Statistics 291: Lecture 12 (February 29, 2024)
Shattering I

Instructor: Mark Sellke

Scribe: Peter Luo

1 Logistics

HW 3 is posted. It will be due on March 22, but you currently have everything you need to work on the HW, so the recommended due date is March 8. You should choose your project topic/group by March 8 as well.

2 Remark from last time

Last time we looked at the case of high temperature and nonzero external field. The procedure for bounding $F_N(\beta)$ was as follows:

1. Find the global maximum $x_*(q)$ on $\sqrt{q}S_N$.
2. Argue by the second moment method on $\mathrm{Band}(x_*(q)) = \{y \in S_N : \langle y - x_*(q), x_*(q)\rangle = 0\}$.

Recent work on various models in these conditions is linked on the course website:

- Sherrington-Kirkpatrick Model: [Bolthausen 18], [Brennecke-Yau 22]
- Ising Perceptron: [Ding-Sun 18]
- Linear Regression: [Qiu-Sen 22]

3 Shattering

Today's topic is shattering. So far in this course, we have only this phenomenon qualitatively. We now give a more precise definition. We use $NBHD_r(S)$ to denote the r-neighborhood of a set $S \subseteq \mathbb{R}^N$.

Definition 3.1. Given a probability measure μ_β on S_N, a **shattering decomposition** with parameters (c, r, b, s) (in our construction, $s \gg r$) is a family of disjoint connected subsets ("clusters") $\mathscr{C}_1, \ldots, \mathscr{C}_M \subseteq S_N$ such that

(a) Clusters have small probability: $\max_{1\le m\le M} \mu_\beta(\mathscr{C}_m) \le e^{-cN}$.

(b) Clusters have small diameter: $\max_{1\le m\le M} \mathrm{diam}(\mathscr{C}_m) \le r\sqrt{N}$.

(c) Clusters are surrounded by a bottleneck: $\mu(\mathrm{NBHD}_{b\sqrt{n}}(\mathscr{C}_m)) \le (1 + e^{-cN})\mu_\beta(\mathscr{C}_m)$ for all $m \le M$.

(d) Separation: $\min_{\substack{m \neq m' \\ x_m \in \mathscr{C}_m \\ x'_m \in \mathscr{C}_{m'}}} \|x_m - x'_m\| \geq s\sqrt{N}$.

(e) Clusters carry approximately all Gibbs mass: $\mu_\beta\left(\bigcup_{m=1}^M \mathscr{C}_m\right) \geq 1 - e^{-cN}$.

Note that by 1 and 5, the total number of clusters is exponentially large: $M \geq e^{c'N}$.

Definition 3.2. A Gibbs measure $\mu_\beta(H_{N,p})$ **shatters** if it admits, with high probability, a shattering decomposition with (c, r, b, s) not depending on N.

Theorem 3.3 (Alaoui-Montanari-Sellke 23)**.** *If $p \geq p_0$ is large, and $\beta \in [C, \beta_{crit}(p))$, then $\mu_\beta(H_{N,p})$ shatters. The critical inverse temperature $\beta_{crit}(p)$ is defined by $\mathbb{E}F_N(\beta) \overset{N\to\infty}{\to} \frac{\beta^2}{2} \iff \beta \leq \beta_{crit}(p) \approx (1+o(1))\sqrt{\log p}$ and $C = 10^6$.*

We will prove this for $\beta \in \left[C, (1+o(1))\sqrt{\frac{\log p}{2}}\right)$. First, we develop a proposition.

Proposition 3.4. *If $\beta \leq \sqrt{\left(\frac{1}{2} - \varepsilon\right)\log p}$ then $\mathbb{E}[Z_N(\beta)^2] \leq \exp(N\beta^2 + o(N))$ (which implies $F_N(\beta) \to \beta^2/2$).*

Proof. Recall that

$$\frac{1}{N}\log\mathbb{E}[Z_N(\beta)^2] = \max_{-1\leq R\leq 1}\left(\beta^2(1+R^p)\frac{1}{2}\log(1-R^2)\right) \overset{?}{\leq} \Phi(0) = \beta^2.$$

The inequality holds if $\beta^2 R^p + \frac{\log(1-R^2)}{2} \leq 0$ for all R. It turns out that the worst case is $R \approx 1 - \frac{1}{p\log p} \implies \beta^2 \approx \frac{\log p}{2}$. □

We expect shattering to happen when $\beta \in (\beta_d(p), \beta_{\text{crit}}(p))$, where $\beta_d(p) = \sqrt{\frac{(p-1)^{p-1}}{p(p-2)^{p-2}}}$ is the "dynamic threshold" (above which slow mixing/hardness of sampling should appear), and $\beta_{\text{crit}}(p)$ is the "static threshold" above which replica symmetry-breaking (RSB) appears.

Proposition 3.5. *Shattering implies exponentially slow mixing of Langevin dynamics due to bottlenecks.*

This is explored more on HW 3. Physicists have the following natural belief:

Proposition 3.6. *No shattering implies fast mixing from a random start.*

Theorem 3.7 (Gamarnik-Jagannath-Kizildag 23)**.** *For Ising p-spin models, if $\beta \in \left(\sqrt{\log 2}, \sqrt{2\log 2}\right)$, and p is large depending on β then a similar shattering behavior occurs with high probability.*

We now prove Theorem 3.3.

Proof of Theorem 3.3. We will prove this for $\beta \in \left[c, (1+o(1))\sqrt{\frac{\log p}{2}}\right)$. The idea is to use contiguity at exponential scale, which was conveniently on HW 2. This idea originates from [Achlioptas-Coja Oghlan 2008] about the k-SAT problem. We first define two distributions on $(H_{N,p}, \boldsymbol{\sigma})$, as on HW 2.

- The "null" model's is given by $\mathbb{P}_N$: First, $H_{N,p}$ is a pure p-spin model, and then $\boldsymbol{\sigma} \sim \mu_\beta = \frac{e^{\beta H_{N,p}(\boldsymbol{\sigma})}}{Z_N(\beta)}\, d\mu$
- The "planted" model is given by $\mathbb{Q}_N$: First sample $\boldsymbol{\sigma} \sim \text{Unif}(S_N)$, and then let $H_{N,p} = \tilde{H_{N,p}} + \frac{\beta}{N^{(p-1)/2}}\boldsymbol{\sigma}^{\otimes p}$, where $\tilde{H}_{N,p}$ is an independent p-spin Hamiltonian.

Note that

$$\mathrm{Law}_{\mathbb{Q}_N}(H_{N,p}|\boldsymbol{\sigma}) \propto \exp\left(-\frac{1}{2}\left\|\mathbf{G}_N^{(p)} - \frac{\beta}{N^{(p-1)/2}}\boldsymbol{\sigma}^{\otimes p}\right\|_2^2\right) \propto \exp\left(-\frac{\left\|\mathbf{G}_N^{(p)}\right\|_2^2}{2} + \beta H_{N,p}(\boldsymbol{\sigma})\right).$$

From HW, the Radon-Nikodym derivative is

$$\frac{d\mathbb{Q}_N(H_{N,p},\boldsymbol{\sigma})}{d\mathbb{P}_N(H_{N,p},\boldsymbol{\sigma})} = \frac{Z_N(\beta)}{\mathbb{E}Z_N(\beta)}.$$

For $\beta \le \beta_2$, then $\frac{\mathbb{Q}_N}{\mathbb{P}_N} = e^{o(N)}$ with exponentially good $\mathbb{P}_N$ probability. This implies $\mathbb{P}_N \triangleleft \mathbb{Q}_N$. In other words $\mathbb{Q}_N$-exponentially likely events are $\mathbb{P}_N$-exponentially likely. □

We will use the following proposition, which we will prove next time. For now, we state it and assume it is true.

Proposition 3.8. *Suppose $p \ge p_0$ and $\beta \in [C, \beta_2)$. Then for N large,*

$$\frac{d}{dq}\mathbb{E}^{\mathbb{Q}_N} F_\beta(q\boldsymbol{\sigma}) \ge \frac{\beta^2 p}{10} > 0$$

for all $q \in \left[1-\frac{1}{2p}, 1-\frac{100}{\beta p}\right]$, where

$$F_\beta(Band(q\boldsymbol{\sigma})) = \frac{1}{N}\log\int_{NBHD_{1/N}(Band(q\boldsymbol{\sigma}))} e^{\beta H_{N,p}(x)}\,dx.$$

Definition 3.9. The quantity $F_\beta(\mathrm{Band}(q\boldsymbol{\sigma}))$ is known as the **Franz-Parisi potential**.

Graphing this potential, we see that the function strictly decreases for $\beta \le 1-\frac{1}{2p}$, but then increases in the range $\left[1-\frac{1}{2p}, 1-\frac{100}{\beta p}]\right]$.

The intuition here is that $\frac{d}{dq}\mathbb{E}F_\beta(\mathrm{Band}(q\boldsymbol{\sigma}))$ is dominated by $\frac{\beta\boldsymbol{\sigma}^{\otimes p}}{N^{(p-1)/2}}$, which implies

$$\frac{d}{dq}(\beta^2 q^p) = \beta^2 p q^{p-1},$$

which is large for $q = 1 - \frac{O(1)}{p}$.

Returning back to our proof, we let $\Delta = \frac{\sqrt{p}}{100}$. In $\mathbb{P}_N$ take $\boldsymbol{\sigma} \sim \mu_\beta$.

Claim 3.10. *With probability $\ge 1 - e^{-cN}$,*

$$\mu_\beta\left(Ball_{10\Delta\sqrt{N}}(\boldsymbol{\sigma})\right) \le \left(1+e^{-cN}\right)\mu_\beta\left(Ball_{\Delta\sqrt{N}}(\boldsymbol{\sigma})\right) \approx \sqrt{2/p}.$$

This claim is immediate by contiguity at exponential scale, and our proposition. Next, let

$$S_{\mathrm{good}} = \left\{\boldsymbol{x} \in S_N : \mu_\beta\left(\mathrm{Ball}_{10\Delta\sqrt{N}}(\boldsymbol{x})\right) \le \left(1+e^{-cN}\right)\mu_\beta\left(\mathrm{Ball}_{\Delta\sqrt{N}}(\boldsymbol{x})\right)\right\} \le \exp\left(\frac{N\beta^2}{2} - cN\right).$$

Then,

$$\mathbb{E}^{H_{N,p}}\mu_\beta(S_{\mathrm{good}}) \ge 1 - e^{-cN}.$$

Because our second moment computation succeeded, $\Phi(q) < \beta^2$ for all $q \in (0,1]$, while $Z_N(\beta) = e^{N\beta^2/2+o(N)}$ with high probability. Then

$$\mathbb{E}^{\boldsymbol{\sigma}\sim\mu_\beta}[Z_\beta(\mathrm{Band}(q\boldsymbol{\sigma})] = \frac{1}{Z_N(\beta)}\int_{S_N\times S_N} e^{\beta(H_{N,p}(\boldsymbol{\sigma})+H_{N,p}(\boldsymbol{\sigma}')}\cdot \mathbb{1}_{\{R(\boldsymbol{\sigma},\boldsymbol{\sigma}')\approx q}\,d\boldsymbol{\sigma}\,d\boldsymbol{\sigma}'$$

$$
\begin{aligned}
&\le \exp(N\Phi(q) + N\varepsilon) \\
&\le \exp(N\beta^2 - N\varepsilon),
\end{aligned}
$$

where the inequality follows Markov with high probability. But, if $\boldsymbol{x}, \boldsymbol{x}' \in S_{\text{good}}$, then

$$
\left\|\boldsymbol{x} - \boldsymbol{x}'\right\| \not\in \left[2\Delta\sqrt{N}, 8\Delta\sqrt{N}\right].
$$

Therefore, we can view the condition $\left\|\boldsymbol{x} - \boldsymbol{x}'\right\| \le 2\Delta\sqrt{N}$ as an equivalence relation on S_{good}, so it induces a partition $A_1, \dots, A_m$ with $S_{\text{good}} = \bigcup_{i=1}^m A_i$. Take $\mathscr{C}_m = \text{Ball}_{2\Delta\sqrt{N}}(\boldsymbol{x}_m)$ for arbitrary $\boldsymbol{x}_m \in A_m$. Note that $\mathscr{C}_m$ are spherical caps. We now check all of the shattering conditions:

1. Clusters have small probability by definition of S_{good}
2. Clusters have small diameter, also by definition of S_{good}.
3. Clusters are surrounded by bottlenecks by definition
4. We have separation because $\left\|\boldsymbol{x}_m - \boldsymbol{x}_m'\right\| \ge 8\Delta\sqrt{N}$.
5. Since $\bigcup_{i=1}^m \mathscr{C}_m \supseteq S_{\text{good}}$ and S_{good} covers μ_β, then clusters cover μ_β.