STOR 565 Spring 2018 Homework 3

Due on 02/09/2018 in Class

Remark. This homework aims to help you further understand the model selection techniques in linear model. 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 3 (d) is optional for extra credits. At most 100 pt can be earned in Hw3.

1. (Textbook 6.3 + 6.4, *20 pt*) Consider the ridge regression

$$\min_{\boldsymbol{\beta}} \|\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 + \lambda \sum_{k=1}^{p} \beta_k^2 \qquad \text{(ridge regression)}$$

with tuning parameter $\lambda \geqslant 0$ and the constrained version of LASSO

$$\begin{aligned} \min_{\boldsymbol{\beta}} \quad & \|\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 \\ \text{s.t.} \quad & \sum_{k=1}^{p} |\beta_k| \leqslant s \end{aligned} \qquad \text{(LASSO)}$$

with tuning parameter $s \geqslant 0$. As λ and s increases from 0 respectively, indicate which of (i) to (v) is correct for: (a) training RSS, (b) test RSS, (c) variance, and (d) (squared) bias. Justify your answer.

   (i) Increase initially, and then eventually start decreasing in an inverted U shape.

   (ii) Decrease initially, and then eventually start increasing in a U shape.

   (iii) Steadily increase.

   (iv) Steadily decrease.

   (v) Remain constant.

2. (*25 pt*) This problem illustrates the estimator property in the shrinkage methods. Let Y be a single observation. Consider Y regressed on an intercept

$$Y = 1 \cdot \beta + \epsilon.$$

   (a) Using the formulation as shown in class, write down the optimization problem of general linear model, ridge regression and LASSO in estimating β respectively.

   (b) For fixed tuning parameter λ, solve for $\hat{\beta}$ (general linear model), $\hat{\beta}_\lambda^R$ (ridge regression) and $\hat{\beta}_\lambda^L$ (LASSO) respectively.

Hint. For the LASSO problem, show that the objective function is convex even though not everywhere differentiable, hence any local minima is also a global minima. Argue that it suffices to solve for $\beta \geqslant 0$ and $\beta < 0$ respectively. Carefully discuss all possible situations (where the minima locates and what's the optimal value in both cases), and then pick the better one as $\hat{\beta}^L_\lambda$.

(c) Represent $\hat{\beta}^R_\lambda$ and $\hat{\beta}^L_\lambda$ by $\hat{\beta}$ and create corresponding plots respectively with $\lambda = 1, 5, 10$. What can you tell?

3. (Textbook 6.5, *5 pt + 10 optional pt*) It is well-known that ridge regression tends to give similar coefficient values to correlated/collinear variables, whereas the LASSO may give quite different coefficient values to correlated/collinear variables. We will now explore this property in a very simple setting. Suppose that we have two observations (X_1, Y_1) and (X_2, Y_2), where $X_1 \neq 0$, $X_1 + X_2 = Y_1 + Y_2 = 0$. Consider the linear model Y_i artificially regressed on (X_i, X_i) without intercept:

$$\begin{cases} Y_1 = X_1\beta_1 + X_1\beta_2 + \epsilon_1 \\ Y_2 = X_2\beta_1 + X_2\beta_2 + \epsilon_2 \end{cases}$$

(a) Write out the ridge regression optimization problem in this setting.

(b) Argue that the ridge coefficient estimates satisfy $\hat{\beta}^R_{\lambda,1} = \hat{\beta}^R_{\lambda,2}$.

(c) Write out the LASSO optimization problem in this setting.

(d) (*Optional, 10 pt*) Argue that in this setting, the LASSO coefficients $\hat{\beta}^L_{\lambda,1}$ and $\hat{\beta}^L_{\lambda,2}$ are not unique. Describe these solutions.

Hint. Starting with an optimal coefficients $(\hat{\beta}^L_{\lambda,1}, \hat{\beta}^L_{\lambda,2})$, indicate that you can find another one. Use the relationship of a usual LASSO problem and its constrained version. Investigate into the relationship between the contour of the objective function and the constraint set in the constrained version.