Stat450 - 1 | 2024 Fall
Final Exam

Tuorui Peng[1]

Problem 1

1.(a)

Since X_i are i.i.d. r.v. with finite variance, we have

$$\mathbb{E}\left[\bar{X}_g\right] = \mu, \quad \mathrm{Var}\left[\bar{X}_g\right] \leq \frac{\sigma^2}{k}$$

$$\mathbb{P}\left(\bar{X}_g - \mu \geq \frac{2\sigma}{\sqrt{k}}\right) \leq \mathbb{P}\left((\bar{X}_g - \mu)^2 \geq \frac{4\sigma^2}{k}\right) \leq \frac{var(\bar{X}_g)}{4\sigma^2/k} \leq \frac{1}{4}$$

1.(b)

With $\hat{\mu} := \mathrm{median}(\bar{X}_1, \ldots, \bar{X}_G)$, we have

$$\mathbb{P}\left(\hat{\mu} - \mu \geq 2\sigma/\sqrt{k}\right) = \mathbb{P}\left(\text{less than half of } \bar{X}_g \geq \mu + 2\sigma/\sqrt{k}\right)$$

thus we consider the events $A_g := \{\bar{X}_g \geq 2\sigma/\sqrt{k}\}$ with $\mathbf{1}(A_g) \sim \mathrm{Bin}(p)$ for some $p \in [0,1]$. Using the result from (a), we further have $p \leq 1/4$.

Since $\bar{X}_g$ are i.i.d., we have

$$\mathbb{P}\left(\hat{\mu} - \mu \geq 2\sigma/\sqrt{k}\right) \leq \mathbb{P}\left(\mathrm{Bin}(G,p) \geq G/2\right) \leq \mathbb{P}\left(\mathrm{Bin}(G,1/4) \geq G/2\right)$$

1.(c)

Note that $\mathrm{Bin}(G,1/4)$ is sub-gaussian with σ at most $\sqrt{G}/2$, we thus have

$$\mathbb{P}\left(\mathrm{Bin}(G,1/4) \geq G/2\right) = \mathbb{P}\left(\mathrm{Bin}(G,1/4) - G/4 \geq G/4\right) \leq \exp\left(-\frac{G^2}{8G}\right) = \exp\left(-\frac{G}{8}\right) = \delta$$

then we substitute $k = n/G = \dfrac{n}{8\log 1/\delta}$ to result in (b) to obtain

$$\mathbb{P}\left(\hat{\mu} - \mu \geq 4\sigma\sqrt{\frac{2\log(1/\delta)}{n}}\right) \leq \delta$$

[1] TuoruiPeng2028@u.northwestern.edu

1.(d)

Denote $t = 4\sigma\sqrt{\dfrac{2\log(1/\delta)}{n}}$, which yields

$$\mathbb{P}\left(\hat{\mu} - \mu \geq t\right) \leq \delta = e^{-\frac{t^2}{2(4\sigma/\sqrt{n})^2}}$$

and also we should have a similar lower bound of the same form. So $\hat{\mu}$ is indeed sub-gaussian with parameter at most $\sim 4\sigma/\sqrt{n}$.

Problem 2

It suffice to bound the card of packing set $\mathcal{M} := \mathcal{M}(P, \varepsilon, \|\cdot\|_2)$ since we have the relation between packing and covering set:

$$|\mathcal{M}(P, 2\varepsilon, \cdot)| \leq |\mathcal{N}(P, \varepsilon, \cdot)| \overset{(i)}{\leq} |\mathcal{M}(P, \varepsilon, \cdot)|$$

We do so by the following steps:

1. We first prove the following lemma (MJW Exercise 2.11 (b) but just up to constant)

   Lemma 2.1. *Given a sequence of i.i.d. $N(0,1)$ guassian r.v. of size n: $\Xi = \{\xi_i\}_{i=1}^n$, we have*

   $$\begin{aligned}\mathbb{E}\left[\sup_{\xi_i \in \Xi} \xi_i\right] &\geq \sqrt{\log n}\,\mathbb{P}\left(\exists\, \xi_i \in \Xi \ s.t.\ \xi_i \geq \sqrt{\log n}\right) \\ &\geq \sqrt{2\log n}\left(1 - \Psi(\sqrt{\log n})\right)^n \\ &\gtrsim \sqrt{\log n}(1 - 1/n)^n \\ &\gtrsim \sqrt{\log n}\end{aligned}$$

   in which $\Psi(\cdot)$ is the c.d.f. of standard normal distribution. And we used the property of mills ratio

   $$m(\delta) := \frac{1 - \Psi(\delta)}{\phi(\delta)} \sim \frac{1}{\delta} \Rightarrow 1 - \Psi(\sqrt{\log n}) \gtrsim \frac{1}{\sqrt{n \log n}} \geq \frac{1}{n}$$

   (From MJW Exercise 2.2 or refer to Wikipedia)

   And if ξ_i are $N(0, \sigma^2)$, the bound would be $\sigma\sqrt{\log n}$ (trivially by scaling).

2. For $\{p\}_{p \in \mathcal{M}} \subset \mathbb{R}^n$ we construct the following two gaussian processes:

   $$\begin{aligned}X_p &:= \langle p, w \rangle\,, \quad w \sim N(0, I) \\ Y_p &\overset{i.i.d.}{\sim} N(0, \varepsilon^2/2)\end{aligned}$$

   Using the construction we have $\forall p, q \in \mathcal{M}$:

   $$\mathbb{E}\left[(X_p - X_q)^2\right] = \|p - q\|_2^2 \geq \varepsilon^2 = \mathbb{E}\left[(Y_p - Y_q)^2\right]$$

   since $\mathcal{M}$ is a packing set.

Then we note the following inequality:

$$\mathbb{E}\left[\sup_{p\in\mathcal{M}} X_p\right] \overset{(ii)}{\geq} \mathbb{E}\left[\sup_{p\in\mathcal{M}} Y_p\right] \overset{(iii)}{\gtrsim} \varepsilon\sqrt{\log|\mathcal{M}|}$$

in which (ii) is due to Sudakov-Fernique inequality, and (iii) is due to Lemma 2.1.

3. Now we upper bound $\mathbb{E}\left[\sup_{p\in\mathcal{M}} X_p\right]$ as follows: Since P is a polytope and $\mathcal{M}\subset P$, we have

$$\mathbb{E}\left[\sup_{p\in\mathcal{M}} X_p\right] \leq \mathbb{E}\left[\sup_{p\in P} X_p\right] = \mathbb{E}\left[\sup_{p\in\text{vertices of } P} \langle p, w\rangle\right] \overset{(iv)}{\lesssim} \sqrt{\log N}$$

in which (iv) used the maximal inequality by noticing that $\langle p, w\rangle$ are sub-gaussian with parameter 1.

Putting the above steps together, we have

$$\varepsilon\sqrt{\log|\mathcal{N}|} \overset{(i)}{\leq} \varepsilon\sqrt{\log|\mathcal{M}|} \overset{(iii)}{\lesssim} \mathbb{E}\left[\sup_{p\in\mathcal{M}} Y_p\right] \overset{(ii)}{\leq} \mathbb{E}\left[\sup_{p\in\mathcal{M}} X_p\right] \overset{(iv)}{\lesssim} \sqrt{\log N}$$

i.e. for some absolute constant C, we have

$$|\mathcal{N}(P,\varepsilon,\|\cdot\|_2)| \leq N^{C/\varepsilon^2}$$

Problem 3

3.(a)

Note that we have

$$\left\|y - X\hat{\theta}\right\|_2^2 \leq \|y - X\theta^*\|_2^2$$

which gives

$$\left\|X(\hat{\theta}-\theta^*)\right\|_2^2 \leq 2w'\left(X(\hat{\theta}-\theta^*)\right) \Rightarrow \left\|X(\hat{\theta}-\theta^*)\right\|_2 \leq 2w'\frac{X(\hat{\theta}-\theta^*)}{\left\|X(\hat{\theta}-\theta^*)\right\|_2}$$

3.(b)

Note that $\hat{\theta}-\theta^*$ should be $2s$-sparse, and $X(\hat{\theta}-\theta^*)\in\mathrm{col}(X_{\mathrm{span}(\hat{\theta}\cup\theta^*)})$, we thus have

$$w'\frac{X(\hat{\theta}-\theta^*)}{\left\|X(\hat{\theta}-\theta^*)\right\|_2} \leq \sup_{|S|\leq 2s}\ \sup_{v_S\in\mathbb{S}^{n-1}, v_S\in\mathrm{col}(X_S)} w'v_S$$

3.(c)

We have for $w \mapsto \sup_{v_S\in\mathbb{S}^{n-1}, v_S\in\mathrm{col}(X_S)} w'v_S := g(w)$:

$$g(w) = g(w_S + w_{S^\complement}) = g(w_S) = \|w_S\|_2$$

in which w_S is the projection of w onto $\text{col}(X_S)$, and $w_{S^\complement}$ is its orthogonal complement. Further since $w \sim N(0, \sigma^2 I)$, we can consider

$$w = P\tilde{w} = P_S\tilde{w}_S + P_{S^\complement}\tilde{w}_{S^\complement}, \quad P \in \text{SO}(n)$$

i.e. decompose w into w_S and $w_{S^\complement}$, with each following gaussian distribution.

Then for $w \sim N(0, \sigma^2 I)$ we have

$$\mathbb{E}\left[g(w)\right] = \mathbb{E}\left[\|w_S\|_2\right] = \mathbb{E}\left[\|P_S\tilde{w}_S\|_2\right] = \sigma\mathbb{E}\left[\sqrt{\chi^2_{2s}}\right] \leq \sigma\sqrt{\mathbb{E}\chi^2_{2s}} = \sigma\sqrt{2s}$$

Then we note that for any $w, y \in \mathbb{R}^d$:

$$|g(w) - g(y)| = \|w_S - y_S\|_2 \leq \|w - y\|_2$$

which means that $g(w)$ is 1-Lipschitz. Then we can apply the concentration inequality for Lipschitz function to obtain that

$$\mathbb{P}\left(|g(w) - \mathbb{E}g(w)| \geq t\right) \leq 2\exp\left(-\frac{t^2}{2\sigma^2}\right) \Rightarrow \mathbb{P}\left(\sup_{v_S\in\mathbb{S}^{n-1}, v_S\in\text{col}(X_S)} w'v_S \geq \sigma\sqrt{2s} + t\right) \leq 2\exp\left(-\frac{t^2}{2\sigma^2}\right)$$

letting $\delta = 2\exp\left(-\dfrac{t^2}{2\sigma^2}\right)$ we have:

$$\sup_{v_S\in\mathbb{S}^{n-1}, v_S\in\text{col}(X_S)} w'v_S \leq \sigma\big(\sqrt{2s} + \sqrt{2\log(2/\delta)}\big), \quad \text{w.p. } \geq 1 - \delta.$$

3.(d)

Using union bound on $\{S \subset [d] : |S| \leq 2s\}$ we have:

$$\sup_{|S|\leq 2s}\ \sup_{v_S\in\mathbb{S}^{n-1}, v_S\in\text{col}(X_S)} w'v_S \leq \sigma\big(\sqrt{2s} + \sqrt{2\log(2/\delta)}\big)$$

with probability at least

$$1 - \delta\sum_{j=1}^{2s}\binom{d}{j} \geq 1 - \delta\big(\frac{ed}{2s}\big)^{2s}$$

then we obtain the bound

$$\left\|X(\hat{\theta} - \theta^*)\right\|_2 \leq 2w'\frac{X(\hat{\theta} - \theta^*)}{\left\|X(\hat{\theta} - \theta^*)\right\|_2} \leq 2\sigma\big(\sqrt{2s} + \sqrt{2\log(2/\delta)}\big), \quad \text{w.p. at least } 1 - \delta\big(\frac{ed}{2s}\big)^{2s}$$

Remark: We may also make re-labelling $\delta(ed/2s)^2s \mapsto \delta$ to obtain a probability bound of $1 - \delta$:

$$\begin{aligned}\left\|X(\hat{\theta} - \theta^*)\right\|_2 \leq& 2\sigma\big(\sqrt{2s} + \sqrt{2\log(2/\delta) + 2s\log(ed/2s)}\big) \\ \lesssim& \sigma\left(\sqrt{s} + \sqrt{\log\frac{1}{\delta}} + \sqrt{s\log\frac{d}{s}}\right), \quad \text{w.p. at least } 1 - \delta\end{aligned}$$