STA3000F: Homework 2

Due: November 8, 2025, 11:59pm on Quercus

Q1: convex loss function

Suppose that $\theta \mapsto f(\theta; x)$ is a convex function in $\theta \in \mathbb{R}^d$, for any x. Let $X_1, X_2, \cdots X_n \overset{\text{i.i.d.}}{\sim} \mathbb{P}$ and let $F(\theta) := \mathbb{E}[f(\theta, X)]$, where the function F is uniquely minimized at θ^* (over $\mathbb{R}^d$). Assume furthermore that there exists function M, such that $\mathbb{E}[M(X)] < +\infty$, and for any θ_1, θ_2, x, we have

$$|f(\theta_1; x) - f(\theta_2; x)| \leq M(x) \left\|\theta_1 - \theta_2\right\|_2 .$$

Define the M-estimator

$$\widehat{\theta}_n := \arg\min_{\theta \in \mathbb{R}^d} \frac{1}{n} \sum_{i=1}^{n} f(\theta; X_i).$$

Prove that $\widehat{\theta}_n \xrightarrow{p} \theta^*$.

Q2: minimax testing for covariance matrix

Suppose that we have n i.i.d. samples $X_1, \cdots, X_n \sim \mathcal{N}(0, \Sigma)$, where Σ is an unknown covariance matrix in $\mathbb{R}^{d\times d}$. Consider the testing problem

$$H_0 : \Sigma = I_d \quad \text{vs} \quad H_1 : \Sigma = I + \alpha vv^\top,$$

where $v \in \mathbb{S}^{d-1}$ is an unknown unit vector and $\alpha > 0$ is a known constant. Find the smallest value of α such that there exists a test with sum of two types of error probabilities at most $1/4$. Your answer should depend on n and d. You do not need to provide an explicit constant factor.

Hint: you may find the following formula useful: for $X \sim \mathcal{N}(0, I_d)$ and a matrix $A \in \mathbb{R}^{d\times d}$ with $\|A\|_{\mathrm{op}} < 1$, we have

$$\mathbb{E}\Big[\exp\big(\frac{1}{2}X^\top AX\big)\Big] = \frac{1}{\sqrt{\det(I - A)}}.$$

Q3: contraction lemma of Rademacher complexity

Let $\mathcal{F}$ be a class of functions mapping from $\mathbb{X}$ to $\mathbb{R}$, and let $\phi : \mathbb{R} \mapsto \mathbb{R}$ be an L-Lipschitz function, i.e., for any $a, b \in \mathbb{R}$, we have $|\phi(a) - \phi(b)| \leq L|a - b|$. Let $X_1, \cdots, X_n$ be i.i.d. samples from distribution $\mathbb{P}$ on $\mathbb{X}$. Prove that

$$\mathbb{E}\Big[\sup_{f\in\mathcal{F}} \frac{1}{n}\sum_{i=1}^{n} \varepsilon_i \phi(f(X_i))\Big] \leq 2L\mathbb{E}\Big[\sup_{f\in\mathcal{F}} \frac{1}{n}\sum_{i=1}^{n} \varepsilon_i f(X_i)\Big],$$

where $\{\varepsilon_i\}_{i=1}^n$ are i.i.d. Rademacher variables independent of $\{X_i\}_{i=1}^n$.

Q4: Rademacher complexity bounds

Given $x_1, x_2, \cdots, x_n \in [-1,1]^d$ for some $d \geq 1$, bound the empirical Rademacher complexity

$$\widehat{\mathcal{R}}_n(\mathcal{F}) := \mathbb{E}_\varepsilon\Big[\sup_{f\in\mathcal{F}} \big|\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i f(x_i)\big|\Big],$$

for the following function classes:

- Two-layer neural networks

$$\mathcal{F}_{\mathrm{NN}}(B_1, B_2) := \Big\{f : \mathbb{R}^d \mapsto \mathbb{R} \mid f(x) = \sum_{j=1}^{m} w_j \sigma(u_j^\top x), \|w\|_1 \leq B_1, \forall j, \|u_j\|_1 \leq B_2\Big\},$$

  where $\sigma(z) = \max\{0, z\}$ is the ReLU activation function. Note that the number of neurons m can be arbitrarily large.

- Bounded Lipschitz functions

$$\mathcal{F}_{\mathrm{BL}} := \Big\{f : \mathbb{R}^d \mapsto \mathbb{R} \mid \sup_{x\in[-1,1]^d} |f(x)| \leq 1, |f(x) - f(y)| \leq \|x - y\|_2, \forall x, y \in [-1,1]^d\Big\}.$$

Try to provide the sharpest bound possible.