STA3000F: Final Exam

December 15th, 2025

- **Submission:** Please submit your solutions by 23:59 at Dec 16th 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. Please do not use generative AI tools to generate the solutions directly (it is fine to use them for auxiliary purposes such as formatting or proofreading).
- **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: probabilistic method and generalized Ramsey number

A graph $\mathcal{G} = (V, E)$ consists of a set of vertices V and a set of edges $E \subseteq V \times V$. Given a subset $V' \subseteq V$, the subgraph $\mathcal{G}[V']$ induced by V' is the graph $\mathcal{G}' = (V', E')$ where $E' = \{(u, v) \in E : u, v \in V'\}$.

Given $m \geq 2$ graphs $\mathcal{G}_1 = (V_1, E_1), \mathcal{G}_2 = (V_2, E_2), \ldots, \mathcal{G}_m = (V_m, E_m)$, with $|V_i| = k$ for each $i \in [m]$. Show that when $n < 2^{\frac{k-1}{2}} m^{-1/k}$, there exists a graph $\mathcal{G} = (V, E)$ with $|V| = n$ such that for every $V' \subseteq V$ with $|V'| = k$, the induced subgraph $\mathcal{G}[V']$ is not isomorphic to any of the graphs $\mathcal{G}_1, \mathcal{G}_2, \ldots, \mathcal{G}_m$.

[Hint: generate a random graph and show that it satisfies the requirement with positive probability.]

Q2: Binary classification

Consider a class $\mathcal{F}$ of binary classifiers $f : \mathcal{X} \to \{0, 1\}$. Let $\mathcal{D}$ be a distribution over $\mathcal{X} \times \{0, 1\}$. Suppose that there exists a classifier $f^* \in \mathcal{F}$ such that $Y = f^*(X)$ almost surely when $(X, Y) \sim \mathcal{D}$. Given an i.i.d. sample $\{(X_i, Y_i)\}_{i=1}^n$ drawn from $\mathcal{D}$, we define the empirical risk minimizer (ERM)

$$\widehat{f} \in \arg\min_{f \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{f(X_i) \neq Y_i},$$

where we choose an arbitrary minimizer if there are multiple ones.

Assume that the VC dimension of the class $\mathcal{F}$ is $d < \infty$. Show that with high probability, the classifier $\widehat{f}$ satisfies

$$R(\widehat{f}) \leq \frac{cd \log n}{n},$$

where $c > 0$ is a constant depending only on the error probability and $R(f) = \mathbb{P}_{(X,Y) \sim \mathcal{D}}(f(X) \neq Y)$ is the risk of the classifier f.

[Hint: you may adapt the rate theorem using localization argument in the lecture notes.]

Q3: Singular location models

Consider a one-dimensional density function

$$p_0(x) = Z^{-1} \frac{1}{\sqrt{|x|}} e^{-x^2},$$

where Z is the normalizing constant. For $\theta \in \mathbb{R}$, we define the location family

$$p_\theta(x) = p_0(x - \theta).$$

Suppose we observe i.i.d. samples $\{X_i\}_{i=1}^n$ drawn from p_{θ^*} for some $\theta^* \in \mathbb{R}$. Consider the testing problem

$$H_0 : \theta = \theta_0, \quad \text{vs.} \quad H_1 : |\theta - \theta_0| \geq \varepsilon,$$

for some known $\theta_0 \in \mathbb{R}$. Find the minimax testing radius ε_n such that there exists a test ϕ_n that makes both type-I and type-II errors less than $1/4$.

[Bonus question: use the result to construct an estimator $\widehat{\theta}$ with high-probability convergence rate guarantees.]

Q4: localized Rademacher complexity of ellipsoids

Let $(\phi_j)_{j=1}^{+\infty}$ be an orthonormal basis of $\mathbb{L}^2(\mathcal{X}, \mu)$, where $\mathcal{X}$ is a measurable space and μ is a probability measure on $\mathcal{X}$. Given a non-increasing sequence of positive numbers $(\lambda_j)_{j=1}^{+\infty}$, we define the ellipsoid

$$\mathcal{E} = \Big\{ f \in \mathbb{L}^2(\mathcal{X}, \mu) : f = \sum_{j=1}^{+\infty} \theta_j \phi_j, \quad \sum_{j=1}^{+\infty} \frac{\theta_j^2}{\lambda_j} \leq 1 \Big\}.$$

Given $r > 0$, we define the localized set

$$\mathcal{E}(r) = \{ f \in \mathcal{E} : \|f - f^*\|_{\mathbb{L}^2(\mu)} \leq r \},$$

where $f^* \in \mathcal{E}$ is a fixed function. Consider the localized Rademacher complexity

$$\mathcal{R}_n(r) = \mathbb{E}\Big[\sup_{f \in \mathcal{E}(r)} \frac{1}{n} \sum_{i=1}^{n} \zeta_i f(X_i) \Big],$$

where $\{X_i\}_{i=1}^n$ are i.i.d. samples drawn from μ and $\{\zeta_i\}_{i=1}^n$ are i.i.d. Rademacher variables independent of $\{X_i\}_{i=1}^n$.

Assuming that $\sum_{j=1}^{+\infty} \lambda_j < +\infty$, show that there exists a constant $c > 0$ such that

$$\mathcal{R}_n(r) \leq c \sqrt{\frac{1}{n} \sum_{j=1}^{+\infty} \min\{\lambda_j, r^2\}}.$$

Q5: fixed point estimation

Denote by $\mathcal{X}$ the set of bounded measurable functions $f : [0,1] \to \mathbb{R}$ equipped with the supremum norm $\|\cdot\|_\infty$, and let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space.

Consider a nonlinear operator $\mathcal{T} : \mathcal{X} \times \Omega \to \mathcal{X}$. Suppose that $\mathcal{T}$ is a contraction operator in expectation, i.e., there exists a constant $\gamma \in (0,1)$ such that for every $f, g \in \mathcal{X}$,

$$\|\mathbb{E}[\mathcal{T}(f,\omega)] - \mathbb{E}[\mathcal{T}(g,\omega)]\|_\infty \leq \gamma \|f - g\|_\infty .$$

Assume that for every $f \in \mathcal{X}$ and $\omega \in \Omega$, we have

$$\|\mathcal{T}(f,\omega)\|_\infty \leq B, \qquad \|\mathcal{T}(f,\omega)\|_{\mathrm{Lip}} \leq L,$$

where $B, L > 0$ are constants and $\|\cdot\|_{\mathrm{Lip}}$ is the Lipschitz norm defined as

$$\|f\|_{\mathrm{Lip}} = \sup_{x \neq y} \frac{|f(x) - f(y)|}{|x - y|}.$$

Suppose that $\mathcal{T}$ is known, and we observe i.i.d. samples $\{\omega_i\}_{i=1}^n$ drawn from $\mathbb{P}$. Our goal is to estimate the fixed point $f^* \in \mathcal{X}$ satisfying

$$f^* = \mathbb{E}_{\omega \sim \mathbb{P}}[\mathcal{T}(f^*, \omega)].$$

Find an estimator $\widehat{f}$ based on the samples $\{\omega_i\}_{i=1}^n$ such that with high probability,

$$\left\|\widehat{f} - f^*\right\|_\infty \leq c\sqrt{\frac{\log n}{n}},$$

where $c > 0$ is a constant depending only on γ, B and the error probability.

[Hint: you may consider fixed-point iteration.]