STA3000F: Final Exam

December 11, 2023

- **Submission:** Please submit your solutions by noon at Dec 12th through Quercus. Under emergent situation, you may also submit the solution directly to `wenlong.mou@utoronto.ca`, and you'll typically have a one-hour grace period.
- **Policy:** Please work on the problem set by yourself. Collaboration or resorting to external help are 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 welcome to also 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. Please try to allocate your time wisely. Besides, partially-solved questions may get partial credits.
- **Have fun!**

Q1. Singularities in location models

Given $\alpha \in (0, 1)$, consider the one-dimensional probability density function

$$p(x) := Z^{-1} |x|^{-\alpha} \cdot \exp(-x^2), \quad \text{for } x \in \mathbb{R},$$

where $Z > 0$ is a normalization constant to make sure that $\int_{\mathbb{R}} p(x) dx = 1$.

Consider the location class $p_\theta(x) = p(x - \theta)$ with one-dimensional parameter $\theta \in \mathbb{R}$, and let $(X_i)_{i=1}^n \overset{\text{i.i.d.}}{\sim} p_\theta$. For $h > 0$, consider the following simple vs. simple testing problem:

$$H_0 : \theta = 0 \quad \text{vs.} \quad H_1 : \theta = h.$$

1. Given $\gamma \in (0, 1)$, find the UMP level-γ test (you don't need to calculate the cutoff values explicitly).
2. Find a threshold h_n such that when $h \leq h_n$, there is $\mathbb{E}_0[\phi] + \mathbb{E}_1[1 - \phi] \geq 1/4$ for any possible test ϕ. Please try to find the largest possible h_n (up to constant factors) you can. You will get full grade if you get the optimal lower bound (though you don't need to prove it's unimprovable), and partial grades if you get a sub-optimal one.

Q2. M-estimation with convex loss

Suppose that $\theta \mapsto f(\theta; x)$ is a convex and differentiable 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 exist functions L_1 and L_2, such that $\mathbb{E}[L_1(X)^2] + \mathbb{E}[L_2(X)^2] < +\infty$, and for any θ_1, θ_2, x, we have

$$|f(\theta_1; x) - f(\theta_2; x)| \leq L_1(x) \left\|\theta_1 - \theta_2\right\|_2, \quad \text{and} \quad \left\|\nabla f(\theta_1; x) - \nabla f(\theta_2; x)\right\|_2 \leq L_2(x) \left\|\theta_1 - \theta_2\right\|_2,$$

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

Q3. Binary classification

Consider the supervised learning problem, with i.i.d. data $(X_i, Y_i) \overset{\text{i.i.d.}}{\sim} \mathbb{P}$, where $X_i \in \mathcal{X}$ and $Y_i \in \{-1, 1\}$. Let $\mathcal{F}$ be a class of functions mapping from $\mathcal{X}$ to $\{-1, 1\}$. Define the ERM estimator

$$\widehat{f}_n := \arg\min_{f \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^n \mathbf{1}\big[Y_i \neq f(X_i)\big].$$

Define the Rademacher complexity

$$\mathcal{R}_n(\mathcal{F}) := \mathbb{E}\Big[\sup_{f \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^n \varepsilon_i f(X_i)\Big], \quad \text{for i.i.d. Rademacher random variables } (\varepsilon_i)_{i=1}^n.$$

Consider the loss functional $L(f) := \mathbf{1}[Y \neq f(X)]$. Show that there exists a universal constant $c > 0$, such that

$$L(\widehat{f}_n) \leq \inf_{f \in \mathcal{F}} L(f) + c\mathcal{R}_n(\mathcal{F}),$$

with probability 1/4.

Let $X_1, X_2, \cdots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, I_d)$, and consider the class $\mathcal{F} := \{x \mapsto 2\mathbf{1}[x^\top \theta > 0] - 1 : \theta \in \mathbb{R}^d\}$ of linear classifiers. Show that when $n \leq d$, we have $\mathcal{R}_n(\mathcal{F}) = 1$.

Q4. Estimating the derivatives

Given a scalar $\beta > 1$, let p be a probability density function on $\mathbb{R}$ such that $p \in \Sigma(\beta)$ (i.e., β-th order Hölder class). We are interested in nonparametric estimation of the derivative p'.

Given a kernel function $K : \mathbb{R} \to \mathbb{R}$ supported on $[-1, 1]$ satisfying the conditions

$$\int_{\mathbb{R}} u^j K(u) du = \begin{cases} 1 & j = 1, \\ 0 & j = 0, 2, \cdots, \lfloor \beta \rfloor. \end{cases}$$

Let $X_1, X_2, \cdots, X_n \overset{\text{i.i.d.}}{\sim} p$. Given bandwidth $h > 0$, consider the kernel-based estimator

$$\widehat{d}_n(x) := \frac{1}{nh^2} \sum_{i=1}^{n} K\Big(\frac{X_i - x}{h}\Big)$$

For any x_0, and prove the MSE bound

$$\mathbb{E}\big[|\widehat{d}_n(x_0) - p'(x_0)|^2\big] \le n^{\frac{-2(\beta-1)}{n+2\beta}}.$$

with an optimal bandwidth $h = h_n$

Q5. Lower bound for isotonic regression

Consider the function class

$$\mathcal{F} := \big\{f : [0,1] \to [0,1],\ f \text{ is non-decreasing}\big\}.$$

In the homework, we have proved a (near-optimal) upper bound for estimating f. Now we turn to the lower bound.

Given $f^* \in \mathcal{F}$, let the observations be $Y_i = f^*(x_i) + \varepsilon_i$ with $x_i = i/n$ for $i = 1, 2, \cdots, n$, where the noises $(\varepsilon_i)_{i=1}^n$ are i.i.d. standard normal random variables.

Prove that

$$\inf_{\widehat{f}} \sup_{f^* \in \mathcal{F}} \mathbb{E}\Big[\Big\|\widehat{f} - f^*\Big\|_n^2\Big] \ge cn^{-2/3},$$

for a universal constant $c > 0$.