Neural Networks and Biological Modeling

Professeur Wulfram Gerstner
Laboratory of Computational Neuroscience

Correction Question Set 6

Exercise 1: Hopfield network with probabilistic upate

1.1 We first split the neurons into two groups: those that *should be active*, $(p_i^3 = +1)$ and those that should not be active. From $h_i(t_0) = p_i^3 m^3(t_0))$ we see that all neurons in the same group get the same input potential h.

We then get the following four expressions for the update dynamics:
Those that *should* be active and are active:
$P\{S_i(t+1) = +1|h_i(t_0), p_i^3 = +1\} = g(h_i(t)) = g(p_i^3 m^3(t_0)) = g(+1m^3(t_0)) = g(m^3(t))$
Those that *should* be active and are not active:
$P\{S_i(t+1) = -1|h_i(t_0), p_i^3 = +1\} = 1 - g(m^3(t))$
Those that *should not* be active and are active:
$P\{S_i(t+1) = +1|h_i(t_0), p_i^3 = -1\} = g(h_i(t)) = g(p_i^3 m^3(t_0)) = g(-1m^3(t_0)) = g(-m^3(t))$
Those that *should not* be active and are not active:
$P\{S_i(t+1) = -1|h_i(t_0), p_i^3 = -1\} = 1 - g(-m^3(t))$

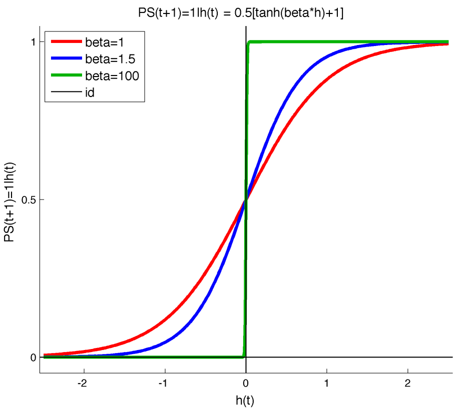
The expected number of neurons in each of the four groups is N_+^3 (or N_-^3) times the probabilities from above.

We start from
$m(t+1) = \frac{1}{N}\sum_i^N p_i^3 S_i(t+1)$
and split the sum into the four groups. Given $p \in \{-1, +1\}$ and $S_i(t+1) \in \{-1, +1\}$ note that we are just "counting". We can get an estimate of that sum by splitting it and replacing those "counts" by the expected number of neurons in each of the four groups (large N):

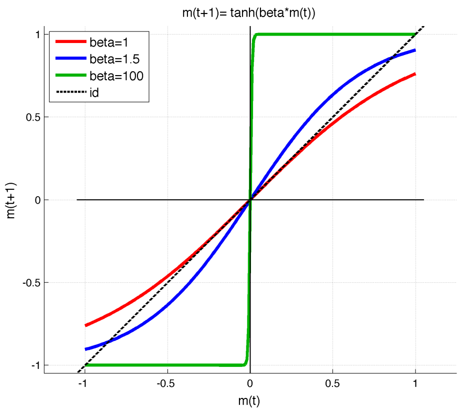
$$m(t+1) = \frac{1}{N}\left[N_+^3 g(m^3(t) \;-\; N_+^3(1 - g(m^3(t))\right] - \frac{1}{N}\left[\; N_-^3 g(-m^3(t) \;-\; N_-^3(1 - g(-m^3(t))\right]$$
$$= \frac{N_+^3}{N}\left[2g(m^3(t)) - 1\right] \;-\; \frac{N_-^3}{N}\left[2g(-m^3(t)) - 1\right]$$
$$= g(m^3(t)) - g(-m^3(t))$$

Where we have used the assumption $P\{p_i^3 = 1\} = 0.5$ and that for a large network $N \to \infty$ we have $N_+^3 = N_-^3$ and $\frac{N_+^3}{N} = \frac{1}{2}$.

1.2 g maps the input potential h onto a probability.
g: $\mathbb{R} \to [0, 1]$, monotonically increasing, symmetric around $g(0) = 0.5$
We plug $g(h) = \frac{1}{2}(tanh(\beta h) + 1)$ into $g(m^3(t)) - g(-m^3(t))$ and simplify it:
$m(t+1) = ... = tanh(\beta m^3(t))$.
See figure 1 to see the effect of β.



(a) **Stochastic update dynamics** for different levels of the inverse temperature β



(b) Evolution of the initial **overlap** in one time step. For $\beta \leq 1$ the fixed point is at 0 and any initial overlap will decrease. For $\beta = 1.5$ the overlap does not go beyond ≈ 0.85: using a stochastic update, the network can only retrieve noisy versions of the pattern.

Figure 1: Update dynamics and overlap. Note the different domain and range of each graph. The lower the temperature, the more deterministic the update becomes and the fixed point of $m(t) \mapsto m(t+1)$ goes to 1 as $\beta \to \infty$.

Exercise 2: Hopfield, the energy picture

In each time step only one neuron is updated (asynchronous dynamics). Let us assume that neuron k has changed. The energy is given by:

$$E := -\sum_{i}^{N}\sum_{j}^{N} w_{ij} S_i S_j \tag{1}$$

We split that sum such that the contribution of the neuron k to the energy is isolated from the other neurons:

$$\begin{aligned} E(t) &= \underbrace{-\sum_j w_{kj} S_k(t) S_j(t)}_{\text{i=k}} - \underbrace{\sum_i w_{ik} S_i(t) S_k(t)}_{\text{j=k}} - \textstyle\sum_{i\neq k}^{N} \sum_{j\neq k}^{N} w_{ij} S_i(t) S_j(t) \\ &= -2S_k(t) \textstyle\sum_j w_{kj} S_j(t) - \sum_{i\neq k}^{N} \sum_{j\neq k}^{N} w_{ij} S_i(t) S_j(t) \end{aligned}$$

The last equation comes from the symmetry of the weights and the fact that the first two sums run over the same range (1 to N).
For $E(t+1)$ we get the same expression but with the neuron k having a different state $S'_k = -S_k$ (all other neurons keep their value $S'_j = S_j$ for $j \neq k$.). When we look at the change in energy, all terms that do not depend on k cancel out. Therefore we have:

$$\begin{aligned} \Delta E = E(t+1) - E(t) &= -2S'_k \textstyle\sum_j w_{kj} S_j - (-2S_k \sum_j w_{kj} S_j) \\ &= -2(S'_k - S_k) \textstyle\sum_j w_{kj} S_j \end{aligned}$$

Because of the update of neuron k, we have $S'_k - S_k = 2S'_k$. Also, $\sum_j w_{kj} S_j \equiv h_k$. Thus, so far we have $\Delta E = -4S'_k h_k$.

Finally, due to the dynamics of the network, $S'_k = sign(h_k)$, the change in energy is
$\Delta E = -4\ h_k\ sign(h_k) < 0$.

In other words, the energy E is a Liapunov function of the deterministic Hopfield network which decreases along trajectories. This yields the valuable insight that the network dynamics necessarily converge towards the minima of the energy function E.

Exercise 3: Binary codes and spikes

3.1 We do a change of variable and specify the Hopfield model:

- The **state** of a neuron i is $\sigma_i \in \{0,1\}$. It relates to S_i by $S_i = 2\sigma_i - 1 \Leftrightarrow \sigma_i = \frac{1}{2}(S_i + 1)$.
- The **weights** are the same. They depend on the patterns, not on the state.
- For the **update dynamics** we rewrite eq. 2 in terms of σ:

$$S_i(t+1) = g(h_i(t)) = sign\left(\sum_{j=1}^{N} w_{ij} S_j(t)\right) \tag{2}$$

$$\begin{aligned}
2\sigma_i(t+1) - 1 &= sign\left(\sum_{j=1}^{N} w_{ij}(2\sigma_j(t) - 1)\right) \\
\sigma_i(t+1) &= \frac{1}{2}\left[sign\left(\sum_{j=1}^{N} w_{ij}(2\sigma_j(t) - 1)\right) + 1\right]
\end{aligned}$$

3.2 The property $\sum_{i=1}^{N} p_i = 0$ means that the patterns are balanced: they have the same number of active and inactive pixels. When patterns have specific statistical properties, these properties may translate into statistical properties of the weights. For that reason, we take the expression from the previous question, insert the definition of the weights and try to simplify it using the given property:

$$\begin{aligned}
h_i(t) &= \sum_{j=1}^{N} w_{ij}(2\sigma_j(t) - 1) \\
&= 2\sum_{j=1}^{N} w_{ij}\sigma_i(t) - \sum_{j=1}^{N} w_{ij}
\end{aligned}$$

$$\begin{aligned}
\sum_{j=1}^{N} w_{ij} &= \frac{1}{N}\sum_{j=1}^{N}\sum_{\mu}^{M} p_i^{\mu} p_j^{\mu} \\
&= \tfrac{1}{N}\textstyle\sum_{\mu}^{M} p_i^{\mu} \sum_{j=1}^{N} p_j^{\mu} \\
&= \tfrac{1}{N}\textstyle\sum_{\mu}^{M} p_i^{\mu} 0 \\
&= 0
\end{aligned}$$

$$h_i(t) = 2\sum_{j=1}^{N} w_{ij}\sigma_j(t)$$

$$\begin{aligned}
\sigma_i(t+1) &= \tfrac{1}{2}\left[sign\left(2\textstyle\sum_{j=1}^{N} w_{ij}\sigma_j(t)\right) + 1\right] \\
&= \frac{1}{2}\left[sign\left(\sum_{j=1}^{N} w_{ij}\sigma_j(t)\right) + 1\right]
\end{aligned}$$

It's interesting to note that for balanced patterns, the weights also sum up to 0.

3.3 We consider the starting point of the previous question:

$$h_i(t) \quad = \quad 2\sum_{j=1}^{N} w_{ij}\sigma_j(t) - \sum_{j=1}^{N} w_{ij}$$

For low-activity patterns we have:

$$\begin{aligned}
\sum_{j=1}^{N} w_{ij} &= c\sum_{j=1}^{N}\sum_{\mu}(\xi_i^\mu - b)(\xi_j^\mu - a) \\
&= c\sum_{\mu}(\xi_i^\mu - b)\sum_{j=1}^{N}(\xi_j^\mu - a) \\
&= c\sum_{\mu}(\xi_i^\mu - b)0 \\
&= 0
\end{aligned}$$

Where, as in the previous question, we used the condition of balanced patterns, that in this case would be: $\sum_j^\mu \xi_j^\mu = Na$.

To simplify our arguments we will choose $b = 0$. For the total input to neuron i we now have:

$$\begin{aligned}
h_i(t) &= 2c\sum_{j=1}^{N} w_{ij}\sigma_j(t) \\
&= 2c\sum_{j=1}^{N}\sum_{\mu}\xi_i^\mu(\xi_j^\mu - a)\sigma_j(t) \\
&= 2c\sum_{j=1}^{N}\sum_{\mu}\xi_i^\mu\xi_j^\mu\sigma_j(t) - 2c\sum_{j=1}^{N}\sum_{\mu}\xi_i^\mu a\sigma_j(t)
\end{aligned}$$

Since $\xi_i^\mu \in \{0,1\}$, $\sigma_i \in \{0,1\}$ and $a > 0$ we observe the first term on the right-hand side is always positive and the second term is always negative (since $2c\sum_{j=1}^{N}\sum_{\mu}\xi_i^\mu a\sigma_i(t) > 0$). We have managed to separate excitatory from inhibitory contributions. So now we can interpret the network as follows: all neurons are excitatory, and they excite each other as well as a hypothetical group of inhibitory neurons, that in return inhibit the former. In that way Dale's law is preserved (i.e. all connections starting from the same neuron are either excitatory or inhibitory, not both) and an effective inhibition is achieved. For more details on the above please refer to the book 'Neuronal Dynamics', Chapter 17.3.2, available online: `http://neuronaldynamics.epfl.ch/online/Ch17.S3.html`.