Math 281C Homework 6 Solutions

1. Let $R_n : \Theta \to \mathbb{R}$ be a sequence of random functions and $R(\theta) = \mathbb{E}R_n(\theta)$. Let $d : \Theta \times \Theta$ be some distance on Θ. Denote $\theta_0 = \mathrm{argmin}_\theta R(\theta)$, and for $0 < \delta < \infty$, define $\Theta_\delta = \{\theta : d(\theta, \theta_0) \le \delta\}$. For $\alpha \in (0, 2), \sigma < \infty$, and $D > 0$, assume we have the continuity bound

$$\mathbb{E}\left[\sup_{\theta \in \Theta_\delta} |(R_n(\theta) - R(\theta)) - (R_n(\theta_0) - R(\theta_0))|\right] \le \frac{\sigma\delta^\alpha}{\sqrt{n}}$$

for all $\delta \le D$. Assume in addition that for some parameters $\beta \in [1, \infty]$ and $v > 0$, we have

$$R(\theta) \ge R(\theta_0) + vd(\theta, \theta_0)^\beta$$

for $d(\theta, \theta_0) \le D$. Let $\widehat{\theta}_n = \mathrm{argmin}_\theta R_n(\theta)$ and assume that $\widehat{\theta}_n$ is consistent for θ_0. Give the largest rate r_n you can for which

$$r_n d(\widehat{\theta}_n, \theta_0) = O_P(1) \quad \text{as} \quad n \to \infty.$$

Solution: If $\beta \le \alpha$, we can't find the largest convergence rate, and if $\beta > \alpha$, $r_n = n^{1/(2(\beta-\alpha))}$. The proof can be done following the peeling argument in Lecture 11.

2. Suppose that we have $X_i \in \mathbb{R}^d$ and observe

$$Y_i = \langle X_i, \theta_0 \rangle + \epsilon_i, \quad \text{where} \quad \epsilon_i = B_i Z_i$$

for $i = 1, \dots, n$. Here $B_i \in \{0, 1\}$ is independent of Z_i and X_i, and $\mathbb{P}(B_i = 0) = p > 1/2$. The variable Z_i has arbitrary distribution, independent of X_i and $\mathbb{E}[|Z_i|] < \infty$. We decide to estimate θ_0 using the absolute loss $\ell(\theta; x, y) = |y - \langle x, \theta \rangle|$. In other words,

$$\widehat{\theta}_n = \underset{\theta \in \mathbb{R}^d}{\mathrm{argmin}} \frac{1}{n} \sum_{i=1}^n \ell(\theta; X_i, Y_i).$$

Let $R_n(\theta)$ and $R(\theta)$ denote the empirical risk and population risk respectively.

(a) Show that for any $\theta \in \mathbb{R}^d$, we have

$$R(\theta) - R(\theta_0) \ge (2p - 1)\mathbb{E}[|\langle X, \theta - \theta_0 \rangle|].$$

Solution: By the tower property, we have

$$R(\theta) = p\mathbb{E}|\langle X, \theta_0 - \theta \rangle| + (1 - p)\mathbb{E}|\langle X, \theta_0 - \theta \rangle + Z|,$$
$$R(\theta_0) = (1 - p)\mathbb{E}|Z|.$$

Thus, it's equivalent to show

$$\mathbb{E}|\langle X, \theta_0 - \theta \rangle + Z| + \mathbb{E}|\langle X, \theta_0 - \theta \rangle| \ge \mathbb{E}|Z|,$$

which is a consequence of triangle inequality.

Now for $v \in \mathbb{S}^{d-1}$, denote $\sigma_v^2 = \mathbb{E}[(\langle v, X \rangle)^2]$, and assume there exist two constants c_1 and c_2 such that

$$\mathbb{P}(|\langle v, X \rangle| \ge c_1\sigma_v) \ge c_2 > 0.$$

Moreover, assume there is a constant $D < \infty$ such that $\|X\|_2 \le D$ with probability 1, and $\mathbb{E}[XX^\top] = \Sigma \succ 0$.

(b) Show that for any $v \in \mathbb{R}^d$,

$$\mathbb{E}[|\langle v, X\rangle|] \geq \rho \|v\|_2,$$

where ρ is a constant that depends on the distribution of X but is independent of v.

Solution: For any $v \in \mathbb{R}^d$, denote $w = v/\|v\|_2$,

$$\begin{aligned}
\mathbb{E}[|\langle w, X\rangle|] &= \int_0^\infty \mathbb{P}(|\langle w, X\rangle| \geq t)\mathrm{d}t \\
&\geq \int_0^{\sigma_w c_1} \mathbb{P}(|\langle w, X\rangle| \geq t)\mathrm{d}t \\
&\geq \int_0^{\sigma_w c_1} \mathbb{P}(|\langle w, X\rangle| \geq \sigma_w c_1)\mathrm{d}t \\
&\geq c_1 c_2 \sigma_w \geq c_1 c_2 \sqrt{\lambda},
\end{aligned}$$

where $\lambda > 0$ is the minimal eigenvalue of Σ.

(c) Show that there exists a constant $\sigma < \infty$ which may depend on (D, d) such that for any $\delta > 0$,

$$\mathbb{E}\left[\sup_{\theta:\|\theta-\theta_0\|_2 \leq \delta} |(R_n(\theta) - R(\theta)) - (R_n(\theta_0) - R(\theta_0))|\right] \leq \frac{\sigma\delta}{\sqrt{n}}.$$

Solution: Denote

$$\begin{aligned}
\Delta(\theta) &= (\ell(\theta; x, y) - R(\theta)) - (\ell(\theta_0; x, y) - R(\theta_0)); \\
\Delta_n(\theta) &= (R_n(\theta) - R(\theta)) - (R_n(\theta_0) - R(\theta_0)).
\end{aligned}$$

First, it can be found that for any θ and θ',

$$|\ell(\theta; x, y) - \ell(\theta'; x, y)| \leq \|x\|_2 \|\theta - \theta'\|_2,$$

so $\ell(\cdot)$ is $\|x\|_2$-Lipschitz continuous in θ. For any $\lambda \in \mathbb{R}$, we have

$$\mathbb{E}[\exp(\lambda(\Delta(\theta) - \Delta(\theta'))] \leq \exp\left(\frac{\lambda^2 D^2 \|\theta - \theta'\|_2^2}{2}\right),$$

where the sub-Gaussian-type bound comes from the boundedness of $\|x\|_2$, and it further implies

$$\mathbb{E}\exp\left(\lambda\frac{\sqrt{n}}{D}(\Delta_n(\theta) - \Delta_n(\theta'))\right) \leq \exp\left(\frac{\lambda^2}{2}\|\theta - \theta'\|_2^2\right),$$

which means $\sqrt{n}\Delta_n(\theta)/D$ is a sub-Gaussian process. Denote $\Theta_\delta = \{\theta : \|\theta - \theta_0\|_2 \leq \delta\}$ as the local ball, and let $N(\Theta_\delta, \|\cdot\|_2, \epsilon)$ be the minimal covering number of Θ_δ, we have

$$\log N(\Theta_\delta, \|\cdot\|_2, \epsilon) \leq d \log(1 + 2\delta/\epsilon).$$

Combining these results with Dudley's entropy bound (materials before Section 1 of Lecture 8), there is a constant c such that

$$\begin{aligned}
\mathbb{E}\left[\sup_{\theta\in\Theta_\delta} |\Delta_n(\theta)|\right] &\leq c\frac{D}{\sqrt{n}} \int_0^\delta \sqrt{\log N(\Theta_\delta, \|\cdot\|_2, \epsilon)}\mathrm{d}\epsilon \\
&\leq c\frac{D\sqrt{d}}{\sqrt{n}} \int_0^\delta \sqrt{\log(1 + 2\delta/\epsilon)}\mathrm{d}\epsilon \leq \mathrm{c}\frac{\mathrm{D}\sqrt{\mathrm{d}}}{\sqrt{\mathrm{n}}}\delta.
\end{aligned}$$

(d) At what rate does $\widehat{\theta}_n$ converge to θ_0? You may assume that $\widehat{\theta}_n$ is consistent for θ_0.

Solution: Combining parts (a)–(c), conditions in Question 1 are satisfied with $\alpha = \beta = 1$. For any $0 < \epsilon < 1$, since

$$\frac{\sigma\delta}{\sqrt{n}} \leq \frac{\sigma\delta^\epsilon}{\sqrt{n}},$$

we can prove convergence with rate $r_n = n^{1/(2(1-\epsilon))}$ using peeling argument.