Lecture Notes in Systems Biology

Lecture 8 - Hopfield networks

Omer Karin

Hopfield networks are a type of artificial neural network introduced by John Hopfield in 1982. They were inspired by the theoretical understanding of how associative memory might work in the brain. The main purpose of Hopfield networks is to store and retrieve patterns or memories using distributed representations. The stored patterns are fixed-point attractors of the network and can be retrieved even when presented with partial or noisy inputs.

We consider a collection of I ”neurons”, each of which can be in an active state $x_i = 1$ or inactive state $x_i = -1$. The network is fully connected with symmetric weights w_{ij} and there are no self-connections.

We will first consider the activity rule $x(a) = 1$ when a is positive and $x(a) = -1$ otherwise.

The rules are activated according to:

$$a_i = \sum_j w_{ij} x_j \tag{1}$$

either synchronously, or asynchronously, one neuron at a time.

The idea is that this network can store memory patterns, which we can think of as binary vectors of length I. Let us consider a collection of N such vectors $x^1, x^2, ..., x^N$.

We store them by adjusting the weights w_{ij} to capture the correlation between the i-th and j-th elements along the patterns, in a matter reminiscent of Hebbs' rule:

$$w_{ij} = \eta \sum_n x_i^{(n)} x_j^{(n)} \tag{2}$$

where the choice of η will not be important for our conclusions.

Hopfield networks have an associated decreasing energy function

We will associate with the Hopfield network the (Lyapunov) energy function:

$$E = -\frac{1}{2}\eta \sum_i \sum_j w_{ij} x_i x_j \tag{3}$$

Consider the case of async updates. If $a_i(t) = a_i(t-1)$ then $E(t) = E(t-1)$. Otherwise, there is a flip of neuron i. Then:

$$\begin{aligned} E(t) - E(t-1) &= -x_i(t)\eta \sum_{j \neq i} w_{ij} x_j + x_i(t-1)\eta \sum_{j \neq i} w_{ij} x_j \\ &= (x_i(t-1) - x_i(t)) \sum_{j \neq i} w_{ij} x_j \end{aligned} \tag{4}$$

Now, if there was a flip and $x_i(t) = 1$ and $x_i(t-1) = -1$, then also $a_i(t) > 0$ and thus:

$$E(t) - E(t-1) = -2a_i(t) < 0 \tag{5}$$

And if $x_i(t) = -1$ and $x_i(t-1) = 1$ then $a_i(t) \leq 0$ and:

$$E(t) - E(t-1) = 2a_i(t) \leq 0 \tag{6}$$

Thus, for the asynchronous updates, the energy decreases with each update, and the decrease is strict when $a_i \neq 0$. For a dynamical system where an energy function is bounded from below and has a decreasing energy function, it will settle to a stable fixed point or a limit cycle.

Stored memories are fixed-point attractors

Consider now that the network is in one of the stored memory patterns $x^{(l)}$. We will assume for simplicity that it is the *only* stored pattern. Then:

$$a_i = \sum_j w_{ij} x_j = \eta x_i^{(l)} \sum_{j \neq i} \left(x_j^{(l)} \right)^2 = \eta (I-1) x_i^{(l)} \tag{7}$$

Thus, $\text{sign}(a_i) = \text{sign}(x_i)$ and the stored memory pattern is a fixed-point attractor.

What happens when there is more than one stored pattern? In this case:

$$w_{ij} = \eta x_i^{(l)} x_j^{(l)} + \eta \sum_{m \neq l} x_i^{(m)} x_j^{(m)} \tag{8}$$

And:

$$a_i = \eta \sum_j w_{ij} x_j = \eta (I-1) x_i^{(l)} + \eta \sum_{j \neq i} \sum_{m \neq l} x_i^{(m)} x_j^{(m)} x_j^{(l)} \tag{9}$$

The term $C_i^{(l)} = \sum_{j \neq i} \sum_{m \neq l} x_i^{(m)} x_j^{(m)} x_j^{(l)}$ is a noise term. In fact, it is the sum of $(N-1)(I-1)$ random variables with mean 0 and variance 1. Thus, for large N, a_i/η is a normally distributed random variable with mean $(I-1)x_i^{(l)}$ and variance $(N-1)(I-1)$. Recall that the CDF at point y of a normal distribution with mean μ and std σ is $\frac{1}{2}\left[1 + \text{erf}\left(\frac{y-\mu}{\sigma\sqrt{2}}\right)\right]$. Thus, the probability that a_i has a different sign compared with x_i is:

$$P = \frac{1}{2}\left[1 + \text{erf}\left(-\sqrt{\frac{I-1}{2(N-1)}}\right)\right] \approx \frac{1}{2}\,\text{erfc}\left(\sqrt{\frac{I}{2N}}\right) \tag{10}$$

Extension to continuous case

So far we discussed discrete neuronal activations and asynchronous updating. All of our analysis can be extended to the continuous case, where now:

$$\dot{x}_i = \tanh a_i - x_i \tag{11}$$

and

$$a_i = \sum_j w_{ij} x_j \tag{12}$$

Everything else stays the same, but now the energy function becomes:

$$E = -\frac{1}{2}\eta \sum_i \sum_j w_{ij} x_i x_j - \sum_i H_2^{(e)}\left(\frac{1}{2}(1 + x_i)\right) \tag{13}$$

with $H_2^{(e)})(q) = -q \log q - (1 - q) \log(1 - q)$.