Math 281 – Qualifying Exam – Spring 2020

Define any symbol you use unless its meaning is clear from context. Name any result you use if it has a name. Be concise and clear. Justify all your answers.

Problem 1. Suppose that we observe data in pairs $(X, Y) \in \mathbb{R}^d \times \mathbb{R}$, where the data come from a linear model

$$Y = X^{\mathrm{T}}\beta^* + \varepsilon,$$

where $\beta^* \in \mathbb{R}^d$ is the (unknown) vector of regression coefficients, and ε is the noise variable. Assume that ε and X are independent, ε is symmetric around 0 and $\mathbb{E}(\varepsilon^2) = \sigma^2 > 0$. Given $\tau > 0$, consider a loss function

$$\ell_\tau(x) = \tau^2 \ell(x/\tau),\ x \in \mathbb{R},$$

where $\ell : \mathbb{R} \to [0, \infty)$ is a function chosen from the following three: (i) $\ell(x) = \sqrt{1 + x^2} - 1$, (ii) $\ell(x) = \log((e^x + e^{-x})/2)$, and (iii) $\ell(x) = (x^2/2 - |x|^3/6)1(|x| \le 1) + (|x|/2 - 1/6)1(|x| > 1)$. Given independent observations $(X_1, Y_1), \ldots, (X_n, Y_n)$ from (X, Y), let $\hat{\beta}_n$ be an M-estimator of β^* that minimizes the empirical loss

$$L_n(\beta) = \frac{1}{n}\sum_{i=1}^n \ell_\tau(Y_i - X_i^{\mathrm{T}}\beta) = \frac{\tau^2}{n}\sum_{i=1}^n \ell\big((Y_i - X_i^{\mathrm{T}}\beta)/\tau\big).$$

That is, $\hat{\beta}_n \in \arg\min_{\beta \in \mathbb{R}^d} L_n(\beta)$. Assume in addition that $\Sigma = \mathbb{E}(XX^{\mathrm{T}})$ is positive definite and $\mathbb{E}\|X\|_2^4 < \infty$.

(a) Let $L(\beta) = \mathbb{E}L_n(\beta)$ be the population loss function. Describe conditions under which β^* is the unique minimizer of $\beta \mapsto L(\beta)$, that is, $\beta^* \in \arg\min_{\beta \in \mathbb{R}^d} L(\beta)$.

(b) Under the conditions from part (a), provide a rigorous proof of the consistency of $\hat{\beta}_n$, that is, $\hat{\beta}_n \to \beta^*$ in probability as $n \to \infty$ (while d is fixed). You can either pick a specific ℓ function from the above three candidates, or provide a generic proof that applies to all three cases.

(c) Provided the consistency of $\hat{\beta}_n$ holds, describe the asymptotic covariance matrix of $\hat{\beta}_n$. Provide a heuristic argument to justify your findings.

Problem 2. Assume the same conditions of Problem 2 hold, but consider a different loss function $\ell_\tau(x) = (x^2/2)1(|x| \le \tau) + (\tau|x| - \tau^2/2)1(|x| > \tau)$ for some $\tau > 0$. Given an iid sample $\{(X_i, Y_i)\}_{i=1}^n$, consider the empirical risk minimization procedure

$$\hat{\beta}_n \in \arg\min_{\beta\in\Theta} \underbrace{\frac{1}{n}\sum_{i=1}^n \ell_\tau(Y_i - X_i^{\mathrm{T}}\beta)}_{L_n(\beta)},$$

where Θ is a subset of $\mathbb{R}^d$. The following result, known as the *Ledoux-Talagrand Rademacher contraction inequality*, may be useful for this question. Let $\phi \circ \mathcal{F} = \{h : h(x) = \phi(f(x)), f \in \mathcal{F}\}$ denote the composition of $\phi(\cdot)$ with functions in $\mathcal{F}$. If $\phi(\cdot)$ is L-Lipschitz, then $R_n(\phi\circ\mathcal{F}) \le LR_n(\mathcal{F})$, where R_n denotes the Rademacher complexity.

(a) Assume $\Theta = \{\beta \in \mathbb{R}^d : \|\beta\|_2 \le R\}$ for some $R > 0$, and $\mathbb{E}\|X\|_2^2 < \infty$. Describe conditions under which

$$\sup_{\beta\in\Theta} |L_n(\beta) - L(\beta)| \to 0 \text{ in probability}$$

as $n \to \infty$, where $L(\beta) := \mathbb{E}L_n(\beta)$ is the population loss.

(b) Let $\Theta = \{\beta \in \mathbb{R}^d : \|\beta\|_2 \le R\}$ for some $R > 0$, and let X be supported on an ℓ_2-ball $\{x \in \mathbb{R}^d : \|x\|_2 \le M\}$. Give the smallest $\epsilon_n(\delta, d, \tau, R, M)$ you can (ignoring the constants) such that

$$\mathbb{P}\left\{\sup_{\theta\in\Theta} |L_n(\beta) - L(\beta)| \ge \epsilon_n(\delta, d, \tau, R, M)\right\} \le \delta.$$

(c) Let $\Theta = \{\beta \in \mathbb{R}^d : \|\beta\|_1 \le R\}$ for some $R > 0$, and let X be supported on an ℓ_∞-ball $\{x = (x_1, \ldots, x_d)^{\mathrm{T}} \in \mathbb{R}^d : \|x\|_\infty = \max_{1\le j\le d} |x_j| \le M\}$. Give the smallest $\epsilon_n(\delta, d, \tau, R, M)$ you can (ignoring the constants) such that

$$\mathbb{P}\left\{\sup_{\theta\in\Theta} |L_n(\beta) - L(\beta)| \ge \epsilon_n(\delta, d, \tau, R, M)\right\} \le \delta.$$

Refer by number any result from the reference sheet that you use.

Problem 3. Consider a setting where $X_1, \dots, X_n \overset{\text{iid}}{\sim} \mathcal{N}(\theta, \sigma^2)$. We consider the problem of estimating θ under square error loss. The variance σ is assumed known except when otherwise specified.

(a) Consider the average risk with respect to the $\mathcal{N}(0, \tau^2)$ prior (where τ is known). Derive the best estimator for that measure of risk and compute its risk.

(b) Prove that the sample mean is minimax. Is the sample mean minimax if it is known that $\theta \in [-1, 1]$? Is the sample mean minimax when σ is unknown? Is the sample mean minimax when σ is unknown but it is known that $\sigma \in [1/10, 10]$?

(c) (Assume again that σ^2 is known.) Prove that $\bar{X}$ is admissible. For what values of $a, b \in \mathbb{R}$ is $a\bar{X} + b$ admissible?