

HW3

STAT-340 Applied Regression Methods

Yerim Oh

Details

Due Date

This assignment is due at 11:59 PM on November 13th.

Grading

All of the problems will be graded for correctness. In grading these problems, an emphasis will be placed on full explanations of your thought process. You usually won't need to write more than a few sentences for any given problem, but you should write complete sentences! Understanding and explaining the reasons behind your decisions is more important than making the "correct" decision.

Collaboration

You are allowed to work with others on this assignment, but you must complete and submit your own write up. You should not copy large blocks of code or written text from another student.

Conceptual Problems

Problem 1: Adapted from ISLR Exercise 6.2

For parts (a) and (b), indicate which of i. through iv. is correct. **Justify your answer.**

(a) The lasso, relative to least squares, is:

- i. More flexible and hence will give improved prediction accuracy when its increase in bias is less than its decrease in variance.
 - ii. More flexible and hence will give improved prediction accuracy when its increase in variance is less than its decrease in bias.
 - iii. Less flexible and hence will give improved prediction accuracy when its increase in bias is less than its decrease in variance.
 - iv. Less flexible and hence will give improved prediction accuracy when its increase in variance is less than its decrease in bias.
- (iii) is correct
 - By performing shrinkage, the lasso can reduce the number of coefficients in the model. This penalty can result in some coefficients being shrunk nearly to zero, for performing an effective variable selection. Since the lasso simplifies the model, it is less flexible compared to the least squares regression.
 - By reducing flexibility, the lasso increases bias but decrease variance because the model becomes simpler and may not capture all the data patterns, and less sensitive to noise in the data. Since the lasso is making the model less complex, the tradeoff between bias and variance is optimal when

the increase in bias is less than the decrease in variance according to the bias-variance tradeoff. This will give improved prediction accuracy.

(b) Repeat (a) for ridge regression relative to least squares.

- i. More flexible and hence will give improved prediction accuracy when its increase in bias is less than its decrease in variance.
 - ii. More flexible and hence will give improved prediction accuracy when its increase in variance is less than its decrease in bias.
 - iii. Less flexible and hence will give improved prediction accuracy when its increase in bias is less than its decrease in variance.
 - iv. Less flexible and hence will give improved prediction accuracy when its increase in variance is less than its decrease in bias.
- (iii) is correct
 - Like the lasso, the ridge regression adds a penalty to the model's coefficients (but uses a different method for regularization: squared magnitude of the coefficients). This reduces the flexibility of the model by shrinking the coefficients less likely to overfitting. Thus, the ridge regression is less flexible compared to the least squares regression.
 - Same with the lasso, the ridge regression increases the bias by shrinking the coefficients and decreases the variance. Therefore, improved prediction accuracy occurs when its increase in bias is less than its decrease in variance.

Problem 2: Adapted from ISLR Exercise 6.5

Ridge regression tends to give similar coefficient values to correlated variables. We will now explore this property in a very simple setting. Suppose that:

- $n = 2$
- $p = 2$
- $x_{11} = x_{12}$, and $x_{21} = x_{22}$
- $x_{11} + x_{21} = 0$ and $x_{12} + x_{22} = 0$
- $y_1 + y_2 = 0$

(a) Show that the least square estimate for the intercept is zero under those conditions, that is, $\hat{\beta}_0^{LS} = 0$.

The least square estimate is to find β 's that minimizes RSS

$$\begin{aligned} RSS &= \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 \\ \sum_{i=1}^2 \left(y_i - \beta_0 - \sum_{j=1}^2 \beta_j x_{ij} \right)^2 &= (y_1 - \beta_0 - \sum_{j=1}^2 \beta_j x_{1j})^2 + (y_2 - \beta_0 - \sum_{j=1}^2 \beta_j x_{2j})^2 \\ &= y_1 - \beta_0 - (\beta_1 x_{11} + \beta_2 x_{12}) + y_2 - \beta_0 - (\beta_1 x_{21} + \beta_2 x_{22}) \\ &= (y_1 + y_2) - 2\beta_0 - (\beta_1(x_{11} + x_{21}) + \beta_2(x_{12} + x_{22})) \\ &= 0 - 2\beta_0 - (\beta_1 \cdot 0 + \beta_2 \cdot 0) \\ &= -2\beta_0 \end{aligned}$$

Since we are looking for the least square estimate for β_0 that minimizes RSS , the minimum value of $-2\beta_0$ is 0.

$$-2\beta_0 = 0 \rightarrow \beta_0 = 0$$

Therefore, $\hat{\beta}_0^{LS} = 0$.

(b) Write out the ridge regression optimization problem under those conditions, that is, adapt equation on slide 30:

$$\sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^p \beta_j^2$$

for the values/assumptions stated in this problem (n , p , x_{ij} , and y_i). You may also use that $\hat{\beta}_0^R = 0$.

Since we can consider $\beta_0 = 0$, $x_1 = x_{11} = x_{12}$, and $x_2 = x_{21} = x_{22}$,

$$\begin{aligned}
& \sum_{i=1}^2 \left(y_i - \sum_{j=1}^2 \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^2 \beta_j^2 \\
&= (y_1 - \sum_{j=1}^2 \beta_j x_{1j})^2 + (y_2 - \sum_{j=1}^2 \beta_j x_{2j})^2 + \lambda \sum_{j=1}^2 \beta_j^2 \\
&= (y_1 - \beta_1 x_{11} - \beta_2 x_{12})^2 + (y_2 - \beta_1 x_{21} - \beta_2 x_{22})^2 + \lambda(\beta_1^2 + \beta_2^2) \\
&= (y_1^2 - y_1 \beta_1 x_1 - y_1 \beta_2 x_1 - y_1 \beta_1 x_1 + \beta_1^2 x_1^2 + \beta_1 \beta_2 x_1^2 - y_1 \beta_2 x_1 + \beta_1 \beta_2 x_1^2 + \beta_2^2 x_1^2) \\
&+ (y_2^2 - y_2 \beta_1 x_2 - y_2 \beta_2 x_2 - y_2 \beta_1 x_2 + \beta_1^2 x_2^2 + \beta_1 \beta_2 x_2^2 - y_2 \beta_2 x_2 + \beta_1 \beta_2 x_2^2 + \beta_2^2 x_2^2) \\
&+ \lambda \beta_1^2 + \lambda \beta_2^2 \\
&= y_1^2 - 2y_1 \beta_1 x_1 - 2y_1 \beta_2 x_1 + \beta_1^2 x_1^2 + 2\beta_1 \beta_2 x_1^2 + \beta_2^2 x_1^2 \\
&+ y_2^2 - 2y_2 \beta_1 x_2 - 2y_2 \beta_2 x_2 + \beta_1^2 x_2^2 + 2\beta_1 \beta_2 x_2^2 + \beta_2^2 x_2^2 \\
&+ \lambda \beta_1^2 + \lambda \beta_2^2 \\
&= y_1^2 - 2y_1(\beta_1 + \beta_2)x_1 + (\beta_1 + \beta_2)^2 x_1^2 + y_2^2 - 2y_2(\beta_1 + \beta_2)x_2 + (\beta_1 + \beta_2)^2 x_2^2 + \lambda(\beta_1^2 + \beta_2^2)
\end{aligned}$$

(c) Prove that under those conditions, the ridge coefficient estimates satisfy $\hat{\beta}_1^R = \hat{\beta}_2^R$.

Since we are trying to minimize $\sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^p \beta_j^2$, we can use partial differentiation the minimum of β_1 and β_2 .

- the estimate of β_1 :

$$\frac{\partial}{\partial \beta_1} = -2y_1 x_1 + 2\beta_1 x_1^2 + 2\beta_2 x_1^2 - 2y_2 x_2 + 2\beta_1 x_2^2 + 2\beta_2 x_2^2 + 2\lambda \beta_1$$

To find β_1 that minimizes, the derivative should be equal to 0.

$$\begin{aligned}
0 &= -2y_1 x_1 + 2\beta_1 x_1^2 + 2\beta_2 x_1^2 - 2y_2 x_2 + 2\beta_1 x_2^2 + 2\beta_2 x_2^2 + 2\lambda \beta_1 \\
0 &= -y_1 x_1 + \beta_1 x_1^2 + \beta_2 x_1^2 - y_2 x_2 + \beta_1 x_2^2 + \beta_2 x_2^2 + \lambda \beta_1
\end{aligned}$$

$$\begin{aligned}
\lambda \hat{\beta}_1^R &= y_1 x_1 - \beta_1 x_1^2 - \beta_2 x_1^2 + y_2 x_2 - \beta_1 x_2^2 - \beta_2 x_2^2 \\
&= y_1 x_1 + y_2 x_2 - \beta_1 (x_1^2 + x_2^2) - \beta_2 (x_1^2 + x_2^2) \\
&= y_1 x_1 + y_2 x_2 - (\beta_1 + \beta_2)(x_1^2 + x_2^2)
\end{aligned}$$

- the estimate of β_2 :

$$\frac{\partial}{\partial \beta_2} = -2y_1 x_1 + 2\beta_1 x_1^2 + 2\beta_2 x_1^2 - 2y_2 x_2 + 2\beta_1 x_2^2 + 2\beta_2 x_2^2 + 2\lambda \beta_2$$

To find β_2 that minimizes, the derivative should be equal to 0.

$$0 = -2y_1x_1 + 2\beta_1x_1^2 + 2\beta_2x_1^2 - 2y_2x_2 + 2\beta_1x_2^2 + 2\beta_2x_2^2 + 2\lambda\beta_2$$

$$0 = -y_1x_1 + \beta_1x_1^2 + \beta_2x_1^2 - y_2x_2 + \beta_1x_2^2 + \beta_2x_2^2 + \lambda\beta_2$$

$$\begin{aligned}\lambda\hat{\beta}_2^R &= y_1x_1 - \beta_1x_1^2 - \beta_2x_1^2 + y_2x_2 - \beta_1x_2^2 - \beta_2x_2^2 \\ &= y_1x_1 + y_2x_2 - \beta_1(x_1^2 + x_2^2) - \beta_2(x_1^2 + x_2^2) \\ &= y_1x_1 + y_2x_2 - (\beta_1 + \beta_2)(x_1^2 + x_2^2)\end{aligned}$$

Since $\lambda\hat{\beta}_1^R = \lambda\hat{\beta}_2^R$, $\hat{\beta}_1^R$ and $\hat{\beta}_2^R$ are also equal.