Econometrics and Machine Learning: Assignment 1

Jui-Chung Yang

Due Date: March 23, 2023

1. A *bridge estimator* is a penalized least squares estimator with a penalty function as the sum of a L_γ norm to the power of γ, i.e.,

$$\hat{\beta} = \arg\min_{\beta} \left[(\mathbf{y} - \mathbf{X}\beta)^\top (\mathbf{y} - \mathbf{X}\beta) + \lambda \|\beta\|_\gamma^\gamma \right].$$

When $\gamma = 1$, the estimator is known as the *least absolute shrinkage and selection operator* (*LASSO*). When $\gamma = 2$, the estimator is known as the *ridge* estimator. Establish the closed form solution of the ridge estimator.

2. Let's conduct an Monte Carlo experiment. Consider the following two data generating process (DGPs):

$$\begin{aligned} &\text{DGP 1: } y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i, \\ &\text{DGP 2: } y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \cdots + \beta_{20} x_{20i} + u_i. \end{aligned}$$

Where in DGP 1, $\beta_0 = \beta_1 = \beta_2 = 1$, and in DGP 2, $\beta_0 = \beta_1 = \beta_2 = \cdots = \beta_2 0 = 1$. $x_{1i}, x_{2i}, \ldots, x_{50i}$ and u_i are independent, and identically distributed standard normal random variable ($\mathcal{N}(0, 1)$.) The sample size is 500. However, we only use the first 400 observations as the *training set*, and the remaining 100 observations are the *testing set*. For each DGP, we use the training set and estimate two models with 25 and 50 regressors, i.e.,

$$\begin{aligned} &\text{Model 1: } y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \cdots + \beta_{25} x_{25i} + u_i, \\ &\text{Model 2: } y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \cdots + \beta_{50} x_{50i} + u_i, \end{aligned}$$

using the *ordinary least squares*, *ridge* and *LASSO*, where the tuning parameter is selected using the 10-fold cross-validation. The number of replications is 2000.

a. Report the average, standard deviation, and root mean squared errors of your estimates for β_1. That is, when your data is generated according to DGP j, and you are estimating Model k using method ℓ, you should have 2000 estimates for β_1: $\{\hat{\beta}_1^{j,k,\ell,(1)}, \hat{\beta}_1^{j,k,\ell,(2)}, \ldots, \hat{\beta}_1^{j,k,\ell,(2000)}\}$. Report

$$\text{Mean}_{\hat{\beta}_1}^{j,k,\ell} = \frac{1}{2000} \sum_{r=1}^{2000} \hat{\beta}_1^{j,k,\ell,(r)}, \ \text{RMSE}_{\hat{\beta}_1}^{j,k,\ell} = \sqrt{\frac{1}{2000} \sum_{r=1}^{2000} \left(\hat{\beta}_1^{j,k,\ell,(r)} - \beta_1 \right)^2},$$

$$\text{and SD}_{\hat{\beta}_1}^{j,k,\ell} = \sqrt{\frac{1}{2000} \sum_{r=1}^{2000} \left(\hat{\beta}_1^{j,k,\ell,(r)} - \text{Mean}_{\hat{\beta}_1}^{j,k,\ell} \right)^2}.$$

Note that the true value of $\beta_1 = 1$ in both DGPs.

b. Report the average, standard deviation, and root mean squared errors of your estimates for β_{21}. Note that the true value of $\beta_{21} = 0$ in both DGPs.

c. Report the mean squared prediction error of your predictions on the testing set. That is, when your data is generated according to DGP j, and you are estimating Model k using method ℓ, you should have 2000 sets of prediction for $\mathbf{y}$: $\{\hat{\mathbf{y}}^{j,k,\ell,(1)}, \hat{\mathbf{y}}^{j,k,\ell,(2)}, \ldots, \hat{\mathbf{y}}^{j,k,\ell,(2000)}\}$ where $\mathbf{y}^{(r)} = \left[y_1^{(r)}, y_2^{(r)}, \ldots, y_{100}^{(r)}\right]^\top$ are the observations in the testing set in the rth replication. Report

$$\mathrm{MSPE}^{j,k,\ell} = \frac{1}{2000}\sum_{r=1}^{2000}\left[\frac{1}{100}\left(\hat{\mathbf{y}}^{j,k,\ell,(r)} - \mathbf{y}^{(r)}\right)^\top \left(\hat{\mathbf{y}}^{j,k,\ell,(r)} - \mathbf{y}^{(r)}\right)\right].$$