Redes de Hopfield

Verónica E. Arriola-Rios

Facultad de Ciencias, UNAM

22 de febrero de 2021

Intro

Temas

1. Intro
   - Definición
   - Paisaje energético

Definición

Definición (Red de Hopfield)

Las redes de *Hopfield* son redes neuronales recurrentes no dirigidas (simétricas) cuya activación y entrenamiento está controlado por una **función de energía**.

- Los valores de activación para la neuronas son binarios:

$$s_i \in \{0, 1\} \text{ ó } \{-1, 1\} \tag{1}$$

- Las aristas tienen pesos $w_{ij} \in \mathbb{Z}$ asociados, que describen el tipo de conexión entre las neuronas correspondientes: **inhibitoria** si $w < 0$ o **exitatoria** si $w > 0$.
- Cada neurona también puede tener asociado un **sesgo**, que indica la tendencia inicial de la neurona a estar encendida o apagada.

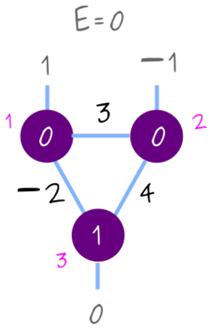
Energía

- Se define una **función de energía** que controla la dinámica de la red:

$$E = -\sum_{i} s_i b_i - \sum_{i<j} s_i s_j w_{ij} \tag{2}$$

  donde:
  - s_i es el valor de activación de la i-ésima neurona,
  - b_i es el sesgo, y cada neurona puede tener uno,
  - w_{ij} es el peso de la conexión en las neuronas i y j.
- La **bondad** es el negativo de la energía $-E$.

Cálculo de la energía



$$E(0,0,0) = 0$$
$$E(0,0,1) = -(0)1 + (0)(-1) + (1)0 - 0x0x3 + 0x1x(-2) + 1x0x4 = 0$$
$$E(0,1,0) = 0$$
$$E(0,1,1) = 1 - 4 = -3$$
$$E(1,0,0) = -1$$
$$E(1,0,1) = -1 + 2 = 1$$
$$E(1,1,0) = -1 + 1 - 3 = -3$$
$$E(1,1,1) = -1 + 1 - 3 + 2 - 4 = -5$$

Temas

1. Intro
   - Definición
   - Paisaje energético

Paisaje energético

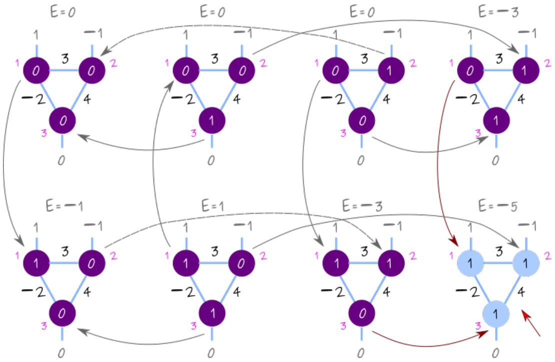


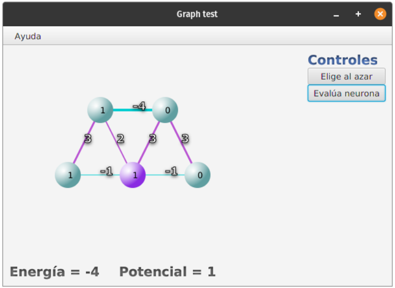
Figura: Posibles estados para una red de Hopfield con tres nodos. Pesos positivos favorecen que nodos adyacentes tengan los mismos valores de activación. Pesos negativos favorecen que tomen valores contrarios.

Red de Hopfield

- Las redes de *Hopfield* son redes neuronales recurrentes no dirigidas (simétricas) cuya activación y entrenamiento está controlado por una **función de energía**.
- Los valores de activación de cada neurona se determinan al aplicar una función que busca colocar a la red en un estado de mínima energía.
- Controlando los valores de los pesos de las conexiones entre las neuronas es posible convertir algunos de sus estados de activación posibles en **atractores estables** (mínimos de la energía a los que tiende el sistema).
- Estos estados pueden ser utilizados para almacenar memorias como patrones distribuidos de actividad. G. Hinton 2012
- El proceso de almacenamiento es la aplicación del algoritmo que permite asignar los pesos que provocarán que los estados deseados sean los mínimos de la energía de la red.

Mínimo local y global

G. Hinton 2012 ilustra el principio detrás de la recuperación de las memorias con un ejemplo sencillo:



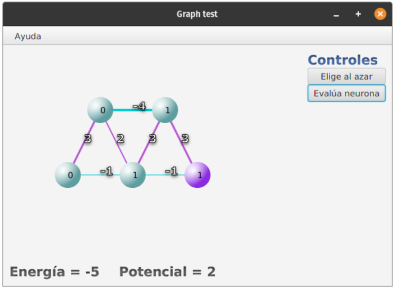


Figura: **Izquierda:** mínimo local. **Derecha:** mínimo global. En ambos casos ya ninguna neurona cambia su valor cuando se usa la regla del perceptrón.

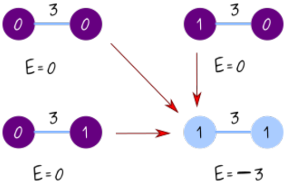
Sistemas dinámicos

Sistemas dinámicos

La dinámica de las **redes recurrentes**, en general, es un sistema dinámico, cuyo comportamiento puede caer en alguno de los siguientes casos:

- Puede evolucionar hasta estabilizarse en un estado estable.
- Oscilar entre estados.
- Seguir trayectorias caóticas de transiciones de estados, que no pueden ser predichos a menos que se conozca su estado inicial con precisión infinita.

Las redes de Hopfield se estabilizan en un **estado estable**, donde cualquier transición desde ese estado, regresa a la red al mismo estado.



Evolución

- El cambio en la energía, al modificar el valor de activación de una neurona se calcula diferenciando E:

$$\begin{aligned}\Delta E_i &= E(s_i = 0) - E(s_i = 1) \\ &= 0 - \left[-\sum_i (1) b_i - \sum_j (1) s_j w_{ij} \right] \\ \Delta E_i &= b_i + \sum_j s_j w_{ij} \end{aligned} \tag{3}$$

  Para decidir si una neurona se activa o no, se aplica la función de decisión **umbral binaria** (escalón) sobre ΔE_i.

- Una vez que la red ha memorizado un patrón, si se le inicia en un estado con algunas neuronas con los valores de activación correctos, la evolución de su dinámica hará que las demás neuronas converjan al valor memorizado.

En forma matricial

Considerando que donde no haya arista $w_{ij} = 0$:

$$\vec{s}(t+1) = \vec{b} + \begin{bmatrix} w_{11} & \cdots & w_{1n} \\ & \ddots & \\ w_{n1} & \cdots & w_{nn} \end{bmatrix} \vec{s}(t) \tag{4}$$

$$= \vec{b} + \boldsymbol{W}\vec{s}(t) \tag{5}$$

CUIDADO:

- G. Hinton 2012 adiverte que relizar la actualización en forma simultánea para todas las neuronas no garantiza el descenso de la energía y se puede caer en oscilaciones.
- Es mejor actualizar sus valores uno por uno en forma aleatoria.

Entrenamiento

Entrenamiento

Si $s_i \in \{-1, 1\}$:

- Para cada patrón ξ^μ que queramos que la red memorize, calculamos su contribución:

$$\Delta w_{ij} = \xi_i \xi_j$$

  En forma matricial:

$$M_{ij}^{\mu} = \frac{1}{N} \xi_i^{\mu} \xi_j^{\mu}$$

- Para memorizarlos todos los patrones, sumamos todas las contribuciones:

$$M_{ij} = \frac{1}{N} \sum_{\mu=0}^{P-1} \xi_i^{\mu} \xi_j^{\mu}$$

Entrenamiento

Si $s_i \in \{0, 1\}$:

- Para cada patrón ξ^μ que queramos que la red memorize, calculamos su contribución:

$$\Delta w_{ij} = 4\left(\xi_i - \frac{1}{2}\right)\left(\xi_j - \frac{1}{2}\right)$$

Memoria

- El entrenamiento provoca que los patrones ξ indicados se conviertan en mínimos de la energía.
- Las memorias son recuperables mientras estos patrones sean ortogonales entre sí.
- Cuando los patrones comienzan a parecerse, se crean mínimos espúreos con mezclas entre las memorias involucradas y éstas comienzan a mezclarse.
- Por lo anterior, la máxima capacidad de memorias almacenables en la red es de alrededor de $0.15N$, donde N es el número de neuronas en la red.

Desaprendizaje

- Cuando, dado un estado inicial, la red converge a un minímo no deseado, se modifican los pesos aplicando la operación contraria a la de aprendizaje.
- Crick y Mitchison propusieron que este método podría modelar la función y funcionamiento de los sueños en el cerebro humano.

Perceptrón/Pseudo-verosimilitud I

Elizabeth Gardiner propuso almacenar un vector a la vez (en estadística se le conoce como ajustar la pseudo-verosimilitud). El procedimiento es el siguiente:

1. Se asignan todos los valores de activación a los valores del patrón que se desea aprender.
2. Para cada neurona se decide, dados los pesos y valores de activacion de sus vecinas, si tomaría el valor correcto o no. Si la respuesta es no, se usa la regla de entrenamiento de perceptrón para modificar los pesos adyacentes.
    - Si el resultado es correcto no hacer nada.
    - Si sacó 0 erróneamente,

$$\vec{w}+ = \vec{x} \tag{6}$$

    - Si sacó 1 erróneamente,

$$\vec{w}- = \vec{x} \tag{7}$$

Perceptrón/Pseudo-verosimilitud II

3. Se repite el procedimiento para cada memoria hasta que haya convergencia (los pesos ya no cambien).
4. Aunque se pueden almacenar un poco más de memorias así, no habrá convergencia si ya se ha superado la capacidad de la red.

Referencias I

- Fee, Michale (2021). *Hopfield Networks*. MIT. URL: https://ocw.mit.edu/courses/brain-and-cognitive-sciences/9-40-introduction-to-neural-computation-spring-2018/lecture-videos/20.-hopfield-networks.
- Hinton, Geoffrey (2012). *Neural Networks for Machine Learning*. English. University of Toronto. URL: https://class.coursera.org/neuralnets-2012-001/.
- Hinton, Geoffrey E. (2007). *Boltzmann machine*. English. URL: http://www.scholarpedia.org/article/Boltzmann_machine.
- Hopfield, John J. (2007). *Hopfield network*. English. URL: http://www.scholarpedia.org/article/Hopfield_network.

Licencia