STOR 767 Spring 2019 Hw1: Theoretical Part

Due on 01/23/2019 in Class

Remark. This homework aims to help you review the necessary preliminary material from linear regression.

Instruction.

- Homework 1 includes **Theoretical Part** (50%) and **Computational Part** (50%).
- Submission of handwritten homework for **Theoretical Part** is allowed.

1. (*20 pt*) Suppose you are given data $\{(\boldsymbol{X}_i, Y_i)\}_{i=1}^n$, where $\boldsymbol{X}_i \in \mathbb{R}^p$ is a p-dimensional covariate vector and $Y_i \in \mathbb{R}$ is the associated response. The linear regression model is as follows

$$Y_i = \boldsymbol{X}_i^T \boldsymbol{\beta} + \epsilon_i \quad (1 \leqslant i \leqslant n),$$

where $\boldsymbol{\beta} \in \mathbb{R}^p$ is an unknown regression coefficient vector and $\{\epsilon_i\}_{i=1}^n$ are (unobservable) uncorrelated random errors with mean 0 and unknown variance parameter $\sigma^2 > 0$.

(i) Write down the matrix form of the regression model.

(ii) What's the least square estimate (LSE) of $\boldsymbol{\beta}$ and σ^2? Solve the optimization problem, and derive the closed-form solution $\left(\hat{\boldsymbol{\beta}}_{\mathrm{LS}}, \hat{\sigma}^2_{\mathrm{LS}}\right)$.

(iii) Write down the "hat matrix" $\mathbf{H}$ and provide its interpretations in terms of covariates $\{\boldsymbol{X}_i\}_{i=1}^n$ and in terms of responses $\{Y_i\}_{i=1}^n$ respectively.

(iv) Compute $\mathbb{E}\left(\hat{\boldsymbol{\beta}}_{\mathrm{LS}}\right)$, $\mathbb{E}\left(\hat{\sigma}^2_{\mathrm{LS}}\right)$ and $\mathbf{Cov}\left(\hat{\boldsymbol{\beta}}_{\mathrm{LS}}\right)$.

(v) Now assume that $\{\epsilon_i\}_{i=1}^n \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$. Derive the maximum likelihood estimate (MLE) $\left(\hat{\boldsymbol{\beta}}_{\mathrm{ML}}, \hat{\sigma}^2_{\mathrm{ML}}\right)$, and prove that $\hat{\boldsymbol{\beta}}_{\mathrm{ML}} \perp\!\!\!\perp \hat{\sigma}^2_{\mathrm{ML}}$.

(vi) What's the difference and connection between $\left(\hat{\boldsymbol{\beta}}_{\mathrm{LS}}, \hat{\sigma}^2_{\mathrm{LS}}\right)$ and $\left(\hat{\boldsymbol{\beta}}_{\mathrm{ML}}, \hat{\sigma}^2_{\mathrm{ML}}\right)$?

(vii) Let $(\boldsymbol{X}, Y)$ be another independent sample to be predicted, that is,

$$(\boldsymbol{X}, Y) \perp\!\!\!\perp \{(\boldsymbol{X}_i, Y_i)\}_{i=1}^n.$$

Compute the prediction mean square error (MSE) of $\hat{\boldsymbol{\beta}}_{\mathrm{LS}}$

$$\mathbf{MSE}(\hat{Y}_{\mathrm{LS}}) = \mathbb{E}\left(Y - \hat{Y}_{\mathrm{LS}}\right)^2$$

where

$$\hat{Y}_{\mathrm{LS}} = \boldsymbol{X}^T \hat{\boldsymbol{\beta}}_{\mathrm{LS}}.$$

Hint. Decompose $\mathbf{MSE}(\hat{Y}_{\text{LS}})$ into bias of $\hat{Y}_{\text{LS}}$, variance of $\hat{Y}_{\text{LS}}$, and variance of exogenous noise from Y. Hat matrix will be useful in simplifying algebras.

2. (*15 pt*) Suppose in the same setup as in Problem 1,

$$\boldsymbol{X}_i = \begin{pmatrix} \boldsymbol{X}_{i1} \\ \boldsymbol{X}_{i2} \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix}, \quad Y_i = \boldsymbol{X}_{i1}^T \boldsymbol{\beta}_1 + \boldsymbol{X}_{i2}^T \boldsymbol{\beta}_2 + \epsilon_i,$$

where $\boldsymbol{X}_{i1}, \boldsymbol{\beta}_1 \in \mathbb{R}^{p_1}$, $\boldsymbol{X}_{i2}, \boldsymbol{\beta}_2 \in \mathbb{R}^{p_2}$ with $p_1 + p_2 = p$. Let $\hat{Y}_{\text{LS}}^{(1)}$ be predicted outcome of Y when performing regression of $\{Y_i\}_{i=1}^n$ on $\{\boldsymbol{X}_{i1}\}_{i=1}^n$ only, and $\hat{Y}_{\text{LS}}^{(1,2)}$ be the prediction when performing regression on the full/saturated covariate vectors $\{\boldsymbol{X}_i\}_{i=1}^n$.

(i) If data are generated from a linear model only involving $\boldsymbol{X}_{i1}$, that is, $\boldsymbol{\beta}_2 = \mathbf{0}$, compare $\mathbf{MSE}\left(\hat{Y}_{\text{LS}}^{(1)}\right)$ and $\mathbf{MSE}\left(\hat{Y}_{\text{LS}}^{(1,2)}\right)$.

(ii) If data are generated from a full/saturated linear model, that is, $\boldsymbol{\beta}_2 \neq 0$, compare $\mathbf{MSE}\left(\hat{Y}_{\text{LS}}^{(1)}\right)$ and $\mathbf{MSE}\left(\hat{Y}_{\text{LS}}^{(1,2)}\right)$. In particular, provide the condition under which

$$\mathbf{MSE}\left(\hat{Y}_{\text{LS}}^{(1,2)}\right) \leqslant \mathbf{MSE}\left(\hat{Y}_{\text{LS}}^{(1)}\right).$$

3. (*15 pt*) Matrix Analysis Exercises

(i) Let $\boldsymbol{x} \in R^p$ be a random vector with known mean $E(\boldsymbol{x}) = \boldsymbol{\mu}$ and known covariance $cov(\boldsymbol{x}) = \boldsymbol{\Sigma}$, where $\boldsymbol{\Sigma}$ is positive definite. Find an affine transformation of $\boldsymbol{x}$ such that the transformed variable, $\tilde{\boldsymbol{x}}$, satisfies that $E(\tilde{\boldsymbol{x}}) = \mathbf{0}$ and $cov(\tilde{\boldsymbol{x}}) = \mathbf{I}$.

(ii) Let $\boldsymbol{x}, \boldsymbol{y} \in R^n$. Find the eigenvalues and eigenvectors of $\boldsymbol{x}\boldsymbol{y}^T$.

(iii) Let $\mathbf{A} = \{a_{ij}\}$ be a real symmetric $p \times p$ matrix with eigenvalues $\lambda_1, \ldots, \lambda_p$. Show that $\sum_{i=1}^p \sum_{j=1}^p a_{ij}^2 = \sum_{i=1}^p \lambda_i^2$.