CPSC 664, Spring 2023 — Scribe: Curtis McDonald
April 11, 2023 — Lecturer: Tim Kunisky

Lecture 21: Rigorous Analysis of SK Model

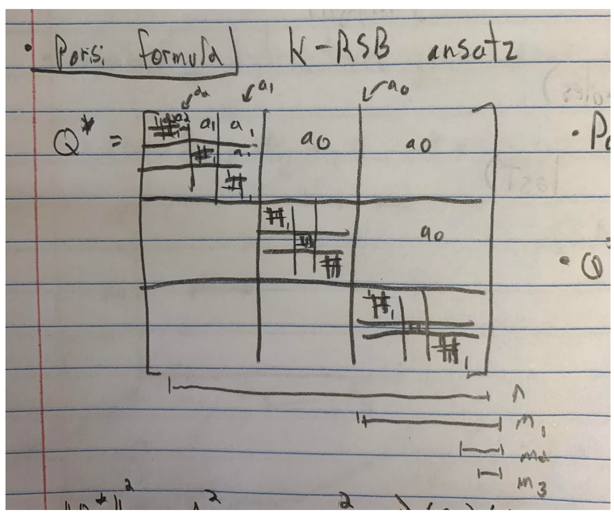
1 Proof Ideas for SK Model

More about Parisi Formula

Overlap Density

Last class, we discussed the Parisi formula for analyzing the limit of the free energy as $n \to 0$. We saw this unusually "flipping" of the order of indexes m_j to now be between 0 and 1. Here we review the conclusions:

Recall the block replica symmetry breaking ansatz matrix $Q^* \in \mathbb{R}^{n\times n}$ with parameters $n \geq m_1 \geq \cdots, m_k > 1$ and $a_0 \leq a_1 \leq a_k \leq \cdots 1$. The entry a_j appears $n(m_j - m_{j+1})$ times, thus the proportion of the matrix which is entry a_j is $\frac{n(m_j-m_{j+1})}{n(n-1)}$. As $n \to 0$, this approaches $m_{j+1} - m_j$, i.e. the "probability" of drawing a_j from the matrix approaches $m_{j+1} - m_j$.



Thus, consider $n = 0 = m_0 \leq m_1 \leq \cdots, \leq m_{k+1} = 1$ and define the step function

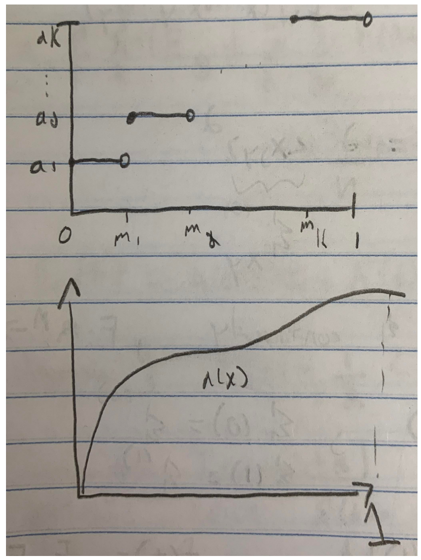
$$
\begin{aligned}
a : [0,1] &\to [0,1] \\
x &\mapsto \sum_{j=1}^{k+1} 1\{m_{j-1} \leq x < m_j\} a_j
\end{aligned}
$$

that is $a(x)$ is constant value a_j on interval $[m_{j-1}, m_j)$.

Consider this as the cumulative distribution function (CDF) of the overlap distribution μ which is a probability measure over the interval $[0, 1]$ defined by

$$\mu([B, A]) = a^{-1}(B) - a^{-1}(A)$$

As the number of replica symmetry breaking terms $k \to \infty$, the step function approaches a continuous CDF function which defines the overlapp density μ via this CDF.



Connections to Heat Formula

In the Parisi formula, we have second derivative terms $\frac{\partial^2}{\partial h^2}$ appearing from the expression $E_{g \sim N(0,1)}[f(g + h)]$. This is a consequence of the heat equation.

The heat equation tells us the solution of the PDE

$$\frac{\partial}{\partial t} u(t, h) = \frac{1}{2} \frac{\partial^2}{\partial h^2} u(t, h) \tag{1}$$

with initial condition $u(0, h) = f(h)$. I.e., the time derivative of u is equal to half the second derivative in space, with initial condition at time 0. The solution to the heat equation is

$$u(t, h) = E_{g \sim N(0,t)}[f(g + h)] \tag{2}$$

thus by the heat equation,

$$\frac{\partial}{\partial t} E_{g \sim N(0,t)}[f(g + h)] = \frac{\partial^2}{\partial h^2} E_{g \sim N(0,t)}[f(g + h)]$$

Talagrand Proof Ideas

Consider $x \in \{\pm \frac{1}{\sqrt{N}}\}^N$ (so hyper cube normalized to be unit vectors) and we are given a random matrix W. For each entry x in unit hypercube, we have value $x^T W x$. Put these together in a vector $G_x^0 \in \mathbb{R}^{2^N}$

$$G_x^0 = \begin{pmatrix} x_1^T W x_1 \\ x_2^T W x_2 \\ \vdots \\ x_{2^N}^T W x_{2^N} \end{pmatrix}$$

can also think of G_x^0 being defined by the choice of W, thus is a mapping from $R^{N\times N} \to R^N$, maps a matrix to a vector. When we write G_x^0, G_y^0 we refer to single entries of the vector G^0 at values x and y. We note a few properties of the mean and covariance of the vector G_x^0,

$$\begin{aligned} E_W[G_x^0] &= 0 \\ E_W[G_x^0 G_y^0] &= E_W[x^T W x y^T W y] \\ &= E_W[\left(\sum_{i=1}^N x_i^2 W_{i,i} + 2\sum_{j>i} W_{i,j} x_i x_j\right)\left(\sum_{i=1}^N y_i^2 W_{i,i} + 2\sum_{j>i} W_{i,j} y_i y_j\right)] \\ &= \frac{2}{N}\sum_{i=1}^N x_i^2 y_i^2 + \frac{4}{N}\sum_{j>i} x_i x_j y_i y_j \\ &= \frac{2}{N}\langle x, y\rangle^2 \end{aligned}$$

The quantity we are interested in studying is the expected maximum of $\frac{1}{N} x^T W x$ over the original hypercube, which is equal to

$$E_W[M(W)] = E_W[\max_{x\in\{\pm\frac{1}{\sqrt{N}}\}} G_x^0] \tag{3}$$

that is, the expected maximum entry of the G_x^0 vector. We will provide an upper bound on this quantity.

2 Gaussian Interpolation

Gaussian interpolation is a method to bound the expected value of a function over a Gaussian by comparing the value to an expectation over a simpler Gaussian which we can compute.

For some function $F : \mathbb{R}^M \to \mathbb{R}$, we want to study $E_{g\sim N(0,\Sigma^0)}[F(g)]$ that is the expected value of F over a Gaussian with covariance Σ^0. We construct a continuous evolution from covariance matrix $\Sigma^{(0)} \to \Sigma^{(1)}$ by $\Sigma : [0,1] \to R_{\geq 0}^{M\times M}$. Define the function of time

$$f(t) = E_{g\sim N(0,\Sigma(t))}[F(g)]$$

by fundamental theorem of calculus,

$$f(1) - f(0) = \int_0^1 f'(t) dt$$
$$f'(t) = \frac{\partial}{\partial t} E_{g\sim N(0,\Sigma(t))}[F(g)]$$

if we can show this derivative $f'(t) > 0$ for all t, then $f(1) > f(0)$ and solving $f(1)$ provides an upper bound on the quantity $f(0)$ we want to study. Pick a matrix $A(t)$ such that the covariance matrix is the square of this matrix $\Sigma(t) = A(t)A(t)$. Then

$$f(t) = E_{g\sim N(0,I)}[F(A(t)g)]$$

all t dependence is inside F function and we can take derivatives more easily.

$$\begin{aligned}
f'(t) &= \frac{\partial}{\partial t} E_{g\sim N(0,I)}[F(A(t)g)] \\
&= E_{g\sim N(0,I)}[\frac{\partial}{\partial t} F(A(t)g)] \\
&= E_g[\langle \nabla F(A(t)g), A'(t)g\rangle] \\
&= E_g[\sum_{i=1}^{M} \partial_i F(A(t)g) \sum_j A'(t)_{i,j} g_j] \\
&= \sum_j E_g[\sum_{i=1}^{M} \partial_i F(A(t)g) A'(t)_{i,j} g_j] \\
&= \sum_j E_g[\sum_{i=1}^{M}\sum_{k=1}^{M} \partial_i\partial_k F(A(t)g) A(t)_{k,j} A'(t)_{i,j}] \quad \text{(Integration by parts)} \\
&= \sum_i \sum_k E_g[\partial_i\partial_k F(A(t)g)(A(t)A'(t))_{i,k}] \\
&= E_g[\langle \nabla^2 F(A(t)g), A(t)A'(t)\rangle]
\end{aligned}$$

that is, expected sum of entry wise multiplication of hessian of F and $A(t)A'(t)$. We can "symmetrize" the inner product, for any matrix H

$$\langle H, A(t)A'(t)\rangle = \langle H, \frac{A(t)A'(t) + A'(t)^T A(t)^T}{2}\rangle$$

thus define the symmetrized matrix

$$\Sigma'(t) = A(t)A'(t) + A'(t)^T A(t)^T$$

which yields the following lemma.

Lemma 2.1.

$$f'(t) = \frac{1}{2}\langle \Sigma'(t), E_{g\sim N(0,I)}[\nabla^2 F(g)]\rangle$$

Consequences of Lemma

We want to deduce situations where the above inner product value is positive, thus $f(1)$ provides an upper bound on $f(0)$. Say we use Gaussian interpolation

$$\Sigma(t) = (1-t)\Sigma^{(0)} + t\Sigma^{(1)}$$
$$\Sigma'(t) = \Sigma^{(1)} - \Sigma^{(0)}$$

constant time change for any t. Thus as a corollary

Corollary 2.2. *If* $\langle \nabla^2 F(x), \Sigma^{(1)} - \Sigma^{(0)} \rangle \geq 0 \ \forall\ x$*, then we have that* $f(0) \leq f(1)$*.*

Note that we only want the expected value to be positive, so ensuring the inner product is positive at all locations is a more restrictive condition then we need but implies the expectation is positive. Furthermore, if $F(x)$ is convex, it's Hessian is positive definite, thus we would only require $\Sigma^{(1)} - \Sigma^{(0)}$ to be PSD as well.

Corollary 2.3. *If* F *is convex and* $\Sigma^{(1)} - \Sigma^{(0)}$ *is PSD, then* $f(1) \geq f(0)$*.*

Finally, as a most restrictive condition of all, we can ensure that entry wise, the multiplication of $\nabla^2 F$ and $\Sigma^{(1)} - \Sigma^{(0)}$ is positive

Corollary 2.4. *If*

$$\begin{cases} \partial_i \partial_j F(x) \geq 0 & \Sigma^{(1)}_{i,j} \geq \Sigma^{(0)}_{i,j} \\ \partial_i \partial_j F(x) \leq 0 & \Sigma^{(1)}_{i,j} < \Sigma^{(0)}_{i,j} \end{cases}$$

for all x*, then* $f(1) \geq f(0)$*.*

3 Application of Gaussian Interpolation to Maximum Value Problem

Our goal is to study the function

$$F(g) = \max_{x \in \{\pm \frac{1}{\sqrt{N}}\}} g_x$$

but instead we study the differentiable soft max function and take $\beta \to \infty$,

$$F_\beta(g) = \frac{1}{\beta} \log(\sum_x e^{\beta g_x})$$

called the "Free Energy."

$$\partial_i F_\beta(g) = \frac{e^{\beta g_i}}{\sum_j e^{\beta g_j}}$$

denote the value $p_i(g) = \frac{e^{\beta g_i}}{\sum_j e^{\beta g_j}}$. Note p as a vector is positive and sums to one, can be thought of as a probability density or weighting over values i. Second derivative is then

$$\partial_i \partial_j F_\beta(g) = \begin{cases} p_i(g) - p_i(g)^2 & i = j \\ -p_i(g) p_j(g) & i \neq j \end{cases}$$

since the $p_i(x)$ are all positive and less than 1, we have that $p_i(g) - p_i(g)^2 \geq 0$ and $-p_i(g)p_j(g) \leq 0$. Thus, by element wise corollary 2.4, we want a matrix $\Sigma^{(1)}$ which is larger than $\Sigma^{(0)}$ on all diagonal entries, and smaller on all off diagonal entries.

Lemma 3.1 (Weak Sudakov-Fernique Inequality). *If* $\Sigma^{(1)}_{i,i} \geq \Sigma^{(0)}_{i,i}$, $\Sigma^{(1)}_{i,j} \leq \Sigma^{(0)}_{i,j} i \neq j$ *then*

$$E_{\Sigma^{(0)}}[\max_x g_x] \leq E_{\Sigma^{(1)}}[\max_x g_x]$$

For example, consider a rank 1 matrix $\Sigma = 11^T$. By above result, max value under this covariance is less than max value under identity covariance $\Sigma = I_{M\times M}$. Furthermore, expected value is even larger for matrix with diagonal entries 1 and non-diagonal entries constant $-\frac{1}{M-1}$. Thus weak Sudakov inequality gives a convenient way to compare expected max values simply by properties of covariance matrix entries.

4 Preview of Next Class

We will cover the Strong Sudakov Fernique inequality, which covers the covariance matrices defined by

$$\Sigma^{(0)}_{xy} = \frac{2}{N}\langle x, y\rangle^2$$
$$\Sigma^{(1)}_{xy} = \frac{4}{N}\langle x, y\rangle$$

turns out $\Sigma^{(1)}$ above is covariance matrix of Gaussian process

$$G^{(1)}_x = \frac{2}{\sqrt{n}}\langle g, x\rangle$$

under these results

$$\begin{aligned} E(\frac{M(W)}{N}) \leq E(\max_{x\in\{\pm\frac{1}{\sqrt{N}}\}}\langle g, x\rangle] &= \frac{2}{N}E[\|g\|_1] \\ &= 2E_{g\sim N(0,1)}[|g|] \\ &= 2\sqrt{\frac{2}{\pi}} \end{aligned}$$

which gives same upper bound provided by the replica-symmetric solution of Sherrington and Kirkpatrick,

$$2P^* \leq 2\sqrt{\frac{2}{\pi}}$$

but as an upper bound and with a more rigorous formulation.