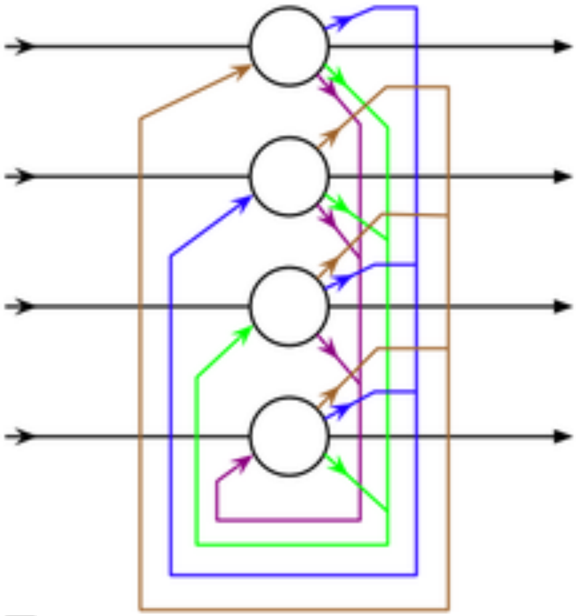
A **Hopfield network** is a form of recurrent artificial neural network invented by John Hopfield. Hopfield nets serve as content-addressable memory systems with binary threshold units. They are guaranteed to converge to a local minimum, but convergence to one of the stored patterns is not guaranteed. Furthermore, Hopfield networks provide a model for understanding human memory.

Structure



A Hopfield net with four nodes.

The units in Hopfield nets are binary threshold units, i.e. the units only take on two different values for their states and the value is determined by whether or not the units' input exceeds their threshold. Hopfield nets can either have units that take on values of 1 or -1, or units that take on values of 1 or 0. So, the two possible definitions for unit *i*'s activation, a_i, are:

(1)
$$a_i \leftarrow \begin{cases} 1 & \text{if } \sum_j w_{ij}s_j > \theta_i, \\ -1 & \text{otherwise.} \end{cases}$$

(2)
$$a_i \leftarrow \begin{cases} 1 & \text{if } \sum_j w_{ij}s_j > \theta_i, \\ 0 & \text{otherwise.} \end{cases}$$

Where:

- w_{ij} is the strength of the connection weight from unit j to unit i (the weight of the connection).
- s_j is the state of unit j.
- θ_i is the threshold of unit i.

The connections in a Hopfield net typically have the following restrictions:

- $w_{ii} = 0, \forall i$ (no unit has a connection with itself)
- $w_{ij} = w_{ji}, \forall i, j$ (connections are symmetric)

The requirement that weights be symmetric is typically used, as it will guarantee that the energy function decreases monotonically while following the activation rules, and the network may exhibit some periodic or chaotic behaviour if non-symmetric weights are used. However, Hopfield found that this chaotic behaviour is confined to relatively small parts of the phase space, and does not impair the network's ability to act as a content-addressable associative memory system.

Hopfield nets have a scalar value associated with each state of the network referred to as the "energy", E, of the network, where:

$$E = -\frac{1}{2} \sum_{i,j} w_{ij} s_i s_j - \sum_i \theta_i \ s_i$$

This value is called the "energy" because the definition ensures that if units are randomly chosen to update their activations the network will converge to states which are local minima in the energy function (which is considered to be a Lyapunov function). Thus, if a state is a local minimum in the energy function it is a stable state for the network. Note that this energy function belongs to a general class of models in physics, under the name of Ising models; these in turn are a special case of Markov networks, since the associated probability measure, the Gibbs measure, has the Markov property.