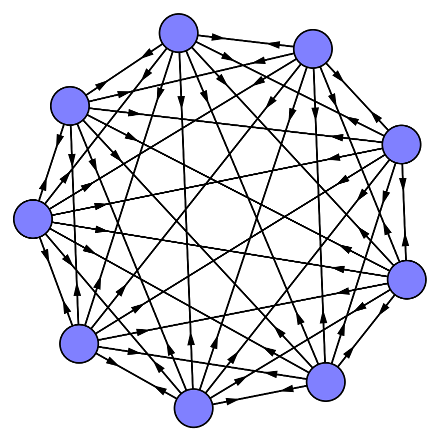
Hopfield Networks

Hopfield Network (discrete version)



- Recurrent network
- Binary units $(-1, 1)$
- Symmetric connections

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

- No self-connections

$$w_{ii} = 0$$

- Information stored in fixpoint attractors

Network Dynamics

- Units are updated according to a thresholding rule:

$$s_i \leftarrow \begin{cases} 1 & \text{when} \quad \sum_j s_j w_{ij} \geq 0 \\ -1 & \text{when} \quad \sum_j s_j w_{ij} < 0 \end{cases}$$

- Update can be made in any order
  - Sequential
  - Random
- Parallel updates possible

Energy

A global measure of the state of the network

$$E = -\frac{1}{2}\sum_i \sum_j w_{ji} s_j s_i$$

- Energy is useful for understanding network dynamics

How does the energy change when a node is updated?

$$E = -\frac{1}{2}\sum_i \sum_j w_{ji} s_j s_i$$

Suppose that s_k changes state from -1 to 1
Observation: Only row k and column k are affected

$$\Delta E = -\sum_i w_{ki} s_i - \sum_j w_{jk} s_j = -2\sum_i w_{ik} s_i$$

$$s_i \leftarrow \begin{cases} 1 & \text{when} \quad \sum_j s_j w_{ij} \geq 0 \\ -1 & \text{when} \quad \sum_j s_j w_{ij} < 0 \end{cases}$$

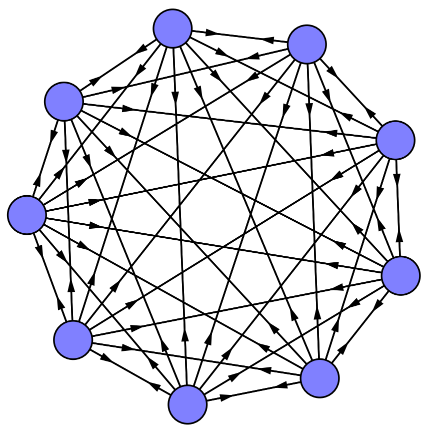
This change only happens when $\sum_j s_j w_{ij} \geq 0$
Thus: $\Delta E < 0$

Convergence property

When connections are symmetric, the network will always converge to a fixpoint attractor

- Holds for sequential updating
- Holds for random updating
- Almost holds for parallel updating

Hopfield Network (continuous version)



- Graded unit values
- First-order differential equation
- Similar to membrane potential in real neurons

Network dynamics of the continuous version

$$\tau \frac{dv_i}{dt} = -v_i + \sum_i \phi(v_j) w_{ji}$$

- v_i is the "inner state" of node i
- $\phi(v)$ is a squashing function

$$s_i = \phi(v_i)$$

- Often used squashing function

$$\phi(v) = \tanh(v) = \frac{e^v - e^{-v}}{e^v + e^{-v}}$$

Energy function needs to be more complicated

$$E = -\frac{1}{2} \sum_i \sum_j w_{ji} s_j s_i + \sum_j \int_0^{s_j} \phi^{-1}(x) dx$$

Lyapunov Function

$$E = -\frac{1}{2} \sum_i \sum_j w_{ji} s_j s_i + \sum_j \int_0^{s_j} \phi^{-1}(x) dx$$

$$\begin{aligned}
\frac{dE}{dt} &= -\sum_j \sum_i w_{ji} s_i \frac{ds_j}{dt} + \sum_j \phi^{-1}(s_j) \frac{ds_j}{dt} \\
&= -\sum_j \left(\sum_i w_{ji} s_i - \phi^{-1}(s_j) \right) \frac{ds_j}{dt} \\
&= -\sum_j \left(\sum_i w_{ji} \phi(v_i) - v_j \right) \frac{ds_j}{dt} \\
&= -\sum_j \tau \frac{dv_i}{dt} \frac{ds_j}{dt} = -\tau \sum_i \frac{dv_i}{ds_i} \left(\frac{ds_i}{dt} \right)^2 \leq 0
\end{aligned}$$

- Energy can only go down ($\frac{dE}{dt} \leq 0$)
- $\frac{dE}{dt} = 0$ only happens when there is no change at all
- There is a lower limit to E

The continuous Hopfield Network will always converge to a fixpoint attractor

Spurious Attractors

Fixpoints that have not been explicitly stored

- Mirror patterns
- Mixtures of stored patterns

Storage Capacity

- Number of patterns: $\approx 0.13N$
- Similarities between patterns reduce this capacity
- Sparse patterns give better capacity