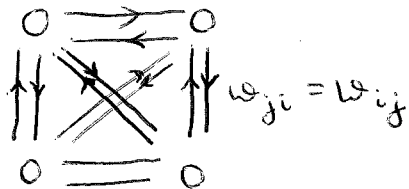


Hopfield networks

Lecture 24

A fully connected feedback network,
symmetric weights:

(no loops)



Associative memories.

Hebbian learning: Donald Hebb, 1949
weights between neurons

$$\frac{dw_{ij}^{\leftarrow}}{dt} = \frac{dw_{ji}^{\rightarrow}}{dt} \sim \text{Corr}(x_i, x_j) \quad (*)$$

↑ ↑
neuron activities

If a stimulus is present \Rightarrow neuron i activated

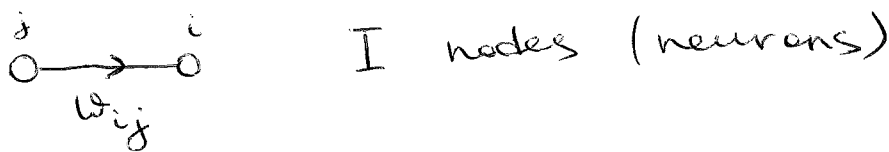
If another stimulus is present \Rightarrow neuron j activated

If the two stimuli are correlated \Rightarrow

\Rightarrow according to (*), w_{ij} & w_{ji} will increase with time & become large.

Now, if neuron i is stimulated, $\overset{\text{neuron}}{j}$ is activated too \Rightarrow associative memory.

Binary Hopfield network



Fully connected, $w_{ij} = w_{ji}$; $w_{ii} = 0, \forall i$
Biases may be included as weights w_{i0}
from neuron \emptyset with $\underbrace{x_0 = 1}_{\text{permanently on}}$
 $x_i = \text{activity of neuron } i$

① Activity rule: $x_i(d_i) = \begin{cases} 1 & d_i \geq 0 \\ -1 & d_i < 0 \end{cases}$
(state update rule)

Synchronous updates: $d_i = \sum_j w_{ij} x_j$, $\forall i$
compute

Then update all neuron states: $x_i = x_i(d_i), \forall i$

asynchronous updates: compute d_i & x_i for
one neuron at a time, continue in a fixed
or random sequence of neurons.

② Learning rule: goal \Rightarrow make a set of
memories $\{\vec{x}^{(n)}\}$ stable states of the
Hopfield network. $\vec{x}^{(n)} = (-1, -1, +1, \dots, +1)$
N memories I entries

Use Hebbian learning: sum over memories

$$w_{ij} = \eta \sum_{n=1}^N x_i^{(n)} x_j^{(n)},$$

$\eta > 0$ is a constant [e.g. $\eta = \frac{1}{N}$]

Biological motivation:

- ① Biological memories are associative, recalled spontaneously
Moscow \leftrightarrow Russia, Oslo \leftrightarrow Norway
- ② ~~Bio.~~ Bio. memories are error-tolerant & robust
Oslo \leftrightarrow Norway \Rightarrow Oslo \leftrightarrow Norway
O. lo \leftrightarrow N. rway \Rightarrow Oslo \leftrightarrow Norway
- ③ Bio. memories are distributed

Continuous Hopfield network

Same as binary but with

$$x_i = \tanh(d_i) \text{ or } x_i = \tanh(\beta d_i)$$

Binary Hopfield network is the $\beta \rightarrow \infty$ limit of the continuous Hopfield network.

—○—
Hopfield networks are basically spin glasses:

$$E = -\frac{1}{2} \sum_{ij} J_{ij} x_i x_j - \sum_i h_i x_i$$

\uparrow
energy

$J_{ij} \leftrightarrow w_{ij}, h_i \leftrightarrow w_{i0}$

[although x_i are not necessarily ± 1]

Spin glasses can be treated using a mean-field approach.

Mean-field approximation

Consider $P(\vec{x}) = \frac{1}{Z} e^{-\beta E(\vec{x})}$, where

$$E(\vec{x}) = -\frac{1}{2} \sum_{ij} J_{ij} x_i x_j - \sum_i h_i x_i$$

↑ state couplings
↓ fields

$$Z = \sum_{\vec{x}} e^{-\beta E(\vec{x})}$$

↑ part'n f'n

We want to approximate $P(\vec{x})$ with $Q(\vec{x}, \vec{\theta})$ which is simpler to compute.

↑ adjustable prms

Introduce

$$\beta \tilde{F}(\vec{\theta}) = \beta \underbrace{\langle E(\vec{x}) \rangle_Q}_{\text{average energy}} - \underbrace{S_Q}_{\text{entropy}} \quad \textcircled{=}$$

↑ mf free energy

$$\textcircled{=} \beta \sum_{\vec{x}} Q(\vec{x}, \vec{\theta}) E(\vec{x}) + \sum_{\vec{x}} Q(\vec{x}, \vec{\theta}) \log Q(\vec{x}, \vec{\theta}) =$$

$$= \sum_{\vec{x}} Q \log \underbrace{\frac{Q}{e^{-\beta E}}}_{ZP} = \underbrace{\sum_{\vec{x}} Q \log \frac{Q}{P}}_{D(Q||P), \text{ KL distance } \geq 0} - \underbrace{\log Z}_{\sum_{\vec{x}} Q(\vec{x}) = 1} \quad \textcircled{=}$$

$$\textcircled{=} D(Q||P) + \beta F.$$

↑ $Z = e^{-\beta F}$

So, \bar{F} is bounded below by exact free energy $F \Rightarrow$ vary $\vec{\theta}$ to minimize \bar{F} .

Consider $Q(\vec{x}, \vec{a}) = \frac{1}{Z_Q} e^{\sum_n a_n x_n}$

[uncoupled spins]
- S_n , entropy of spin n

Then $S_Q = - \sum_{\vec{x}} Q \log Q = - \sum_n \left[q_n \log q_n + (1-q_n) \log(1-q_n) \right]$, where

sum over spins

$$\left\{ \begin{array}{l} q_n = \frac{e^{a_n}}{e^{a_n} + e^{-a_n}} \\ 1 - q_n = \frac{e^{-a_n}}{e^{a_n} + e^{-a_n}} \end{array} \right. \begin{array}{l} \Leftarrow \text{prob. that spin } n = +1 \\ \Leftarrow \text{prob. that spin } n = -1 \end{array}$$

Furthermore,

$$\langle E(\vec{x}) \rangle_Q = \sum_{\vec{x}} Q \left[-\frac{1}{2} \sum_{i,j} J_{ij} x_i x_j - \sum_i h_i x_i \right] =$$

$$= -\frac{1}{2} \sum_{i,j} J_{ij} \langle x_i \rangle \langle x_j \rangle - \sum_i h_i \langle x_i \rangle,$$

where $\langle x_i \rangle = \frac{e^{a_i} - e^{-a_i}}{e^{a_i} + e^{-a_i}} = \tanh(a_i)$

Finally,

$$\beta \tilde{F}(\vec{a}) = -\beta \left[\frac{1}{2} \sum_{\substack{i,j \\ i \neq j}} J_{ij} \langle x_i \rangle \langle x_j \rangle + \sum_i h_i \langle x_i \rangle \right] - \sum_n S_n$$

Minimize \tilde{F} :

$$\begin{aligned} \frac{\partial}{\partial q_n} S_n(q_n) &= -1 - \log q_n + 1 + \log(1 - q_n) = \\ &= \log \frac{1 - q_n}{q_n} = -2a_n \end{aligned}$$

Note that $q_n = \frac{1}{1 + e^{-2a_n}}$,

$$e^{-2a_n} = \frac{1}{q_n} - 1 = \frac{1 - q_n}{q_n}$$

Moreover,

$$\langle x_i \rangle = \frac{1 - e^{-2a_i}}{1 + e^{-2a_i}} = \frac{1 - \left(\frac{1}{q_i} - 1\right)}{1 + \left(\frac{1}{q_i} - 1\right)} = 2q_i - 1$$

so that $\frac{\partial}{\partial q_n} \langle x_n \rangle = 2$.

Then
$$\beta \frac{\partial \tilde{F}}{\partial a_m} = -\beta \left[\sum_j J_{mj} \underbrace{\frac{\partial \langle x_m \rangle}{\partial q_m}}_2 \langle x_j \rangle \right] \left(\frac{\partial q_m}{\partial a_m} \right) - \underbrace{\left(\frac{\partial q_m}{\partial a_m} \right)}_2 \frac{\partial S_m}{\partial q_m} \ominus - 2a_m$$

$$\ominus \underbrace{2 \left(\frac{\partial \phi_m}{\partial a_m} \right)}_{*0} \left\{ -\beta \left[\sum_j J_{mj} \langle x_j \rangle + h_m \right] + a_m \right\}$$

$$\frac{\partial \tilde{F}}{\partial a_m} = 0 \Rightarrow a_m = \beta \left[\sum_j J_{mj} \langle x_j \rangle + h_m \right]$$

Recall that $\langle x_j \rangle = \tanh(a_j)$.

We can see that there is an analogy between Hopfield networks & mean-field approach to spin glasses:

Hopfield

$$x_i = \tanh(\beta a_i)$$

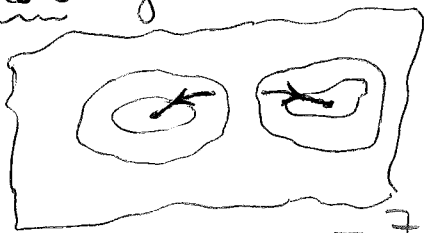
$$a_i = \sum_j w_{ij} x_j + w_{i0}$$

Mean-field

$$\langle x_i \rangle = \tanh(a_i)$$

$$a_i = \beta \left[\sum_j \underbrace{J_{ij}}_{w_{ij}} \langle x_j \rangle + \underbrace{h_i}_{w_{i0}} \right]$$

Thus \tilde{F} plays the role of free energy in the Hopfield network. On the free energy landscape defined by \tilde{F} , there are distinct basins of attraction corresponding to each memory. Asynchronous update: changing 1 spin at a time. Synchronous update: changing all spins (a global move, is not guaranteed to decrease \tilde{F}).

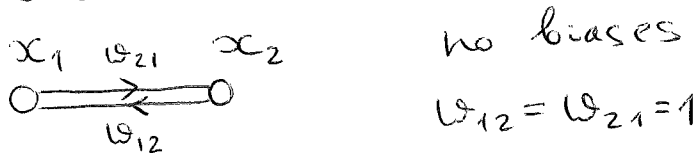


It can be shown that asynchronous updates are guaranteed to decrease \tilde{F} monotonously, similarly to steepest descent.

Moreover, \bar{F} is convex in the vicinity of the minima \Rightarrow the dynamics converges to a stable fixed point (which one depends on the ~~boundary~~^{initial} conditions) & there are no limit cycles (& no chaotic behavior).

Examples

① Binary Hopfield network



t=0: $x_1 = 1, x_2 = -1$ $\underline{E = -x_1 x_2 = 1}$

Synchronous update: $\begin{cases} a_1 = w_{12} x_2 = x_2, \\ a_2 = w_{21} x_1 = x_1. \end{cases}$

t=1: $\begin{cases} a_1 = -1 \\ a_2 = 1 \end{cases} \Rightarrow \begin{cases} x_1 = -1 \\ x_2 = 1 \end{cases}$ both spins flipped $\underline{E = 1}$

t=2: $\begin{cases} a_1 = 1 \\ a_2 = -1 \end{cases} \Rightarrow \begin{cases} x_1 = 1 \\ x_2 = -1 \end{cases}$ both spins flipped again, etc.
no convergence, limit cycle

Asynchronous update:

Update spin 1, then spin 2 (& continue)

t=1: $a_1 = -1 \Rightarrow \begin{cases} x_1 = -1, \\ x_2 = -1 \end{cases}$ spin 1 flipped $E = -1$

t=2: $a_2 = -1 \Rightarrow \begin{cases} x_1 = -1, \\ x_2 = -1 \end{cases}$ spin 2 not flipped $E = -1$

$(-1, -1)$ is a stable fixed point
 $(1, 1)$ is another] -8-

② Continuous Hopfield network

2 nodes, no biases : $h_1 = h_2 = 0$

$\beta = 1$, $J_{12} = J_{21} = 1$:

$$\tilde{F} = - \underbrace{\bar{x}_1}_{2q_1-1} \underbrace{\bar{x}_2}_{2q_2-1} + q_1 \log q_1 + (1-q_1) \log(1-q_1) + q_2 \log q_2 + (1-q_2) \log(1-q_2)$$

t=0: start with $\bar{x}_1 = 1, \bar{x}_2 = -1$:

$$\begin{cases} a_1 = +\infty, \\ a_2 = -\infty \end{cases} \quad \begin{cases} q_1 = 1, \\ q_2 = 0 \end{cases}$$

Then $\tilde{F} = 1$

t=1:

(asynchronous 121212... updates)

$$a_1 = \bar{x}_2 = -1 \Rightarrow \begin{cases} \bar{x}_1 = \tanh(a_1) = \frac{e^{-1} - e}{e^{-1} + e} \approx -0.76 \\ \bar{x}_2 = -1 \end{cases}$$

Then $\tilde{F} = \frac{e^{-1} - e}{e^{-1} + e} \oplus \begin{cases} q_1 = \left(\frac{e^{-1} - e}{e^{-1} + e} + 1 \right) \frac{1}{2} = \frac{1}{1+e^2} \approx 0.12 \\ q_2 = 0 \end{cases}$

$\oplus \frac{1}{1+e^2} \log \frac{1}{1+e^2} + \frac{e^2}{1+e^2} \log \frac{e^2}{1+e^2} \approx -1.13$

t=2: $a_2 = \bar{x}_1 \approx -0.76 \Rightarrow \begin{cases} \bar{x}_1 = -0.76, \\ \bar{x}_2 = \tanh(a_2) \approx -0.64 \end{cases}$

$\begin{cases} q_1 = 0.12 \\ q_2 = \frac{1+\bar{x}_2}{2} \approx 0.18 \end{cases} \Rightarrow \tilde{F} \approx -1.32$ etc.