STOR 565 Spring 2018 Homework 2

Due on 01/31/2018 in Class

Remark. This homework aims to help you go through the necessary preliminary from linear regression. Credits for **Theoretical Part** and **Computational Part** are in total 100 pt. For **Theoretical Part**, you can summit a hand-writing homework.

Theoretical Part

Note: Problem 1 is optional for extra credits. At most 100 pt can be earned in Homework 2.

1. (*20 pt, Optional*) Suppose you are given sample $\{(\boldsymbol{X}_i, Y_i)\}_{i=1}^n$ 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 setup as

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

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.

Hint. Stack $\{Y_i\}$ and $\{\epsilon_i\}$ as n-dimensional vectors $\boldsymbol{Y}$ and $\boldsymbol{\epsilon}$, and $\{\boldsymbol{X}_i\}$ as an $n \times p$ matrix $\mathbf{X}$.

(ii) Describe the least square estimate (LSE) of $\boldsymbol{\beta}$ and σ^2. Then derive the closed-form solution $\hat{\boldsymbol{\beta}}_{LS}$ and $\hat{\sigma}^2_{LS}$.

Hint. For $\hat{\boldsymbol{\beta}}_{LS}$, write down an optimization problem associated with the least-square procedure in matrix form, and then take multivariate derivative to obtain the first-order conditions. For $\hat{\sigma}^2_{LS}$, remember to use an unbiased estimate of σ^2.

(iii) What's the "hat matrix" $\mathbf{H}$? Derive it and provide an interpretation.

Hint. Consider the interpretations of $\mathbf{H}\boldsymbol{Y}$ in terms of $\mathbf{X}$ and $\boldsymbol{Y}$ respectively.

(iv) Derive $\mathbb{E}\left(\hat{\boldsymbol{\beta}}_{LS}\right)$ and $\mathbf{Cov}\left(\hat{\boldsymbol{\beta}}_{LS}\right)$.

Hint. Randomness comes from $\{\epsilon_i\}$. You can represent the above quantities in $\boldsymbol{\epsilon}$ and then calculate it.

(v) Now assume that $\{\epsilon_i\} \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$. Derive the maximum likelihood estimate (MLE) $\hat{\boldsymbol{\beta}}_{ML}$ and $\hat{\sigma}^2_{ML}$.

Hint. Write down the likelihood function of $\epsilon_i = Y_i - \boldsymbol{X}_i^T \boldsymbol{\beta}$ and then the joint likelihood. Solve the maximization problem with respect to the unknown parameters $\boldsymbol{\beta}$ and σ^2. It might help with calculation to take log of your joint likelihood function and write in matrix form.

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

2. (*30 pt*, Textbook[1] Exercises 3.4) I collect a set of data ($n = 100$ observations) containing a single predictor and a quantitative response. I then fit a linear regression model to the data, as well as a separate cubic regression, i.e. $Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + \epsilon$.

   (a) Suppose that the true relationship between X and Y is linear, i.e. $Y = \beta_0 + \beta_1 X + \epsilon$. Consider the training residual sum of squares (RSS) for the linear regression (noticing that $MSE = \frac{SSE}{n}$ in this case), and also the training RSS for the cubic regression. Would we expect one to be lower than the other, would we expect them to be the same, or is there not enough information to tell? Justify your answer.

   (b) Answer (a) using test rather than training RSS.

   (c) Suppose that the true relationship between X and Y is not linear, but we don't know how far it is from linear. Consider the training RSS for the linear regression, and also the training RSS for the cubic regression. Would we expect one to be lower than the other, would we expect them to be the same, or is there not enough information to tell? Justify your answer.

   (d) Answer (c) using test rather than training RSS.

References

[1] Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An introduction to statistical learning*, volume 112. Springer, 2013. 2