ASSIGNMENT-3

PART 1

Question 2

a)Total number of observations=n

Probability that the first bootstrap observation is the jth observation from the original sample=1/n.

Probability that the first bootstrap observation is not the jth observation from the original sample is (1-1/n).

b) Total number of observations=n

Probability that the second bootstrap observation is the jth observation from the original sample=1/n.

Probability that the second bootstrap observation is not the jth observation from the original sample is (1-1/n).

c)The probabilities are independent of each other.

Probability that the jth observation is not in the first bootstrap sample is (1-1/n).

Probability that the jth observation is not in the second bootstrap sample is (1-1/n).

Probability that the jth observation is not in the nth bootstrap sample is (1-1/n).

Hence, Probability that the jth observation is not in the bootstrap sample is (1-1/n)\* (1-1/n)\*…..n times.= (1-1/n)n.

d) Probability that the jth observation is not in the bootstrap sample is (1-1/n)n.

Probability that the jth observation is in the bootstrap sample is 1-(1-1/n)n.

Here n=5, So 1-(1-1/5)5 = 0.67.

e) Probability that the jth observation is not in the bootstrap sample is (1-1/n)n.

Probability that the jth observation is in the bootstrap sample is 1-(1-1/n)n.

Here n=100, So 1-(1-1/100)100 = 0.634.

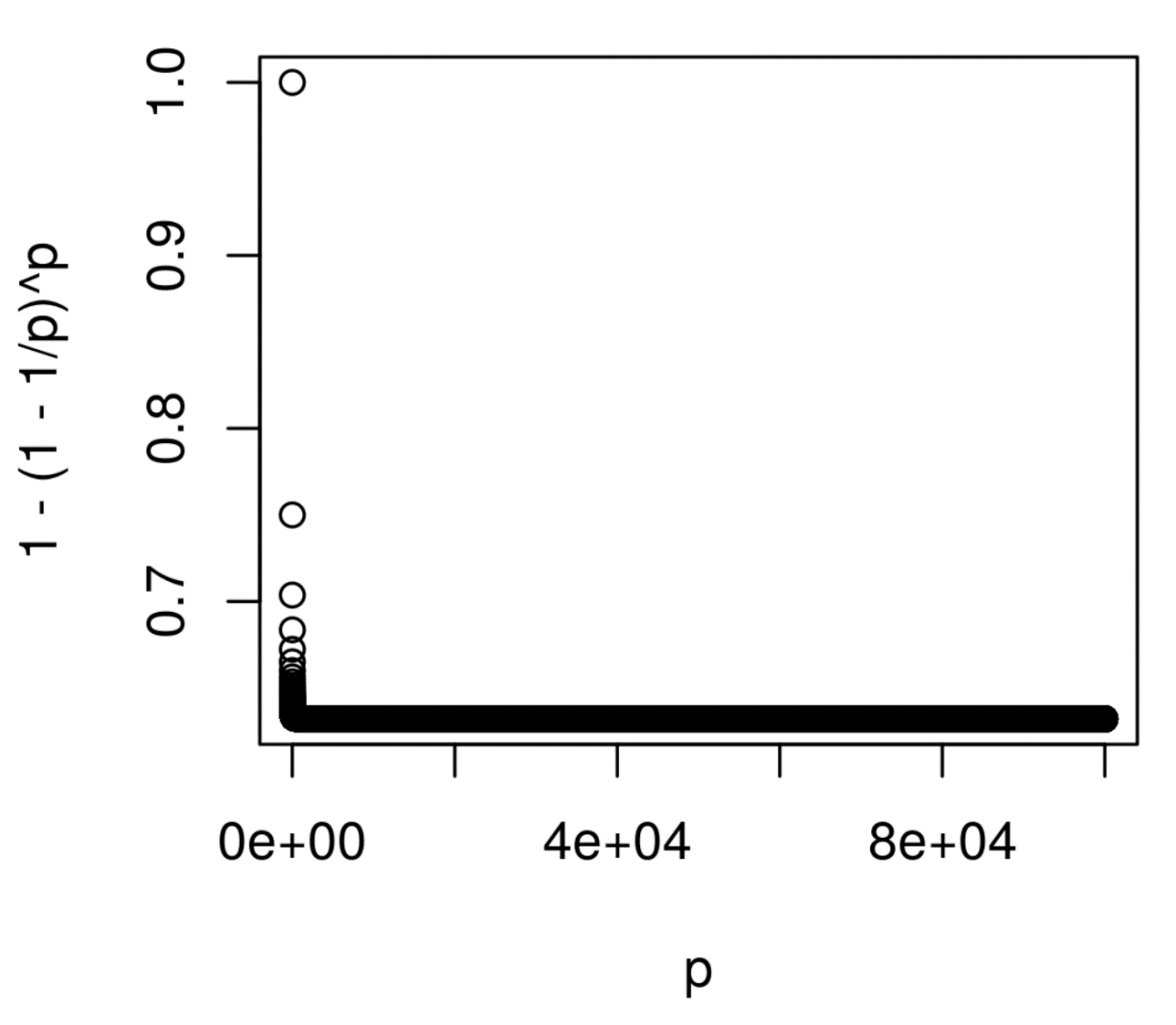
f) Probability that the jth observation is not in the bootstrap sample is (1-1/n)n.

Probability that the jth observation is in the bootstrap sample is 1-(1-1/n)n.

Here n=10000, So 1-(1-1/10000)10000 = 0.632.

g) p <- 1:100000

plot(p, 1 - (1 - 1/p)^p)



It is observed that from the value of 0.60 the plot decreases asymptotically.

h) store=rep(NA,10000)

for(i in 1:10000){

store[i]=sum(sample(1:100,rep=TRUE)==4)>0

}

mean(store)

->0.6308

The probability of jth observation occurring in the bootstrap sample is same for any of the j value.So, for n=100 and j=4 the probability is (1-1/100)100=0.634.

We observe that for both the cases the probability is very close to each other.

Question 3

a)K-fold cross validation-

This method of cross validation is the process of dividing the set of observations into k equal parts. The parts which are divided are called as folds. Leaving one of the fold where this is the hold-out set and the model is fit on the remaining observations. We then compute the MSE on the hold-out set. Until we reach k MSE, we repeat these steps using different fold each time. The result of the k-fold cross validation estimate is the average of the set of MSE values.

b)1)Disadvantage of k-fold cross validation is computational cost where it is consuming time for some of the models. When we take the case of validation approach ,the model will learn the data only once and then this model is tested once. A validation set has a chance that it will overestimate the error as we have used less data for training. When we have a limited number of observations i.e few number of observations then it is better to use k-fold rather than validation approach as it is difficult to separate the validation set. k-fold has less variance and high bias as compared to validation set approach. Validation set approach is easier and simpler to implement. Advantage is that it has high accuracy as there are more observations used for training. A validation set has a chance that it will overestimate the error as we have used less data for training. The test error rate varies a lot as the data points are part of training and validation sets.

2)The computational cost is worst for LOOCV as compared to k-fold as we need to test and train n times instead of k. As LOOCV is highly correlated ,it has high variance and k-fold has high bias and less variance. So we prefer k-fold CV.

Advantages of k-fold cross validation relative to LOOCV are: The test error rate estimates are accurate for k-fold CV as compared to LOOCV. If n is large then k-fold produces a accurate result with accurate estimates. It is less variable as compared to LOOCV.

PART-2

Question -1

a) Best subset selection model will have the least training RSS because after we consider all the possible models with k parameters then we will select the final model among all the models.

b) We can’t predict the correct model. If we have large p predictors then best subset approach may overfit the model. The same model is being used by the other two models and therefore it is hard to predict which is the best model for smallest test RSS.

c)i) TRUE- The k+1 variable model will contain all the k features from the k variable model and the best additional feature.

ii) TRUE-The k variable model will contain all except one feature in the k+1 best model and minus the single feature which will result in smallest gain in RSS.

iii) FALSE-There is a possibility for the disjoint sets where these are identified by backward and forward selection.

iv) FALSE-There is a possibility for the disjoint sets where these are identified by backward and forward selection.

v)FALSE-By selecting the available (k+1) predictor models, then the model with (k+1) predictors is obtained.

Exercise 2

a) Option iii is TRUE. Lasso is less flexible when compared to least squares. When the increase in bias is less than decrease in its variance then we get a more improved prediction accuracy.

b) Option iii is correct. The reason is same as the above one.

c)Option ii is correct. As the non-linear methods are more flexible than least square and this may result in more prediction accuracy.

Exercise 3

a)Steadily decrease.

When we increase s from 0, then the restriction on the coefficients is reduced and thus this may result in an increase in least square estimates. Therefore the model may become more flexible and this may result in training RSS.

b) Decrease initially, and then eventually start increasing in a U shape.

When we increase s from 0, then the restriction on the coefficients is reduced and thus this may result in an increase in least square estimates. Therefore, the model may become more flexible and this may result in less test RSS and then starts increasing in a U-shape.

c)Steadily increase.

When we increase s from 0, then the restriction on the coefficients is reduced and thus this may result in an increase in least square estimates. Therefore the model may become more flexible and this may result in increase in variance.

d) Steadily decrease.

When we increase s from 0, then the restriction on the coefficients is reduced and thus this may result in an increase in least square estimates. Therefore the model may become more flexible and this may result in decrease in bias of the model.

e) Remain Constant.

Irreducible errors are always independent of the model. Therefore, it does not depend on the value of s.

Exercise 4

a)Steadily increase.

When we increase from 0, then the restriction on the coefficients is increased and thus this may result in an deviation of least square estimates. Therefore the model may become less flexible and this may result in increase in training RSS.

b)