Notes for Hopfield theory lecture

Shaul Druckmann

1 Signal-to-Noise Analysis of Hopfield Network Pattern Capacity

Let us try and develop some intuition regarding the choice of the particular shape of the connectivity matrix of the Hopfield model. Assuming a network of size N with *one* memory pattern let us examine the utility of choosing the connectivity matrix of the following form:

$$J_{ij} \propto \xi_i \xi_j \tag{1}$$

A minimal condition for the network to perform its function is that if the network state is already at one of the memory patterns then it should remain at this state. In other words, at the memory pattern, the output for each neuron should be equal to the memory pattern. Before we do that, we need to define the dynamics. Since we are working with binary neurons, and assuming asynchronous dynamics:

$$\begin{aligned} &Output(t+1) = sign(input(t)) \\ &S_i(t+1) = \Theta(h_i(t)) \;\; ; \;\; h_i(t) = \sum_j J_{ij} S_j(t) \end{aligned} \tag{2}$$

Or: $\forall i \;\; : sign(h_i) = \xi_i$. Now, a minimal requirement from the Hopfield model, as a model of associative memory, would be that if we start the network from one of the memories, that would be the memory that is the final result of the dynamics. Or in other words, that this network state will be stable. Let us check this notion:

$$\begin{aligned} &\forall i \;\; : S_i = \xi_i \quad \Rightarrow \quad \Theta(h_i) = \xi_i \\ &h_i = \sum_j J_{ij}\xi_j = \sum_j [\xi_i \xi_j]\xi_j = \xi_i \sum_j \xi_j \xi_j \\ &= \xi_i \sum_j \xi_j^2 = \xi_i \sum_j 1 = \xi_i N \end{aligned} \tag{3}$$

Our result is that the input to the ith neuron is *always* of the same sign as our one memory pattern, which means that it will always be stable. In addition, it is of size N. The fact that the pattern is always stable is an intuitive justification for Hopfield's choice.

Now consider starting from a network state that is not quite the correct memory pattern, but a corrupted version of it. By that we mean that the network state is equal to the memory pattern, except at a small number of neurons, k, in which it differs. Since we are working with binary neurons, different just means that the sign is flipped.

$$S_i = \begin{cases} \xi_i & \text{for } i = 1...(N-k) \\ \\ -\xi_i & \text{for } i = (N-k)...N \end{cases} \tag{4}$$

Let us look at the dynamics again:

$$\begin{aligned} &\forall i \ : S_i = \xi_i \quad \Rightarrow \quad \Theta(h_i) = \xi_i \\ &h_i = \sum_j J_{ij}\xi_j = \sum_j [\xi_i\xi_j]\xi_j = \xi_i \sum_j \xi_j\xi_j \\ &= \xi_i \sum_j \xi_j^2 = \xi_i \sum_j 1 = \xi_i N \end{aligned} \tag{5}$$

What happens if we would like to try and store more than one memory pattern? We extend the definition in the same manner to the following form:

$$J_{ij} \propto \sum_{\mu=1}^{P} \xi_i^\mu \xi_j^\mu \tag{6}$$

We saw that in equation (3) we ended up with an input of size N. In order to normalize this we will divide by N (not by p as would be intuitively guessed from the sum over p). In addition, we will set the diagonal terms to zero for reasons that will become evident in the future. Thus, our connectivity matrix will take the form:

$$J_{ij} = \frac{1}{N} \sum_{\mu=1}^{P} \xi_i^\mu \xi_j^\mu \ ; \ J_{ii} = 0 \tag{7}$$

Again, the most basic check of our model will be whether a network started from an initial condition of one of the memory patterns, say ξ_j^ν, will be stable for these initial conditions. Let us check:

$$\begin{aligned} &\forall i \ : S_i = \xi_i^\nu \quad \Rightarrow \quad h_i = \sum_j J_{ij}\xi_j^\nu = \sum_j [\frac{1}{N} \sum_{\mu=1}^{P} \xi_i^\mu \xi_j^\mu]\xi_j^\nu \\ &= \frac{1}{N}\xi_i^\nu \sum_j (\xi_j^\nu)^2 + \frac{1}{N} \sum_j \sum_{\mu \neq \nu}^{P} \xi_i^\mu \xi_j^\mu \xi_j^\nu = \frac{N-1}{N}\xi_i^\nu + \frac{1}{N} \sum_j \sum_{\mu \neq \nu}^{P} \xi_i^\mu \xi_j^\mu \xi_j^\nu \end{aligned} \tag{8}$$

Where the transition that took place between the first and second line of the equation was the splitting of the sum into the term were $\mu = \nu$ and the rest of the terms.

The first term, which is approximately of size 1, is of the same sign as the initial memory pattern and thus contributes towards the stability of the memory in the *ith* site. Due to this beneficial property it is termed the *signal* term. However, the second term can interfere with that stability and is hence termed the *noise* term. What is the size of this term? It depends on the particular choice of the sign of the neurons in the *ith* site in all memory patterns. Recall that these are determined randomly. Thus, we would like to now the *statistics* of the size of this term.

Importantly, though the sign of the memory patterns is a random variable, it is not of the same type of randomness as we have previously encountered in our stochastic analysis of neuron dynamics. The type we have encountered was a moment to moment randomness in the transitions of the neurons from a silent to an active state. This new type of randomness is chosen, randomly, *only once* (per network) and then left constant. Random variables like these are called *quenched random variables*. Their statistics in which we are interested are statistics over all possible choices of such networks. Averaging over quenched random variables is often denoted by double brackets <<>> as opposed to single brackets for averaging over standard random variables. For instance, what is the average of the sign of the *ith* neuron of a certain memory pattern in the unbiased case?

$$<< \xi_i^\mu >>= (1) * (P(\xi_i^\mu) = 1) + (-1) * (P(\xi_i^\mu) = -1) = 1 * \frac{1}{2} - 1 * \frac{1}{2} = 0 \quad (9)$$

As was expected by the term *unbiased*. Note that these averages are exactly like standard averages and have the same properties. Namely the two properties we will employ:

$$\begin{aligned} &<< X_1 + X_2 >>=<< X_1 >> + << X_2 >> \\ &if X_1 \text{ and } X_2 \text{ independent} : << X_1 * X_2 >>=<< X_1 >><< X_2 >> \end{aligned} \quad (10)$$

In other cases in the context of averaging over signs of neurons in memory patterns we will be performing averages over vector random variables:

$$<< y >>= \sum_{\vec{x} \in S(\vec{x})} P(\vec{x}) y(\vec{x}) \; ; \; P(\vec{x}) = P(X_1 = x_1, \ldots, X_n = x_n) \quad (11)$$

Now let us turn to calculate the average of the term we were interested in:

$$\begin{aligned} &<< \frac{1}{N} \sum_{\mu \neq \nu}^{P} \xi_i^\mu \xi_j^\mu \xi_j^\nu >>= \frac{1}{N} \sum_{\mu \neq \nu}^{P} << \xi_i^\mu >><< \xi_j^\mu >><< \xi_j^\nu >> \\ &= \frac{1}{N} \sum_{\mu \neq \nu}^{P} 0 * 0 * 0 = 0 \end{aligned} \quad (12)$$

Note that the transition in the first line is justified as the sign of any neuron in a memory pattern is independent of the sign of any other neuron, both within

the same memory pattern and between different memory patterns. The mean of the noise term is thus zero. This is good news, but we still have to address the other moments of the noise term. As the noise term is the sum over a large number of random variables we can assume it is a gaussian random variable due to the central limit theorem. Let us then calculate the variance of the noise term:

$$Var[noise] = E[noise^2] - E[noise]^2 = E[noise^2] - 0$$
$$=<< [\frac{1}{N}\sum_j \sum_{\mu\neq\nu}^{P} \xi_i^\mu \xi_j^\mu \xi_j^\nu]^2 >> = \frac{1}{N^2} << \sum_j \sum_{\mu\neq\nu}^{P} (\xi_i^\mu \xi_j^\mu \xi_j^\nu) \sum_{j'} \sum_{\mu'\neq\nu}^{P} (\xi_i^{\mu'} \xi_{j'}^{\mu'} \xi_{j'}^{\nu'}) >> \tag{13}$$

Can we now conclude that all the terms of the r.h.s are uncorrelated and that it will be just a multiplication of zeros as before? No! We must remember that the sign of a neuron in a memory pattern is correlated with *itself*. Thus, let us separate the above sum into those terms where $\mu = \mu', j = j'$ and the rest.

$$\frac{1}{N^2} << \sum_j \sum_{\mu\neq\nu}^{P} (\xi_i^\mu \xi_j^\mu \xi_j^\nu) \sum_{j'} \sum_{\mu'\neq\nu}^{P} (\xi_i^{\mu'} \xi_{j'}^{\mu'} \xi_{j'}^{\nu'}) >>$$
$$= \frac{1}{N^2} << [\sum_j \sum_{\mu\neq\nu}^{P} (\xi_i^\mu)^2 (\xi_j^\mu)^2 (\xi_j^\nu)^2 + \sum_j \sum_{\mu\neq\nu}^{P} (\xi_i^\mu \xi_j^\mu \xi_j^\nu) \sum_{j'\neq j} \sum_{\mu'\neq\nu\neq\mu}^{P} (\xi_i^{\mu'} \xi_{j'}^{\mu'} \xi_{j'}^{\nu'})] >>$$
$$= \frac{1}{N^2} << \sum_j \sum_{\mu\neq\nu}^{P} 1 >> + \frac{1}{N^2} << \sum_j \sum_{\mu\neq\nu}^{P} \sum_{j'\neq j} \sum_{\mu'\neq\nu\neq\mu}^{P} (\xi_i^{\mu'} \xi_{j'}^{\mu'} \xi_{j'}^{\nu'})(\xi_i^\mu \xi_j^\mu \xi_j^\nu) >>$$
$$= \frac{(N-1)(P-1)}{N^2} + 0 \simeq \frac{P}{N} \tag{14}$$

Were the last transition was justified by the fact that we will ultimately be interested in large N and large P. In summary we find that the noise term acts as a gaussian random variable with a mean of 0 and standard deviation $\sqrt{\frac{P}{N}}$.

Turning back to the original question of the stability of the memory pattern, we know that the signal is of size 1 and we know the statistics of the noise term. Thus, we can calculate the probability of the noise term destabilizing the *ith* neuron. If the memory pattern sign is positive then in order for the noise term to flip the sign it has to be lesser than -1. If the memory pattern sign is negative it has to be greater than 1. Note that due to the symmetry of the gaussian distribution we can consider just one of these cases. Let us consider that of the positive memory sign. In order for the memory to be *stable* we need the noise not to be lesser than 1, or just greater than 1. Let us calculate this probability:

$$P(stability) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-1}^{\infty} exp\Big(-\frac{x^2}{2\sigma^2}\Big) \mathrm{d}x = \frac{1}{2}\Big[1 + erf\Big(\sqrt{\frac{1}{2\sigma^2}}\Big)\Big] \tag{15}$$

We will now use the following approximation for the error function:

$$erf(x) \approx 1 - \frac{1}{\sqrt{\pi}x} e^{-x^2} \tag{16}$$

Plugging it in we get:

$$P(stability) = \frac{1}{2}\Big[1 + erf\Big(\sqrt{\frac{1}{2\sigma^2}}\Big)\Big] = \frac{1}{2}\Big[1 + 1 - \frac{1}{\sqrt{\pi}} \frac{1}{\sqrt{\frac{1}{2\sigma^2}}} exp\Big(-\sqrt{\frac{1}{2\sigma^2}}^2\Big)\Big]$$

$$= 1 - \frac{1}{2}\frac{1}{\sqrt{\pi}}\sqrt{\frac{1}{\frac{1}{2\pi\sigma^2}}} exp\Big(-\frac{1}{2\pi\sigma^2}\Big) = 1 - \frac{1}{2}\frac{1}{\sqrt{\pi}}\sqrt{2\sigma^2} exp\Big(-\frac{1}{2\sigma^2}\Big)$$

$$= 1 - \frac{\sigma}{\sqrt{2\pi}} \exp\Big(-\frac{1}{2\sigma^2}\Big) \tag{17}$$

This is the probability of stability, the probability of *instability* is the complementary probability:

$$P(instability) = 1 - P(stability) = 1 - \Big[1 - \frac{\sigma}{\sqrt{2\pi}} \exp\Big(-\frac{1}{2\sigma^2}\Big)\Big]$$

$$= \frac{\sigma}{\sqrt{2\pi}} \exp\Big(-\frac{1}{2\sigma^2}\Big) \tag{18}$$

This is the probability of a single neuron being unstable and ultimately yielding the wrong sign, or an error. The mean number of errors can be empirically approximated simply by the number of neurons times the probability of an error in each individual neuron:

$$N_{error} = NP(instability) = N\frac{\sigma}{\sqrt{2\pi}} \exp\Big(-\frac{1}{2\sigma^2}\Big) \tag{19}$$

We are interested in seeing when we will have exactly one error:

$$1 = N\frac{\sigma}{\sqrt{2\pi}} \exp\Big(-\frac{1}{2\sigma^2}\Big) \quad \Rightarrow \quad ln\Big[1\Big] = ln\Big[N\frac{\sigma}{\sqrt{2\pi}} \exp\Big(-\frac{1}{2\sigma^2}\Big)\Big]$$

$$0 = ln(N) + ln(\sigma) - \frac{1}{2}ln(2\pi) - \frac{1}{2\sigma^2} \;\Rightarrow \frac{N}{2P} = ln(N) + ln(\frac{P}{N}) - \frac{ln(2\pi)}{2} \tag{20}$$

Where in the last transition we plugged in our previous calculation regarding the noise, namely that $\sigma^2 = \frac{P}{N}$. Now we will leave only the leading terms. We are interested in cases were N is large and P is large but smaller than N. Thus, we will eliminate the second and third terms of the r.h.s of the final equation as they are smaller than the remaining two terms. Thus, we end up with:

$$\frac{N}{2P_{max}} = ln(N) \quad \Rightarrow \quad \frac{N}{2} = P_{max}ln(N) \quad \Rightarrow \quad P_{max} = \frac{N}{2ln(N)} \tag{21}$$

This is the estimate, based on statistical signal-to-noise analysis, of the maximal number of memory patterns one can embed in an unbiased Hopfield network. Note that this number grows with N (as N is larger than $ln(N)$) but it is a slightly disappointing result that the number of memory patterns that can be embedded grows not even linearly in N.