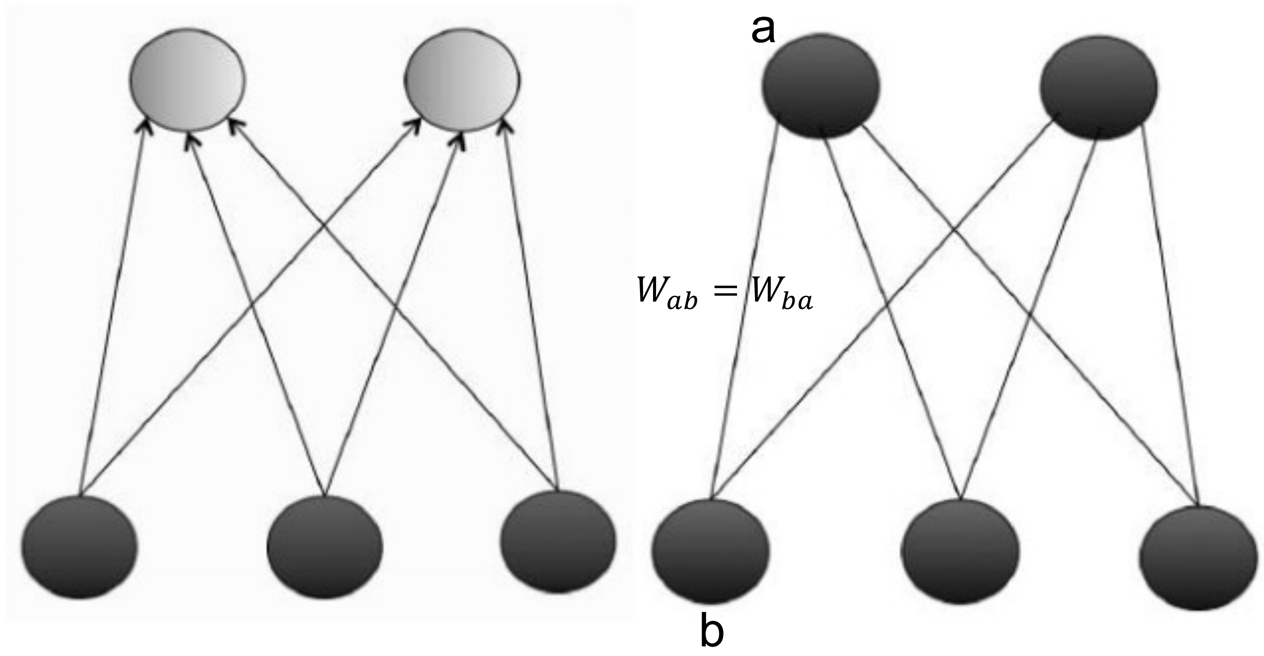
Chapter 17

Hanchi Ren

08.07.2021

Symmetric Weights and Deep Belief Networks



On the left the weighted links are directional so that whether or not the lighter-coloured nodes fire affects the firing of the darker-coloured nodes, but not vice-versa, while on the right the two layers are symmetric, so that the firing of the upper layer can change the firing of the lower layer in the same way as from bottom to top.

ENERGETIC LEARNING: THE HOPFIELD NETWORK

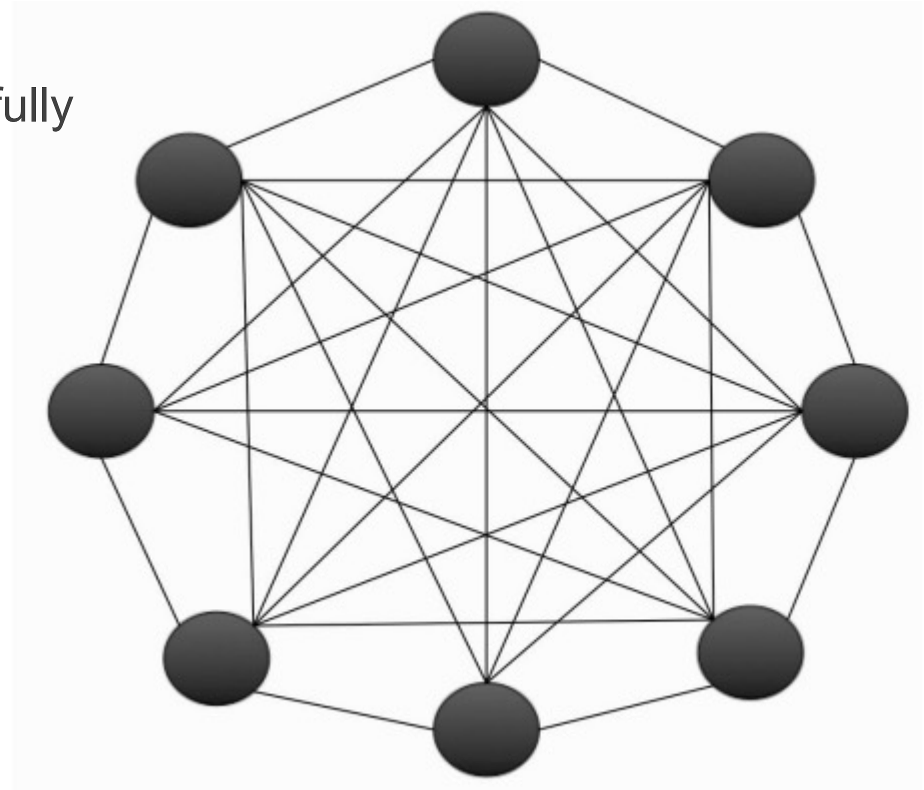
- <<Neural networks and physical systems with emergent collective computational abilities>>, 1982, John Hopfield
- It is a kind of Energy Based Model (EBM). The conclusion of statistical mechanics shows that any probability distribution can be transformed into an EBM.
- An energy function is sign to each weight, and the state of the model is presented by the energy value.

ENERGETIC LEARNING: THE HOPFIELD NETWORK

1. Associative Memory
    a) It is one of the most useful memory types and also known as contxt-addressable memory.
    b) Once we learned many patterns, if there comes a new pattern, the memory reproduces whichever of the learned patterns most closely resembles it.
    c) Compare a new person with those have been in memory.
    d) Face blindness often happens when you were seeing a foreign movie.

ENERGETIC LEARNING: THE HOPFIELD NETWORK

1. Associative Memory in Hopfield Network
   a) The Hopfield network consists of a set of neurons what are fully connected.
   b) Hebb's rule: $W_{ij} = S_i S_j = \frac{1}{N}\sum_{n=1}^{N} S_i(n) S_j(n)$
   c) Synchronous Update: $S_i^t = sign(\sum_j W_{ij} S_j^{(t-1)})$
   d) Asynchronous Update: $S_i = sign(\sum_j W_{ij} S_j)$



ENERGETIC LEARNING: THE HOPFIELD NETWORK

The Hopfield Algorithm

- **Learning**
  - take a training set of N d-dimensional inputs $\mathbf{x}(1), \mathbf{x}(2), \ldots, \mathbf{x}(N)$ with elements ± 1
  - create a set of d neurons (or $d+1$ including a bias node that is permanently set to 1) and set the weights to:

$$w_{ij} = \begin{cases} \frac{1}{N} \sum_{n=1}^{N} x_i(n) x_j(n) & \forall i \neq j \\ 0 & \forall i = j \end{cases} \tag{17.5}$$

- **Recall**
  - present the new input $\mathbf{x}$ by setting the states s_i of the neurons to x_i
  - **repeat**
    * update the neurons using:

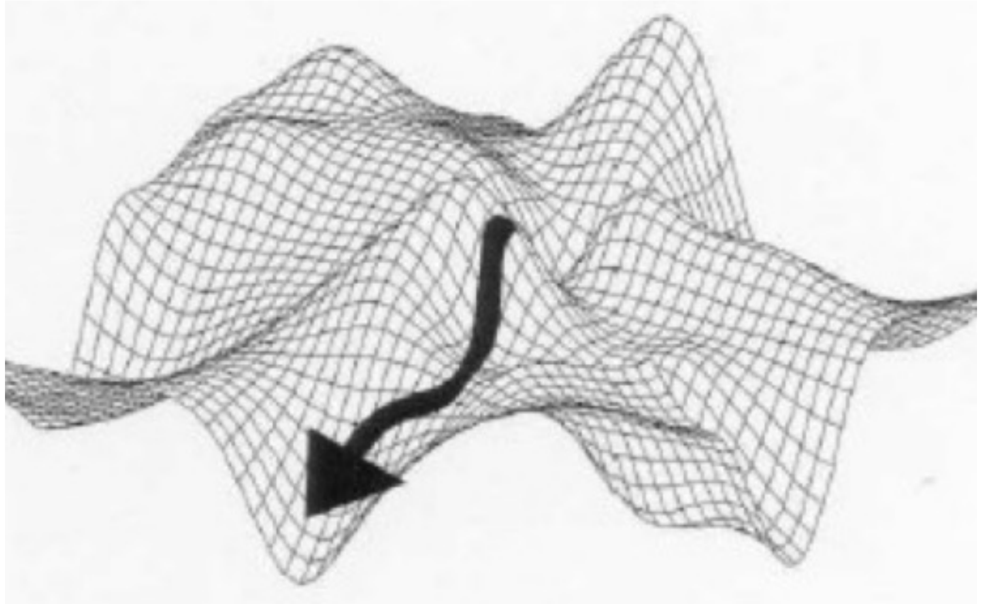
$$s_i^{(t)} = \text{sign}\left(\sum_j w_{ij} s_j^* \right), \tag{17.6}$$

    where s_j^* is $s_j^{(t)}$ if that neuron has been updated already, and $s_j^{(t-1)}$ if it has not. This means that for synchronous update $s_j^* = s_j^{(t-1)}$ for every node, while for asynchronous update it could be either value.
  - **until** the network stabilises
  - read off the states s_i of the neurons as the output

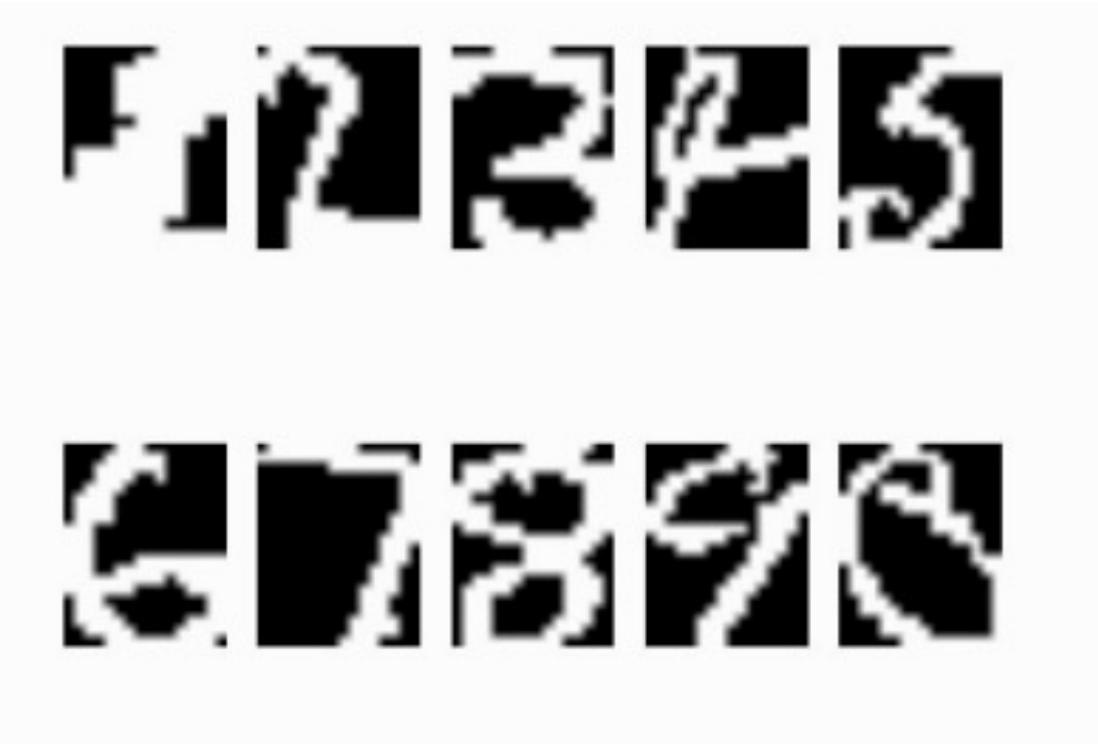
ENERGETIC LEARNING: THE HOPFIELD NETWORK

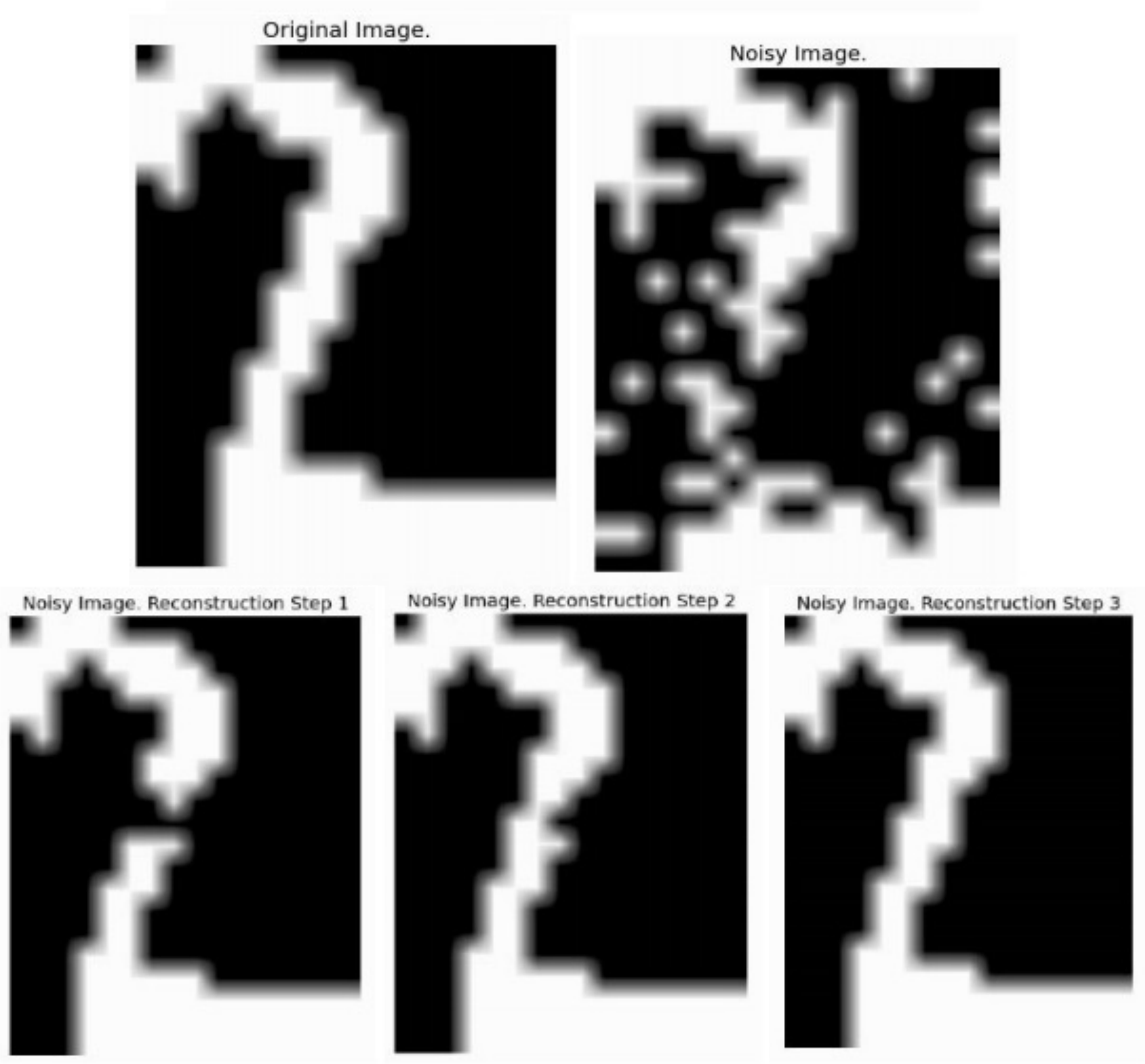
$$
\begin{aligned}
H &= -\frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d} w_{ij} s_i s_j \\
&= -\frac{1}{2}\mathbf{s}\mathbf{W}\mathbf{s}^T
\end{aligned}
$$

2. Energy Function
   a) Energy functions are used in physics to compute how much energy a system has.
   b) Lower energies is better.
   c) It is used for explaining how and why the Hopfield Network works.
   d) We can imagine the change in energy as the network learns as an **energy landscape**.

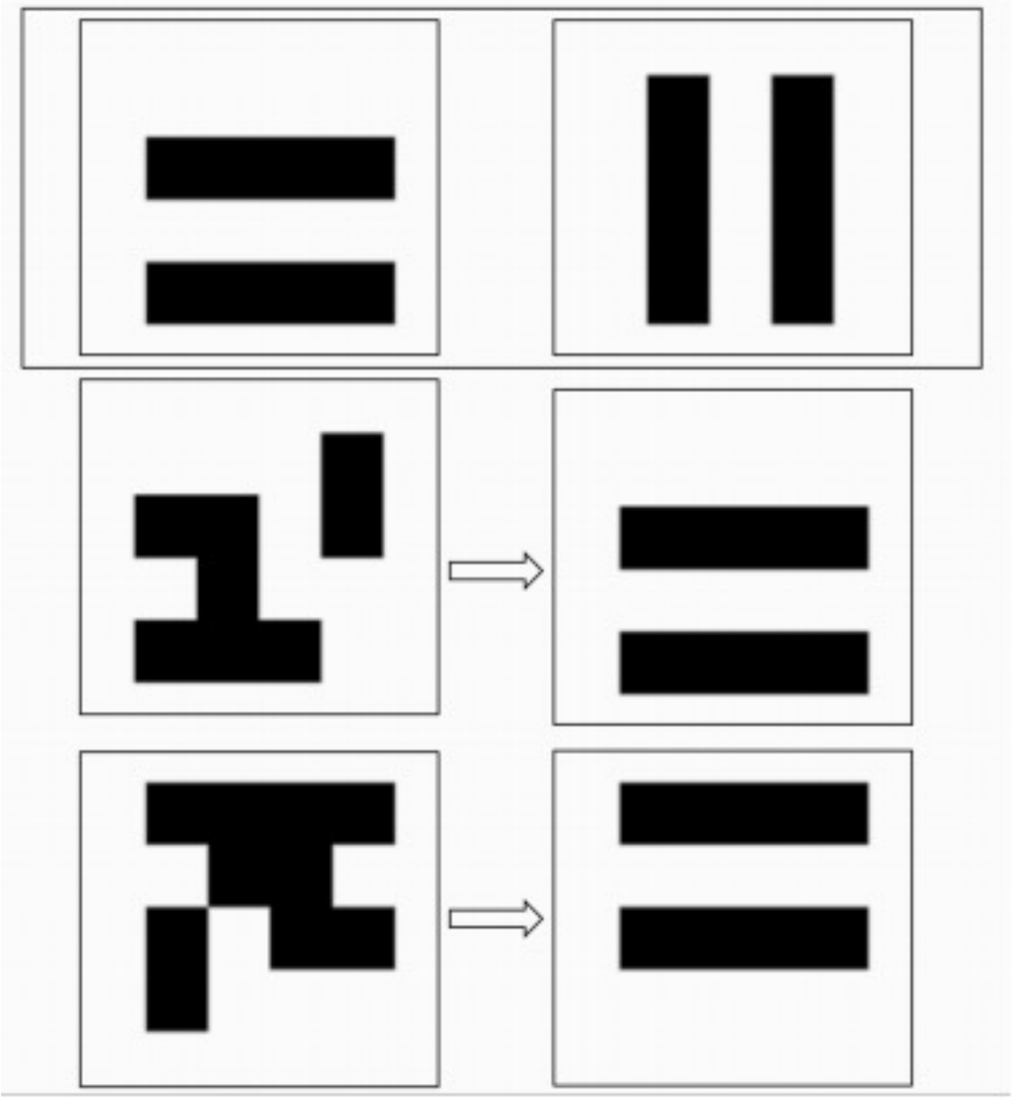


ENERGETIC LEARNING: THE HOPFIELD NETWORK





ENERGETIC LEARNING: THE HOPFIELD NETWORK



ENERGETIC LEARNING: THE HOPFIELD NETWORK

3. Capacity of the Hopfield Network

$$s_i = \sum_{j=1}^{d} w_{ij} x_j(n) \Leftarrow w_{ij} = \frac{1}{N}\sum_{n=1}^{N} s_i(n) s_j(n)$$

$$= \sum_{j=1, j\neq i}^{d} \sum_{n=1}^{N} x_i(n) s_j(n) \; x_j(n)$$

$$= \sum_{j=1, j\neq i}^{d} [x_i(n) x_j(n) + \sum_{m=1, m\neq n}^{N} x_i(m) x_j(m)] \; x_j(n)$$

$$= \sum_{j=1, j\neq i}^{d} [[x_i(n) x_j(n)] x_j(n) + [\sum_{m=1, m\neq n}^{N} x_i(m) x_j(m)] \; x_j(n)]$$

$$= \sum_{j=1, j\neq i}^{d} [x_i(n) + \sum_{m=1, m\neq n}^{N} x_i(m) x_j(m) x_j(n)]$$

$$= x_i(n)(d-1) + \sum_{j=1, j\neq i}^{d} \sum_{m=1, m\neq n}^{N} x_i(m) x_j(m) x_j(n)$$

$N \approx 0.18d$?

ENERGETIC LEARNING: THE HOPFIELD NETWORK

4. The Continuous Hopfield Network
   a) Binary values -> continuous values
   b) Sigmoid or Hyperbolic Tangent function as activation function
   c) Compute the probability distribution instead of energy values.

$$
\begin{aligned}
H &= -\frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d} w_{ij} s_i s_j \\
&= -\frac{1}{2}\mathbf{s}\mathbf{W}\mathbf{s}^T
\end{aligned}
$$

Eneergy

$$p(\mathbf{x}|\mathbf{W}) = \frac{1}{Z(\mathbf{W})}\exp\left[\frac{1}{2}\mathbf{x}^T\mathbf{W}\mathbf{x}\right]$$

$$Z(\mathbf{W}) = \sum_{\mathbf{x}} \exp\left(\tfrac{1}{2}\mathbf{x}^T\mathbf{W}\mathbf{x}\right)$$

Probability Distribution

Thank you!