El modelo de Hopfield a partir de la función energía

Volvamos a la idea de una única memoria almacenada y miremos la distancia de Hamming entre el estado $\bar{S}(t)$ y la memoria

$$d^1 = \frac{1}{4N} \sum_{i=1}^{N} (S_i - \xi_i^1)^2$$

$$d^1 = \frac{1}{4N} \sum_{i=1}^{N} \left(S_i^2 - 2 S_i \xi_i^1 + (\xi_i^1)^2\right) = \frac{1}{4N} \cdot 2 \sum_i (1 - S_i \xi_i^1)$$

$$= \frac{1}{2N} \sum_{i=1}^{N} 1 - \frac{1}{2N} \sum_{i=1}^{N} S_i \xi_i^1$$

$$= \frac{1}{2} \left(1 - \frac{1}{N} \sum_{i=1}^{N} S_i \xi_i^1\right)$$

$$d^1 = \frac{1}{2} (1 - m^1)$$

donde $m^1 = \frac{1}{N} \sum_{i=1}^{N} S_i \cdot \xi_i^1$

m^1 es la "superposición" entre la memoria y el estado de la red $\bar{S}(t)$

$$d^1(t) = \frac{1}{2} (1 - m^1(t))$$

$$d' = 0 \quad \Longrightarrow \quad m' = 1$$

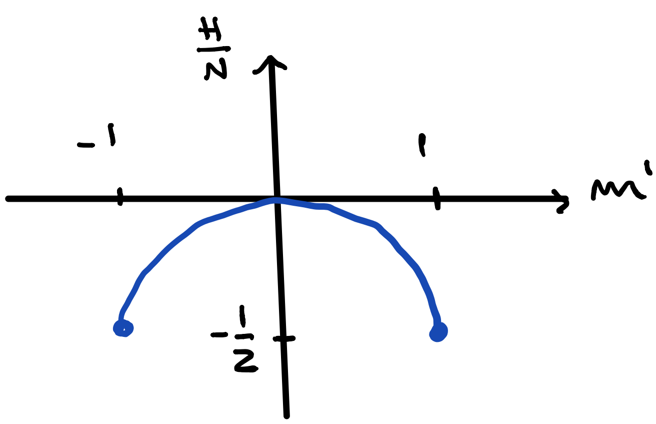
$$d' = 1 \quad \Longrightarrow \quad m' = -1$$

$$d' = 1/2 \quad \Longrightarrow \quad m' = 0$$

Nosotros usamos de ahora en más el parámetro $m'(t)$ para medir el desempeño. Como no nos importa si va a la memoria o a su negativo, en general nos ocuparemos de $|m'|$.

Si en el atractor $|m'|=1$ el sistema llegó a la memoria.

Definamos $$\frac{H}{N} = -\frac{1}{2}(m')^2$$



$$H = -\frac{N}{2}(m')^2 = -\frac{N}{2}\left(\sum_i S_i(t)\xi_i'\right)\left(\sum_j S_j(t)\xi_j'\right)$$

$$= -\frac{N}{2}\frac{1}{N^2}\sum_i \sum_j \xi_i' \xi_j' \; S_i(t)\, S_j(t)$$

$$= -\frac{1}{2}\sum_i \sum_j \frac{\xi_i' \xi_j'}{N} S_i(t) S_j(t)$$

$$= -\frac{1}{2}\sum_i \sum_j \frac{\xi_i' \xi_j'}{N} S_i(t) S_j(t) +$$

$$\frac{1}{2}\sum_i \frac{\xi_i' \xi_i'}{N} S_i(t) S_i(t)$$

$$= -\frac{1}{2}\sum_i \sum_j W_{ij} S_i(t) S_j(t) + \frac{1}{2}$$

Salvo una constante

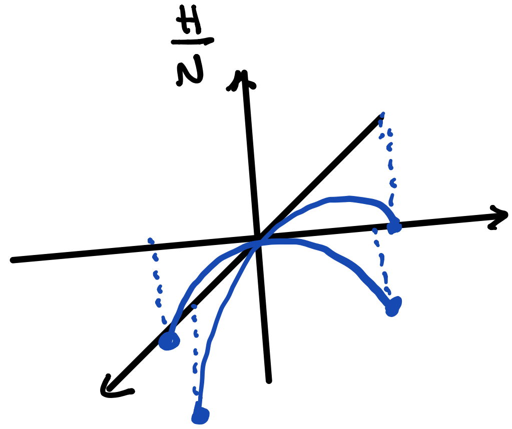
$$\boxed{H = -\frac{1}{2}\sum_i \sum_j W_{ij} S_i(t) S_j(t)}$$

Vemos que H es mínima cuando $m' = 1$ o $m' = -1$ y máxima en $m' = 0$.

¡Es justo lo que queremos!

Si $P=2$, $W_{ij} = \frac{1}{2}\left(\xi_i^1 \xi_j^1 + \xi_i^2 \xi_j^2\right)$

$$\frac{H}{N} = -\frac{1}{2}(m^1)^2 - \frac{1}{2}(m^2)^2$$



Si $S_i = \xi_i^1$, entonces $m^1 = 1$

$m^2 = O(1/\sqrt{N})$

$S_i = \xi_i^2$, entonces $m^1 = O(1/\sqrt{N})$

$m^2 = 1$

Se puede ver que

$$H(t) = -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \frac{1}{N}\left(\xi_i^1 \xi_j^1 + \xi_i^2 \xi_j^2\right) S_i(t) S_j(t)$$

$$= -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} W_{ij}\, S_i(t)\, S_j(t)$$

Ahora podemos generalizar esta expresión:

$$H(t) = -\frac{N}{2} \sum_{\mu=1}^{P} \left(m^{\mu}(t)\right)^2$$

$$= -\frac{N}{2} \sum_{\mu=1}^{P} \left(\frac{1}{N} \sum_{i} \xi_i^{\mu} S_i(t) \right)^2$$

$$= -\frac{N}{2} \sum_{\mu=1}^{P} \frac{1}{N^2} \sum_{i}^{N} \xi_i^{\mu} S_i(t) \sum_{j}^{N} \xi_j^{\mu} S_i(t)$$

$$= -\frac{N}{2} \frac{1}{N^2} \sum_{i}^{N} \sum_{j} \xi_i^{\mu} \xi_j^{\mu} S_i(t) S_j(t)$$

$$= -\sum_{i \neq j} \left(\frac{1}{N} \sum_{\mu}^{P} \xi_i^{\mu} \xi_j^{\mu} \right) S_i(t) S_j(t)$$

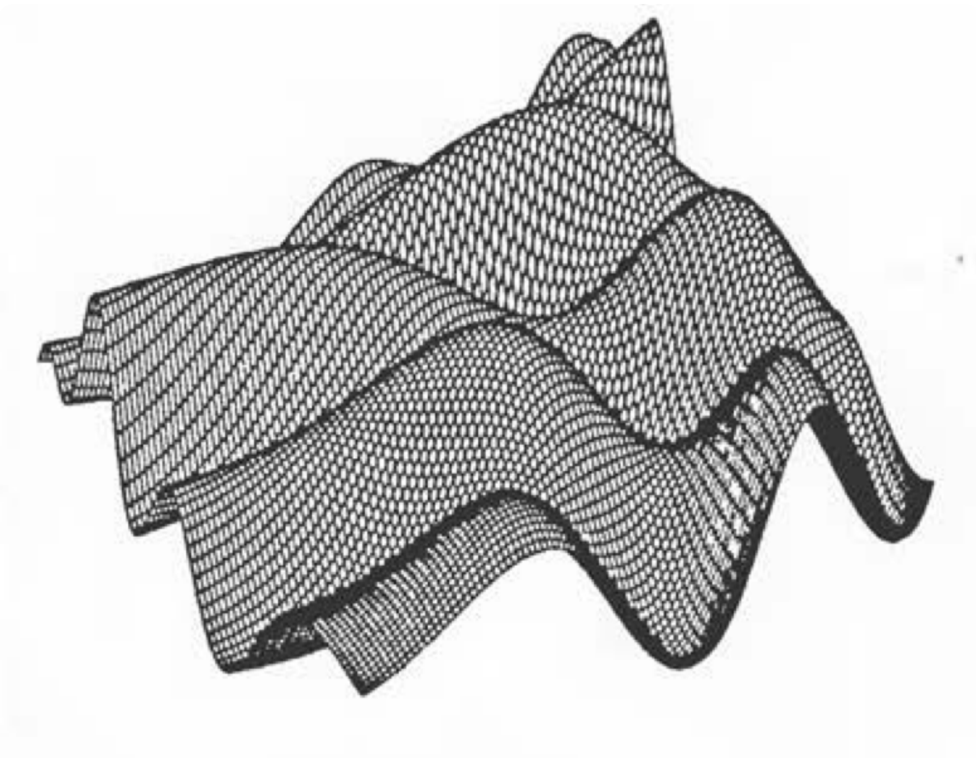
$$+ \underbrace{\frac{1}{N} \sum_{\mu=1}^{P} \sum_{i=1}^{N} \underbrace{\xi_i^{\mu} \xi_i^{\mu}}_{+1} \underbrace{S_i(t) S_i(t)}_{+1}}_{\frac{NP}{N}}$$

$$= -\sum_{i \neq j}^{N} w_{ij} S_i(t) S_j(t) + P$$

Como sumar un término constante no afecta a una definición de la energía, función que debemos "minimizar".

Este es un método para construir reglas.

Cuando ponemos más de dos memorias la función energía tiene otros mínimos.



En el caso más general

$$H = -\frac{m_1^2}{2} - \frac{m_2^2}{2} - \cdots - \frac{m_p^2}{2}$$

$$H(t) = -\frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij} S_i(t) S_j(t)$$

con $$w_{ij} = \frac{1}{N} \sum_{\mu=1}^{p} \xi_i^{\mu} \xi_j^{\mu} \quad , \quad w_{ii} = 0$$

Ahora veremos que a medida que evoluciona la red, $H(t)$ disminuye o se estaciona (no puede crecer).

Supongamos que actualizamos la neurona k.

$$S_k(t+\Delta t) = S'_k = -S_k(t) = -S_k$$

$$H(t+\Delta t) - H(t) = -\frac{1}{2} \sum_i \sum_j w_{ij} S'_i S'_j + \frac{1}{2} \sum_i \sum_j w_{ij} S_i S_j$$

$-$

$$\Delta H = -\frac{1}{2}\sum_{i\neq k}\sum_{j\neq k} \cancel{w_{ij}\left(S_i' S_j' - S_i S_j\right)} = 0$$

$$-\frac{1}{2}\sum_{j\neq k} w_{kj}\left(S_k' S_j' - S_k S_j\right)$$

$$-\frac{1}{2}\sum_{i\neq k} w_{ik}\left(S_i' S_k' - S_i S_k\right)$$

Para $i \neq k$ y $j \neq k$ $\quad S_i' = S_i \quad S_j' = S_j$

$$\Delta H = -\frac{1}{2}\sum_{j\neq k} w_{kj}\left(S_k' S_j - S_k S_j\right)$$

$$-\frac{1}{2}\sum_{i\neq k} w_{ki}\left(S_k' S_i - S_k S_i\right)$$

$$= -\sum_{i\neq k} w_{ki} S_i \left(S_k' - S_k\right)$$

$$= -\left(S_k' - S_k\right)\sum_{i\neq k} w_{ki} S_i$$

$$= +2 S_k \sum_{i\neq k} w_{ki} S_i$$

$$= 2 S_k \underbrace{\sum_{i} w_{ki} S_i}_{\text{esto tiene el signo de } S_k'} - 2\overline{w}_{kk}$$

Entonces, como S_k tiene signo opuesto a S_k'

Entonces el primer término es negativo por definición, ya que

$$S_k' = \text{signo}\left(\sum w_{ki} S_i\right)$$

y

$$S_k' S_k < 0$$

Luego, si se cambia el valor de S_k, la energía H disminuye. Si $S_k = S_k$ queda inalterada. Así

$$\boxed{\Delta H(t+\Delta t) \leq 0}$$

Esto nos dice que la red evoluciona desde la configuración inicial decreciendo el valor de H. Si tenemos suerte, llegaremos al mínimo $m^{\nu} = \pm 1$.

Pero yo les dije, y más adelante lo mostraremos, que si $P > 2$ (si guardamos más de 2 configuraciones) se generan mínimos.

Pensemos un poco sobre las consecuencias de estos nuevos mínimos.

- Si son mínimos locales no es grave, pero si son globales sí, pues cuanto más profundo es el mínimo, mayor es su "cuenca".

- Estos mínimos pueden estar totalmente descorrelacionados de las memorias y sus negativos. Estos mínimos serán estados de NO MEMORIA.

- Estos mínimos pueden estar correlacionados con una o más memorias almacenadas. Estos serán estados de CONFUSIÓN.

Los parámetros fundamentales aquí son los m's, o sea, el vector de parámetros

$$\vec{m} = (m^1, m^2, m^3, \ldots\ldots, m^p)$$

Nosotros "presentamos" a la red una configuración inición inicial

$$\vec{S}(0) = (S_1(0), S_2(0), \ldots\ldots, S_N(0))$$

y la dejamos evolucionar con la dinámica de Hopfield hasta que se "clave", o sea, se estacione en un mínimo de H. Recuerden que H depende de $\bar{S}(t)$ y de la matriz W_{ij}.

En el mínimo medimos $\bar{m}$.

Si llegó a una de las memorias almacenadas $\bar{m}$ al final será

$$\bar{m}(\infty) = \left(\underset{\mu=1}{0},\ \underset{\mu=2}{0},\ \dots\ \underset{\mu=\nu}{\pm 1},\ \dots,\ \underset{\mu=P}{0}\right)$$

Si pasa esto, decimos que la red asoció $\bar{S}(0)$ con la memoria ν, pues si $m^{\nu}(\infty) = \pm 1$ la configuración final estacionaria será la memoria ν (si es $+1$) o a su negativo (si es -1).

H está acotada por abajo, y decrece en el tiempo, lo cual nos garantiza que en un tiempo finito se "clavará" en algún mínimo.

Ahora bien, una solución de la forma:

$$\bar{m}(\infty) = \left(0, 0, \ldots \underset{i}{\pm 1}, \ldots, 0\right)$$

solo es posible si las memorias son ortonormales

$$\frac{1}{N} \sum_{i=1}^{N} \xi_i^\mu \xi_i^\nu = \delta^{\nu\mu} = \begin{cases} 1 & \text{si } \mu = \nu \\ 0 & \text{si } \mu \neq \nu \end{cases}$$

Pero como son aleatorias las memorias, entonces

$$\frac{1}{N} \sum_{i=1}^{N} \xi_i^\mu \xi_i^\nu = \mathcal{O}\left(1/\sqrt{N}\right)$$

Esto dice que las soluciones serán de la forma

$$\bar{m}(\infty) = \left(m^1(\infty), m^2(\infty), \ldots 1, \ldots m^p(\infty)\right)$$

Por ejemplo, si $N = 1000$, $\sqrt{N} = 100$

a $1/\sqrt{N} = 0{,}01$. Una solución

puede ser

$$\bar{m}(\nu) = \Big(0{,}01,\ -0{,}02,\ \ldots\ \underset{\mu=\nu}{\underset{\downarrow}{1}},\ \ldots,\ -0.015\Big)$$

O sea, esperamos $m^\mu = O(1/\sqrt{N})$ si

$\mu \neq \nu$

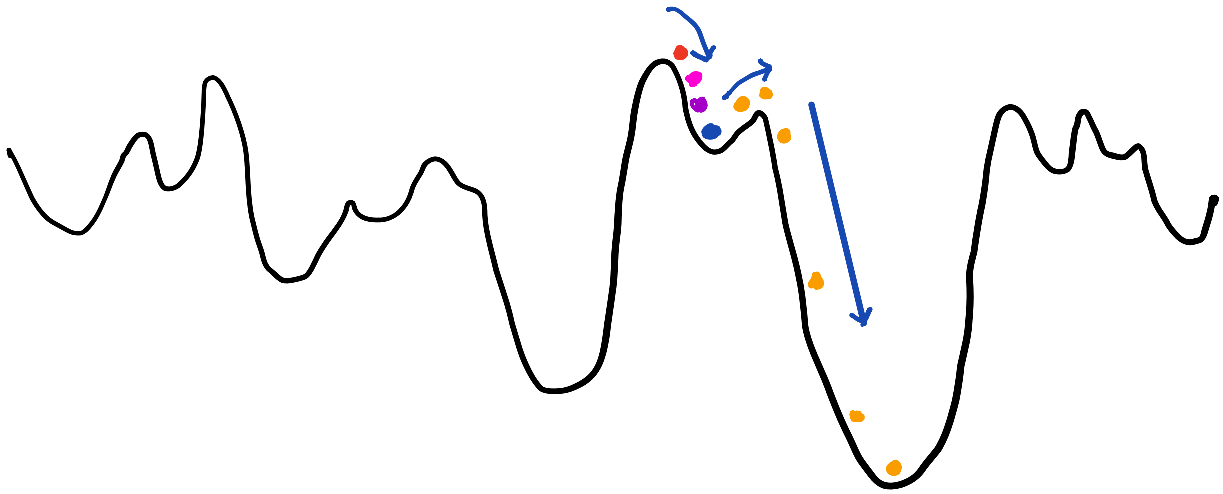
Neuronas Aleatorias.

Vimos que la dinámica

$$S_i(t+\Delta t) = \text{signo}\left(h_i(t)\right)$$

hace que baje siempre H, o quede constante, si $S_i(t+\Delta t) = S_i(t)$.

Para evitar que el sistema se quede "preso" en un mínimo local, vamos a permitir "eventualmente" que la dinámica incremente el valor de H.



Este método permite saltar por los mínimos locales.

Recordemos las neuronas biológicas.
Tres elementos de electricidad importantes son

- el número de vesículas en el axón.
- el número de neurotransmisores en la vesícula
- el número de neurotransmisores que llegan a la dendrita post-sináptica

Entonces h_i es una variable aleatoria cuyo valor medio es

$$\bar{h}_i(t) = \sum_j w_{ij} S_j(t)$$

en tiempo t, y con varianza δ^2

Como son muchas las dendritas la distribución de probabilidades es gaussiana con buena aproximación.

$$P(h_i = h) = \frac{1}{\sqrt{2\pi \delta^2}} e^{-\frac{(h-\bar{h})^2}{2\delta^2}}$$

la neurona dispara si $h_i > 0$

$$P(S_i(t+\Delta t) = +1) = \frac{1}{\sqrt{2\pi\delta^2}} \int_0^\infty e^{-\frac{(h-\bar{h}_i)^2}{2\delta^2}} dh$$

Si $u = \frac{h - \bar{h}_i}{\sqrt{2\pi\delta^2}}$ $\quad du = \frac{dh}{\sqrt{2\pi\delta^2}}$

$$P(S_i(t+\Delta t) = 1) = \int_{\frac{\bar{h}_i}{\sqrt{2\delta^2}}}^\infty \frac{e^{-u^2} du}{\sqrt{\pi}}$$

$$P(S_i(t+\Delta t)=+1) = \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{\bar{h}_i}{\sqrt{2}\,\delta}\right)\right]$$

$$P(S_i(t+\Delta t)=-1) = \frac{1}{2}\left[1 - \mathrm{erf}\left(\frac{\bar{h}_i}{\sqrt{2}\,\delta}\right)\right]$$

o, mejor

$$P(S_i(t+\Delta t)=\pm 1) = \frac{1}{2}\left[1 \pm \mathrm{erf}\left(\frac{\bar{h}_i}{\sqrt{2}\,\delta}\right)\right]$$

$\mathrm{erf}(x)$ es muy bien aproximada, con error menor a 1%, por la función $\tanh(x)$.

$$P(S_i(t+\Delta t)=\pm 1) = \frac{1}{2}\left[1 \pm \tanh(\beta h_i)\right]$$

con $\beta = \dfrac{1}{2\sqrt{2}\,\delta} = \dfrac{1}{T}$

o

$$P(S_i) = \frac{1}{2}\left[1 + S_i \tanh(\beta h_i)\right]$$

$$\boxed{P(S_i) = \frac{1}{2}\left[1 + \tanh(\beta h_i S_i)\right]}$$

Si $T \to 0^+$, $\beta \to \infty$, $\frac{1}{2}\left[1 + \tanh(\beta h_i S_i)\right] \to \text{signo}(h_i)$

Si $T \to \infty$, $\beta \to 0^+$, $\frac{1}{2}\left[1 + \tanh(\beta h_i S_i)\right] \to \frac{1}{2}$

Estados Espurios

Pueden surgir, cuando p aumenta, que se creen estados espurios, o sea, mínimos de H, o si quieren, atractores de la dinámica de la red, que no coinciden con ninguna memoria. Estos pueden ser mínimos no correlacionados con memoria alguna

$$\bar{m} = (0, 0, 0, \ldots, 0)$$

o que mezclan varias memorias

$$\bar{m} = (m_1, m_2, m_3, 0, \ldots \quad 0)$$

En este caso podemos decir que el atractor está confundido entre los 3 primeros memorias.

Ahora tenemos una dinámica aleatoria regida por un parámetro β. En física, $\beta = \frac{1}{T}$. Si $T \to 0$ $\beta \to \infty$, recuperamos la dinámica discreta. El sistema puede "subir" por la función energía con cierta probabilidad.

Calculemos el valor medio de S_i

$$\langle S_i \rangle = \text{Prob}(+1) \cdot 1 + \text{Prob}(-1) \cdot (-1)$$

$$= \frac{1}{1+e^{-2\beta h_i}} - \frac{1}{1+e^{2\beta h_i}}$$

$$= \frac{e^{\beta h_i}}{e^{\beta h_i}+e^{-\beta h_i}} - \frac{e^{-\beta h_i}}{e^{\beta h_i}+e^{-\beta h_i}}$$

$$= \tanh(\beta h_i)$$