IEMS 402 Statistical Learning - 2025 Winter
HW7

Tuorui Peng[1]

Exercise 1 Dvoretzky-Kiefer-Wolfowitz inequality via Uniform Bounds

1.(a)

For each given P, consider the $\{\varepsilon, 2\varepsilon, \ldots, \lfloor 1/\varepsilon \rfloor \varepsilon\}$ quantiles of P, denoted by $q_1, q_2, \ldots, q_m$, where $m = \lfloor 1/\varepsilon \rfloor$. With this notation, we have

$$\{\mathbf{1}\{x \leq q_i\}\}_{i=1}^m$$

being a ε covering of $\mathcal{F} = \{\mathbf{1}\{x \leq t\}\}_{t\in\mathbb{R}}$. Thus we have

$$\sup_P \log N(\mathcal{F}, L_2(P), \varepsilon) \leq \log m \lesssim \log(1 + \frac{1}{\varepsilon})$$

1.(b)

When upgrading to n points, the above covering number bound becomes

$$\log N(\mathcal{F}^n, L_2(P), \varepsilon) \lesssim n \log(1 + \frac{1}{\varepsilon})$$

With the covering number bound, we apply the following:

$$\begin{aligned}
R_n(\mathcal{F}) =& \mathbb{E}\left[\sup_{t\in\mathbb{R}} \left| \frac{1}{n}\sum_{i=1}^n \mathbf{1}\{X_i \leq t\} - P(t) \right|\right] \\
\leq& \frac{2}{n}\mathbb{E}\left[\sup_{F\in\mathcal{F}} \left|\left\langle \varepsilon, F_{X_1^n}(t) \right\rangle\right|\right] \\
\overset{\text{Dudley}}{\lesssim}& \frac{1}{n}\int_0^1 \sqrt{\log N(\mathcal{F}^n, L_2(P), u)}\,\mathrm{d}u \\
\lesssim& \int_0^1 \sqrt{n \log(1 + \frac{1}{u})}\,\mathrm{d}u \\
\lesssim& \frac{1}{\sqrt{n}}
\end{aligned}$$

1.(c)

Note that $P_n(\,\cdot\,)$ is sub-Gaussian, we can apply the following:

$$\left|\sup_{t\in\mathbb{R}} |P_n(X \leq t) - P(X \leq t)| - \mathbb{E}\left[\sup_{t\in\mathbb{R}} |P_n(X \leq t) - P(X \leq t)|\right]\right| \leq \varepsilon \quad \text{w.p. at least } 1 - 2\exp\left(-\frac{2n\varepsilon^2}{C^2}\right)$$

i.e. with the above bound on expectation, we have

$$\sup_{t\in\mathbb{R}} |P_n(X \leq t) - P(X \leq t)| \geq \frac{C}{\sqrt{n}} + \varepsilon \quad \text{w.p. at most } 2\exp\left(-cn\varepsilon^2\right)$$

[1] TuoruiPeng2028@u.northwestern.edu

Exercise 2

- Assume we have a set that form a 2ε packing $\mathcal{S}(2\varepsilon)$ and we take any two points $m_1, m_2 \in \mathcal{S}(2\varepsilon)$ and any $h \in \mathbb{B}(h_i, \varepsilon)$, then

$$\|h - h_i\| \geq \|h_i - h_j\| - \|h - h_i\| > 2\varepsilon - \varepsilon = \varepsilon \quad \forall h \in \mathbb{B}(h_i, \varepsilon)$$

  which means that $\mathcal{S}(2\varepsilon)\backslash\{h_i\}$ is not a ε-covering. By ranging over all such $\mathcal{S}$ we would eventually have some set satisfying the maximal packing, however in any of the setting we see that $|\mathcal{S}| \leq |\mathcal{N}(\varepsilon)|$. And we can conclude that

$$M(2\varepsilon) \leq N(\varepsilon)$$

- Consider the maximal packing $\mathcal{M}\varepsilon$. For any $h \in \mathcal{H}$ we have

$$\exists h_i \in \mathcal{M},\, st.\, \|h - h_i\| \leq \varepsilon$$

  (otherwise we have $\mathcal{M} \cup \{h\}$ as a larger packing). Thus we directly see that this set forms a ε-covering, and we have

$$N(\varepsilon) \leq M(\varepsilon)$$

To summarize, we have

$$M(2\varepsilon) \leq N(\varepsilon) \leq M(\varepsilon)$$

Exercise 3

- For the minimal covering set $\mathcal{N}(\Theta)$ we have:

$$\mathrm{vol}(\Theta) \leq |\mathcal{N}(\Theta)| \cdot \mathrm{vol}(B(\varepsilon)) \Rightarrow |\mathcal{N}(\Theta)| \geq \frac{\mathrm{vol}(\Theta)}{\mathrm{vol}(B(\varepsilon))}$$

- For the maximal packing set $\mathcal{M}(\Theta)$ we have:

$$\mathrm{vol}(\Theta + B(\varepsilon/2)) \geq |\mathcal{M}(\Theta)| \cdot \mathrm{vol}(B(\varepsilon)) \Rightarrow |\mathcal{M}(\Theta)| \leq \frac{\mathrm{vol}(\Theta + B(\varepsilon/2))}{\mathrm{vol}(B(\varepsilon))}$$

So we have

$$\frac{\mathrm{vol}(\Theta)}{\mathrm{vol}(B(\varepsilon))} \leq |\mathcal{N}(\Theta)| \leq |\mathcal{M}(\Theta)| \leq \frac{\mathrm{vol}(\Theta + B(\varepsilon/2))}{\mathrm{vol}(B(\varepsilon))}$$

Exercise 4 Covering Number of Sobolev Ellipsoid

4.(a)

In this setting, we have for any pair $\theta^1 \in \tilde{\mathcal{E}}$ and any $\theta \in \mathcal{E}$:

$$\begin{aligned}\left\|\theta-\theta^1\right\|_2^2 =& \sum_{i=1}^{t}(\theta_i-\theta_i^1)^2+\sum_{i=t+1}^{d}\theta_i^2\\ \leq&\delta^2+\sum_{i=t+1}^{d}\theta_i^2 \leq \qquad\qquad \delta^2+\delta^2\sum_{i=t+1}^{d}\frac{1}{\mu_j}\\ \leq& 2\delta^2\end{aligned}$$

i.e. by choosing $t:\mu_t\leq\delta^2$, we have the δ-covering of $\tilde{\mathcal{E}}$ satisfying $\sqrt{2}\delta$-covering of $\mathcal{E}$.

4.(b)

As stated above, it suffices to bound the truncated ellipsoid $\tilde{\mathcal{E}}$, which has "finite" dimension t and we know that the covering number of $\tilde{\mathcal{E}}$ is bounded by

$$N(\tilde{\mathcal{E}},\delta)\lesssim\left(\frac{1}{\delta}\right)^t$$

in which t satisfies $\mu_t=t^{-2\alpha}\asymp\delta^2\Rightarrow t\asymp\delta^{-1/\alpha}$.

Adn we have

$$N(\tilde{\mathcal{E}},\delta)\leq N(\tilde{\mathcal{E}},\delta/\sqrt{2})\lesssim\left(\frac{\sqrt{2}}{\delta}\right)^{\delta^{-1/\alpha}}$$

i.e.

$$\log N(\tilde{\mathcal{E}},\delta)\lesssim\left(\frac{1}{\delta}\right)^{1/\alpha}\log\frac{1}{\delta}$$