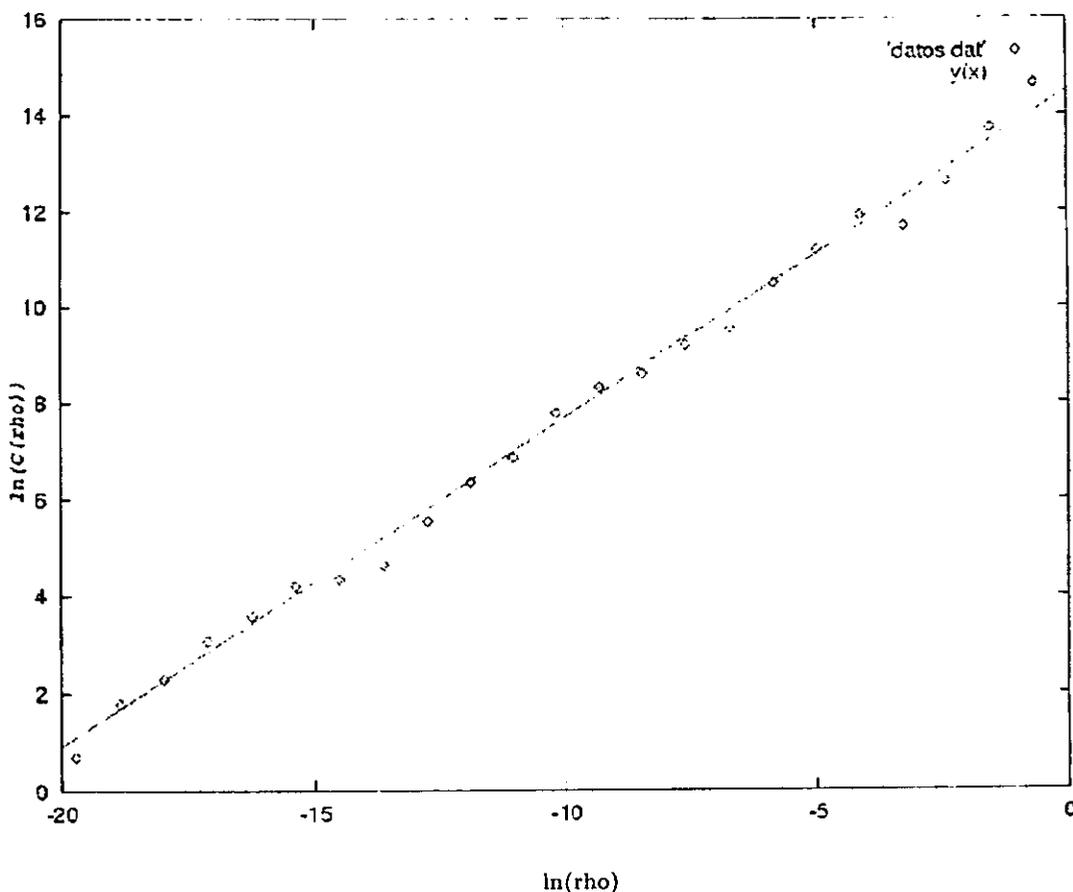


fig 5.18. Gráfica de la función de correlación  $\ln(C(\rho))$  y la curva ajustada  $y(x) = m * x + b$ , En este caso el ajuste se realizó para el caso de  $R = 0.9$ . Se observa que la recta se ajusta bastante bien a los datos y se observa una reducción en la dimensión del atractor.

#### Resultados de la regresión exponencial.

El modelo es  $\ln(y) = m * x + \ln(b)$ .

Pendiente  $m = 0.118257$ .

Ordenada  $\ln(b) = 3.019227$ .

Coefficiente de correlación  $r = 0.925735$ .

Nuevamente la dimensión del atractor es menor que la dimensión del sistema por lo cual es muy probable que la dinámica si sea caótica.

A continuación presentaremos los resultados para  $R = 1.0$  y cien mil iteraciones (fig5.19).

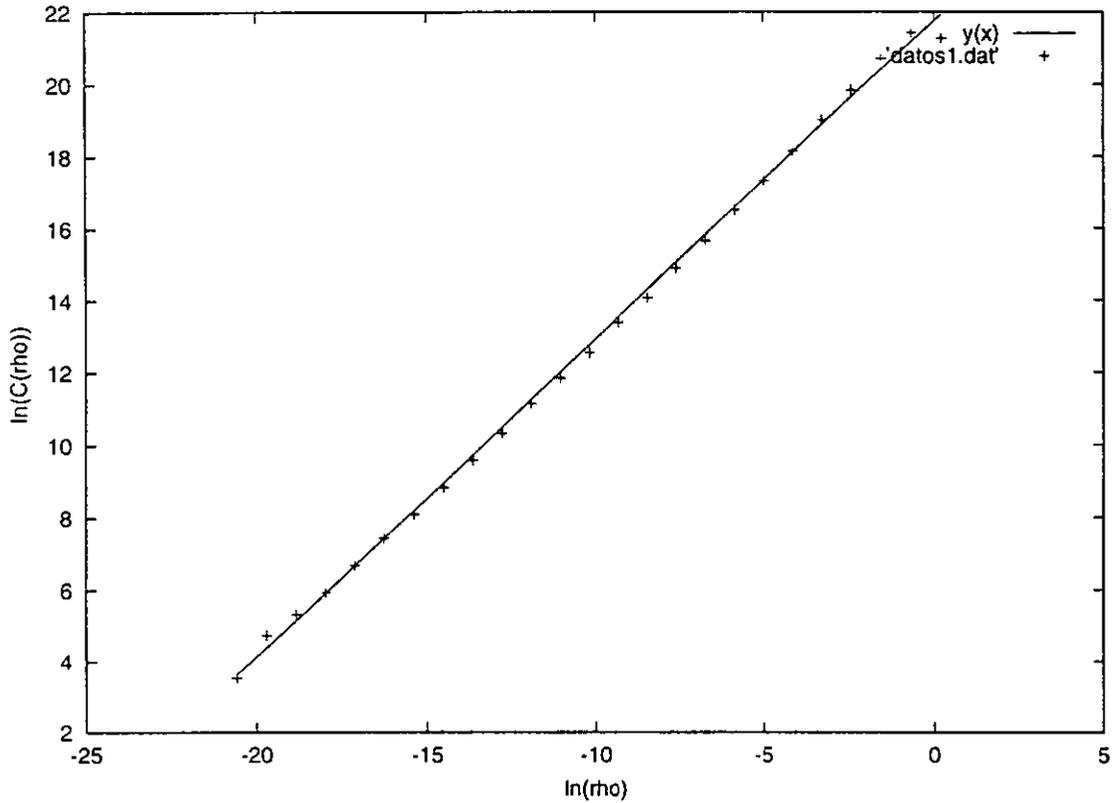


fig 5.19. Gráfica de la función de correlación  $\ln(C(\rho))$  y la curva ajustada  $y(x) = m * x + b$ . Esta gráfica corresponde a  $R = 1.0$ . Se observa que el ajuste es muy bueno y presenta una sensible reducción del atractor.

El número de datos para la regresión es: 23.

Resultados de la regresión lineal.

El modelo es  $y(x) = m * x + b$ .

Pendiente  $m = 0.883266$ .

Ordenada  $b = 21.749366$ .

Coefficiente de correlación  $r = 0.999446$ .

Dimensión del atractor  $D_a = 0.883266$ .

Constante  $e^b = 2790163414.406901$ .

Resultados de la regresión exponencial.

El modelo es  $\ln(y) = m * x + \ln(b)$ .

Pendiente  $m = 0.076373$ .

Ordenada  $\ln(b) = 3.232435$ .

Coefficiente de correlación  $r = 0.986062$ .

A pesar de que existe una reducción en la dimensión del atractor esta no es una condición suficiente para poder asegurar que la dinámica es caótica.

### ANÁLISIS DE LA SENSIBILIDAD DE LAS CONDICIONES INICIALES

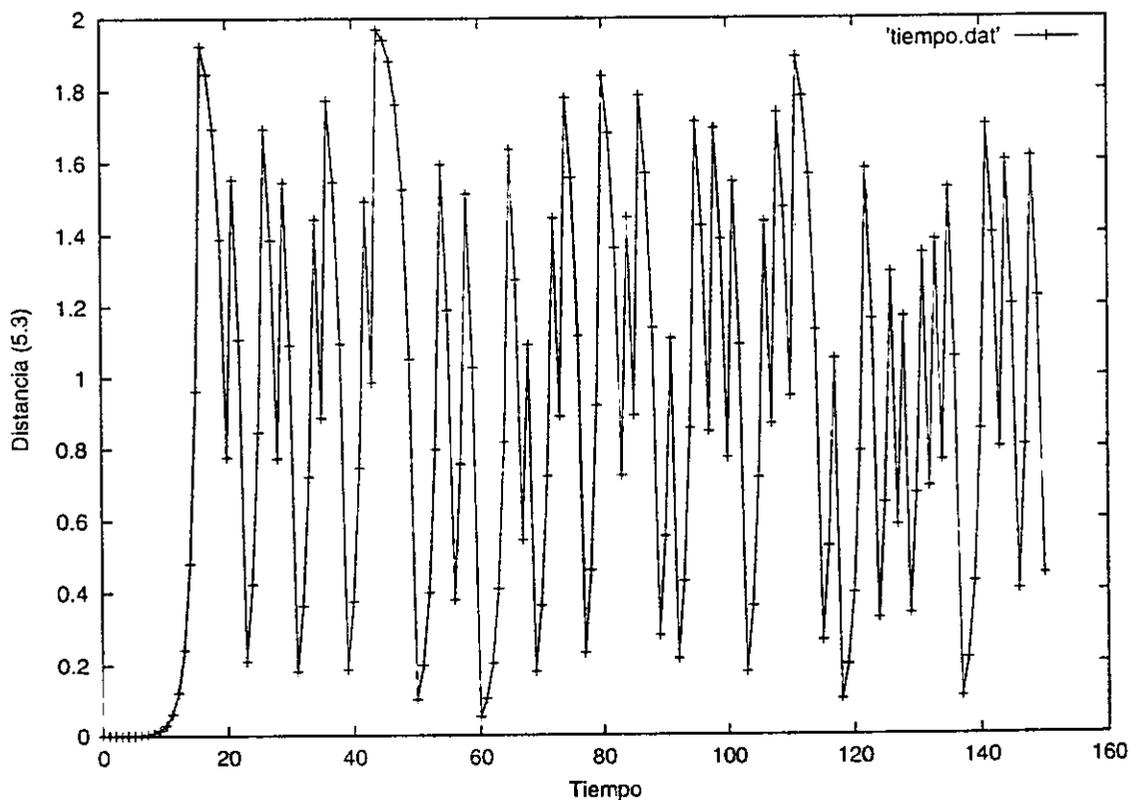


fig 5.20. Comparación de las trayectorias entre  $S_0$  y  $S'$ . Esta gráfica corresponde al análisis de la sensibilidad a las condiciones iniciales para el caso de  $R = 0.8$ . Se puede observar que inicialmente las trayectorias se encuentran muy cercanas y después de unas cuantas iteraciones estas se separan bastante, esto nos indica que la red es sensible a las condiciones iniciales.

A continuación mostraremos el análisis de las condiciones iniciales para el modelo estilo Crisanti con memoria. Como primera parte de este análisis nos enfocaremos en encontrar un estado cercano  $S_0$  a  $S'$ , el cual es nuestro estado fijo.

El estado más cercano a  $S'$  para la simulación de  $R = 0.8$  y cien mil iteraciones es el estado  $S_0 = 81554$  cuya distancia entre  $d(S_0, S') = 0.000059$ . Ahora compararemos la distancia entre las trayectorias de  $S_0$  y  $S'$  la gráfica (fig5.20) nos mostrará como evolucionan las trayectorias.

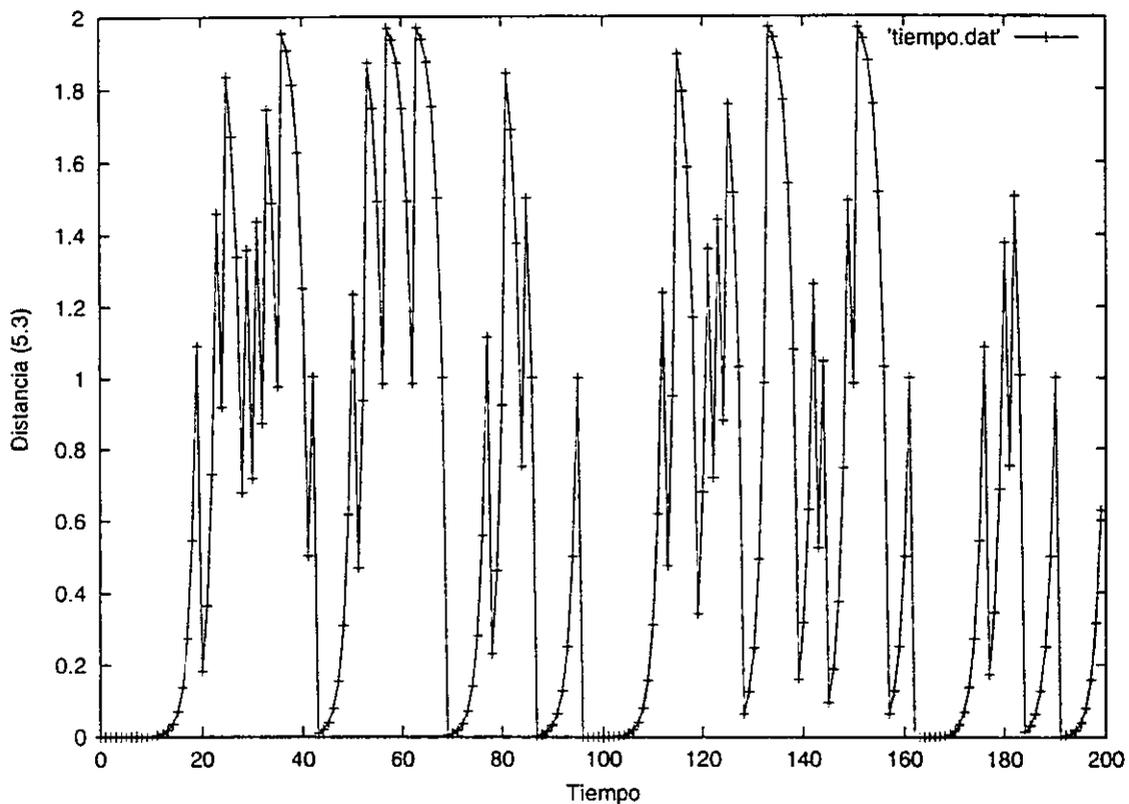


fig 5.21. Comparación de las trayectorias entre  $S_0$  y  $S'$ . Este caso corresponde a  $R = 0.9$ . podemos observar que la red sí presenta sensibilidad a las condiciones iniciales ya que en ciertas iteraciones la distancia entre las trayectorias llega a ser considerablemente grande.

Como se puede observar en la gráfica las trayectorias se separan bastante esto nos da un indicio de que la red puede presentar sensibilidad a las condiciones iniciales. para el caso de  $R = 0.8$ . A continuación compararemos las trayectorias

entre dos estados cercanos para el caso de  $R = 0.9$  con cien mil iteraciones (fig5.21).

En este caso parecería que las trayectorias se aproximan demasiado en ciertos estados sin embargo esos puntos en la realidad no están tan cerca de nuestro estado  $S'$ . En este caso la distancia inicial entre  $S_0 = 37700$  y  $S'$  fue de  $4 \times 10^{-6}$  y en los puntos de la gráfica que aparecen con distancia cero en general tienen una distancia mayor a  $10^{-3}$ , la cual es relativamente grande.

Por último presentaremos el análisis de las trayectorias de  $S_0$  y  $S'$  para el caso de  $R = 1.0$  y cien mil iteraciones, en este caso la distancia inicial fue de  $3 \times 10^{-6}$  (fig5.22).

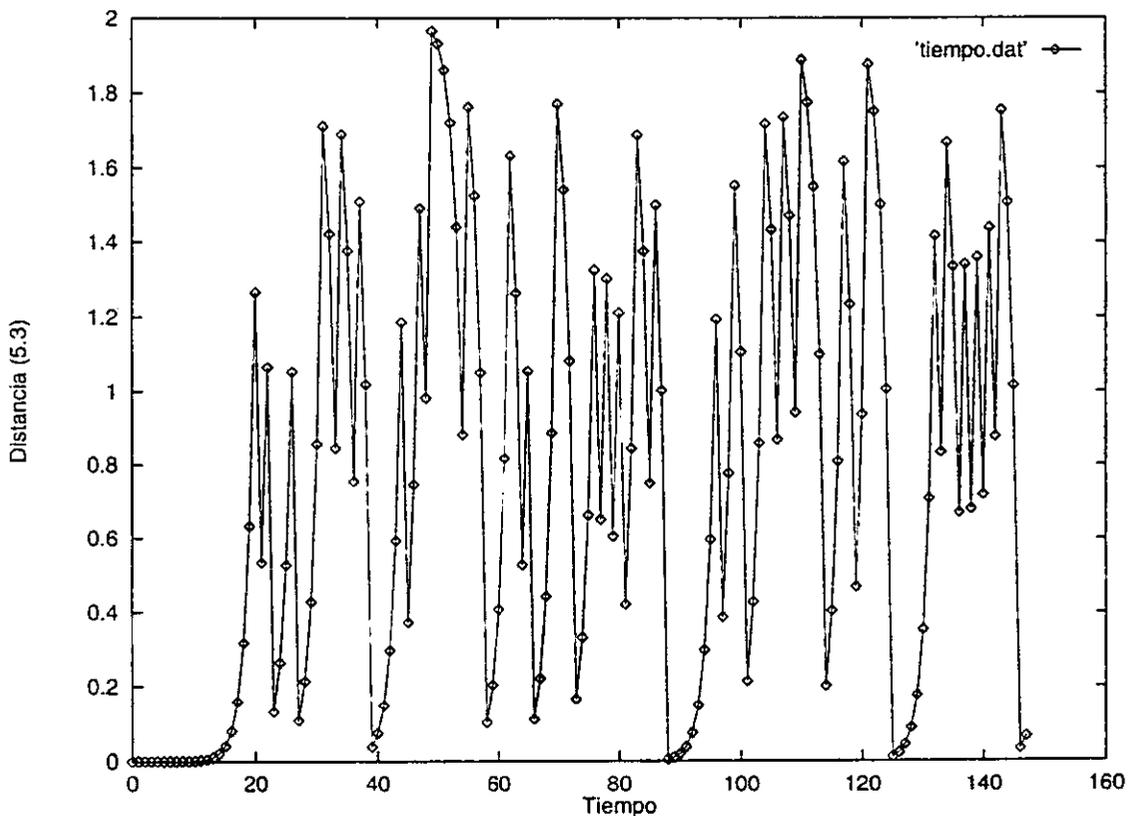


fig 5.22. Comparación de las trayectorias entre  $S_0$  y  $S'$ . Este caso corresponde a  $R = 1.0$ , este caso es el que presento una menor reducción en el atractor, sin embargo se puede observar en la gráfica que la red sí presenta sensibilidad a las condiciones iniciales.

En este caso la gráfica es más clara, ya que en general las trayectorias se mantienen más separadas.

Obviamente hasta el momento las gráficas anteriores nos dan un indicio de la sensibilidad a las condiciones iniciales de la red, sin embargo no hemos visto si la red realmente es sensible a las condiciones iniciales.

Nuevamente generaremos dos estados tan cercanos que la diferencia entre ellos estará en el último renglón, como resultado de esto los dos estados iniciales tendrán una distancia inicial menor que  $10^{-24}$ .

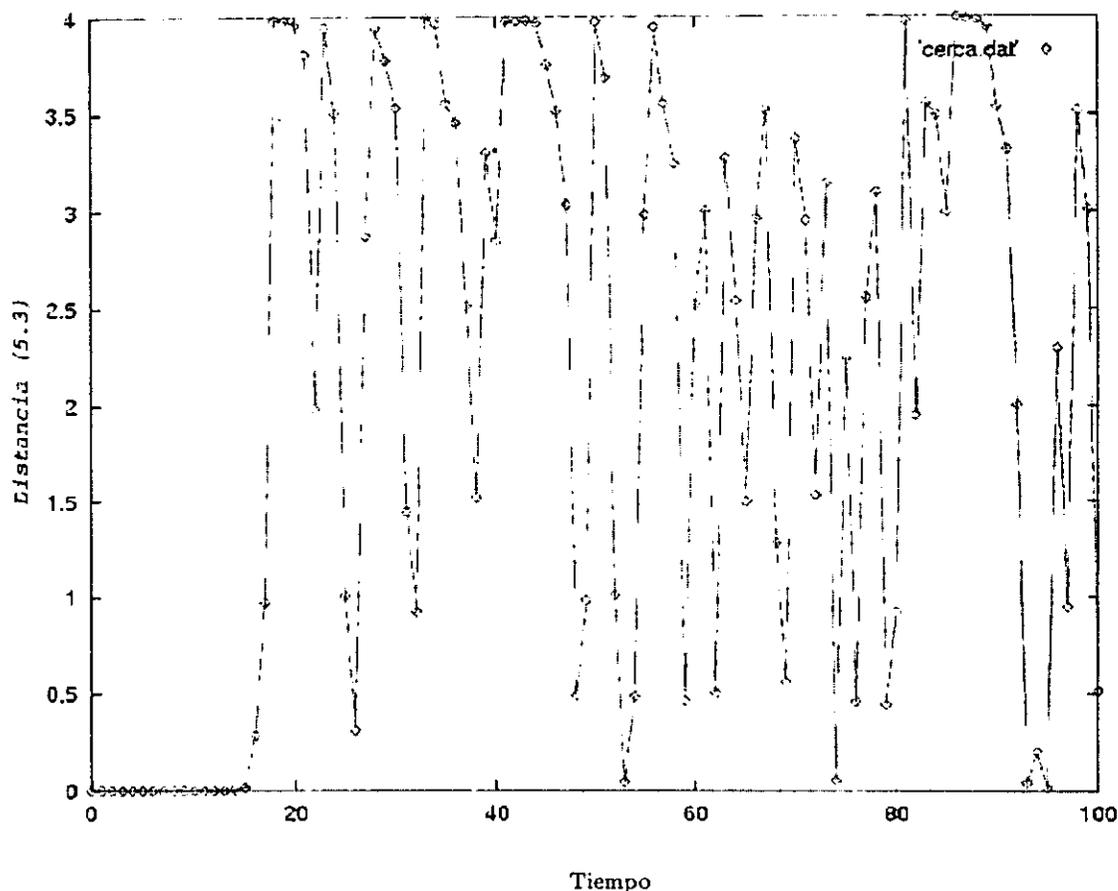


fig 5.23. Comparación de las trayectorias entre  $S_0$  y  $S'$ . Para finalizar el análisis de la sensibilidad a las condiciones iniciales generaremos un estado muy cercano a nuestro estado inicial y estudiaremos las distancias entre las trayectorias. En esta gráfica el valor de  $R = 0.8$  y se puede observar que la red sí tiene sensibilidad a las condiciones iniciales.

Como se puede observar en la gráfica (fig 5.23) la red sí presenta una clara evidencia de sensibilidad a las condiciones iniciales para el caso de  $R = 0.8$ . A continuación presentaremos los resultados de  $R = 0.9$  (fig 5.24).

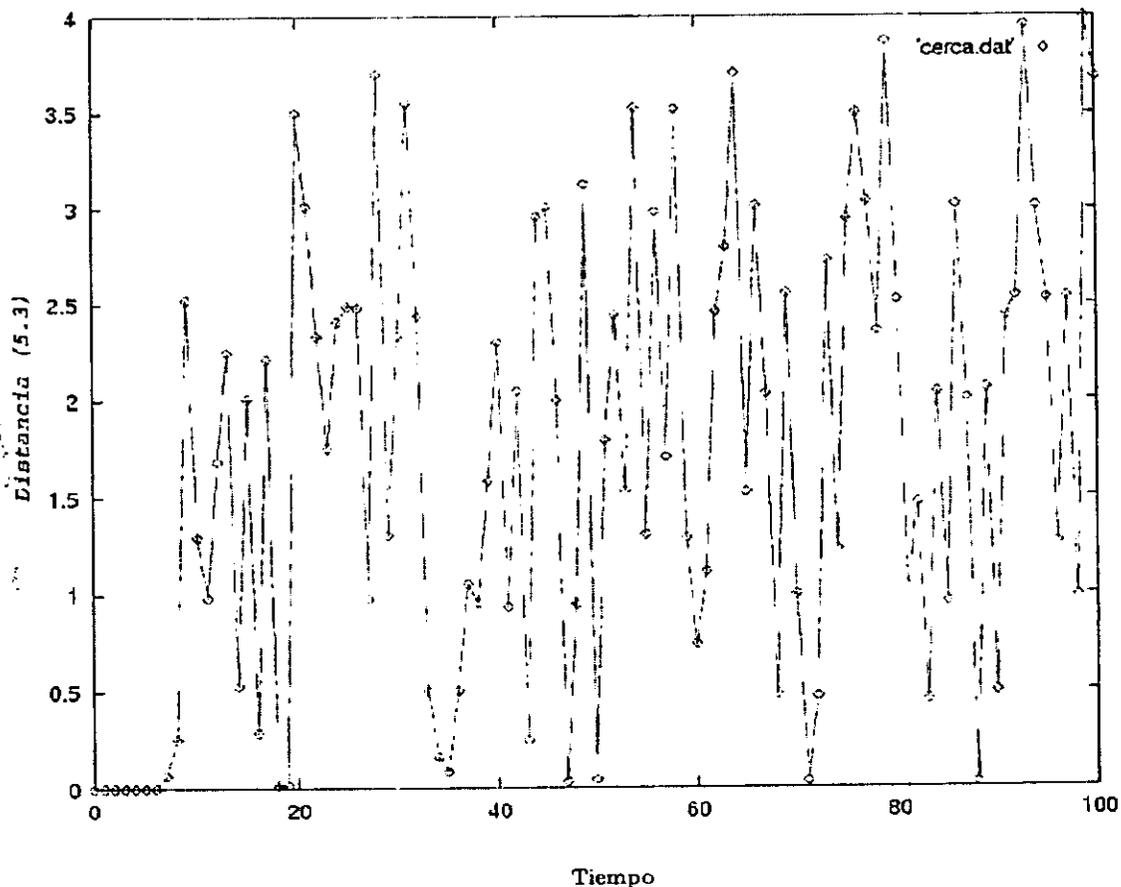
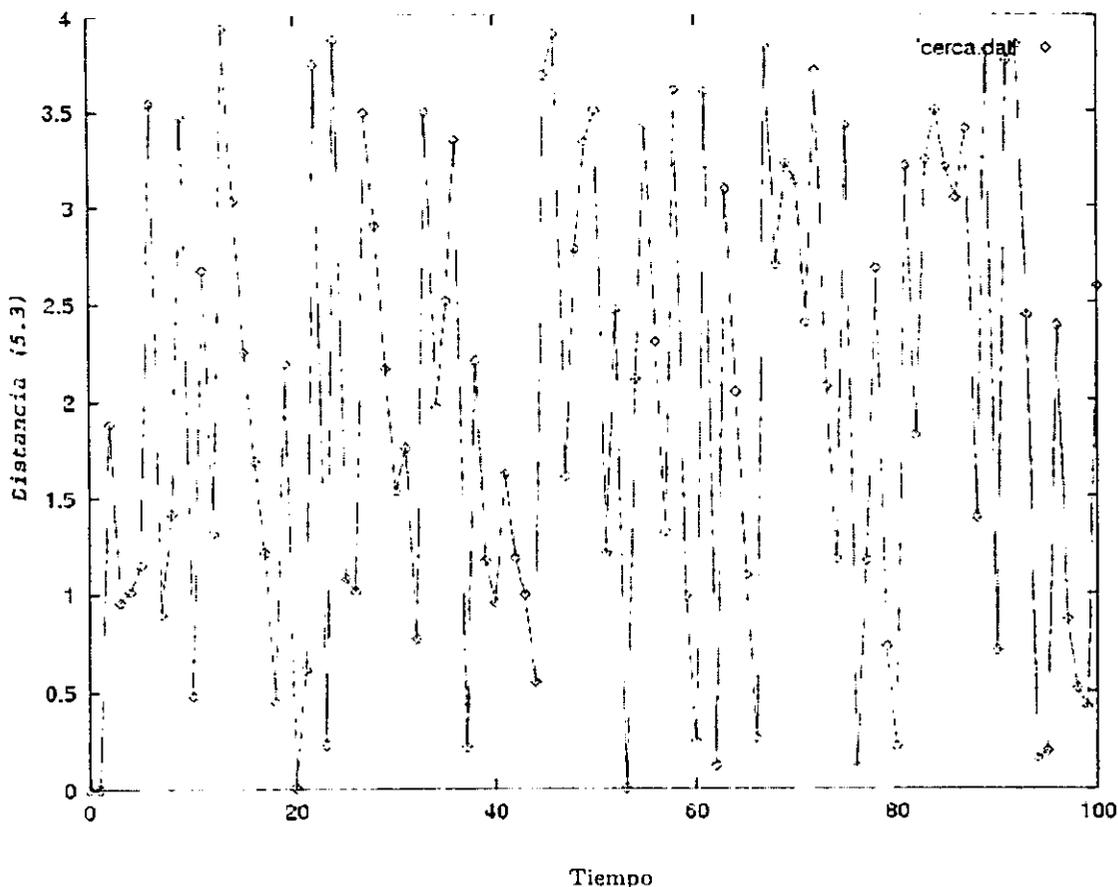


fig 5.24. Comparación de las trayectorias entre  $S_0$  y  $S'$ . En esta gráfica se tiene que  $R = 0.9$ , nuevamente se generó un estado muy cercano al estado inicial y en la gráfica podemos observar las distancias entre las trayectorias con lo que podemos asegurar que la red sí presenta sensibilidad a las condiciones iniciales.

Nuevamente la red muestra sensibilidad a las condiciones iniciales. Para terminar el estudio de este modelo presentaremos el análisis de las condiciones iniciales para  $R = 1.0$  (fig 5.25).

En esta gráfica vemos que para  $R = 1.0$  la red es aparentemente sensible a las condiciones iniciales.

En conclusión la dinámica de la red neuronal presenta evidencia de sensibilidad a las condiciones iniciales para los tres parámetros estudiados  $R = 0.8, 0.9, 1.0$ .



ESTA TESTA NO DEBE  
SALIR DE LA BIBLIOTECA

fig 5.25. Comparación de las trayectorias entre  $S_0$  y  $S'$ , en esta gráfica el valor de  $R = 1.0$ . Nuevamente generamos un estado aleatorio muy cercano al estado inicial y graficaremos la distancia entre las trayectorias, lo que nos permite decir que la red sí presenta sensibilidad a las condiciones iniciales.

En el caso de  $R = 0.8$  y  $R = 0.9$  nos confirma que en principio la red sí presenta una dinámica caótica. En el caso de  $R = 1.0$  la dimensión del atractor no nos indica que exista una reducción significativa y por otro lado el análisis a la sensibilidad de las condiciones iniciales nos muestra que la red en principio es altamente sensible a las condiciones iniciales. Como resultado de los análisis no podemos concluir que la red presente una dinámica caótica para el caso de  $R = 1.0$ .







## CAPÍTULO 6

## CONCLUSIONES

A lo largo de esta tesis hemos estudiado diversos modelos de redes neuronales. El primer modelo que estudiamos fue el modelo de Hopfield<sup>[1,2]</sup>, el cual nos ha servido como introducción formal al estudio de las redes neuronales. Durante el estudio del modelo de Hopfield<sup>[1,2]</sup> primero estudiamos el caso de tener una red totalmente determinística la cual presenta una estabilidad si se cumplen ciertas condiciones, la estabilidad de la red quiere decir que la red tiene una buena memoria asociativa. La introducción de la función de energía en estos sistemas evidenció la analogía entre las redes neuronales y los vidrios de spin, que a su vez permitió la introducción del concepto de temperatura en las redes neuronales, a partir de ese momento el estudio de las redes neuronales se enfocó en las propiedades térmicas de estos sistemas. En general la analogía con los vidrios de spin propició una solución de la capacidad de la red neuronal en el límite de temperatura cero. Vimos que las redes neuronales tienen una cierta capacidad  $\alpha$  para el correcto almacenamiento de los patrones, si esta capacidad es rebasada entonces la red pierde la estabilidad y surgen los estados espúreos. En el caso del modelo de Hopfield el surgimiento de los estados espúreos propicia que la red pierda la memoria asociativa.

En el capítulo tres estudiamos una red neuronal sin memoria tipo Hopfield<sup>[1,2]</sup>, con una matriz de pesos asimétrica. El estudio de esta red se enfocó principalmente en los resultados obtenidos de la simulación en la cual encontramos que los ciclos de la red se asemejan a un proceso aleatorio de una función gaussiana. El ajuste de la dinámica de la red a la curva gaussiana es muy bueno en algunos casos, a pesar de eso. este proceso muestra una ligera diferencia con respecto a un proceso meramente aleatorio. También se pudo observar que la red no presenta indicios de una dinámica caótica ya que la red no mostró indicios de transitividad topológica (Definición 5.2.). En las gráficas de los ciclos de la red se puede observar que la red no presenta transitividad topológica ya que estas gráficas no muestran estados cuya distancia de Hamming al estado de referencia  $S'$  sea pequeña\*.

En el capítulo cuatro estudiamos una red neuronal asimétrica sin memoria propuesta por Crisanti et.al...<sup>[5]</sup>. La dinámica de esta red neuronal en principio

---

\* Un estado con distancia de Hamming pequeña al estado de referencia es un estado cuya distancia de Hamming es 1, 2, 3.

ha sido vista como una dinámica caótica<sup>[5]</sup>, debido a que los ciclos límite de la red son de período  $e^N$ , permitiendo que la dinámica de la red antes de alcanzar el ciclo sea descrita como compleja. Sin embargo en nuestros estudios la red no puede presentar caos ya que nuevamente la dinámica no es topológicamente transitiva (en este caso no existen estados con distancias de Hamming menores a  $N/4$  a  $S'$ ). En un estudio posterior realizado por Crisanti, Falcioni, Mantica y Vulpiani<sup>[15]</sup> se estudia el comportamiento de las trayectorias en los sistemas dinámicos finitos, en especial se estudia este modelo. Las trayectorias se relacionan con la sensibilidad a las condiciones iniciales y a partir de ellas se puede realizar una aproximación del exponente de Lyapunov. En general si el exponente de Lyapunov es positivo entonces se dice que el sistema es caótico. Al realizar un estudio de las trayectorias de los estados de una red neuronal de este estilo, se observa que estas pueden llegar a ser muy regulares. Es decir que las trayectorias no muestran indicios de la presencia de sensibilidad a las condiciones iniciales, además el cálculo del exponente de Lyapunov indica que este es positivo. A pesar del problema de las trayectorias de la red neuronal este modelo exhibe un comportamiento muy semejante al de un proceso caótico ya que la función de correlación decae de manera exponencial<sup>[5]</sup>. Todas estas situaciones llevan a considerar a Crisanti, et al... que el modelo es muy complejo ya que sufre de un cambio de fase alrededor del valor crítico  $k_c$ .

En el capítulo cinco se estudiaron redes neuronales asimétricas con memoria. En estos modelos la topología del espacio cambia, ya que para poder definir correctamente los conceptos de caos de  $\mathbb{R}^n$  para sistemas dinámicos discretos<sup>[4]</sup> se definió el espacio topológico  $\Omega$  en el cual la distancia relevante está dada por la ecuación (5.3). El primer modelo que estudiamos de las redes neuronales con memoria fue propuesto por H.Waelbroeck y F.Zertuche<sup>[4]</sup> ahí se estudian los ciclos de la red concluyendo que la dinámica de ese modelo en principio presenta caos, ya que la dimensión del atractor muestra una reducción y la red presenta transitividad topológica. En esta tesis se han ampliado esos estudios dando como resultado que la dinámica de ese modelo presenta un comportamiento caótico a partir de la definiciones de un caos discreto<sup>[4]</sup>, en esta tesis se realizó el análisis a la sensibilidad de las condiciones iniciales como se puede observar en las gráficas *fig 5.9-5.13*. se revisó la transitividad topológica como se puede observar en las gráficas *fig 5.1, 5.3, 5.5*. y se calculó la dimensión del atractor, dando una dimensión aproximada de  $D_a \approx 2.0$ . la dimensión del atractor fue calculada para dos distancia distintas. permitiendos asegurar que la red presenta una reducción en la dimensión del atractor, recordemos que una característica de los sistemas dinámicos caóticos es la reducción de la dimensión del atractor.

El último modelo es una red neuronal que consta de una neurona y  $T = 100$  estados de memoria. En este modelo se estudió la dimensión del atractor la cual presenta una reducción de aproximadamente  $D_a = 0.68$ . el cual es el valor

menor del atractor que encontramos (*fig5.18*). También llevamos a cabo un estudio sobre la sensibilidad a las condiciones iniciales del modelo y ciertamente la red presenta indicios de sensibilidad a las condiciones iniciales. En este modelo podemos concluir que en principio la red sí presenta un comportamiento caótico, a reserva de realizar un análisis para ver si la red es topológicamente transitiva.



## BIBLIOGRAFÍA

- [1].- Hertz, J. Krogh, A. Palmer, R. *Introduction to the theory of neural computation*. (Addison Wesley) (1991).
- [2].- Müller, B. Reinhardt, J. *Neural networks. an introduction*. (Springer-Verlag) (1991).
- [3].- Little, W.A. *The existence of persistent states in the brain*. Mathematical biosciences. **39** (1974) 101-120; Little, W.A. Shaw, G. *Analytic study of the memory storage capacity of a neural network*. Mathematical biosciences. **39** (1978) 281-290.
- [4].- Waelbroeck, H. Zertuche, F. *Discrete chaos*. J.Phys. A:Math. Gen. **32** (1999) 175-189.
- [5].- Crisanti, A. Falcioni, M. Vulpiani, A. *Transition from regular to complex behaviour in a discrete deterministic asymmetric neural network*. J.Phys. A:Math. Gen. **26** (1993) 3441-34553.
- [6].- Beale, R. *Neural computing: an introduction*. (Bristol:A. Hilger) (1990).
- [7].- Canavos, G. *Probabilidad y estadística Aplicaciones y métodos*. (Mc Graw Hill) (1986).
- [8].- Feller, W. *Introducción a la teoría de probabilidades y sus aplicaciones*. (Limusa-Wiley) (1973).
- [9].- Selby, S. *Standard mathematical tables*. (The Chemical Rubber Co) (1967).
- [10].- Arfken, G. *Mathematical Methods for Physicist*. (Academic) (1974).
- [11].- Zertuche, F. López-Peña, R. Waelbroeck, H. *Recognition of temporal sequences of patterns using state-dependent synapses*. J.Phys. A: Math. Gen. **27** (1994) 5879-5887.
- [12].- Montgomery, D. Peck, E. *Introduction to linear regression analysis*. (New York: J. Wiley) (1992).
- [13].- Grassberger, P. Procaccia, I. *Characterization of strange attractors*. Phys. Rev. Lett. **50** (1983) 346-349; Physica. D **9** (1983) 189-208.
- [14].- Zertuche, F. López-Peña, R. Waelbroeck, H. *Storage capacity of neural networks with state-dependent synapses*. J.Phys. A:Math. Gen. **27** (1994) 1575-1583.

[15].- Crisanti, A. Falcioni, M. Mantica, G. Vulpiani, A. *Applying algorithmic complexity to define chaos in the motion of complex systems* Physical Review E. **50** n 3 (1994) 1959-1967.

[16].- Rojas, P. *Neural networks: a systematic introduction*. (New York:Springer) (1996).

[17].- Courant, R. John. F. *Introducción al cálculo y al análisis matemático*. (Limusa) (1987).

[18].- Prabodh Shukla. *Theory of the dynamics of the Hopfield model of associative memory*. J. Stat. Phys. **71** (1993) 705-717.

[19].- Klaus Nützel. *The length of attractors in asymmetric random neural networks with deterministic dynamics*. J.Phys. A:Math. Gen. **24** (1991) L151-L157.