Neural Networks and Biological Modeling

Professeur Wulfram Gerstner
Laboratory of Computational Neuroscience

Correction Question Set 5

Exercise 1

1.1 Each neuron connects to eight other neurons (there are no self-connections). That makes 72 connections in total. The weights follow $w_{ij} = p_i^\mu p_j^\mu$ where p^μ is the prototype. That is, weights between "black" neurons are $+1$, weights between "white" neurons are also $+1$, and weights between neurons receiving two opposite colors are -1. See figure 1.

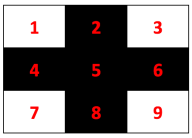
Given the symmetry of the problem, let us assume that the central bit has been flipped. The dynamics of the activity $S_{\text{center}}(t)$ for this bit follows

$$S_{\text{center}}(t+1) = sgn\left(\sum_{j\neq\text{center}} w_{ij} S_j(t)\right).$$

Therefore, all other "black neurons" will contribute by $1 \times 1 = 1$ while all other "white neurons" will bring $(-1) \times (-1) = 1$ as well. The resulting activity is 8 which has positive sign. In one iteration, the central bit is corrected. See figure 1(c) for an example where two more bits are flipped.

Similarly, the flipped bit is the only one to bring a "bad" unit signal (-1 if the neuron is black, 1 otherwise) to the activity of other neurons, but this isn't enough to make the sign of the resulting activities change. Therefore, other bits do not fall in the dark side (or bright side): the memory is fully recovered.

1.2 Using the same reasoning, having more and more bits flipped iteratively, you can convince yourself that less than half of the bits can be flipped to recover the pattern, i.e. 4.



(a) arbitrary neuron indexing

Pattern		-1	1	-1	1	1	1	-1	1	-1
	Neuron	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
-1	**1**	0	-1	1	-1	-1	-1	1	-1	1
1	**2**	-1	0	-1	1	1	1	-1	1	-1
-1	**3**	1	-1	0	-1	-1	-1	1	-1	1
1	**4**	-1	1	-1	0	1	1	-1	1	-1
1	**5**	-1	1	-1	1	0	1	-1	1	-1
1	**6**	-1	1	-1	1	1	0	-1	1	-1
-1	**7**	1	-1	1	-1	-1	-1	0	-1	1
1	**8**	-1	1	-1	1	1	1	-1	0	-1
-1	**9**	1	-1	1	-1	-1	-1	1	-1	0

(b) Weigths w_{ij}. Note: we have no recurrent connections

$$[\,-1\ \ 1\ \ -1\ \ 1\ \ 0\ \ 1\ \ -1\ \ 1\ \ -1\,] \bullet \begin{bmatrix} 1 \\ 1 \\ -1 \\ 1 \\ -1 \\ 1 \\ -1 \\ -1 \\ -1 \end{bmatrix} = 4$$

(c) Weights w_{5j}, a pattern $S(t)$ with 3 errors and the resulting input potential $h_5 = 4$

Figure 1: Pattern, weights w_{ij} and an example of error correction: The input potential h to neuron $i = 5$, $h_5 = \sum_j^9 w_{5j} \cdot S_j$, is strong enough to correct neuron 5.

Exercise 2

We recall the definitions for the weights w_{ij} and the overlap $m^\mu(t)$. $S(t)$ denotes the current activity pattern of the network:

$$w_{ij} = c \sum_{\mu=1}^{M} p_i^\mu p_j^\mu \text{ with } c = \frac{1}{N} \tag{1}$$

$$m^\mu(t) = \frac{1}{N} \sum_{i=1}^{N} p_i^\mu S_i(t) \tag{2}$$

The network dynamics are given by the state equation:

$$S_i(t+1) = g(h_i(t)) = g\left(\sum_{j=1}^{N} w_{ij} S_j(t)\right) \tag{3}$$

We express $S_i(t+1)$ using equations 1 and 2:

$$\begin{aligned}
S_i(t+1) &= g(h_i(t)) \\
h_i(t) &= \sum_{j=1}^{N} w_{ij} S_j(t) \\
&= \sum_{j=1}^{N} c \sum_{\mu=1}^{M} p_i^\mu p_j^\mu S_j(t) \\
&= c \sum_{\mu=1}^{M} p_i^\mu \sum_{j=1}^{N} p_j^\mu S_j(t) \\
&= cN \sum_{\mu=1}^{M} p_i^\mu m^\mu(t) \\
&= \sum_{\mu \neq 3} p_i^\mu m^\mu(t) + p_i^3 m^3(t) \text{ (split the sum)} \\
&= p_i^3 m^3(t) \text{ (there's no overlap with other patterns)} \\
S_i(t+1) &= g\left(p_i^3 m^3(t)\right) \\
&= p_i^3\, g\left(m^3(t)\right)
\end{aligned}$$

For the last equivalence we used $p_i^3 \in \{-1, +1\}$ and that we can pull out a sign from an odd function: $g(p_i^3 \times \cdot) = p_i^3 \times g(\cdot)$

We can now use that expression $S_i(t+1)$ to compute $m^3(t+1)$:

$$\begin{aligned}
m^3(t+1) &= \frac{1}{N} \sum_{i=1}^{N} p_i^3 S_i(t+1) \\
&= \frac{1}{N} \sum_{i=1}^{N} p_i^3 p_i^3 g\left(m^3(t)\right) \\
&= \frac{1}{N} N g\left(m^3(t)\right)
\end{aligned}$$

Therefore, we end up with

$$m^3(t+1) = g(m^3(t)) \tag{4}$$

We can now discuss according to the nature of g:

- if $g = sgn$ then the dynamics stops after the first step: a positive overlap $m > 0$ will immediately increase to $+1$. A negative overlap will decrease to -1 (that's the "negative" pattern: *every* neuron has the opposite state of the pattern).
- for a monotonically increasing g, the situation is more complicated (see figure 2): for $g(m) = tanh(m)$ we "see" that $|tanh(m)| \leq |m|$ on the interval $[-1, 1]$. That means, starting with any overlap m, the dynamics will *reduce* the overlap and finally end up at $m = 0 = tanh(0)$.
- For completeness, we introduce an extra parameter β in $g(m) = tanh(\beta m)$ and show its effect in figure 2. For $\beta \to \infty$ we are back the sign function. For a discussion of this "inverse temperature" parameter β we refer to the book "Neuronal Dynamics", Chapter 17.2.3, available online: http://neuronaldynamics.epfl.ch/online/Ch17.S2.html.

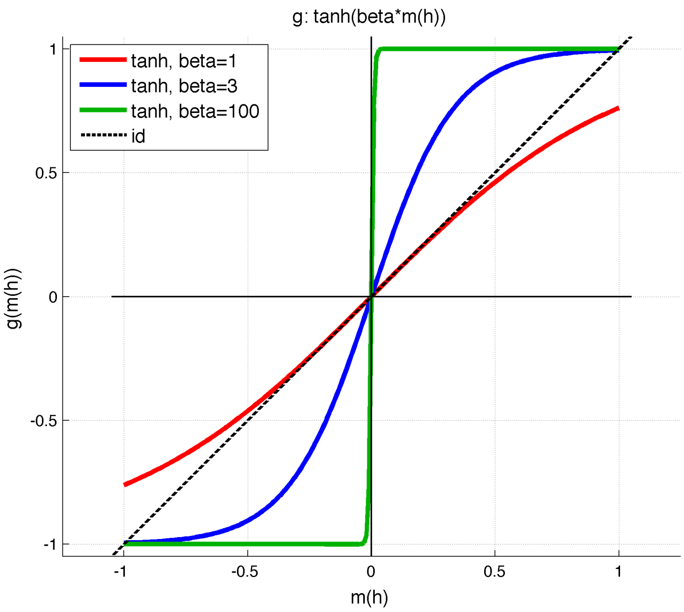


Figure 2: $g(m) = tanh(\beta m)$. For $\beta = 1$ (red curve) the overlap will decrease: $m(t+1) \leq tanh(m(t))$

Exercise 3: Probability of error in the Hopfield model

3.1

$$S_i(t=1) \quad = \quad sign(\sum_j \omega_{ij} S_j(t=0))$$

We insert the definition of w_{ij} (eq. 1) and use $S_j(t=0) = p_j^1$:

$$= \quad sign(\sum_j \left(\frac{1}{N} \sum_\mu p_j^\mu p_i^\mu \right) p_j^1)$$

Multiply with $1 = p_i^1 p_i^1$. Then pull out a factor $p_i^1 = \pm 1$ from the sign function

$$= \quad p_i^1 sign(\sum_j \left(\frac{1}{N} \sum_\mu p_j^\mu p_i^\mu \right) p_j^1 p_i^1)$$

$$= \quad p_i^1 sign(\sum_j \frac{1}{N} (p_j^1)^2 (p_i^1)^2 + \sum_j \sum_{\mu \neq 1} \frac{1}{N} p_j^\mu p_i^\mu p_j^1 p_i^1)$$

$$= \quad p_i^1 sign(1 + \sum_j \sum_{\mu \neq 1} \frac{1}{N} p_j^\mu p_i^\mu p_j^1 p_i^1)$$

3.2 We are given the dynamics $S_i(t=1) = p_i^1 sign(1 + \sum_{\mu \neq 1}^P \sum_j^N \frac{1}{N}\ p_i^1\ p_i^\mu\ p_j^1\ p_j^\mu)$. We see that $S_i(t=1) \neq S_i(t=0)$ is true when
$\displaystyle\sum_{\mu \neq 1}^P \sum_j^N \frac{1}{N}\ p_i^1\ p_i^\mu\ p_j^1\ p_j^\mu) \leq -1.$

3.3 We interpret the double sum as a sequence of random steps $\frac{1}{N}\ p_i^1\ p_i^\mu\ p_j^1\ p_j^\mu$. Then we formulate a probability density that quantifies how that random walk will move away from zero. We can do so with the help of the central limit theorem (CLT). The steps to apply it are: 1) define the random variable X and formulate its mean μ and variance σ^2. 2) define the sum $Z_n = X_1 + X_2 + ... + X_n$. 3) for large n and by the CLT, the sum Z_n is approximately normal with $\mathcal{N}(n\mu, n\sigma^2)$. We now apply these steps:
We define a random variable $X = p_i^1\ p_i^\mu\ p_j^1\ p_j^\mu$. It has a probability mass function $Pr(X=-1) = Pr(X=+1) = 0.5$. The mean is $\bar{X} = 0$. The variance is $Var[X] = 1$ as we can see from the definition: $Var[X] = \mathbb{E}[(X - \bar{X})^2] = \mathbb{E}[X^2] = \mathbb{E}[1] = 1$.
We now sum the random variable X:
$Z'_n = \displaystyle\sum_{\mu \neq 1}^P \sum_j^N X$
In that double sum we are actually summing $n = N(P-1)$ realizations of X. For large n, we can apply the central limit theorem to approximate the distribution of Z'_n:
$Z'_n \sim \mathcal{N}(\ N(P-1) \times 0,\ N(P-1) \times 1)$

We used the symbol Z' because we have left out the scaling factor $1/N$ in the sum. We introduce it now, knowing that $\mathrm{Var}(aX) = a^2 \mathrm{Var} X$.
$Z_n \sim \mathcal{N}(0, \dfrac{1}{N^2} N(P-1)) = \mathcal{N}(0, \dfrac{P-1}{N})$
We now have our final expression:

$$\sum_{\mu \neq 1} \sum_j \frac{1}{N} p_i^\mu p_i^1 p_j^\mu p_j^1 \ \sim \mathcal{N}(0, \sigma^2) \qquad \text{with } \sigma^2 = \frac{P-1}{N}$$

3.4 We use the two previous results to find the probability of a neuron to flip: We make a probabilistic statement about the condition from question 3.2 using the Gaussian PDF found in 3.3. Then, we transform the equation to get an equivalent expression in terms of the error function erf:

$$P_{error} = Prob\left\{\left(\frac{1}{N}\sum_{\mu\neq 1}\sum_{j} p_i^{\mu} p_i^1 p_j^{\mu} p_j^1\right) \leq -1\right\}.$$

$$\begin{aligned}
P_{error} &= \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{-1} e^{-\frac{x^2}{2\sigma^2}}\,dx \\
&= \frac{1}{\sqrt{2\pi\sigma^2}}\int_{1}^{\infty} e^{-\frac{x^2}{2\sigma^2}}\,dx \\
&= \frac{1}{2} - \frac{1}{\sqrt{2\pi\sigma^2}}\int_{0}^{1} e^{-\frac{x^2}{2\sigma^2}}\,dx \\
&= \frac{1}{2}\left[1 - \sqrt{\frac{2}{\pi\sigma^2}}\int_{0}^{1} e^{-\frac{x^2}{2\sigma^2}}\,dx\right] \\
&= \frac{1}{2}\left[1 - \frac{2}{\sqrt{\pi}}\int_{0}^{\sqrt{\frac{N}{2(P-1)}}} e^{-x'^2}\,dx'\right] \\
&= \frac{1}{2}\left[1 - \mathrm{erf}\left(\sqrt{\frac{N}{2(P-1)}}\right)\right],
\end{aligned}$$

where the two first equalities come by symmetry and the third one by a change of variable $x \to \sqrt{2}\sigma x'$.

3.5 We consider the flipping of different pixels as being independent. The expected number of pixels flips is NP_{error}. The maximal number of patterns P^* is the highest number which satisfies the equation,

$$NP_{error}(N, P^*) < 1.$$

This implies $\mathrm{erf}\left(\sqrt{\frac{N}{2(P-1)}}\right) > 1 - 2/N = 0.9998$. Using the hint, $\mathrm{erf}(2.6) = 0.9998$ and the fact that erf is monotonically increasing, we have $\sqrt{\frac{N}{2(P-1)}} > 2.6$, which gives us $P^* = 740$.

3.6 $S_i(1) = p_i^1\ sign(1 + \sum_{\mu\neq 1}^{P}\ p_i^1\ p_i^{\mu}\ m^{\mu 1})$.

3.7 Now the random walk has $P-1$ terms with coefficient 0.1, which is approximated by a random gaussian variable with standard deviation $\sigma = \sqrt{(0.1^2\ (P-1))}$. Following the same steps, we get to $\sqrt{\frac{1}{2\ 0.1^2\ (P^*-1)}} > 2.6 \implies P^* = 8$, much less than before.