6/04/18 DPA Assignment 3 Avinash Vellineni

A20406657

Question 1.1 :-

2) a) $\left(1-\frac{1}{m}\right)$. Total probability of all the observation is 1. So the probability of each observation is $\frac{1}{m}$, therefore the probability that the first bootstrap observation is not the jth observation is $1-\frac{1}{m}$.

b) $\left(1-\frac{1}{m}\right)$

c) In bootstrap model sampling is done with replacement, so each observation is equally likely to be the jth observation with a probability of $\frac{1}{m}$.

Then for all the observations (n) is $\left(1-\frac{1}{m}\right)^{m}$.

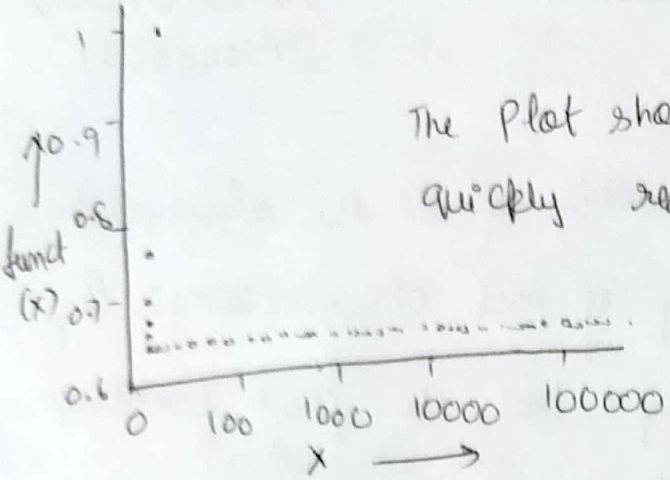
d) $1-\left(1-\frac{1}{m}\right)^{m}$

$\Rightarrow 1-\left(1-\frac{1}{5}\right)^{5} = 1-\left(\frac{4}{5}\right)^{5} = 67.23\%$

e) $1-\left(1-\frac{1}{100}\right)^{100} = 1-\left(\frac{99}{100}\right)^{100} = 63.34\%$

f) $1-\left(1-\frac{1}{10000}\right)^{10000} = 63.2\%$

g)
```
funct = function(n)
        return (1-(1-1/n)^n)
x = 1: 100000
Plot (x, funct(x))
```



The Plot show that probability quickly reaches a value of aroun 63.2%.

h) The fraction of 10,000 samples with j=4 stay close to the predicted probability $1-\left(1-\frac{1}{100}\right)^{100}$ $= 63.34\%$

3)

a) K-fold method takes n observations and randomly splits into K non-overlapping parts. Here one of each these K non-Overlapping parts acts as a test set and remaining as the train set. Test error is calculating by taking the average of the resulting K test MSE's.

b)

(i) In validation set approach divides the data into two halves. Draw backs:-

- Test error rate is highly variable based on which observations are included in train or test se
- The Validation set error rate may tend to overestimate the test error rate. for the model fit on the entire data.

Scanned by CamScanner

ii) LOOCV is same as k-fold method with k = n.

Therefore LOOCV is computationaly expensive since the model fit is done n times. LOOCV has higher variance & low bias than k-fold method.

question 1.2 :-

1) a) Best subset selection method has the lowest training RSS. Because the other methods depend on which predictor they pick at each iteration.

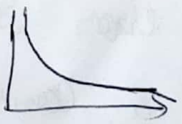
b) Since best subset selection considers all the possible models. it provides the lowest test RSS.

c) i) True ii) True iii) False (iv) false v) false

2) a) iii) less fexible, improved prediction since we have less variance & high bias

b) iii) less flexible, improved prediction

c) ii) More flexible, high variance, low bias

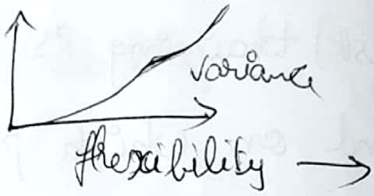
3) a) (iv) steadily decrease 

as s increases β's increase from 0 to OLS β's. Training decreases from maximum value to OLS RSS as β increases from zero.

b) (ii) Decreases initially reaches a minimum & then Increases in a U. shape

When β is 0 we have maximum test error and as β approaches to OLS coefficient values. test error reaches a minimum value & then increases due to overfitting the model using training set.

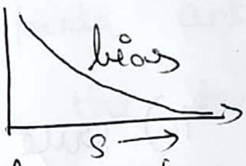
c) (iii) Steadily increase..



when $S=0$ there is no variance.

As S is increased, variance starts increasing. thus we have steady increase in variance as S is increased

d) (iv) Steadily decreased.



when $S=0$ we have high bias & as S is increased bias decreases as the model starts predicting better and better.

e) (v) Remains constant.

Irreducible error is independent of variations in S. So it is a constant.

4) a) (iii) Steadily increase

as λ is increased, all β's decrease from OLS estimate to 0. So since train error is minimum at OLS coefficient + ~~coef~~ maximum at β's equal to 0. So it steadily increases.

b) (ii) Decrease and then start increasing in U shape

As λ increases β's decreases to zero so overfitting is reduced. So the test error initially decreases & then increases due to underfitting as β's approaches to zero

c) (iv) Steadily decreases..

when β's are at OLS estimates we have high variance. As ~~β~~ λ increases β's decreases to zero. So the variance from maximum value decreases to zero.

d) (iii) steadily increases

As λ increases β's decreases from OLS estimates to zero. Bias is low when the β's are at their OLS estimates. So as λ increases, bias increases from minimum to maximum

e) (v) Remains constant

irreducible error is independent of λ so it is a constant.