Econ 241 C.

1. *Ridge regression. (33p)* (a. Bayes perspective) Consider the normal linear regression model

$$Y_i = X_i'\theta_0 + \epsilon_i, \quad i = 1, 2, \ldots, n, \quad \epsilon_i \sim N(0, 1) \quad \text{i.i.d} \tag{1}$$

Suppose $\theta_0 \sim N(0, \tau^2 I_p)$. Derive the posterior distribution for θ_0. Discuss the mean posterior limit when (a) $\tau = \infty, n$ finite and (b) when $n \to \infty, \tau$ finite.

(b. Frequentist perspective) Consider the ridge estimator

$$\widehat{\theta}_\tau^{\text{ridge}} = \arg\min_{\theta \in \mathrm{R}^p} \left\{ \frac{1}{n} \sum_{i=1}^{n} (Y_i - X_i'\theta)^2 + \tau \|\theta\|_2^2 \right\}$$

- Show that for any τ, $\widehat{\theta}_\tau^{\text{ridge}}$ is uniquely defined. Derive the closed-form solution.
- Compute the bias of $\widehat{\theta}_\tau^{\text{ridge}}$ and show that it is bounded in absolute value by $\|\theta_0\|_2$.
- Discuss the connection between (a) and (b).

2. (33p) *Hard and soft thresholding.* Consider a linear model

$$Y_i = X_i'\theta_0 + \epsilon_i, \quad i = 1, 2, \ldots, n, \quad \epsilon_i \sim N(0, 1) \quad \text{i.i.d} \tag{2}$$

where $X_i \in \mathrm{R}^p$ are fixed (i.e., non-random) vectors obeying the condition

$$\frac{1}{n} \sum_{i=1}^{n} X_i X_i' = I_p. \tag{3}$$

Define the MLE/OLS estimator $\widehat{\theta} = \frac{1}{n} \sum_{i=1}^{n} X_i Y_i$, the hard thresholded estimator $\widehat{\theta}^{\text{HRD}} = (\widehat{\theta}_1^{\text{HRD}}, \widehat{\theta}_2^{\text{HRD}}, \ldots, \widehat{\theta}_p^{\text{HRD}})'$ as

$$\widehat{\theta}_j^{\text{HRD}} = \begin{cases} \widehat{\theta}_j & \text{if } |\widehat{\theta}_j| > 2\rho, \\ 0 & \text{if } |\widehat{\theta}_j| \leq 2\rho, \end{cases}, \quad j = 1, 2, \ldots, p$$

and the soft thresholded estimator $\widehat{\theta}^{\text{SFT}} = (\widehat{\theta}_1^{\text{SFT}}, \widehat{\theta}_2^{\text{SFT}}, \ldots, \widehat{\theta}_p^{\text{SFT}})'$ as

$$\widehat{\theta}_j^{\mathrm{SFT}} = \begin{cases} \widehat{\theta}_j - 2\rho, & \text{if } \widehat{\theta}_j > 2\rho \\ 0, & \text{if } |\widehat{\theta}_j| < 2\rho, \\ \widehat{\theta}_j + 2\rho, & \text{if } \widehat{\theta}_j < -2\rho. \end{cases}$$

Show that

$$\widehat{\theta}^{\mathrm{HRD}} = \arg\min_{\theta \in \mathrm{R}^p} \left\{ \frac{1}{n} \sum_{i=1}^{n} (Y_i - X_i'\theta)^2 + 4\rho^2 \|\theta\|_0 \right\} \tag{4}$$

$$\widehat{\theta}^{\mathrm{SFT}} = \arg\min_{\theta \in \mathrm{R}^p} \left\{ \frac{1}{n} \sum_{i=1}^{n} (Y_i - X_i'\theta)^2 + 4\rho \|\theta\|_1 \right\}. \tag{5}$$

Hint: use ORT condition (3) to reduce the p-dimensional optimization problem (4) into p one-dimensional problems.

3. *Finite-sample Confidence Interval. (33p.)* Let $(X_i)_{i=1}^n$ be an i.i.d random sample of σ^2-subGaussian random variables with mean $\mathrm{E}X = \mu$. Let

$$\bar{X} = n^{-1} \sum_{i=1}^{n} X_i.$$

Given a level δ, define

$$CI_{1-\delta} := (\bar{X} - \sigma\sqrt{2\log(2/\delta)/n}, \quad \bar{X} + \sigma\sqrt{2\log(2/\delta)/n}) \tag{6}$$

(a) Show that

$$\Pr(\mu \in CI_{1-\delta}) \geq 1 - \delta$$

in other words, $CI_{1-\delta}$ is a valid **finite-sample!** confidence Interval for μ that is valid for any finite $n = 1, 2, \ldots$. Do NOT use the Central Limit Theorem.

(b) Suppose X is bounded a.s. by some known constant K, i.e., $|X| \leq K$ a.s. . If σ is unknown, the CI (6) is no longer feasible. Propose a modification of $CI_{1-\delta}$, replacing the unknown σ by some expression depending on (known) K.