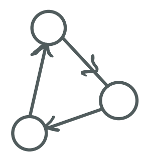
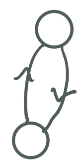
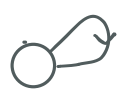
REDES NEURONALES RECURRENTES

Toda red que permite que la información viaje en dos sentidos, es una red recurrente. Esto puede lograrse con

loops: 

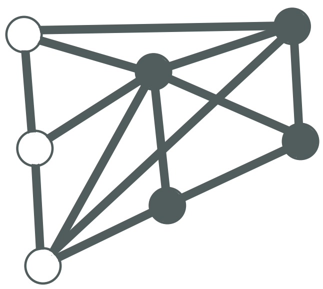
retornos: 

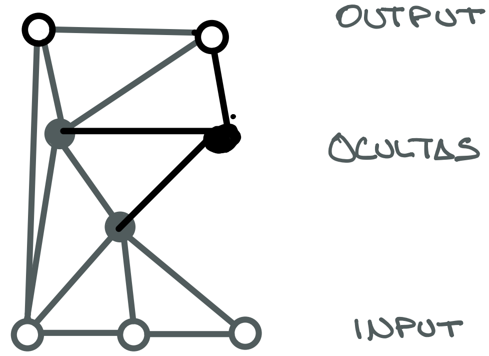
autointeracciones: 

Ya hemos visto la red Hopfield (1982) que dijimos, permite almacenar información holísticamente en la red, y la recupera por asociación. Hay cientos de modificaciones sobre el modelo de Hopfield.

Seguiremos ahora con las llamadas Máquinas de Boltzmann, que son redes neuronales probabilísticas.

MAQUINA DE BOLTZMANN (Hinton 1983)





○ neuronas visibles

● neuronas invisibles

Los acoplamientos entre dos neuronas, cuando existen, son bidireccionales y simétricos:

$$w_{ij} = w_{ji}$$

NEURONAS ESTOCASTICAS

$S_i = +1$ la neurona dispara

$S_i = -1$ la neurona está en reposo

La dinámica es similar a la de Hopfield

$$\text{Prob}(S_i(t+\Delta t)=+1) = g(h_i) = g\left(\sum_j w_{ij} S_j(t)\right)$$

$$\text{Prob}(S_i(t+\Delta t)=-1) = 1 - g(h_i)$$

$$h_i(t) = \sum_j w_{ij} S_j(t)$$

$$\beta \equiv \frac{1}{T}$$

$$g(h) = \frac{1}{1+e^{-2\beta h}}$$

T: medida del ruido

Por ser los acoplamientos simétricos, el sistema tiene una función de Liapunov que cumple que, si $T=0$, esta función solo decrece o se mantiene en el tiempo a medida que avanza el tiempo.

$$H\{S_i\}(t) = -\frac{1}{2}\sum_i \sum_j w_{ij}\, S_i(t)\, S_j(t).$$

El sistema, a un dado valor de T, recorrerá las 2^N posibles configuraciones con probabilidad

$$P\{S_i(t)\} = \frac{e^{-\beta H(\{S_i\})}}{Z}$$

$$Z = \sum_{\{S_i\}}^{2^N} e^{-\beta H(\{S\})}$$

Con esta probabilidad podemos calcular promedios de cualquier cantidad que depende de $\{S\}$

$$\langle X \rangle = \sum_{\{S_i\}}^{2^N} P(\{S_i\})\, X(\{S_i\})$$

¿Qué hace esta red? Nos concentramos en las neuronas visibles solamente, a las que denominamos S_i^α. A las ocultas las llamamos S_j^β.

La red tiene que recorrer las posibles configuraciones de forma tal que las neuronas visibles recorran sus subestados con cierta probabilidades.

Si tenemos 2 neuronas visibles tenemos 4 estados

	S_1^α	S_2^α	Prob
$\alpha=1$	$+1$	$+1$	$0.6 = R_1$
$\alpha=2$	$+1$	-1	$0.1 = R_2$
$\alpha=3$	-1	$+1$	$0.1 = R_3$
$\alpha=4$	-1	-1	$0.2 = R_4$

$2^2 = 4$

$$P_\alpha = \sum_\beta P_{\alpha\beta} = \sum_\beta \frac{e^{-\beta H_{\alpha\beta}}}{Z}$$

$$Z = \sum_{\alpha\beta} e^{-\beta H_{\alpha\beta}}$$

$$H_{\alpha\beta} = -\frac{1}{2}\sum_i \sum_j w_{ij} S_i^{\alpha\beta} S_j^{\alpha\beta}$$

$$E = \sum_\alpha R_\alpha \ln\left(\frac{R_\alpha}{R_\beta}\right)$$

R_α : probabilidad de $\{S_\alpha\}$ deseada

$$W_{ij}^{\text{nuevo}} = W_{ij}^{\text{anterior}} + \Delta W_{ij}$$

$$= W_{ij}^{\text{anterior}} - \eta \frac{\partial E}{\partial w_{ij}}$$

$$= W_{ij}^{\text{anterior}} + \eta \sum_\alpha \left(\frac{R_\alpha}{P_\alpha}\right)\left(\frac{\partial P_\alpha}{\partial w_{ij}}\right)$$

$$\frac{\partial P_\alpha}{\partial w_{ij}} = \frac{\beta \sum_\beta e^{-\beta H_{\alpha\beta}} S_i^{\alpha\beta} S_j^{\alpha\beta}}{Z} - \frac{\left(\sum_\beta e^{-\beta H_{\alpha\beta}}\right)\beta \sum_{\lambda\mu} e^{-\beta H_{\lambda\mu}} S_i^{\lambda\mu} S_j^{\lambda\mu}}{Z}$$

$$= \beta\left(\sum_\beta S_i^{\alpha\beta} S_j^{\alpha\beta} P_{\alpha\beta} - P_e \langle S_i S_j \rangle\right)$$

$$\Delta w_{ij} = \eta\beta\left[\sum_\alpha \frac{R_\alpha}{P_\alpha} \sum_\beta S_i^{\alpha\beta} S_j^{\alpha\beta} P_{\alpha\beta} - \sum R_\alpha \langle S_i S_j \rangle\right]$$

$$= \eta\beta\left[\sum_{\alpha\beta} R_\alpha P_{\beta|\alpha} S_i^{\alpha\beta} S_j^{\alpha\beta} - \sum_\alpha R_\alpha \langle S_i S_j \rangle\right]$$

$$= \eta\beta\left[\langle S_i S_j\rangle_{clamped} - \langle S_i S_j\rangle_{libres}\right]$$

$$\Delta w_{ij} = \eta\beta\left[\overline{\langle S_i S_j\rangle}_{sujetados} - \langle S_i S_j\rangle_{libres}\right]$$

$$\overline{\langle S_i S_j\rangle}_{sujetados} = \sum_{\alpha\beta} R_\alpha P_{\beta|\alpha} S_i^{\alpha\beta} S_j^{\alpha\beta}$$

La red ajusta todos los acoplamientos de la red, que sigue la distribución de Maxwell-Boltzmann, de forma tal que las neuronas visibles recorran sus estados α con probabilidad R_α.

Esta es una red generativa, pues genera configuraciones con probabilidades deseadas.

Para deep learning se usan redes de Boltzmann restringidas (RBM).

Nota: esto no es más que una dinámica Monte Carlo

$$\text{Prob}(S_i \to -S_i) = \frac{1}{1 + e^{\beta \Delta H_i}}$$

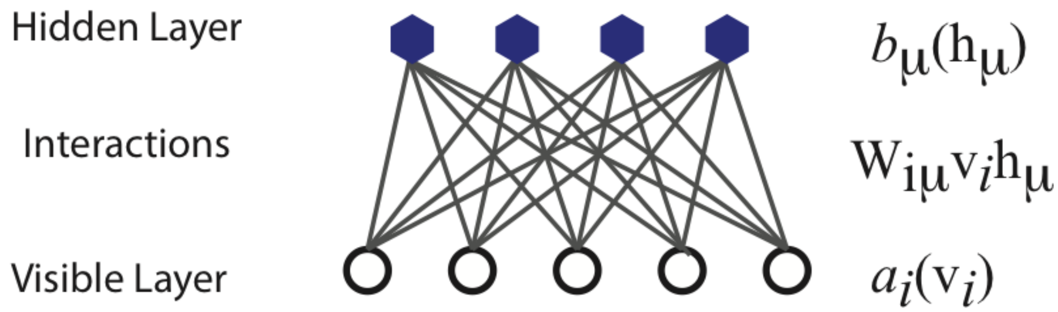
ΔH_i: cambio en H al invertir S_i

Es lento pero poderoso. Se usa simulated annealing (templado simulado) o simulate annealing optimizado (máquina de Cauchy).

- En el nivel más alto, debemos actualizar cada peso muchas veces para que converja, calculando en una simulación Monte Carlo $\langle S_i S_j \rangle_{\text{sujetos}}$ y $\langle S_i S_j \rangle_{\text{libres}}$.
- Luego variamos T, descendiéndola lentamente
- Corregimos las sinapsis.

MÁQUINA DE BOLTZMANN RESTRINGIDA

$$E(v,h) = -\sum a_i(v_i) - \sum b_\mu(h_\mu) - \sum_{i\mu} w_{i\mu} v_i h_\mu$$



Capa Bernoulli

$$a_i(v_i) = \begin{cases} a_i v_i & \text{si } v_i \in \{0,1\} \text{ binarias} \\ \dfrac{v_i^2}{2\sigma_i^2} & \text{si } v_i \in \mathbb{R} \end{cases}$$

capa gaussiana

$$b_\mu(h_\mu) = \begin{cases} b_\mu h_\mu & h_\mu \in \{0,1\} \\ \dfrac{h_\mu^2}{2\sigma_\mu^2} & h_\mu \in \mathbb{R} \end{cases}$$