HW6

Kamiar Asgari

July 2023

1 2. ISLR 6.6.3

(a) Steadily decreases: As we increase s from 0, all β's increase from 0 to their least square estimate values. Training error for all $\beta = 0$ is the maximum and it steadily decreases to the Ordinary Least Square RSS.

(b) Decrease Initially and then eventually starts increasing in a new shape. The test RSS will decrease to a certain level after which it will hit an inflection point and the test residual sum error will start to increase.

(c) Steadily Increases: β is almost zero in the beginning which would mean that there is almost no variance in the beginning. It is only after the module becomes flexible does the variance start to go up.

(d) Steadily decrease: When $s = 0$, the model effectively predicts a constant and hence the prediction is far from actual value. Thus bias is high. As s increases, more βs become non-zero and thus the model continues to fit training data better. And thus, bias decreases.

(e) Remains constant: By definition, irreducible error is model independent and hence irrespective of the choice of s, remains constant.

2 3. ISLR, 6.6.5

Note that we have

$$X = x_{11} = x_{12} = -x_{21} = -x_{22}, \quad Y = y_2 = -y_1 \tag{1}$$

(a)

$$\min \sum_{i=1}^{n} \left(y_i - \hat{\beta}_0 - \sum_{j=1}^{p} \hat{\beta}_j x_j \right)^2 + \lambda \sum_{i=1}^{p} \hat{\beta}_i^2$$

In this case, $\hat{\beta}_0 = 0$ and $n = p = 2$. So, the optimization looks like:

$$\min \left(y_1 - \hat{\beta}_1 x_{11} - \hat{\beta}_2 x_{12} \right)^2 + \left(y_2 - \hat{\beta}_1 x_{21} - \hat{\beta}_2 x_{22} \right)^2 + \lambda \left(\hat{\beta}_1^2 + \hat{\beta}_2^2 \right)$$

or,

$$\min \left(-Y - \hat{\beta}_1 X - \hat{\beta}_2 X\right)^2 + \left(Y + \hat{\beta}_1 X + \hat{\beta}_2 X\right)^2 + \lambda \left(\hat{\beta}_1^2 + \hat{\beta}_2^2\right)$$

Equivalently:

$$\min \left(Y + \hat{\beta}_1 X + \hat{\beta}_2 X\right)^2 + \left(Y + \hat{\beta}_1 X + \hat{\beta}_2 X\right)^2 + \lambda \left(\hat{\beta}_1^2 + \hat{\beta}_2^2\right)$$

(b) Clearly, the problem is symmetric with respect to $\hat{\beta}_1$ and $\hat{\beta}_2$ which means if $\left(\hat{\beta}_1, \hat{\beta}_2\right) = (a, b)$, then $\left(\hat{\beta}_1, \hat{\beta}_2\right) = (b, a)$ is also an answer. However, because this problem is strongly convex w.r.t. the vector $\left(\hat{\beta}_1, \hat{\beta}_2\right)$ (for a non-zero λ), it has a unique answer. This two results in the fact that there is only one answer and in that answer $\hat{\beta}_1 = \hat{\beta}_2$.

(c) Similarly:

$$\min \left(Y + \hat{\beta}_1 X + \hat{\beta}_2 X\right)^2 + \left(Y + \hat{\beta}_1 X + \hat{\beta}_2 X\right)^2 + \lambda \left(\left|\hat{\beta}_1\right| + \left|\hat{\beta}_2\right|\right)$$

(d) Clearly, this problem is also symmetric with respect to $\hat{\beta}_1$ and $\hat{\beta}_2$ but it is NOT strongly convex w.r.t. the vector $\left(\hat{\beta}_1, \hat{\beta}_2\right)$. However, this problem is strongly convex w.r.t. $B = \hat{\beta}_1 + \hat{\beta}_2$. For example, let $\hat{\beta}_1 \geq 0$ and $\hat{\beta}_2 \geq 0$, then we can write the problem as:

$$\min 2(Y + BX)^2 + \lambda B$$

whose answer is $B = \frac{\lambda + 4YX}{-4X^2}$. If this value is positive then we have a set of answers where almost everywhere $\hat{\beta}_1 \neq \hat{\beta}_2$.

3 4. ISLR 8.4.5

Majority Polling In this, probabilities ≥ 0.5 will be true and the classification will be false otherwise. False = 4 and True = 6 and hence, the red samples are larger than green samples and so the classification will be red class.

Average Probability Mean for given probabilities = 0.45 hence the final prediction is green.

4 5. ISLR 9.7.3

notebook is provided