STAT 620: Asymptotic Statistics **Spring 2022**

Lecture: Mar 31

Lecturer: Xianyang Zhang

1 Example: Lipschitz functions

Let $\Theta \subseteq \mathbb{R}^d$ and $\mathcal{F} = \{l(\theta, \cdot)\,;\theta \in \Theta\}$ be a collection of $L(x)$-Lipschitz functions, i.e.,

$$|l(\theta_1, x) - l(\theta_2, x)| \ \le\ L(x)|\theta_1 - \theta_2\|\,,$$

$\forall\, \theta_1, \theta_2 \in \Theta$. Assume that $\mathrm{diam}(\Theta) < \infty$. Then we can find a ball with radius $\mathrm{diam}(\Theta)$ that contains the set Θ. Thus we have

$$\log N(\Theta, \|\cdot\|, \epsilon) \ \le\ d\,\log\left(1 + 2\frac{\mathrm{diam}(\Theta)}{\epsilon}\right)\,.$$

Further, suppose that $\mathcal{F} = -\mathcal{F}$. The goal here is to find a bound for $E\left[\|P_n - P\|_{\mathcal{F}}\right]$.

1.1 Proof

Define the process $Z_f := \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varepsilon_i\, f(X_i)$, where ϵ_i is a sequence of i.i.d Rademacher random variables. By Hoeffding's Lemma, we have

$$\begin{aligned}
E\left[\exp\left(\lambda(Z_f - Z_g)\right)|X_1,\ldots,X_n\right] &= \prod_{i=1}^{n} E\left[\exp\left(\frac{\lambda}{\sqrt{n}}\,\varepsilon_i\left(f(X_i) - g(X_i)\right)\right)\middle| X_1,\ldots,X_n\right] \\
&\le \prod_{i=1}^{n}\exp\left(\frac{\lambda^2}{2n}(f(X_i) - g(X_i))^2\right) \\
&= \exp\left(\frac{\lambda^2}{2}\,\|f - g\|^2_{L_2(P_n)}\right)\,,
\end{aligned}$$

where $\|f - g\|^2_{L_2(P_n)} = n^{-1}\sum_{i=1}^{n}(f(X_i) - g(X_i))^2$.

Define $Z_f = n^{-1/2}\sum_{i=1}^{n}\epsilon_i f(x_i)$ for $f \in \mathcal{F}$, where ϵ_i is a sequence of i.i.d Rademacher random variables. Define $-\mathcal{F} = \{-f : f \in \mathcal{F}\}$ and let $\mathcal{F}^* = \mathcal{F} \cup -\mathcal{F}$. Note that $\sup_{f\in\mathcal{F}}|Z_f| = \sup_{f\in\mathcal{F}^*} Z_f$. By the assumption that $\mathcal{F} = -\mathcal{F}$, and the symmetrization argument, we have

$$\begin{aligned}
E[\|P_n - P\|_{\mathcal{F}}] \le& \frac{2}{\sqrt{n}}E[E[\sup_{f\in\mathcal{F}}|Z_f|\big|X_1,\ldots,X_n]] \\
=& \frac{2}{\sqrt{n}}E[E[\sup_{f\in\mathcal{F}^*} Z_f\big|X_1,\ldots,X_n]] \\
=& \frac{2}{\sqrt{n}}E[E[\sup_{f\in\mathcal{F}} Z_f\big|X_1,\ldots,X_n]],
\end{aligned}$$

where we have used the assumption that $\mathcal{F} = -\mathcal{F}$ to get the last equality. Applying a result in the previous lecture, we have

$$E\left[\|P_n - P\|_{\mathcal{F}}\right] \le \frac{2}{\sqrt{n}}\, E\Big[E\big[\sup_{f\in\mathcal{F}} Z_f \,\big|X_1,\ldots,X_n\big]\Big] \le \frac{8\sqrt{2}}{\sqrt{n}}\, E\left[\int_0^{\mathrm{diam}(\mathcal{F})/2} \sqrt{\log N(\mathcal{F}, L_2(P_n), \epsilon)}\; d\epsilon\right].$$

Notice that

$$\begin{aligned}\|l(\theta_1,\cdot) - l(\theta_2,\cdot)\|_{L_2(P_n)} &= \Big(\frac{1}{n}\sum_{i=1}^n \big(l(\theta_1, X_i) - l(\theta_2, X_i)\big)^2\Big)^{1/2} \le \Big(\frac{1}{n}\sum_{i=1}^n \|\theta_1-\theta_2\|^2\, L(X_i)^2\Big)^{1/2} \\ &= \|L\|_{L_2(P_n)}\, \|\theta_1 - \theta_2\|\,.\end{aligned}$$

Let $\theta_1,\ldots,\theta_M$ form an $\epsilon/\|L\|_{L_2(P_n)}$ cover for Θ. Then $\{l(\theta_i,\cdot) \,:\, i=1,\ldots,M\}$ form an ϵ cover for $\mathcal{F}$. Hence

$$\log N(\mathcal{F}, L_2(P_n), \epsilon) \le \log N\left(\Theta, \|\cdot\|, \frac{\epsilon}{\|L\|_{L_2(P_n)}}\right).$$

Therefore

$$\begin{aligned}E\left[\|P_n - P\|_{\mathcal{F}}\right] &\le \frac{8\sqrt{2}}{\sqrt{n}}\, E\left[\int_0^{\frac{\mathrm{diam}(\Theta)}{2}\,\|L\|_{L_2(P_n)}} \sqrt{\log N(\mathcal{F}, L_2(P_n), \epsilon)}\; d\epsilon\right] \\ &\le \frac{8\sqrt{2}}{\sqrt{n}}\, E\left[\int_0^{\frac{\mathrm{diam}(\Theta)}{2}\,\|L\|_{L_2(P_n)}} \sqrt{\log N\left(\Theta, \|\cdot\|, \frac{\epsilon}{\|L\|_{L_2(P_n)}}\right)}\; d\epsilon\right].\end{aligned}$$

Set $u = \frac{2\epsilon}{\mathrm{diam}(\Theta)\,\|L\|_{L_2(P_n)}}$ for $0 \le u \le 1$. Then we get

$$E\left[\|P_n - P\|_{\mathcal{F}}\right] \le \frac{8\sqrt{2}}{\sqrt{n}}\, E\left[\int_0^1 \sqrt{\log N\left(\Theta, \|\cdot\|, \frac{\mathrm{diam}(\Theta)\, u}{2}\right)}\, \frac{\mathrm{diam}(\Theta)\,\|L\|_{L_2(P_n)}}{2}\, du\right].$$

Notice that

$$\log N\left(\Theta, \|\cdot\|, \frac{\mathrm{diam}(\Theta)\, u}{2}\right) \le d\, \log\left(1 + \frac{2\mathrm{diam}(\Theta)}{\frac{\mathrm{diam}(\Theta)\, u}{2}}\right) = d\, \log\left(1 + \frac{4}{u}\right).$$

Therefore we have

$$\begin{aligned}E\left[\|P_n - P\|_{\mathcal{F}}\right] &\le \frac{4\sqrt{2}}{\sqrt{n}}\, \mathrm{diam}(\Theta)\, E\left[\|L\|_{L_2(P_n)} \int_0^1 \sqrt{d\, \log\left(1+\frac{4}{u}\right)}\, du\right] \\ &\le \frac{4\sqrt{2}}{\sqrt{n}}\, \mathrm{diam}(\Theta) \int_0^1 \sqrt{\frac{4d}{u}}\, du\; E\left[\|L\|_{L_2(P_n)}\right].\end{aligned}$$

As

$$E\left[\|L\|_{L_2(P_n)}\right] \le \left(E\left[\|L\|^2_{L_2(P_n)}\right]\right)^{1/2} = \left(E\left[L(X_1)^2\right]\right)^{1/2},$$

we have

$$E\left[\|P_n - P\|_{\mathcal{F}}\right] \le \frac{c}{\sqrt{n}}$$

for some positive constant c.

2 VC dimension

VC dimension is a measure for the complexity of a collection of sets. Consider a space $\mathcal{X}$. Denote $X_1^n := \{X_1, \ldots, X_n\}$ with $X_i \in \mathcal{X}$. Let $\mathcal{C}$ be a collection of subsets of $\mathcal{X}$. We say $\mathcal{C}$ picks out a certain subset of X_1^n if the subset is of the form $C \cap X_1^n$ for $C \in \mathcal{C}$.

Let $\Delta(\mathcal{C}, X_1^n)$ be the cardinality of $\{C \cap X_1^n \,:\, C \in \mathcal{C}\}$. In other words, $\Delta(\mathcal{C}, X_1^n)$ is the number of subsets of X_1^n that can be picked out by $\mathcal{C}$. When

$$\Delta(\mathcal{C}, X_1^n) = 2^n,$$

we say that X_1^n is shattered by $\mathcal{C}$. The VC dimension of $\mathcal{C}$, denoted by $VC(\mathcal{C})$, is the largest n such that $\exists$ $X_1^n \subseteq \mathcal{X}$ with $\Delta(\mathcal{C}, X_1^n) = 2^n$. In other words,

$$VC(\mathcal{C}) = \sup_n \left\{ n \in \mathbb{N} \,:\, \sup_{X_1^n \subseteq \mathcal{X}} \Delta(\mathcal{C}, X_1^n) \;=\; 2^n \right\}.$$

If there is no set of points $X_1, \ldots, X_{n+1} \in \mathcal{X}$ that $\mathcal{C}$ can shatter, then $VC(\mathcal{C}) < n+1$.

2.1 Examples

1. For $\mathcal{C} = \{(-\infty, x] \,:\, x \in \mathbb{R}\}$, $VC(\mathcal{C}) = 1$.
2. For $\mathcal{C} = \{(-\infty, x_1] \times (-\infty, x_2] \,:\, x_1, x_2 \in \mathbb{R}\}$, $VC(\mathcal{C}) = 2$.