Statistical Methods Part B

Balin Lin (1649159)

Exercise 1

Solve exercises 3, 4, 5 and 10 from CHAPTER 6 of the 1st Edition of the book An Introduction to Statistical Learning and Applications, James, Witten, Hastie, Tibshirani.

3.

a. **(Ans: iv)**. As the model gets more flexible the training RSS will steadily decrease.
b. **(Ans: ii)**. Test RSS will initially decrease and then start increasing again as the model gets more flexible and starts to overfit the training data.
c. **(Ans: iii)**. With increased flexibility variance always increases.
d. **(Ans: iv)**. Squared bias steadily decreases with increased flexibility.
e. **(Ans: v)**. The irreducible error is a constant independent of the model.

4.

a. **(Ans: iii)**. Training error will increase steadily with less and less flexibility in the model.
b. **(Ans: ii)**. Test error decreases initially and then starts to increase again as the model gets less and less flexible.
c. **(Ans: iv)**. Variance decreases with less and less flexibility.
d. **(Ans: iii)**. Bias increases with less and less flexibility.
e. **(Ans: v)**. The irreducible error is a constant independent of the model.

5.

a. $(y_1 - \hat{\beta}_1 x_1 - \hat{\beta}_2 x_1)^2 + (y_2 - \hat{\beta}_1 x_2 - \hat{\beta}_2 x_2)^2 + \lambda(\hat{\beta}_1^2 + \hat{\beta}_2^2)$

b. $\frac{\partial}{\partial \hat{\beta}_1}: (2\hat{\beta}_1 x_{11}^2 - 2x_{11}y_1 + 2\hat{\beta}_2 x_{11}x_{12}) + (2\hat{\beta}_1 x_{21}^2 - 2x_{21}y_2 + 2\hat{\beta}_2 x_{21}x_{22}) + 2\lambda\hat{\beta}_1 = 0$

So we can get the following equation which yields $\hat{\beta}_1 = \hat{\beta}_2$

$\lambda\hat{\beta}_1 = x_1 y_1 + x_2 y_2 + 2\hat{\beta}_1 x_1 x_2 + 2\hat{\beta}_2 x_1 x_2$

$\lambda\hat{\beta}_2 = x_1 y_1 + x_2 y_2 + 2\hat{\beta}_1 x_1 x_2 + 2\hat{\beta}_2 x_1 x_2$

c. $(y_1 - \hat{\beta}_1 x_1 - \hat{\beta}_2 x_1)^2 + (y_2 - \hat{\beta}_1 x_2 - \hat{\beta}_2 x_2)^2 + \lambda(|\hat{\beta}_1| + |\hat{\beta}_2|)$

d. We will use the alternate form of the lasso optimization problem

$(y_1 - \hat{\beta}_1 x_1 - \hat{\beta}_2 x_1)^2 + (y_2 - \hat{\beta}\, 1x_2 - \hat{\beta}_2 x_2)2 \ subject\ to\ |\hat{\beta}_1| + |\hat{\beta}_2| \leq s.$

Geometrically the lasso constraint take the form of a diamond centered at the origin of the plane $(\hat{\beta}_1, \hat{\beta}_2)$ which intersects the axes at a distance s from the origin. By using the setting of this problem ($x_{11} = x_{12} = x_1$, $x_{21} = x_{22} = x_2$, $x_1 + x_2 = 0$ and $y_1 + y_2 = 0$), we have to minimize the expression

$2[y_1 - (\hat{\beta}_1 + \hat{\beta}_2)x_1]^2 \geq 0$

This optimization problem has a simple solution : $\hat{\beta}_1 + \hat{\beta}_2 = y_1/x_1$. Geometrically, this is a line parallel to the edge of the diamond of the constraints. Now, solutions to the lasso optimization problem are contours of the function $[y_1 - (\hat{\beta}_1 + \hat{\beta}_2)x_1]^2$ that intersects the diamond of the constraints. So, the entire edge $\hat{\beta}_1 + \hat{\beta}_2 = s$ (as is the edge $\hat{\beta}_1 + \hat{\beta}_2 = -s$) is a potential solution to the lasso optimization problem. Thus, the lasso optimization problem has a whole set of solutions instead of a unique one : $\{(\hat{\beta}_1, \hat{\beta}_2): \hat{\beta}_1 + \hat{\beta}_2 = s\ with\ \hat{\beta}_1, \hat{\beta}_2 \geq 0\ and\ \hat{\beta}_1 + \hat{\beta}_2 = -s\ with\ \hat{\beta}_1, \hat{\beta}_2 \leq 0\}$.

10. **Please check the file “Exercise 1-10.nb.html”.**

Exercise 2

a. **False positives:** 188
   False negatives: 54
   True positives: 192
   True negatives: 1945
b. **False positive rate:** 0.0881387717
   False discovery rate: 0.494736842
c. **False positive rate:** The false positive rate is calculated as the ratio between the number of negative events wrongly categorized as positive (false positives) and the total number of actual negative events (regardless of classification).
   False discovery rate: The false discovery rate is the ratio of the number of false positive results to the number of total positive test results. The FDR is the rate that features called significant are truly null.
d. **Sensitivity:** 0.780487805
   Specificity: 0.911861228

Exercise 3

Exercise 4