CPSC 664, Spring 2023 Scribe: Jinzhao Wu
Apr 4, 2023 Lecturer: Tim Kunisky

LECTURE 19: Replica Symmetry Breaking

We first continue our calculation of the RS solution of SK model.

1 Replica Analysis of SK Model cont'd

The replica method told us to calculate the term $\mathbb{E}[Z(\beta)^n]$. During the last lecture, In the last lecture, we derived an approximation for this expression as follows:

$$\mathbb{E}[Z(\beta)^n] \approx \mathsf{exp}\left(N \sup_{\substack{Q \succcurlyeq 0 \\ Q_{i,i}=1}} I(Q) \right)$$

Let Q^* be the optimal solution. the optimal solution Q^* corresponds to the renormalized Gram matrix of n i.i.d. vectors, $x_1, x_2, \cdots, x_n$, drawn from the Gibbs measure μ_β. The idea is that suppose the optimal solution Q^* looks like

$$Q^* = \begin{pmatrix} 1 & a & \cdots & a \\ a & 1 & \cdots & a \\ \vdots & \vdots & \ddots & \vdots \\ a & a & \cdots & 1 \end{pmatrix}$$

However, the typical correlation matrix cannot always concentrate around such matrix when we draw any copy. Recall that

$$\mu_\beta(x) \propto \mathsf{exp}\left(\beta H(x)\right), \text{ where } H(x) = x^T W x.$$

In particular, this means that $\mu_\beta(x) = \mu_\beta(-x)$ for any x. Then, for any independent $x_1, x_2, \cdots, x_n$, it is clear that

$$\mu_\beta^{\otimes n}(x_1, \cdots, x_n) = \mu_\beta^{\otimes n}(\pm x_1, \cdots, \pm x_n).$$

This suggests that if we draw $x_1, x_2, \cdots, x_n$, and write down the correlation matrix. The probability of observing Q is always equal to the probability of observing DQD where D is a diagonal matrix.

$$D = \begin{pmatrix} \pm 1 & 0 & \cdots & 0 \\ 0 & \pm 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \pm 1 \end{pmatrix}$$

In this interpretation, the overlap matrix Q does not concentrate around a single specific matrix. Instead, it concentrates around an "orbit," which encompasses all possible 2^n choices of D. This means that there are 2^{n-1} optimal Q which are equal to DQ^*D for some Q^* looks like the matrix above.

Let's first consider the case when $a = 0$. When we draw a bunch of copies, the Gram matrix resembles the identity matrix, suggesting that μ_β is nearly a uniform distribution over the hypercube. However, when $a > 0$, μ_β tends to concentrate around the hypercube, where the inner product of two independent copies typically holds a certain value.

Figure 1: Picture of μ_β looks like

Now let us go back to where we left off last time. The prediction for $f(\beta)$ is as follows:

$$f(\beta) = \lim_{n\to 0} \lim_{N\to\infty} \frac{1}{Nn} \log \mathbb{E}\, Z(\beta)^n$$

Assuming that we can interchange the order of these two limitations, we get that:

$$f(\beta) = \sup_a \inf_b \underbrace{\left\{ \log 2 + \beta^2(1-a^2) + \mathop{\mathbb{E}}_{g\sim N(0,1)} \log\cosh\left(g\sqrt{2b}\right) - b + ab \right\}}_{\stackrel{\text{def}}{=} h(a,b)}$$

By taking the derivative with respect to a, we obtain:

$$\frac{\partial h(a,b)}{\partial a} = -2\beta^2 a + b.$$

Setting the derivative to 0, we find that $b = 2\beta^2 a$. By plugging this result back, we reduce the problem to a one-dimensional optimization.

$$f(\beta) = \sup_a \underbrace{\left\{ \log 2 + \beta^2(1-a)^2 + \mathop{\mathbb{E}}_{g} \log\cosh\left(2 - \beta g\sqrt{a}\right) \right\}}_{\stackrel{\text{def}}{=} g(a)}$$

Now let us again optimize this with respect to a, we get that

$$\begin{aligned} \frac{\partial g(a)}{\partial a} &= -2\beta^2(1-a) + \frac{\beta}{\sqrt{a}} \mathop{\mathbb{E}}_{g}\, g \tanh\left(2\beta g\sqrt{a}\right) \\ &= -2\beta^2(1-a) + \frac{\beta}{\sqrt{a}} 2\beta\sqrt{a} \mathop{\mathbb{E}}_{g} \operatorname{sech}\left(2\beta g\sqrt{a}\right)^2. \end{aligned}$$

where the last equation follows from Gaussian integration by parts, i.e.,

$$\mathop{\mathbb{E}}_{g\sim N(0,1)}\left[g\cdot l(g)\right]=\mathop{\mathbb{E}}_{g\sim N(0,1)}\left[l'(g)\right].$$

Forcing the derivative to be 0, we deduce that that

$$a=\mathop{\mathbb{E}}_{g}\mathsf{tanh}\left(2\beta g\sqrt{a}\right)^2.$$

And we could see that a is a fixed-point of some function. This provides a SK solution of SK model. That is to say,

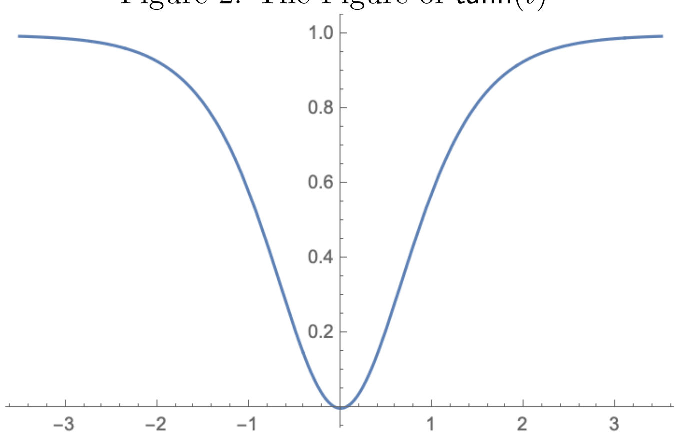
$$f(\beta)=\mathsf{log}\,2+\beta^2(1-a)^2+\mathop{\mathbb{E}}_{g}\mathsf{log}\,\mathsf{cosh}\,(2\beta g)$$

for a solving

$$a=\mathop{\mathbb{E}}_{g}\mathsf{tanh}\left(2\beta g\sqrt{a}\right)^2.$$

We first have a look at the figure of $\mathsf{tanh}(t)^2$. The function value is always between $[0,1]$.

Figure 2: The Figure of $\mathsf{tanh}(t)^2$



Particularly, $\mathsf{tanh}(0)^2=0$. This implies that $a=0$ is always a solution. Meanwhile, there is another solution a_1 exists iff $\beta>\frac{1}{2}$. We call this solution $\alpha_1(\beta)$.

2 Potential Issues with the Prediction

2.1 Ground State

We now look back to the ground state. Recall that in the last lecture, we define

$$2P_*=\lim_{N\to\infty}\frac{1}{N}\mathop{\mathbb{E}}_{W}\max_{x\in\{\pm1\}^n}x^TWx$$

Notice that

$$\begin{aligned}\lim_{N\to\infty}\frac{1}{N}\mathop{\mathbb{E}}_{W}\max_{x\in\{\pm1\}^n} x^TWx &= \lim_{\beta\to\infty}\frac{1}{\beta}f(\beta)\\ &= \lim_{\beta\to\infty}\frac{1}{\beta}\lim_{N\to\infty}\frac{1}{N}\mathop{\mathbb{E}}_{W}\log\left(\sum_x \exp\left(\beta x^TWx\right)\right)\end{aligned}$$

Notice that here is actually taking the softmax without normalization. We use the previous prediction, and observe that as β approaches infinity, $\alpha_1(\beta)$ approaches 1 sufficiently fast. What means that $\beta(1-a_1(\beta)^2))$ goes to 0 as β tends to infinity. Therefore, we could see that

$$\begin{aligned}2P_* &= \lim_{\beta\to\infty}\frac{1}{\beta}\mathop{\mathbb{E}}_{g}\log\cosh\left(2\beta g\right)\\ &= \lim_{\beta\to\infty}\frac{1}{\beta}\mathop{\mathbb{E}}_{g}\log\left(\left(\exp(2\beta g)+\exp(-2\beta g)\right)/2\right)\\ &= \mathbb{E}\max\left\{2g,-2g\right\}\\ &= 2\,\mathbb{E}\,|g|\\ &= 2\sqrt{\frac{2}{\pi}}.\end{aligned}$$

where the third step is derived from that we treat the term as a softmax.

Till now, we have a predicted value for $2P^*$, i.e. $2\sqrt{\frac{2}{\pi}}$, which is roughly 2.0798. However, the empirical value of $2P^*$ is 2.0763. This indicates that the RS prediction of the ground state is too big.

2.2 Entropy

Now we consider about entropy. Let us consider the entropy of the Gibbs measure:

$$\mathsf{Ent}\left(\mu_\beta\right) = -\sum_x \mu_\beta(x)\log\mu_\beta(x).$$

Expanding the term, we get that

$$\begin{aligned}\mathsf{Ent}\left(\mu_\beta\right) &= -\sum_x \mu_\beta(x)\log\mu_\beta(x)\\ &= -\sum_x \frac{1}{Z(\beta)}\exp\left(\beta H(x)\right)\left(\beta H(x)-\log Z(\beta)\right)\\ &= \log Z(\beta) - \frac{\beta}{Z(\beta)}\sum_x \exp\left(\beta H(x)\right)H(x)\\ &= \log Z(\beta) - \frac{\beta}{Z(\beta)}Z'(\beta)\\ &= \log Z(\beta) - \beta\left(\log Z(\beta)\right)'\\ &= F(\beta)-\beta F'(\beta)\end{aligned}$$

Therefore, let us define $\mathsf{ent}(\beta)$ as:

$$\mathsf{ent}(\beta) = \lim_{N\to\infty} \frac{1}{N} \mathsf{Ent}\left(\mu_\beta\right) = f(\beta) - \beta f'(\beta).$$

However, for sufficiently large values of β, it turns out that $\mathsf{ent}(\beta)$ becomes negative. This result is problematic because entropy should always be non-negative.

2.3 Stability of Q^*

We now have a look at the stability of Q^* where

$$Q^* = \begin{pmatrix} 1 & a_1(\beta) & \cdots & a_1(\beta) \\ a_1(\beta) & 1 & \cdots & a_1(\beta) \\ \vdots & \vdots & \ddots & \vdots \\ a_1(\beta) & a_1(\beta) & \cdots & 1 \end{pmatrix}$$

and Q^* should be a maximizer of $I(Q)$. We could see that $\nabla I(Q^*) = 0$. However, its Hessian matrix $\nabla^2 I(Q^*)$ possesses positive eigenvalues for sufficiently large values of β. This observation implies that local improvements may be possible, indicating that Q^* might not be the true maximizer. This further suggests that the optimal solution might exhibit more complex geometrical structures.

3 Replica Symmetry Breaking

The issues raised above suggests that we need less symmetric guess for the optimal matrix Q^*. There is a natural way to guess less symmetrically. That is to divide the matrix into blocks.

$$Q^* = \begin{array}{c|c|c} \begin{matrix} 1 & & a_1 \\ & \ddots & \\ a_1 & & 1 \end{matrix} & a_0 & a_0 \\ \hline a_0 & \begin{matrix} 1 & & a_1 \\ & \ddots & \\ a_1 & & 1 \end{matrix} & a_0 \\ \hline a_0 & a_0 & \begin{matrix} 1 & & a_1 \\ & \ddots & \\ a_1 & & 1 \end{matrix} \\ \underbrace{\hphantom{aaaaaa}}_{m_1} & \underbrace{\hphantom{aaaaaa}}_{m_1} & \underbrace{\hphantom{aaaaaa}}_{m_1} \end{array}$$

One could guess the optimal matrix as above or even with more hierarchies. Upon performing calculations with these assumptions, it appears that the issues are partially mitigated. The predicted P^* becomes more accurate, the threshold β at which $\mathsf{ent}(\beta)$ becomes negative increases, and the eigenvalues approach 0.

With these new optimal matrices, we can come up with new geometries for the distribution. Now, let us shift our focus to the geometry and consider the sphere $\mathbb{S}^{n-1}$ in $\mathbb{R}^n$ instead of the hypercube. In the replica symmetric setting discussed earlier, we identified two distinct cases, as illustrated previously. When $a = 0$, μ_β is approximately uniform over the sphere. However, due to the high-dimensional nature of the space, we observe that for any particular direction, the mass is concentrated near the equator. When $a > 0$, a higher ring emerges, and two independent points drawn from this ring exhibit a typical inner product.

Upon zooming in on the ring, which is $\mathbb{S}^{n-2}$, we observe no further structures. In other words, it appears similar to the uniform case. For the one-step replica symmetry breaking, after zooming in the ring which is $\mathbb{S}^{n-2}$, the probability distribution concentrates on a bunch of sub-spheres. When we look at one of those, there are no further structures. That is to say, we have one more step recursion than the case that $a > 0$.