MATH560 HW6

Jon Griffith and Annabelle Cunningham

April 21, 2025

2. For parts (a) through (c), 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.

   Answer: (iii) Lasso is less flexible than OLS since it sets a restriction on the predictor variables that can only result in their minimization and/or elimination. It therefore retains the same number of fewer variables than OLS, meaning fewer degrees of freedom to explain variance in predictions. When proper cross-validation is performed and test MSE is minimized, an increase in bias less than a decrease in variance will result in variance less than OLS. If an increase in bias is more than a decrease in variance, the CV results would likely suggest using an OLS model.

b. Repeat (a) for ridge regression relative to least squares.
   Answer: (iii) Similar to the above reasoning, since the predictor variables are subject to a constraint that can only result in their minimization, it provides less flexibility to these predictors in predictions compared to OLS. When a decrease in variance outweighs an increase in bias, the CV results will favor a ridge regression model with lower variance than OLS resulting in more accurate predictions.

c. Repeat (a) for non-linear methods relative to least squares.
   Answer: (ii) Non-linear methods are more flexible than OLS because they allow transformations of the predictor variables which add additional

degrees of freedom to the model that can be used to explain additional variation in predictions. This is subject to a proper CV that shows a decrease in bias that outweighs the corresponding increase in variance. That is, the benefit of bias reduction outweighs the cost of variance increase.

3. Suppose we estimate the regression coefficients in a linear regression model by minimizing

$$\sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij}\right)^2 \quad \text{subject to} \quad \sum_{j=1}^{p}|\beta_j| \leq s$$

for a particular value of s. For parts (a) through (e), 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.

Answer: (iv) Since a value of $s = 0$ indicates the null model, as s increases, the model becomes more flexible and so the training RSS will steadily decrease as more variation in predictions can be explained by the predictor variables and their coefficients.

(b) Repeat (a) for test RSS.

Answer: (ii) As the model becomes more flexible, the test RSS will decrease with the training RSS, but at a certain point, the model will start overfitting and so test RSS will increase.

(c) Repeat (a) for variance.

Answer: (iii) With $s = 0$, the null model has no variance in its predictions but as s increases, the variation in predictions will also increase due to more flexibility in predictors.

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

Answer: (iv) With $s = 0$, the null model has very high bias in its predictions but as s increases and the model becomes more flexible, the squared bias will steadily decrease.

(e) Repeat (a) for the irreducible error.

Answer: (v) The irreducible error is unaffected by the model no matter what and so it remains constant.

4. Suppose we estimate the regression coefficients in a linear regression model by minimizing

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

for a particular value of λ. For parts (a) through (e), indicate which of i. through v. is correct. Justify your answer.

(a) As we increase λ 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.

Answer: (iii) An increasing λ results in less flexibility in the coefficients for the predictors, restraining their impact on predictions. Therefore, we expect a steady increase in training RSS.

(b) Repeat (a) for test RSS.

Answer: (ii) As λ increases, we get a less flexible model, so this initially results in a smaller variation in predictions reflected in a decreasing RSS, but then at some point, the bias will overtake the variance decrease resulting in increasing RSS.

(c) Repeat (a) for variance.

Answer: (iv) As λ increases, the model becomes less flexible and the resulting variance in the predictions will steadily decrease as the model will eventually report close to the mean for all predictions as λ gets larger and larger.

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

Answer: (iii) As λ gets larger and the model gets less flexible, the bias in predictions will steadily increase as the resulting model approaches the null model.

(e) Repeat (a) for the irreducible error.

Answer: (v) The irreducible error is unaffected by the model no matter what and so it remains constant.