

Homework 3: STA 721 Fall24

Your Name

September 16, 2024; Due in one week (see Gradescope)

You are encouraged to write up solutions in LaTeX.

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 $\mathbb{E}[\boldsymbol{\epsilon}] = \mathbf{0}_n$ and $\text{Cov}[\boldsymbol{\epsilon}] = \sigma^2 \mathbf{I}_n$.
 - (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 $\mathbf{I}_n - \mathbf{P}_{\mathbf{X}^T}$. 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. For $\mathbf{X} \in \mathbb{R}^{n \times p}$ rank $p < n$ and $\mathbf{V} > 0$, show that $\mathbf{P}_{\mathbf{V}} \equiv \mathbf{X} (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1}$ is idempotent.
3. Show that $\mathbf{P}_{\mathbf{V}}$ (defined above) 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 $\langle \mathbf{P}_{\mathbf{V}} \mathbf{y}, (\mathbf{I}_n - \mathbf{P}_{\mathbf{V}}) \mathbf{y} \rangle_{\mathbf{V}^{-1}} = 0$. Is it always an orthogonal projection using $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T \mathbf{y}$?
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 $\text{var}[\epsilon_i] = \sigma^2 \times v_i$. Compute the variance of $\hat{\beta}_w$, and using calculus or some other method, find the values of w_1, \dots, 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 $\|\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\|_{\mathbf{V}^{-1}}^2$ as $\|(\mathbf{Y} - \mathbf{P}_{\mathbf{V}} \mathbf{Y}) + (\mathbf{P}_{\mathbf{V}} \mathbf{Y} - \mathbf{X}\boldsymbol{\beta})\|_{\mathbf{V}^{-1}}^2$ 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 $\boldsymbol{\mu}$. (again assume $\mathbf{X} \in \mathbb{R}^{n \times p}$ is rank $p < n$ and $\mathbf{V} > 0$.)
6. Corollary from class (assume that \mathbf{X} is of rank $p < n$ as before and matrices \mathbf{V} , $\boldsymbol{\Phi}$, $\boldsymbol{\Psi}$ are all square, positive definite and of the appropriate dimensions.)
 - (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 $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$. Show that the GLS estimator $\hat{\boldsymbol{\beta}}_{\mathbf{V}}$ is equal to the LUE given by

$$\hat{\boldsymbol{\beta}}_{\boldsymbol{\Phi}} = (\mathbf{X}^T \boldsymbol{\Phi}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Phi}^{-1} \mathbf{Y}$$

iff $\text{Cov}[\epsilon] = \sigma^2 \mathbf{V}$, where

$$\mathbf{V} \equiv \mathbf{X}\Psi\mathbf{X}^T + \Phi\mathbf{H}\Omega\mathbf{H}^T\Phi$$

for some \mathbf{H} such that $\mathbf{H}^T\mathbf{X} = \mathbf{0}$ and for some positive definite matrices Ψ , Φ and Ω . Hint: transform the data by $\Phi^{-1/2}$ and reduce to the result of the theorem.

- (c) Show that that $\mathbb{E}[\mathbf{P}_{\mathbf{V}}\epsilon\epsilon^T(\mathbf{I}_n - \mathbf{P}_{\mathbf{V}})^T] = \mathbf{0}$ always for $\text{Cov}[\epsilon] = \sigma^2\mathbf{V}$
- (d) Show that that $\mathbb{E}[\mathbf{P}_{\Phi}\epsilon\epsilon^T(\mathbf{I}_n - \mathbf{P}_{\Phi})^T] = \mathbf{0}$ under the conditions of the theorem/corollary.
- (e) Show that if $\text{Cov}[\epsilon] = \sigma^2\mathbf{V}$ that $\mathbb{E}[\mathbf{P}_{\mathbf{I}_n}\epsilon\epsilon^T(\mathbf{I}_n - \mathbf{P}_{\mathbf{I}_n}^T)] \neq \mathbf{0}$ for general \mathbf{V} , for $\mathbf{P}_{\mathbf{I}_n} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$. Suggestion: try showing for \mathbf{V} diagonal, then consider what happens using the spectral decomposition for a general \mathbf{V} with "whitening".