Homework 3: STA 721 Fall24

Your Name

September 12, 2024; Due in one week (see Gradescope)

1. Suppose we have a linear model with $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ with $\mathbf{X} \in \mathbb{R}^{n \times p}$, rank $r < p < n$ and $\mathsf{E}[\boldsymbol{\epsilon}] = \mathbf{0}_n$ and $\mathsf{Cov}[\boldsymbol{\epsilon}] = \sigma^2 \mathbf{I}_n$.
   (a) Is $\mathbf{P}_{\mathbf{X}^T} = (\mathbf{X}^T\mathbf{X})(\mathbf{X}^T\mathbf{X})^-$ a projection onto $C(\mathbf{X}^T)$?
   (b) Is the expression for $\mathbf{P}_{\mathbf{X}^T}$ unique?
   (c) Is $\mathbf{P}_{\mathbf{X}^T}$ an orthogonal projection in general?
   (d) Is $\mathbf{P}_{\mathbf{X}^T}$ using the Moore-Penrose generalized inverse an orthogonal projection?
   (e) For $\mathbf{X} \in \mathbb{R}^{n \times p}$ rank $p < n$, find $\mathbf{I}_n - \mathbf{P}_{X^T}$. What does this lead you to conclude about whether an arbitrary $\boldsymbol{\lambda}$ leads to a unique unbiased estimator of $\boldsymbol{\lambda}^T\boldsymbol{\beta}$ using the result from class that generalizes Proposition 2.1.6 in Christensen?

2. Show that $\mathbf{P}_\mathbf{V} \equiv \mathbf{X}\left(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{V}^{-1}$ is idempotent (and therfore is a projection).

3. Show that $\mathbf{P}_\mathbf{V}$ is an orthogonal projection under the inner product $\langle \mathbf{x}, \mathbf{y} \rangle_{\mathbf{V}^{-1}} \equiv \mathbf{x}^T\mathbf{V}^{-1}\mathbf{y}$ using the definition in the slides; $\mathbf{P}_\mathbf{V}$ is an orthogonal projection if $\langle \mathbf{P}_\mathbf{V}\mathbf{y}, (\mathbf{I}_n - \mathbf{P}_\mathbf{V})\mathbf{y} \rangle_{\mathbf{V}^{-1}} = 0$

4. Weighted Regression: Consider the simple straight-line through the origin regression model, $Y_i = \beta x_i + \epsilon_i$ where β and x_i are scalars.
   (a) Show that $\hat{\beta}_w = \sum w_i Y_i x_i / \sum w_i x_i^2$ is an unbiased estimator of β as long as $\sum w_i x_i^2 \neq 0$.
   (b) Now suppose the ϵ_i's are uncorrelated but $\mathsf{var}[\epsilon_i] = \sigma^2 \times v_i$. Compute the variance of $\hat{\beta}_w$, and using calculus or some other method, find the values of $w_1, \ldots, w_n$ that minimize this variance.

5. If $\mathbf{P}_\mathbf{V}$ is an orthogonal projection under the inner product $\langle \mathbf{x}, \mathbf{y} \rangle_{\mathbf{V}^{-1}} \equiv \mathbf{x}^T\mathbf{V}^{-1}\mathbf{y}$, show that by writing the loss function $\|\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\|^2_{\mathbf{V}^{-1}}$ as $\|(\mathbf{Y} - \mathbf{P}_\mathbf{V}\mathbf{y}) + (\mathbf{P}_V\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})\|^2_{\mathbf{V}^{-1}}$ and expanding appropriately that the cross-product term is zero and that $\mathbf{P}_\mathbf{V}\mathbf{Y}$ minimizes the generalized loss and is the GLS estimator of $\boldsymbol{\mu}$. Aren the conditions equivalent to $\langle \mathbf{P}_\boldsymbol{\Phi}\mathbf{y}, (\mathbf{I}_n - \mathbf{P}_\boldsymbol{\Phi})\mathbf{y} \rangle_{\mathbf{V}^{-1}} = 0 \quad \forall \mathbf{y} \in \mathbb{R}^n$

6. Corollary from class
   (a) Let $\mathbf{P}$ be a possibly oblique projection onto $C(\mathbf{X})$. Find a LUE $\tilde{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ such that $\mathbf{X}\tilde{\boldsymbol{\beta}} = \mathbf{P}\mathbf{Y}$ for each $\mathbf{Y} \in \mathbb{R}^n$, and show that it is unique.
   (b) Suppose $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ and
   $$\mathsf{Cov}[\boldsymbol{\epsilon}] = \mathbf{V} \equiv \mathbf{X}\boldsymbol{\Psi}\mathbf{X}^T + \boldsymbol{\Phi}$$
   Show the GLS estimator $\hat{\boldsymbol{\beta}}_\mathbf{V}$ is equal to the LUE given by
   $$\hat{\boldsymbol{\beta}}_\boldsymbol{\Phi} = \left(\mathbf{X}^T\Phi^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^T\boldsymbol{\Phi}^{-1}\mathbf{Y}$$
   (Review the steps of the proof outlined in class of the Theorem that showed $\hat{\boldsymbol{\beta}}_\mathbf{V} = \hat{\boldsymbol{\beta}}$ for all $\mathbf{Y}$ iff $\mathbf{V}$ can be written $\mathbf{V} = \mathbf{X}\boldsymbol{\Psi}\mathbf{X}^T + \mathbf{H}\boldsymbol{\Phi}\mathbf{H}^T$ for some $\mathbf{H}$ such that $\mathbf{H}^T\mathbf{X} = 0$ and for some positive definite matrices $\boldsymbol{\Psi}$ and $\boldsymbol{\Phi}$ of the appropriate dimension and Proposition 2.7.5 in Christensen.)

(c) Show that that the $\mathsf{Cov}[\mathbf{P_V}\boldsymbol{\epsilon}, (\mathbf{I}_n - \mathbf{P_V})\boldsymbol{\epsilon}] = \mathbf{0}$. (Hint construct an appropriate decomposition of $\boldsymbol{\epsilon} = \boldsymbol{\epsilon}_{\mathbf{X}} + \boldsymbol{\epsilon}_N$ that has the desired covariance sturture as in class.

(d) Show that that the $\mathsf{Cov}[\mathbf{P}_{\boldsymbol{\Phi}}\boldsymbol{\epsilon}, (\mathbf{I}_n - \mathbf{P}_{\boldsymbol{\Phi}})\boldsymbol{\epsilon}] = \mathbf{0}$.

(e) (Challenging) Show that for $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ and $\mathsf{Cov}[\boldsymbol{\epsilon}] = \mathbf{V} \equiv \mathbf{X}\boldsymbol{\Psi}\mathbf{X}^T + \boldsymbol{\Phi}$, that $\hat{\boldsymbol{\beta}}_{\mathbf{V}} = \hat{\boldsymbol{\beta}}_{\boldsymbol{\Phi}}$ iff the covariance of $\mathbf{P}_{\boldsymbol{\Phi}}\boldsymbol{\epsilon}$ and $(\mathbf{I}_n - \mathbf{P}_{\boldsymbol{\Phi}})\boldsymbol{\epsilon}$ is $\mathbf{0}$.