STA 141C: Homework 3

Due 5/9/2025, 8PM

This homework assignment has to be submitted electronically on Gradescope by the due date. Please submit one PDF file containing all your answers and the code used for your data analysis. It is strongly recommended to type your answers rather than submitting handwritten work. If handwritten, please ensure that it is legible and neat.

Problem 1. (7/25 points)
(Part I) We perform best subset, forward stepwise, and backward stepwise selection on a single data set. For each approach, we obtain $p+1$ models, containing $0, 1, 2, \ldots, p$ predictors. Explain your answers:

(a) Which of the three selection methods with k predictors has the smallest training RSS?

(b) Which of the three selection methods with k predictors has the smallest test RSS?

(c) True or False: the predictors in Model 1 are a subset of the predictors in Model 2:

	Model 1	Model 2	T/F
i.	Forward selection, k variables	Forward selection, $k+1$ variables	
ii.	Backward selection, k variables	Backward selection, $k+1$ variables	
iii.	Backward selection, k variables	Forward selection, $k+1$ variables	
iv.	Forward selection, k variables	Backward selection, $k+1$ variables	
v.	Best subset selection, k variables	Best subset selection, $k+1$ variables	

(Part II) Indicate which of i. through iv. below 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.

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

Problem 2. (4/25 points) Suppose we estimate the regression coefficients in a linear regression model by minimizing

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

subject to $\sum_{j=1}^{p}|\beta_j| \leq s$ for a particular value of s. For parts (a) through (d), indicate which of i. through v. is correct. Justify your answer.

(a) As we increase s from 0, the training RSS will:

  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.

(b) Repeat (a) for test RSS.

(c) Repeat (a) for variance.

(d) Repeat (a) for (squared) bias.

Problem 3. (6/25 points)
It is well-known that ridge regression tends to give similar coefficient values to correlated variables, whereas the lasso may give quite different coefficient values to correlated variables. We will now explore this property in a very simple setting.

Suppose that $n = 2, p = 2, x_{11} = x_{12}, x_{21} = x_{22}$. Furthermore, suppose that $y_1 + y_2 = 0$ and $x_{11} + x_{21} = 0$ and $x_{12} + x_{22} = 0$, so that the estimate for the intercept in a least squares, ridge regression, or lasso model is zero: $\hat{\beta}_0 = 0$.

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

(b) Argue that in this setting, the ridge coefficient estimates satisfy $\hat{\beta}_1 = \hat{\beta}_2$.

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

(d) Argue that in this setting, the lasso coefficients $\hat{\beta}_1$ and $\hat{\beta}_2$ are not unique – in other words, there are many possible solutions to the optimization problem in (c). Describe these solutions.

Problem 4. (3/25 points)
In this exercise, we will predict the number of applications received using the other variables in the `College` data set in R or Python. In R, the data are in the ISLR2 package. In Python, you can download the data here.

(a) Split the data set into a training set and a test set.

(b) Fit a linear model using least squares on the training set, and report the test error obtained.

(c) Fit a ridge regression model on the training set, with λ chosen by cross-validation. Report the test error obtained.

(d) Fit a lasso model on the training set, with λ chosen by cross-validation. Report the test error obtained, along with the number of non-zero coefficient estimates.

Problem 5. (5/25 points)
Suppose that a curve $\hat{g}$ is computed to smoothly fit a set of n points using the following formula:

$$\hat{g} = \arg\min_{g} \left(\sum_{i=1}^{n} (y_i - g(x_i))^2 + \lambda \int \left[g^{(m)}(x) \right]^2 dx \right),$$

where $g^{(m)}$ represents the mth derivative of g (and $g^{(0)} = g$). Provide example sketches of $\hat{g}$ in each of the following scenarios.

(a) $\lambda = \infty, m = 0$.

(b) $\lambda = \infty, m = 1$.

(c) $\lambda = \infty, m = 2$.

(d) $\lambda = \infty, m = 3$.

(e) $\lambda = 0, m = 3$.