MODELOS BASADOS EN ENERGIA

MEMORIA ASOCIATIVA Y ENERGIA
EL MODELO DE HOPFIELD

Problema: *almacenar un conjunto de patrones de información de tal manera que, cada vez que se pone una entrada al sistema,la salida sea el patrón que más información tiene en común con la entrada.*

→ **Procedural:** un programa secuencial en una máquina von Neumann.

Soluciones

→ **Conexionista**: típicamente, una red neuronal.

(existen, claro, otros paradigmas)

Ejemplo:

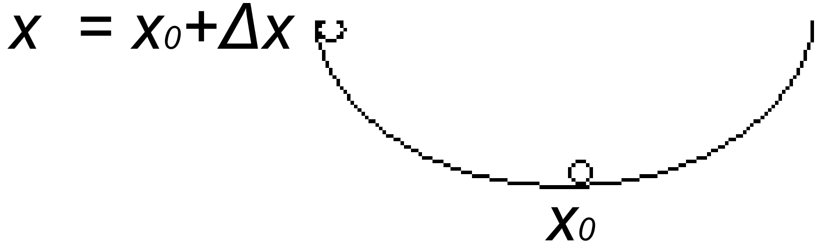
Patrón 1: jorgeBcorrientes1161B49627781
Patrón 2: irmaBBmedranoBBB580BB4873440B
Patrón 3: pabloBcordobaBBB4215BB47751763

(B = espacio en blanco)

Resolución procedural típica: base de datos. Podrá recuperar el patrón 3 con la entrada "cordoba", p.ej., pero nada podrá hacer si la entrada es "bloBcor". La entrada debe ser una unidad de información completa (aunque mínima). La tolerancia a fallas es poca.

Buscaremos asociar una energía a cada posible configuración del sistema y los mínimos de esa energía a las memorias (los patrones a memorizar).

Sistema físico:



x_0 : estado estable (mínima energía: $E_p = 0$, $E_c \rightarrow$ calor))
x : conocimiento parcial de x_0
Δx : error

Muchas memorias:

MODELO DE HOPFIELD

Elemento básico: *neurona de McCulloch y Pitts* (1943)

$$S_i\,(t+\Delta t) = sgn\left(\Sigma w_{ij}\, s_j\,(t) - \theta_i\right)$$

con $sgn(x) = 1 \text{ si } x \geq 0, -1 \text{ si } x < 0$

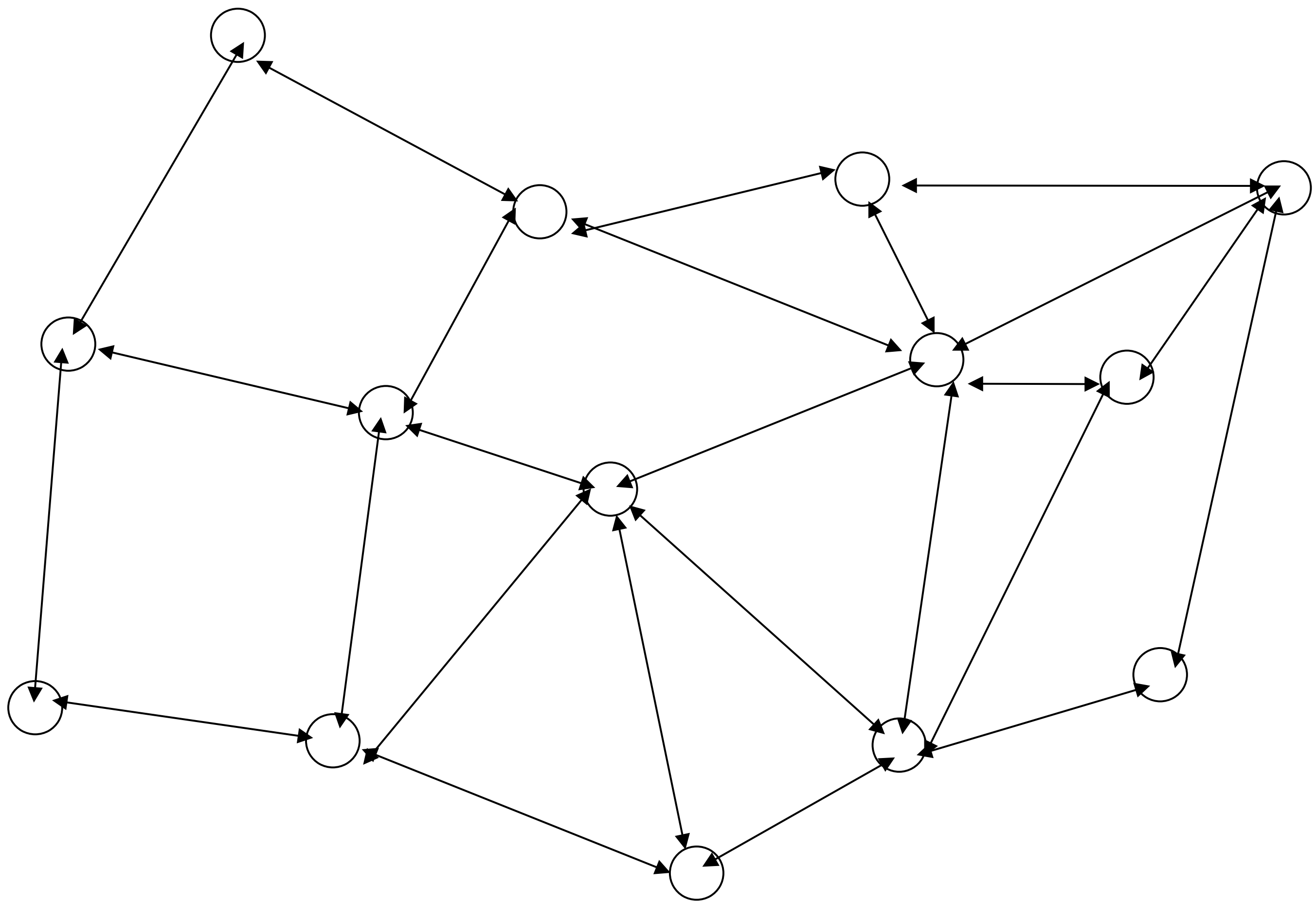
S_i **estado de la unidad i** : 1 ≡ activa , -1 ≡ inactiva

Observación: w_{ij}, θ_i son invariantes en el tiempo, pues en ellos está contenida la información (las memorias).

Formas de actualización:

- **Sincrónica**: *reloj central, supervisor.*
- **Asincrónica**: *implementación paralela, la probabilidad de que dos unidades se actualicen al mismo tiempo es nula.*

Red recurrente (Hopfield)
para memoria asociativa



neurona

conexión sináptica

UNA ENERGIA ADECUADA

¿Qué definición de energía será apropiada para la dinámica definida arriba? (i.e. el sistema evolucione hacia sus mínimos)
Se propone:

$$E(x) = -x\,W\,x = -\sum_{\text{pares } i,j} w_{ij}\, x_i\, x_j = -1/2 \sum_{i,j} w_{ij}\, x_i\, x_j$$

Interpretación: w_{ij} es una restricción que tiende a hacer que *i* y *j* tomen el mismo valor, si $w_{ij} > 0$, o contrario, si $w_{ij} < 0$.

$$E(x) = -1/2 \sum_{i,j\neq k} w_{ij}\, x_i\, x_j - 1/2 \sum_{i} w_{ki}\, x_k\, x_i - 1/2 \sum_{i} w_{ik}\, x_i\, x_k =$$

$$-1/2 \sum_{i,j\neq k} w_{ij}\, x_i\, x_j - \sum_{i} w_{ki}\, x_k\, x_i \qquad \text{siempre que } w_{ik} = w_{ki}$$

$$= S - \sum_{i} w_{ki}\, x_k\, x_i = S - x_k\, \alpha_k \qquad \alpha_k \text{ activación de } x_k$$

Sea $E(x') = S - x'_k\, \alpha_k$ energía después de que k se actualizó

$\rightarrow \Delta E = E(x') - E(x) = (x_k - x'_k)\, \alpha_k = - \Delta x_k\, \alpha_k$ (1)

$\rightarrow \Delta E \leq 0$ (analizando caso $\alpha_k \geq 0$ y $\alpha_k < 0$)

Pero E es acotada inferiormente → alcanzaría un mínimo (local).

A partir de allí: sólo cambios tales que $\Delta x \neq 0$ y $\Delta E = 0$.

→ (por (1)) $\alpha_k = 0$ → cambio de -1 a 1 → a lo sumo N más.

Luego, estabilización.

Notar: esta propiedad es independiente de cómo se definan los w_{ij}.

PESOS

Hopfield propone:

$$w_{ij} = \frac{1}{N} \sum_{\mu=1}^{p} \xi_i^{\mu} \xi_j^{\mu} .$$

Interpretación (semi-intuitiva): correlación entre el estado de la neurona i y la j.

Si <u>frecuentemente</u> (en el conjunto de memorias) $S_i = S_j \longrightarrow$ cada una contribuye positivamente (excitatoriamente) a que la otra tome su mismo estado.

Si <u>frecuentemente</u> (en el conjunto de memorias) $S_i = -S_j \longrightarrow$ cada una está haciendo una contribución para que la otra tome el estado contrario)

INSPIRACION BIOLOGICA: *POSTULADO DE HEBB* (1949)

Cuando un axón de una célula A está suficientemente próximo para excitar a una célula B o toma parte en su disparo de forma persistente, tiene lugar algún proceso de crecimiento o algún cambio metabólico en una de las células, o en las dos, de tal modo que la eficiencia de A, como una de las células que desencadenan el disparo de B, se ve incrementada.

D. Hebb, *The organisation of behaviour*, 1949

Observación: *Hopfield va más allá de la hipótesis de Hebb, ya que cuando ambas memorias (pre- y post-sináptica) están inactivas, su conexión también se incrementa positivamente, lo cual probablemente no sea realista desde un punto de vista biológico.*

ESTABILIDAD

<u>Definición</u>: $\{\xi_\mu\}$ $\mu = 1, . . ., p$ se dicen *pseudoortogonales sii*

$$E\ (\xi_\mu\ \xi_\nu) = 0$$

cuando $\mu \neq \nu$ (ortogonales en valor medio)

- Para un conjunto de patrones pseudoortogonales aprendidos $\{\xi_\mu\}$ $\mu = 1,..., p$, la probabilidad de que un determinado bit de uno de ellos sea inestable decrece cuando *N/p crece.*

$$P(sgn\ (h_{i\nu}) \neq \xi_{i\nu}) = ½ \left\{ 1- 2 / \sqrt{\pi} \int_0^{\sqrt{N/2p}} exp\ (-u^2)\ du \right\}$$

Sólo se refiere a la estabilidad inicial → *es cota superior*.

- Un análisis más fino (desde la teoria de *spin glasses*) muestra que cuando $p \approx 0.138\ N$ se produce una avalancha de cambios.

ESTADOS ESPURIOS

p/N > valor crítico$\rightarrow$puntos fijos no deseados ($\xi \neq \xi(\mu)$ para todo μ)

En general:

- Cuencas de atracción menores

- Energía mayor

- Combinaciones lineales de las memorias

ESTADOS ESPURIOS

I) **Inversos**: el “negativo” del memorizado: $-\xi_i^\mu$ por la perfecta simetría del modelo. _Misma energía que el original_

ii) **Espurios mixtos**: **c**ombinaciones lineales de número impar de patrones

$$\xi_i^{\rm mix} = {\rm sgn}(\pm\xi_i^{\mu_1} \pm \xi_i^{\mu_2} \pm \xi_i^{\mu_3})$$

Observando que

$$h_i^{\rm mix} = \frac{1}{N}\sum_{j\mu}\xi_i^\mu \xi_j^\mu \xi_j^{\rm mix} = \frac{1}{2}\xi_i^{\mu_1} + \frac{1}{2}\xi_i^{\mu_2} + \frac{1}{2}\xi_i^{\mu_3} + \text{cross-terms}$$

la condición de estabilidad se satisface (en valor medio).

Vale para 5, 7, etc. términos, con cuencas decrecientes y energía creciente $\longrightarrow$ *menos estables*

iii) **Spin glasses**: para *p* grande, no correlacionados con ningún número finito de los patrones originales.

IMPORTANCIA DE LA ALINEALIDAD EN EL MODELO

En general, en todas las arquitecturas de RN.

De lo contrario: subespacios de estados estables → *infinitos estados espurios*

memoria 1 memoria 2 → memoria 3

LA MECANICA ESTADISTICA Y EL MODELO DE HOPFIELD

Modelo de Ising

↑ ↓ ↑ ↓ ↑ ↓ ↑ ↓ ↑ ↓ ↑ ↓ ↑ ↓ ↑ ↑ ↓ ↑ ↓ ↑ ↓ ↑ ↓

↑ ↓ ↑ ↑ ↓ ↓ ↑ ↓ ↑ ↓ ↓ ↑ ↓ ↑ ↑ ↓ ↓ ↑ ↓ ↑ ↓ ↑ ↓

↑ ↓ ↑ ↓ ↑ ↓ ↑ ↓ ↑ ↓ ↑ ↓ ↑ ↓ ↑ ↑ ↓ ↓ ↑ ↓ ↑ ↓ ↑

↑ ↑ ↑ ↓ ↑ ↑ ↓ ↓ ↑ ↓ ↑ ↓ ↑ ↓ ↑ ↑ ↓ ↓ ↑ ↓ ↑ ↓ ↑

Postulado básico:

- *a alta temperatura, cada spin está en 1 o en -1 con igual probabilidad*
- *a baja temperatura, cada spin tiende a alinearse paralelo a su campo local*

$h_i = \Sigma w_{ij} S_j + h_e$ campo magnético sobre la partícula i

S_j spin de la partícula j

$w_{ij} = w_{ji}$ fuerzas de interacción

h_e campo externo

Energía:

$$H = -1/2\ \Sigma w_{ij} S_i S_j - he\ \Sigma S_i$$

Analogía:

Modelo de Hopfield Modelo de Ising

pesos sinápticos ↔ fuerzas de interacción

entrada neta a una neurona ↔ campo sobre un spin

umbral ↔ campo externo

UN MODELO BASICO, MUCHAS VARIANTES Y EXTENSIONES

- *Estocástico(* inspirado en la mecánica estadística)
$S_i(t+\Delta t) = +1$ con probabilidad $g(h_i(t))$, -1 con probabilidad $1 - g(h_i(t))$

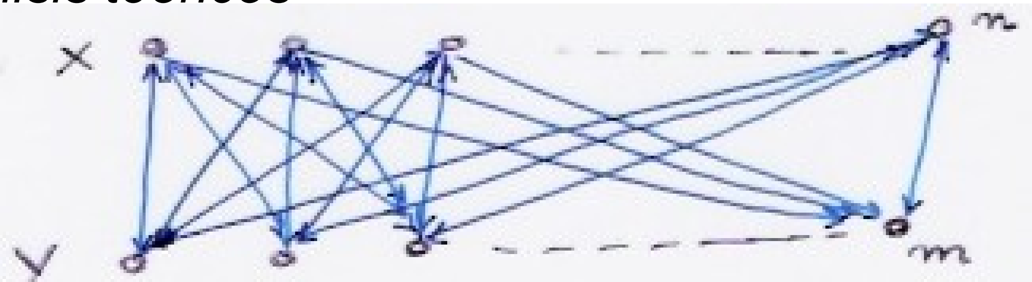
$g(h) = 1 / (1 + \exp(-2\beta h))$ y $\beta = 1/K_B T$

<u>Ventaja</u>: se eluden los estados espurios (mínimos locales) con alta probabilidad.
Desventaja: no podemos determinar a qué memoria, dado un estado inicial, converge el sistema

- *Variable continua* (Hopfield '84): $V_i = g(u_i) = g(\Sigma w_{ij} V_j)$
u_i entrada neta a i V_i salida de *i*

<u>Ventajas</u> (según su autor)
.Mayor realismo biológico
.Posibilidad de aplicaciones mediante implementación circuital
.Mayor facilidad para los análisis teóricos

- *Bidireccional* entrada
(BAM, Kosko)



Problema del Viajante (TSP)

Modelo para el problema (Hopfield-Tank) :

mínimos de energía ↔ mínimos de la función de costo

Restricciones: $\sum_a \eta_{ia} = \sum_i \eta_{ia} = 1 \quad \forall\, i, \forall\, a$

incorporadas a la función de energía.

$$L = \frac{1}{2} \sum_{i,j} d_{ij}\, \eta_{ia} \left(\eta_{j,a+1} + \eta_{j,a-1} \right)$$

$$E = L + \frac{\gamma}{2} \left\{ \sum_a \left(1 - \sum_i \eta_{ia}\right)^2 + \sum_i \left(1 - \sum_a \eta_{ia}\right)^2 \right\}$$

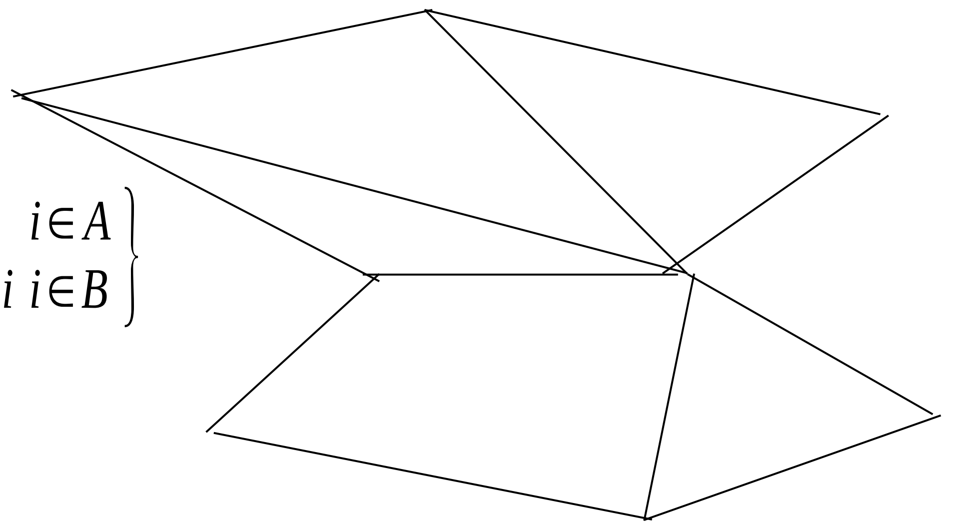
Variables de estado: binarias; ηia = 1 si la ciudad *i* está en la posición *j* del tour, 0 si no.

J.J. Hopfield y D.W. Tank, Computing with Neural Circuits: a Model, Science, ***233****, 625-633.*

Bipartición de grafos

Minimizando:
- Número de arcos que cruzan entre un conjunto y otro.
- Diferencia de cardinalidad entre ambos conjuntos

C_{ij} = 1 si hay un arco entre *i* y *j*, 0 si no



Variables de estado: binarias; $S_i = \begin{cases} +1 & si \;\; i \in A \\ -1 & si \; i \in B \end{cases}$

Restricción: Σs_i
incorporada a la función de energía

Función objetivo (a minimizar): $F = -\sum_{\text{pares}\, i,j} C_{ij} S_i S_j$
sujeto a la restricción anterior

$$\Rightarrow H = F + \mu \left(\sum_i S_i\right)^2 = -\sum_{pares\ i,j} C_{ij} S_i S_j + 2\mu \sum_{pares\ i,j} S_i S_j + N\mu = N\mu - \sum_{pares\ i,j} w_{ij} S_i S_j$$

$$con\ \ w_{ij} = C_{ij} - 2\mu$$

MODELO DE HOPFIELD ESTOCASTICO

Inspirado en la mecánica estadística

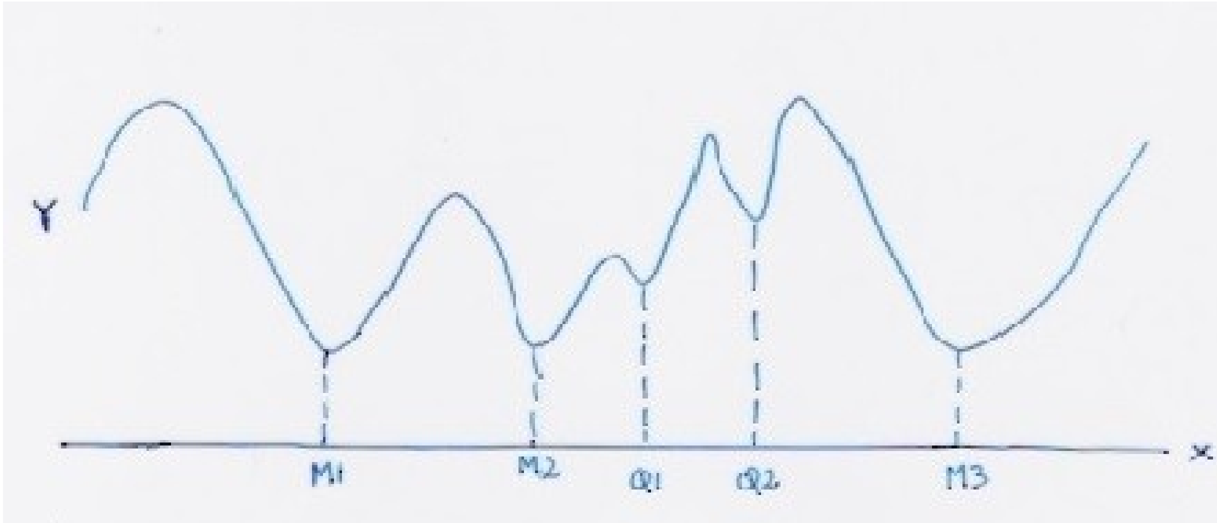
$$\mathrm{Prob}(S_i = \pm 1) = f_\beta(\pm h_i) = \frac{1}{1 + \exp(\mp 2\beta h_i)}$$

función logística, con $\beta = 1/K_B T$

cada vez que la neurona i es elegida para actualización

Ventaja: se eluden los estados espurios (mínimos locales) con alta probabilidad.
Desventaja: no podemos determinar a qué memoria, dado un estado inicial, converge el sistema.

METAFORAS Y PAISAJES



<u>Funciones de memoria asociativa</u>

Reconocimiento: respuesta positiva a la percepción ("está en memoria").

Recuerdo: reconstrucción de una memoria completa basada en un fragmento.

<u>El ruido y sus efectos</u>

Bajo: estados espurios

Medio: desestabilización de espurios, pocos errores en recuperación de memorias

Alto: sin estados espurios, dinámica ergódica

MODELOS NEUROPSICOLOGICOS

Hoffman (I)

Manía: cambios abruptos de un patrón (gestalt) de discurso a otro ↔ comportamiento ruidoso, saltos desde un atractor a otro

Esquizofrenia: múltiples fragmentos combinados, estructura incoherente pero estable. ↔ estados espurios, combinaciones estables de fragmentos atractores

R.E. Hoffman, Computer simulations of neural information processing and the schizophrenia-mania dichotomy, Archives of General Psychiatry, 44, 178

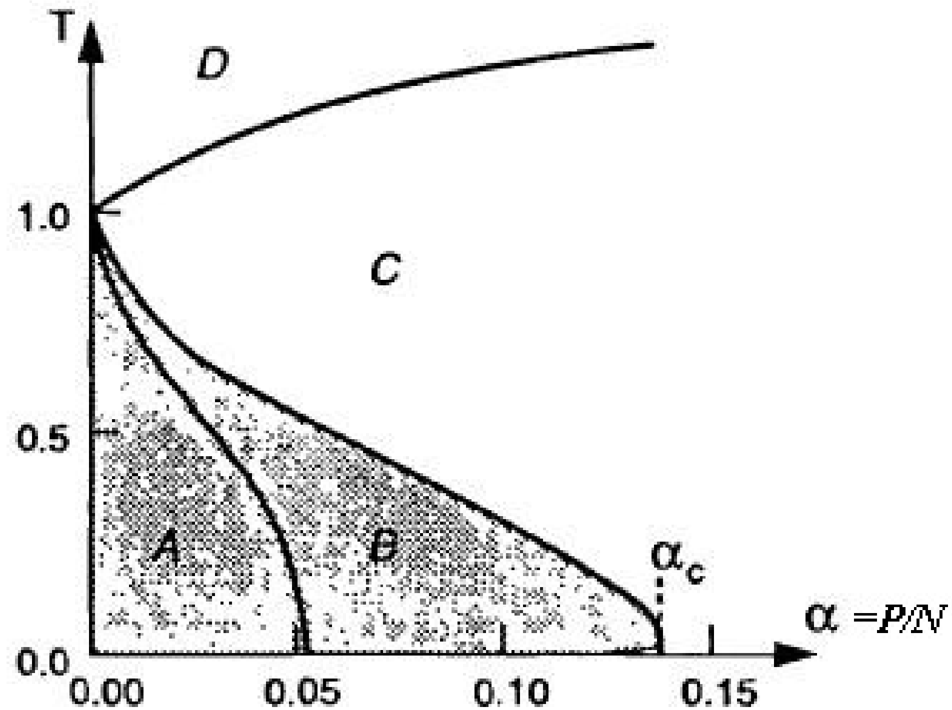
Amit

Epilepsia: reacción rápida a test de reconocimiento; errores ↔ recuperación rápida de un estado espurio

Epilepsia + ritalina: reconocimiento más lento; el número de errores decrece ↔ recuperación de patrón memorizado, pero demandando más tiempo

D.J. Amit, Modeling Brain Function, Cambridge University Press, 1989

DIAGRAMA DE FASES



$\alpha_{c(t)} \approx (1-T)^2$ frontera entre B y C

A: *las memorias son mínimos globales*

B: *E(spin glasses) < E(memorias) → los s.g. son los más estables*

C: *sólo s.g. son estables*

D: *no hay estados estables (salvo* $\langle S_i \rangle = 0$ *) para T suficientemente alta*

ESTABILIDAD

- *Memorias*: A, B

- *Espurios mixtos*: subregión triangular de A U B, decreciente con el número de términos de la combinación lineal, máxima para tres memorias (hasta T_c = 0,46; α_c = 0,03)

- *Spin glasses*: B, C

$\rightarrow$ *0,46 < T < 1* garantizaría $\begin{cases} \text{memorias estables} \\ \text{mixtos inestables y s.g. menos estables si } \alpha \text{ pequeño} \end{cases}$

HOPFIELD ESTOCASTICO

Glauber ('63): dotar a Ising de dinámica estocástica.
→ *sistema en contacto con un baño de calor a temperatura fija*.

Ensamble Canónico:
- *Tamaño* (#partículas) : fijo
- *T* : fija
- *E* : variable

Postulado de Autopromediación (self-averaging): *cualquier propiedad del sistema en equilibrio se puede calcular como sobre un "ensemble" de sistemas idénticos con distribución de probabilidades ≈ exp(-E/kT)*

→ promedio temporal (a intervalos suficientemente grandes) **≈** promedio sobre muchos sistemas (o en diferentes partes) que aparecen en distintos estados (con frecuencia relativa ≈ *exp(-E/kT)* y descorrelacionados entre sí)

Hipótesis: *Balance Detallado* (de la teoría de procesos markovianos)

Sea $W(A/B)$ la matriz de probabilidades de transición de B a A.

$\longrightarrow$ si existe F: $W(A/B)\ F(B) = W(B/A)\ F(A)$ para todo par A, B

entonces $\rho_A(n) \longrightarrow F(A)$ (con $n \rightarrow \infty$) solución estacionaria

Nuestro caso: si $F(A) = exp(-E_A/kT) \longrightarrow$ el proceso se estabiliza en la distribución de Gibbs (Boltzmann):

$\longrightarrow\ \rho_A \approx C\ exp(-E_A/kT)$ $\quad$ C constante de normalización

$C = 1/Z,\ \ Z = \sum_A exp(-E_A/kT)$ *función de partición*

TEOREMA: *si* $w_{ii} = 0$ para todo i , entonces la dinámica de Hopfield-Glauber cumple balance detallado

Demostración: $w_{ii} = 0 \longrightarrow h_i$ independiente de S_i

Supongamos $S_{i,A} = - S_{i,B} \equiv S_i$ i.e. $P(A/B) = Prob\ (S_i \rightarrow -S_i)$

$W(A/B) = (1/N)\ 1/(1 + exp\ (2\beta h_i S_i)$; $W(B/A) = (1/N)\ 1/(1 + exp(-2\beta h_i S_i)$

$\longrightarrow \quad W(A/B) / W(B/A) = exp(-\beta h_i S_i) / exp(\beta h_i S_i) = exp\ (-2h_i S_i /kT)$

$\longrightarrow \quad W(A/B)\ exp\ (-E_B/kT) = W(B/A)\ exp\ (-E_A/kT)$

Ergodicidad: cualquier condición inicial conduce al mismo equilibrio

⟶ *clasificador a baja temperatura*
sorteando espurios (≈ mínimos locales)