Homework 4: STA 721 Fall24

Your Name

September 17, 2024; Due in one week (see Gradescope)

You are encouraged to write up solutions in LaTeX. Please solve/attempt Problem 1 before class on Tuesday.

1. For the following assume that $\mathbf{Y} \mid \mathbf{X} \sim \mathsf{N}(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}_n)$.
   (a) Consider estimation of $\boldsymbol{\beta}$ using quadratic loss $(\boldsymbol{\beta} - \mathbf{a})^T(\boldsymbol{\beta} - \mathbf{a})$ for some estimator $\mathbf{a}$. Find the expected quadratic loss if we use the MLE $\hat{\boldsymbol{\beta}}$ for $\mathbf{a}$ conditional on $\mathbf{X}$. Simplify the expression as a function of the eigenvalues of $\mathbf{X}^T\mathbf{X}$. What happens to the expected loss as the smallest eigenvalue of $\mathbf{X}^T\mathbf{X}$ goes to 0?
   (b) Consider estimation of $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$ at the observed data points $\mathbf{X}$. Find the expected quadratic loss $E[(\boldsymbol{\mu} - \mathbf{X}\hat{\boldsymbol{\beta}})^T(\boldsymbol{\mu} - \mathbf{X}\hat{\boldsymbol{\beta}})]$ conditional on $\mathbf{X}$. What happens as the smallest eigenvalue of $\mathbf{X}^T\mathbf{X}$ goes to 0?
   (c) Consider predicting a new $\mathbf{Y}_*$ at the observed data points $\mathbf{X}$ where $\mathbf{Y}_*$ is independent of $\mathbf{Y}$. Find the expected quadratic loss for $E[(\mathbf{Y}_* - \mathbf{X}\hat{\boldsymbol{\beta}})^T(\mathbf{Y}_* - \mathbf{X}\hat{\boldsymbol{\beta}})]$. What happens as the smallest eigenvalue of $\mathbf{X}^T\mathbf{X}$ goes to 0?
   (d) Consider predicting $\mathbf{Y}_*$'s at new points $\mathbf{X}_*$ with $E[\mathbf{X}_*^T\mathbf{X}_*] = \mathbf{I}_p$. Find the expected quadratic loss $E[(\mathbf{Y}_* - \mathbf{X}_*\hat{\boldsymbol{\beta}})^T(\mathbf{Y}_* - \mathbf{X}_*\hat{\boldsymbol{\beta}})]$ conditional on $\mathbf{X}$ and $\mathbf{X}_*$ and then unconditional on $\mathbf{X}_*$ (but still conditional on $\mathbf{X}$. What happens as the smallest eigenvalue of $\mathbf{X}^T\mathbf{X}$ goes to 0? (If $E[\mathbf{X}_*^T\mathbf{X}_*] = \Sigma_{\mathbf{X}} > 0$ does that change the result)
   (e) Briefly comment on the difference in estimation and prediction at observed data versus new data as $\mathbf{X}$ becomes non-full rank. Which is the most stable? Which is the least?
2. Let $\mathbf{Y} \sim \mathsf{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{V})$ where $\mathbf{X}$ and $\mathbf{V}$ are known but $\boldsymbol{\beta}$ and σ^2 are unknown. Let $\hat{\boldsymbol{\beta}}$ denote the OLS estimator and $\hat{\boldsymbol{\beta}}_{\mathbf{V}}$ denote the GLS estimator.
   (a) Let $\hat{\boldsymbol{\epsilon}} = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}$. Find an unbiased estimator of σ^2 that is a scalar multiple of $\hat{\boldsymbol{\epsilon}}^T\hat{\boldsymbol{\epsilon}}$.
   (b) Let $\hat{\boldsymbol{\epsilon}}_{\mathbf{V}} = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{\mathbf{V}}$. Find an unbiased estimator of σ^2 that is a scalar multiple of $\hat{\boldsymbol{\epsilon}}_{\mathbf{V}}^T\mathbf{V}^{-1}\hat{\boldsymbol{\epsilon}}_{\mathbf{V}}$.
   (c) Do you think one of these estimators is generally better than the other?
3. Let $\mathbf{Y} \sim \mathsf{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{V})$ where $\mathbf{V} > 0$ is known. Find the MLE of $(\boldsymbol{\beta}, \sigma^2)$, and its distribution. Is the MLE of σ^2 unbiased?
4. Let $\mathbf{U} \in \mathbb{R}^{n \times n}$ be an orthogonal matrix so $\mathbf{U}\mathbf{U}^T = \mathbf{U}^T\mathbf{U} = \mathbf{I}_n$. Separate the columns of $\mathbf{U}$ into the three matrices $\mathbf{U}_1 \in \mathbb{R}^{n \times p_1}$, $\mathbf{U}_2 \in \mathbb{R}^{n \times p_2}$, $\mathbf{U}_3 \in \mathbb{R}^{n \times p_3}$ where $p_1 + p_2 + p_3 = n$ and $\mathbf{U}$ is equal to $\mathbf{U}_1$, $\mathbf{U}_2$ and $\mathbf{U}_3$ column-binded together.
   (a) Show that $\mathbf{U}_1\mathbf{U}_1^T + \mathbf{U}_2\mathbf{U}_2^T + \mathbf{U}_3\mathbf{U}_3^T = \mathbf{I}_n$.
   (b) Let $\mathbf{Z} \sim \mathsf{N}_n(\mathbf{0}, \mathbf{I}_n)$. Find the distributions of $x_k = \mathbf{Z}^T\mathbf{U}_k\mathbf{U}_k^T\mathbf{Z}$ for $k \in \{1, 2, 3\}$, and show that x_1, x_2 and x_3 are independent. Find the distribution of $x_1 + x_2 + x_3$ and compare it to the distribution of $x = \mathbf{Z}^T\mathbf{Z}$.

5. Projecting out nuisance factors: Suppose we have the ordinary linear model $\mathbf{Y} = \mathbf{W}\boldsymbol{\alpha} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ where $\mathsf{E}[\boldsymbol{\epsilon}] = 0$, $\mathsf{Cov}[\boldsymbol{\epsilon}] = \sigma^2\mathbf{I}_n$, and $\mathbf{W} \in \mathbb{R}^{n\times q}$ and $\mathbf{X} \in \mathbb{R}^{n\times p}$ are observed model matrices, such that the columns of $\mathbf{W}$ and $\mathbf{X}$ taken together are linearly independent. Suppose we are only interested in estimating $\boldsymbol{\beta}$.

   (a) Let $\mathbf{N}$ be an orthonormal basis for the null space of $C(\mathbf{W})$. Using this matrix, find a transformation $\tilde{\mathbf{Y}}$ of $\mathbf{Y}$ so that the expectation of $\tilde{\mathbf{Y}}$ depends on $\mathbf{X}$ but not $\mathbf{W}$. Write out the linear model for $\tilde{\mathbf{Y}}$, and find the OLS estimator of $\hat{\boldsymbol{\beta}}_{\mathbf{N}}$ of $\boldsymbol{\beta}$ under this model in terms of $\mathbf{Y}, \mathbf{X}$ and $\mathbf{N}$ (or $\mathbf{I} - \mathbf{P}$).

   (b) Let $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})$ be the OLS estimator of $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ for the full linear model $E[\mathbf{Y}] = \mathbf{W}\boldsymbol{\alpha} + \mathbf{X}\boldsymbol{\beta}$. Using results on partitioned matrices, show that $\hat{\boldsymbol{\beta}}_{\mathbf{N}} = \hat{\boldsymbol{\beta}}$.

   (c) Obtain a form for the usual unbiased estimator of σ^2 using the model in part (a), and also for the model in (b), and show they are the same.