(Lesson ? of 9/12/19) Compiled: December 9, 2019

0.1 Hopfield Model - part 2

We arrived at:

$$Z = \sum_{\{S_1,\dots,S_N\}} \exp\left(\frac{\beta}{2}\sum_{i<j} J_{ij} S_i S_j\right) \qquad S_i = \{-1,+1\}$$

and, for the Hopfield model:

$$J_{ij} = \frac{1}{N}\sum_{\mu=1}^{p} \xi_1^\mu \xi_j^\mu$$

where $\boldsymbol{\xi}^\mu = \{\xi_1^\mu, \dots, \xi_N^\mu\}$, with $\mu = 1, \dots, p$ are p *patterns* that are initially stored in the network.
We rewrote the partition function as:

$$Z = \int_{-\infty}^{+\infty} \prod_{\mu=1}^{d} \mathrm{d}q_\mu \exp\big(-\beta N u(q_1, \dots, q_p)\big)$$

$$u(\boldsymbol{q}) = \frac{1}{2}\sum_{\mu=1}^{p} q_\mu^2 - \frac{1}{\beta N}\sum_{i=1}^{N} \underbrace{\ln\left[2\cosh(\beta \boldsymbol{q}\cdot\boldsymbol{\xi}_i)\right]}_{\ln 2 + \ln(\cosh\dots)} \qquad \boldsymbol{q}\cdot\boldsymbol{\xi}_i = \sum_{\mu=1}^{p} q_\mu \xi_i^\mu$$

To compute these integrals we use the *saddle point approximation*. So we look for the configuration $\boldsymbol{q}^* = \{q_1^*, \dots, q_p^*\}$ that *minimizes* $u(\boldsymbol{q})$:

$$\boldsymbol{q}^* = \min_{\boldsymbol{q}} u(\boldsymbol{q}) \Rightarrow \frac{\partial u}{\partial q_1} = 0, \frac{\partial u}{\partial q_2} = 0, \dots, \frac{\partial u}{\partial q_p} = 0 \tag{1}$$

Then, applying the approximation:

$$Z = e^{-\beta N u(q_1^*, \dots, q_p^*)}$$

And finally we can compute the *free energy*:

$$f = -\frac{1}{N\beta}\log(Z) = u(q_1^*, \dots, q_p^*)$$

So all that's left is to solve the p equations in (1):

$$\frac{\partial u}{\partial q_\nu} = q_\nu - \frac{1}{\beta N}\sum_{i=1}^{N} \frac{\sinh(\beta \boldsymbol{q}\cdot\boldsymbol{\xi_i})}{\cosh(\beta \boldsymbol{q}\cdot\boldsymbol{\xi_i})}\beta \xi_i^\nu \overset{!}{=} 0$$

$$\Rightarrow q_\nu = \frac{1}{N}\sum_{i=1}^{N} \tanh(\beta \boldsymbol{q}\cdot\boldsymbol{\xi_i})\xi_i^\nu$$

In vector notation:

$$\boldsymbol{q} = \frac{1}{N}\sum_{i=1}^{N} \tanh(\beta \boldsymbol{q}\cdot\boldsymbol{\xi_i})\boldsymbol{\xi_i}$$

We are interested in the case with *many neurons*, that is the limit for $N \to \infty$. So:

$$\frac{1}{N}\sum_{i=1}^{N} f(\xi) = \langle f\rangle = \int \mathrm{d}\xi\, p(\xi) f(\xi)$$

and we know that $\xi_i = -1, +1$ with the same $1/2$ probability. This means:

$$= \int \mathrm{d}\xi\, p(\xi)\left(\frac{1}{2}\sum_{x=\{\pm 1\}} \delta(x-\xi)\right)$$

leading to:

$$\boldsymbol{q} = \mathbb{E}[\tanh(\beta \boldsymbol{q}\cdot\boldsymbol{\xi})\boldsymbol{\xi}] \Rightarrow q_\mu = \mathbb{E}[\tanh(\beta \boldsymbol{q}\cdot\boldsymbol{\xi_i})\xi^\mu]$$

To find the physical interpretation of $\boldsymbol{q}$, recall, from the previous steps, that:

$$Z = \sum_{\{S_1,\dots,S_N\}} \int_{-\infty}^{+\infty} \prod_{\mu=1}^{p} \mathrm{d}q_\mu \exp\left(-\beta N \tilde{u}(\boldsymbol{q}; S_1,\dots,S_N)\right)$$
$$\tilde{u}(\boldsymbol{q}, S_1,\dots,S_N) = \frac{1}{2}\sum_{\mu=1}^{p} q_\mu^2 - \frac{1}{N}\sum_{\mu=1}^{p} q_\mu \sum_{i=1}^{N} S_i \xi_i^\mu$$

where we have also the *physical* parameters S_i (network's state). If we now repeat the previous steps, looking for the minimum of the exponential, we get:

$$\frac{\partial u}{\partial q_\nu} = 0 \Rightarrow q_\mu = \frac{1}{N}\sum_{i=1}^{N} S_i \xi_i^\mu$$

So we can interpret the q_μ as the *overlap* of the network's state with the μ-th pattern of the neural network.
Suppose now $\boldsymbol{q} = (q_1, 0, \dots, 0)$ (vector with only the first component non-zero). Plugging it in the equations:

$$\text{First eq.}: q_1 = \mathbb{E}[\tanh\big(\beta q_1 \xi^1\big)\xi^1]$$
$$\text{Last } p-1 \text{ eqs.}: q_\nu = \mathbb{E}[\tanh\big(\beta q_i \xi^1\big)\xi^\nu] \qquad \nu \neq 1$$

Note that:

$$q_1 = \frac{1}{2}\tanh(\beta q_1)\cdot 1 - \frac{1}{2}\tanh(\beta q_1)(-1)$$
$$q_\nu = \sum_{\xi_1,\dots,\xi_p} p(\xi_1)\cdots p(\xi_p)\tanh\big(\beta \xi^1\big)\xi^\nu$$

So the last $p-1$ equations are satisfied by $q_\nu = 0$, and we are left with the *first equation* $q = \tanh(\beta q)$, which is similar to the equation for the magnetization in the Ising model: $m = \tanh(\beta m)$. We then observe that, if we sample configurations with a Boltzmann-probability:

$$\exp\left(\beta \sum_{ij} J_{ij} S_i S_j\right)$$

for $T < T_c$, where T_c is a certain *critical temperature*, we have a non-zero probability to sample a network configuration that is *strongly (anti)correlated* with one of the patterns. [Insert fig.1]

0.2 Sherrington-KirkPatrick Model

The SK model is a netweork *without any pattern embedded inside*. We start by recalling the energy function for the Hopfield Model:

$$H = -\sum_{i<j} J_{ij} S_i S_j \qquad S_i = \{-1, +1\}$$

However, we now pick the J_{ij} weights *at random*, according to a Gaussian distribution:

$$p(J_{ij}) = \frac{1}{\sigma}\sqrt{\frac{N}{2\pi}} \exp\left(-\frac{N}{2\sigma^2} J_{ij}^2\right)$$

Each spin S_i interacts with *all other spins* (**long range interaction model**) with a *random strength*. Note that:

$$\langle J^2 \rangle \sim \frac{1}{N}$$

This choice will lead to an *extensive* total free energy, that is:

$$\begin{aligned} F &= \frac{1}{\beta}\log(Z_J) \sim N \\ Z_J &= \sum_{\{S_1,\dots,S_N\}} \exp\left(-\beta H_J[S_1, \dots, S_N]\right) \end{aligned}$$

Note that now the partition function *explicitly depends* on the system's realization (the choice of J_{ij}). Also, the number of *connections* is in the order of $O(N^2)$, while in the Ising's model (local interactions) we had $O(N)$.

We can check that F is linear in N by doing a *high-temperature expansion* (small β expansion) of Z:

$$\begin{aligned} Z_J &= \sum_{\{S_1,\dots,S_N\}} \exp\left(\beta \sum_{i<j} J_{ij} S_i S_j\right) = \\ &\approx \sum_{\{S_1,\dots,S_N\}} \left(1 + \beta \sum_{i<j} J_{ij} S_i S_j + \frac{\beta^2}{2} \sum_{i<j}\sum_{k<l} J_{ij} J_{kl} S_i S_j S_k S_l\right) \end{aligned}$$

Note that if we have a product of $p \in \mathbb{N}$ spins, with an odd number of copies of index k, their sum over *all states* will be 0:

$$\sum_{S_{i_k}=\{\pm 1\}} S_{i_1} S_{i_2} \cdots \overbrace{S_{i_k} \cdots S_{i_k}}^{2m+1} \cdots S_{i_p} = 0$$

Also:

$$\sum_{\{S_1,\dots,S_N\}} 1 = 2^N$$

So we can expand Z_J:

$$\begin{aligned} Z_J &\approx 2^N + \underbrace{\sum_{\{S_1,\dots,S_N\}} \sum_{i<j} J_{ij} S_i S_j}_{=0} + \frac{\beta^2}{2} \underbrace{\sum_{i<j} \sum_{k<l} J_{ij} J_{kl} S_i S_j S_k S_l}_{\sum_{i<j} J_{ij}^2} = \\ &= 2^N \left(1 + \frac{\beta^2}{2} \sum_{i<j} J_{ij}^2 + O(\beta^3) \right) \end{aligned}$$

Taking the logarithm:

$$\begin{aligned} \log(Z_J) &= N \log(2) + \log\left(1 + \frac{\beta^2}{2} \sum_{i<j} J_{ij}^2 + O(\beta^3) \right) = \\ &= N \log(2) + \frac{\beta^2}{2} \underbrace{\sum_{i<j} J_{ij}^2}_{\sim N^2 \langle J^2 \rangle} + O(\beta^2) = \\ &= N \log(2) + \frac{\beta^2}{2} N^2 \langle J^2 \rangle + \cdots \sim F \end{aligned}$$

So to have $F \sim N$ we need $\langle J^2 \rangle = 1/N$, confirming the choice we made before. Now, consider again the free energy:

$$f_J = -\frac{1}{N\beta} \log\left(\sum_{\{\boldsymbol{S}\}} \exp\left(\beta \sum_{i<j} J_{ij} S_i S_j \right) \right) = -\frac{1}{N\beta} \ln(Z_J)$$

And we are interested in the $N \to \infty$ limit. We would like that, in this limit, the result will not depend on the specific choice of J_{ij}, meaning that *averages over disorder* make sense. This happens with free energy, and we say that it is *self-averaging*, that is:

$$\lim_{N \to \infty} \frac{\overline{[F_J^2]} - \overline{[F_J]}^2}{\overline{[F_J]}^2} \sim \frac{1}{\sqrt{N}}$$

where $[\dots]$ denotes an *average over disorder*:

$$\overline{[f_J]} = \int \prod_{i<j} \mathrm{d}J_{ij}\, p(J_{ij}) f(\{J_i j\}_{i<j})$$

In other words, this mean that the *free energy* takes a *more and more* "definite" value (i.e. its distribution $p(f)$ has a smaller width) as we consider a larger and larger system:

$$\lim_{N\to\infty} -\frac{1}{N\beta}\log(Z_J) = \lim_{N\to\infty} -\frac{1}{N\beta}\overline{\log(Z_J)} = f$$

[Insert fig.2] However, if we write that integral:

$$\overline{[f_J]} = \int \prod_{i<j} \mathrm{d}J_{ij}\, p(J_{ij}) \left(-\frac{1}{N\beta}\right) \log\left(\sum_{\{\boldsymbol{S}\}} \exp\left(\beta \sum_{i<j} S_i S_j J_{ij}\right)\right)$$

we note that the J_{ij} appears both as the variables of integration, and as terms of the sum over all states, leading to a very difficult expression to evaluate. To simplify the problem we use the **Replica trick**, that involves rewriting the logarithm in terms of its Taylor expansion:

$$\log(x) = \lim_{n\to 0} \frac{x^n - 1}{n}$$

Then:

$$\log\left(\sum_{\{\boldsymbol{S}\}}\right) \exp\left(\beta \sum_{i<j} S_i S_j J_{ij}\right) = \lim_{n\to 0} \frac{\sum_{\{\boldsymbol{S}\}} \exp\left(\beta \sum_{i<j} J_{ij} S_i S_j\right)^n - 1}{n}$$

Let's focus on the power term:

$$\int_{-\infty}^{+\infty} \prod_{i<j} \mathrm{d}J_{ij}\, p(J_{ij}) \sum_{\substack{\{S_i^\alpha,\dots,S_N^\alpha\}\\ \alpha=1,\dots,n}} \exp\left(\beta \sum_{\alpha=1}^{n} \sum_{i<j} J_{ij} S_i^\alpha S_j^\alpha\right) \tag{2}$$

where $n \in \mathbb{N}$ for all the intermediate steps, but at the end we take $n \to 0$ as if it were a real parameter. The index α *labels* the *replicas* of the system, that is the elements of a set of n copies of the original system.
Note now that:

1. Replicas are uncoupled: there are no products $S_i^\alpha S_j^\beta$ with $\alpha \neq \beta$ (replicas *do not interact*)

2. Spins are coupled: there are products of spins carrying different indexes, such as $S_i^\alpha S_j^\alpha$

By performing a *gaussian integration* we can *move* the coupling from spins to replicas, so that:

1. Replicas become coupled

2. Spins become uncoupled

The idea is that the energy will form a *many valleys landscape*. If we now consider two copies (replicas) evolving in this landscape, they will *behave* as *non-interacting particles* if the temperature is high enough, but will *strongly interact* when the temperature is low. This is the meaning of "coupled replicas", as we will now mathematically derive.

This part may contain errors!

We start by integrating a single term of (2):

$$\int_{-\infty}^{+\infty} \mathrm{d}J_{ij} \exp\left(-\frac{N}{2\sigma^2}J_{ij}^2 + \beta J_{ij}\sum_{\alpha=1}^{n} S_i^\alpha S_j^\alpha\right) = \exp\left(\frac{\beta^2\sigma^2}{2N}\sum_{\alpha,\beta=1}^{n} S_i^\alpha S_i^\beta S_j^\alpha S_j^\beta\right) =$$

$$= \exp\left(\frac{\beta^2\sigma^2}{2\cdot 2N}\sum_{\alpha,\beta=1}^{n}\sum_{i\neq j} S_i^\alpha S_j^\beta\right) =$$

$$\approx \exp\left(\frac{\beta^2\sigma^2}{4N}\sum_{\alpha,\beta=1}^{n}\left(\sum_{i=1}^{N} S_i^\alpha S_i^\beta\right)^2\right)$$

and then:

$$\overline{\log(Z_J)} = \lim_{n\to 0}\frac{\overline{Z^n}-1}{n} \Rightarrow \lim_{n\to 0}\overline{Z^n}$$

and so:

$$\overline{Z^n} = \sum_{\substack{\{S_1^\alpha,\ldots,S_N^\alpha,\}\\ \alpha=1,\ldots,n}} \exp\left(\frac{\beta^2\sigma^2}{4N}\sum_{\alpha,\beta=1}^{n}\left(\sum_{i=1}^{N} S_i^\alpha S_i^\beta\right)^2\right)$$