Computational Statistics II

Homework 5

Subash Kharel

April 23, 2020

1. **(Problem 2 (a –f) on page 197 (Section5.4 –2)) We will now derive the probability that a given observation is part of a bootstrap sample. Suppose that we obtain a bootstrap sample from a set of n observations.**
   a. **What is the probability that the first bootstrap observation is not the jth observation from the original sample? Justify your answer.**
   Answer: The probability that first bootstrap observation is the jth observation from the original sample is **1/n**, where n is the number of original samples. The probability that it is not the jth observation is **(1- 1/n)**
   b. **What is the probability that the second bootstrap observation is not the jth observation from the original sample?**
   Answer: The probability that second bootstrap observation is the jth observation from the original sample is **1/n**, where n is the number of original samples. The probability that it is not the jth observation is **(1- 1/n)**
   c. **Argue that the probability that the jth observation is not in the bootstrap sample is (1 – 1/n)^n.**
   Answer: Probability that first observation is not the jth of the sample is **(1-1/n**)
   Probability that the second observation is not the jth of the sample is **(1-1/n)**
   .
   .
   .
   Probability that the nth observation is not the jth observation of the sample is **(1-1/n)**
   So, using the theorem of probability, probability that second bootstrap observation is not in the bootstrap sample = **(1-1/n)*(1-1/n)……. till n = (1-1/n)^n**
   d. **When n = 5, what is the probability that the jth observation is in the bootstrap sample?**
   Answer: From answer in c, we can derive the probability that the jth observation is in the sample which is = 1 - (1 – 1/n)^n = 1 - (1 - 1/5)^5 = 0.67232
   e. **When n = 100, what is the probability that the jth observation is in the bootstrap sample?**
   Answer: Like in question d, probability = 1 - (1 - 1/100)^100 = 0.63397
   f. **When n = 10, 000, what is the probability that the jth observation is in the bootstrap sample?**
   Answer: Like in question d, probability = 1 - (1 – 1/1000)^1000 = 0.63230

2. **(Problem 1 on page 259 (Section 6.8 - 1)) 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,...,p predictors. Explain your answers:**

a. **Which of the three models with k predictors has the smallest training RSS?**
   Answer: The model with smallest training RSS is the best subset selection method. The forward selection method selects smallest RSS among the p-k models, backward selection selects smallest RSS among k models. But best selection method aims to get the smallest training RSS among all possible combination of k predictors from total p predictors.
b. **Which of the three models with k predictors has the smallest test RSS?**
   Answer: We cannot guarantee that one of the models has the smallest test RSS. But the best subset selection strategy is more likely to have the smallest one since it considers selecting best out of many models compared to other.
c. **True or False:**
   i. **The predictors in the k-variable model identified by forward stepwise are a subset of the predictors in the (k+1)-variable model identified by forward stepwise selection.**
      True: (k + 1) variable model = k variable model + 1 more variable
   ii. **The predictors in the k-variable model identified by backward stepwise are a subset of the predictors in the (k + 1)- variable model identified by backward stepwise selection.**
      True: (K + 1) variable model = k variable model + 1more variable
   iii. **The predictors in the k-variable model identified by backward stepwise are a subset of the predictors in the (k + 1)- variable model identified by forward stepwise selection.**
      False: We cannot generalize this since, model selection for forward and backwards method are done in opposite direction.
   iv. **The predictors in the k-variable model identified by forward stepwise are a subset of the predictors in the (k+1)-variable model identified by backward stepwise selection.**
      False: We cannot generalize this since, model selection for forward and backwards method are done in opposite direction.
   v. **The predictors in the k-variable model identified by best subset are a subset of the predictors in the (k + 1)-variable model identified by best subset selection.**
      False: The best combination of k + 1 predictors may not contain all the best combination of k predictors.

3. **(Problem 2 (a – b) on page 259 (Section 6.8 - 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.**
         Correct
      iv. **Less flexible and hence will give improved prediction accuracy when its increase in variance is less than its decrease in bias.**
   b. **The ridge regression, 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.**
Correct
iv. **Less flexible and hence will give improved prediction accuracy when its increase in variance is less than its decrease in bias.**

4. **(Problem 3 on page 260 (Section 6.8 -3))**
   a. **As we increase s from 0, the training RSS will:**
   Steadily decrease, increasing s from 0, decreases restriction thus making the model more flexible.
   b. **Test RSS will**
   Decrease initially, and then eventually start increasing in a U shape. Since, the model becomes flexible first decreasing the test RSS but when starts to become more flexible, overfitting comes to play.
   c. **Variance will**
   Steadily increase, the more flexible the model becomes, the more the variance.
   d. **Squared Bias will**
   Steadily decrease, the more flexible the model becomes the bias will decrease.
   e. **Irreducible error will**
   Remain constant. No matter what the model is, irreducible error always remains constant and is not reducible.

5. **(Problem 4 on page 260 (Section 6.8 - 4))**
   a. **As we increase λ from 0, the training RSS will**
   Steadily increase, the coefficients are restricted, and the model becomes less flexible.
   b. **Test RSS will**
   Decrease initially and eventually start increasing in U shape. Firstly, the model becomes less flexible which may generalize the model first, but later the model deviates much more, resulting in increase.
   c. **Variance will**
   Steadily decrease. Less flexible model, variance will decrease.
   d. **Squared bias will**
   Steadily increase. Less flexible model squared bias increases.
   e. **Irreducible error will**
   Remains constant. There is no relation of irreducible error with the model.