STA3000F: Final Exam

December 12, 2024

- **Submission:** Please submit your solutions by noon at Dec 12th EST through Quercus. In an emergent situation, you may also submit the solution directly to `wenlong.mou@utoronto.ca`, and you will typically have a one-hour grace period.
- **Policy:** Please work on the problem set by yourself. Collaboration or resorting to external help is not allowed. On the other hand, please feel free to refer to any textbooks, papers, and online materials (or even ChatGPT, if you trust it).
- **Grading:** Each question is worth 20% of the final exam. All these questions can be solved using results from the lectures and the homeworks. You are also welcome to use ideas from other resources (books, papers, etc.). However, you are required to provide self-contained solutions to the problems using only the results from lectures and homeworks. Citing existing results directly as a black box may lead to deductions in the points depending on the nature of these results.
- **Hints:** The difficulties of problems are *not* in ascending (or descending) order. Try to allocate your time wisely. Besides, partially-solved questions may get partial credits.
- **Have fun!**

Q1. Asymptotics can be misleading

Let $X_1, X_2, \cdots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\theta, 1)$ for some unknown $\theta \in \mathbb{R}$. Consider the following estimator

$$\widehat{\theta}_n := \overline{X}_n \mathbf{1}_{|\bar{X}_n| \geq n^{-1/4}}, \quad \text{where } \overline{X}_n := \frac{1}{n} \sum_{i=1}^{n} X_i.$$

- (1/2 credit) For any fixed $\theta \in \mathbb{R}$, compute

$$R_{\text{asym}}(\theta) := \lim_{n \to +\infty} \Big\{ n \cdot \mathbb{E}_\theta \big[|\widehat{\theta}_n - \theta|^2 \big] \Big\}.$$

- (1/2 credit) Show that

$$\lim_{n \to +\infty} \sup_{\theta \in \mathbb{R}} \Big\{ n \cdot \mathbb{E}_\theta \big[|\widehat{\theta}_n - \theta|^2 \big] \Big\} = +\infty.$$

Q2. Gaussian mixture model

Suppose that we observe d-dimensional data from a mixture of two normal distributions

$$X_1, X_2, \cdots, X_n \overset{\text{i.i.d.}}{\sim} \frac{1}{2}\mathcal{N}(\theta^*, I_d) + \frac{1}{2}\mathcal{N}(-\theta^*, I_d)$$

where $\theta^* \in \mathbb{R}^d$ is the underlying unknown parameter.

Prove that there exist universal constants $c, c_1 > 0$, such that for $n \geq c_1 d$, we have

$$\inf_{\widehat{\theta}_n} \sup_{\theta^* \in \mathbb{R}^d} \mathbb{E}\Big[\Big\| \widehat{\theta}_n - \theta^* \Big\|_2^2 \Big] \geq c\sqrt{\frac{d}{n}},$$

Q3. convergence rate for k-means clustering

Let $\mathbb{P}$ be a probability distribution in $\mathbb{R}^d$, satisfying $\mathbb{E}_{\mathbb{P}}\big[\,\|X\|_2^2\,\big] < +\infty$. Given i.i.d. samples from $\mathbb{P}$, we try to divide them into K clusters of symmetric locations. In doing so, we consider the following constrained M-estimator within a constrained set $\Theta \subseteq \mathbb{R}^d$.

$$\big(\widehat{\theta}_n^{(1)}, \widehat{\theta}_n^{(2)}, \cdots, \widehat{\theta}_n^{(K)}\big) := \arg \min_{\theta_1,\theta_2,\cdots,\theta_K \in \Theta} \frac{1}{n} \sum_{i=1}^{n} \min_{1\le j\le K} \|X_i - \theta_j\|_2^2 \,. \tag{1}$$

Clearly, the minimizer may not be unique in general due to symmetry in the problem. In such a case, we take any possible minimizer. We define the population-level loss function

$$F(\theta_1, \theta_2, \cdots, \theta_K) := \mathbb{E}\Big[\min_{1\le j\le K} \|X_i - \theta_j\|_2^2\Big].$$

Assume that Θ is compact. Show that there exists a constant $c > 0$ that depends only on the constraint set Θ and the distribution of X (and of course, the dimension d), such that

$$F\big(\widehat{\theta}_n^{(1)}, \widehat{\theta}_n^{(2)}, \cdots, \widehat{\theta}_n^{(K)}\big) - \min_{\theta_1,\theta_2,\cdots\theta_K \in \Theta} F(\theta_1, \theta_2, \cdots, \theta_K) \le c\sqrt{K/n}$$

with probability 9/10.

Q4. Rademacher complexity of ℓ_1 ball

Given d basis functions $\phi_1, \phi_2, \cdots, \phi_d$ satisfying $\|\phi_j\|_\infty \le 1$ for $j \in [d]$, define the function class

$$\mathcal{F} := \big\{ x \mapsto \theta^\top \phi(x) \mid \theta \in \mathbb{R}^d, \|\theta\|_1 \le 1 \big\}.$$

Let $X_1, X_2, \cdots, X_n$ be i.i.d. data from the underlying distribution $\mathbb{P}$. Show that

- (1/2 credit) For any distribution $\mathbb{P}$ and any set of basis functions under above assumptions, we have

$$\mathbb{E}\Big[\sup_{f \in \mathcal{F}} \frac{1}{\sqrt{n}} \sum_{i=1}^n \varepsilon_i f(X_i) \Big] \le c\sqrt{\log d}.$$

- (1/2 credit) There exists a distribution $\mathbb{P}$ and a class of basis functions satisfying above assumptions, such that

$$\liminf_{n \to +\infty} \ \mathbb{E}\Big[\sup_{f \in \mathcal{F}} \frac{1}{\sqrt{n}} \sum_{i=1}^n \varepsilon_i f(X_i) \Big] \ge c'\sqrt{\log d}.$$

Q5. sup norm convergence rate

Let p be a probability density function supported on the interval $[0,1]$, satisfying the uniform Lipschitz condition

$$|p(x) - p(y)| \le |x - y|, \quad \text{for any pair } x, y \in [0,1].$$

Given samples $X_1, X_2, \cdots, X_n \overset{\text{i.i.d.}}{\sim} p$, construct a kernel density estimator $\widehat{p}_n$ (with your choice of kernel and bandwidth) satisfying

$$\mathbb{E}\Big[\sup_{x\in[0,1]} |\widehat{p}_n(x) - p(x)|^2\Big] \le c\Big(\frac{\log n}{n}\Big)^{2/3}.$$

[Hint: you may use the Bernstein inequality as given.]

$$\mathbb{P}\Big(\Big|\sum_{i=1}^{n}\big(Z_i - \mathbb{E}[Z]\big)\Big| \ge t\Big) \le 2\exp\Big(\frac{-t^2/2}{n\,\mathrm{var}(Z) + t/3}\Big),$$

for i.i.d. random variables $Z_1, Z_2, \cdots, Z_n$ that takes value in $[0,1]$.