Report AML

Prob1)

a)When we compare the performance of inbuilt and our implementation,the inbuilt gives better accuracy in less time. Our implementation takes roughly 3 to 4 times more time and accuracy is nearly 10% less on an average.
For n_estimators =80 and m = 8

```
OUR IMPLEMENTATION

The accuracy is :
0.908695652174
Time in seconds :
1.101469

The accuracy is :
0.868115942029
Time in seconds :
1.398869

The accuracy is :
0.901449275362
Time in seconds :
1.255522
```

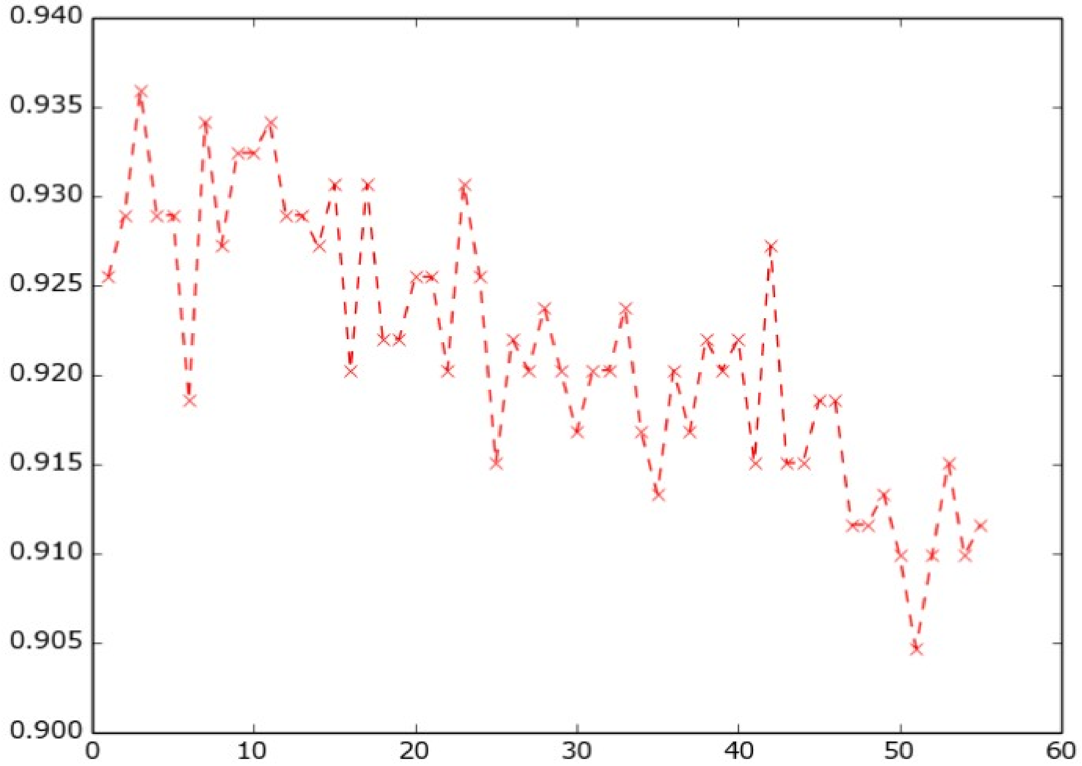
```
Inbuilt


The accuracy is :
0.954380883418
Time in seconds :
0.389075

The accuracy is :
0.955104996379
Time in seconds :
0.420919

The accuracy is :
0.958001448226
Time in seconds :
0.369429
```

b) Comapring the sensitivity vs the m values
We notice that sensitivity decreases as value of m increases



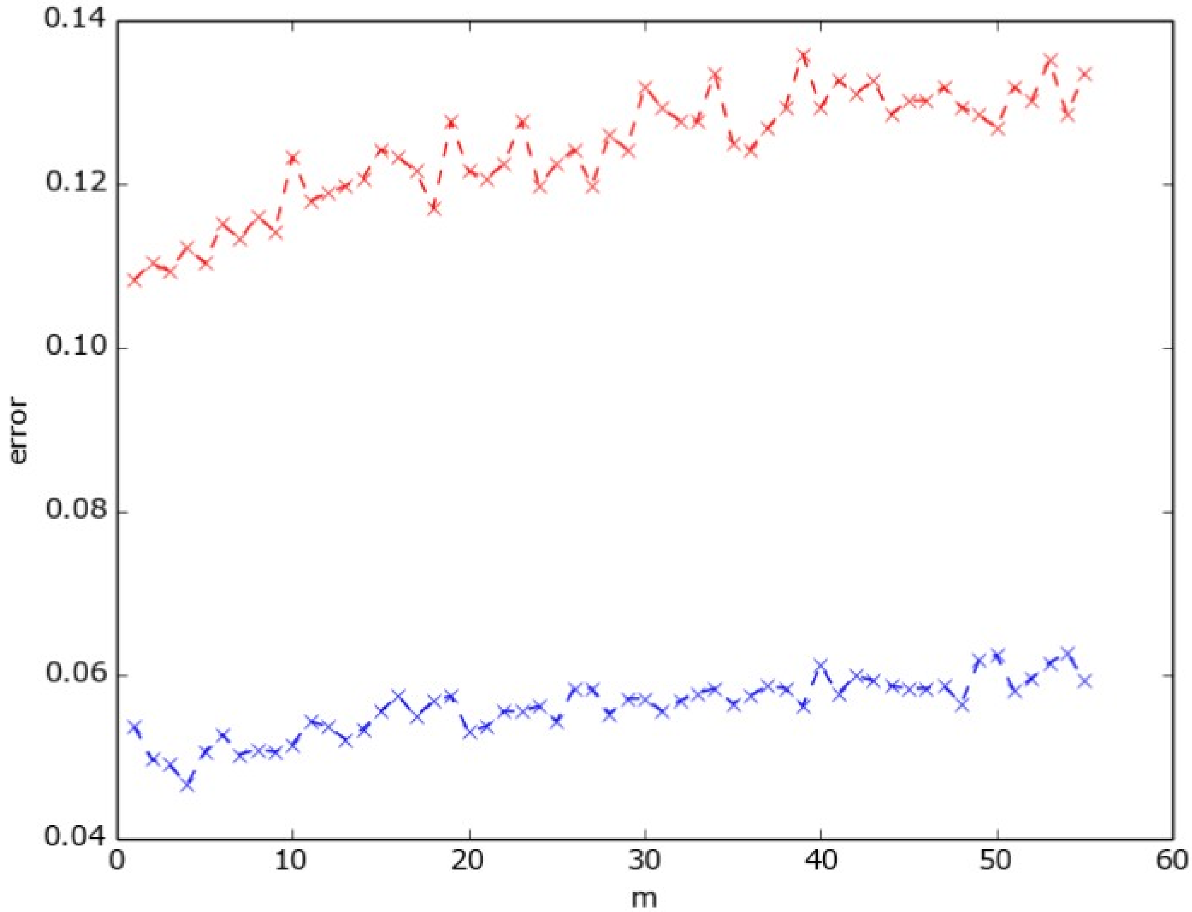
Sensitivity ,also called Recall or True Positive Rate, is the proportion of actual positives which are correctly identified as positives by the classifier.

Sensitivity = TP / (TP +FN)

c) OOB and test error vs m values

Each tree is constructed using a different bootstrap sample from the original data. About one-third of the cases are left out of the bootstrap sample and not used in the construction of the kth tree.

We put each case left out in the construction of the kth tree down the kth tree to get a classification. In this way, a test set classification is obtained for each case in about one-third of the trees. At the end of the run, take j to be the class that got most of the votes every time case n was oob. The proportion of times that j is not equal to the true class of n averaged over all cases is the oob error estimate.



OOB ERROR is BLUE
TEST ERROR is RED

We notice that test error and oob error increases as value of m increases.
We find otpimal split value of m between 0-10

3 prob)

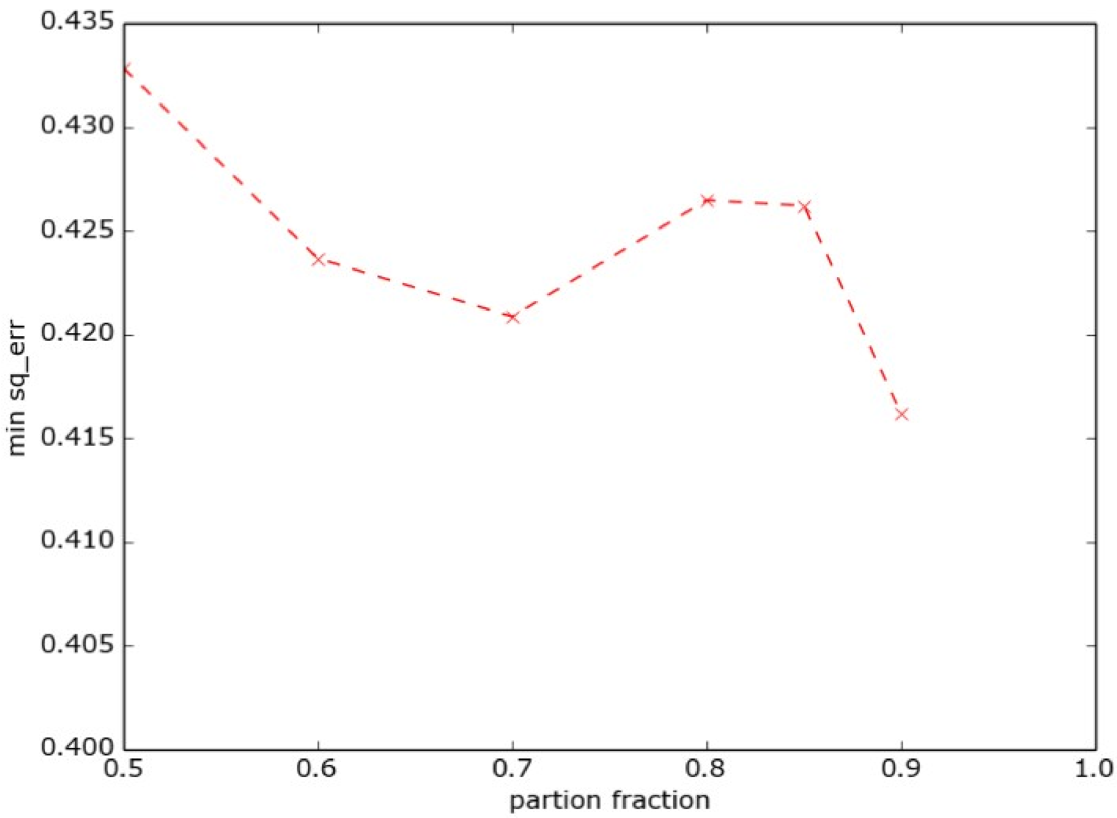
e)We find that lamda around 50-60 is the best value we can get for lamda.
We find that the fifth attribute is the most significant attribute.

When we remove two to three least significant atrributes, we find that sqaure mean error doesnt change much.

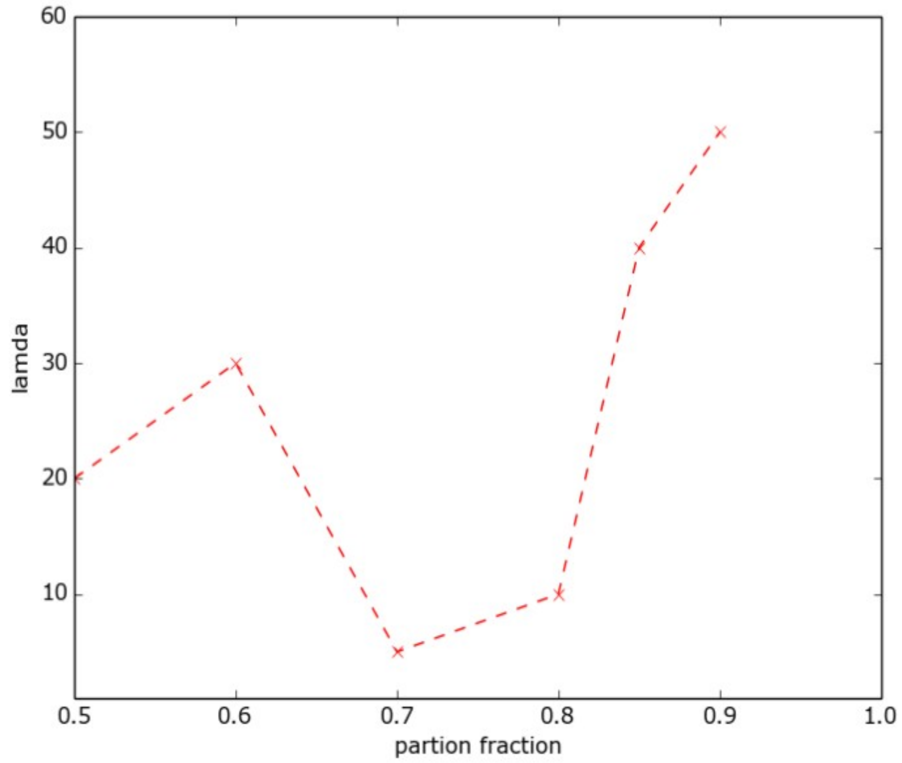
f) Yes the effect of lamda changes for different partitions of data. This can be seen from the graphs plotted which ahve been submitted.

g) minsq_error vs partition fraction graph

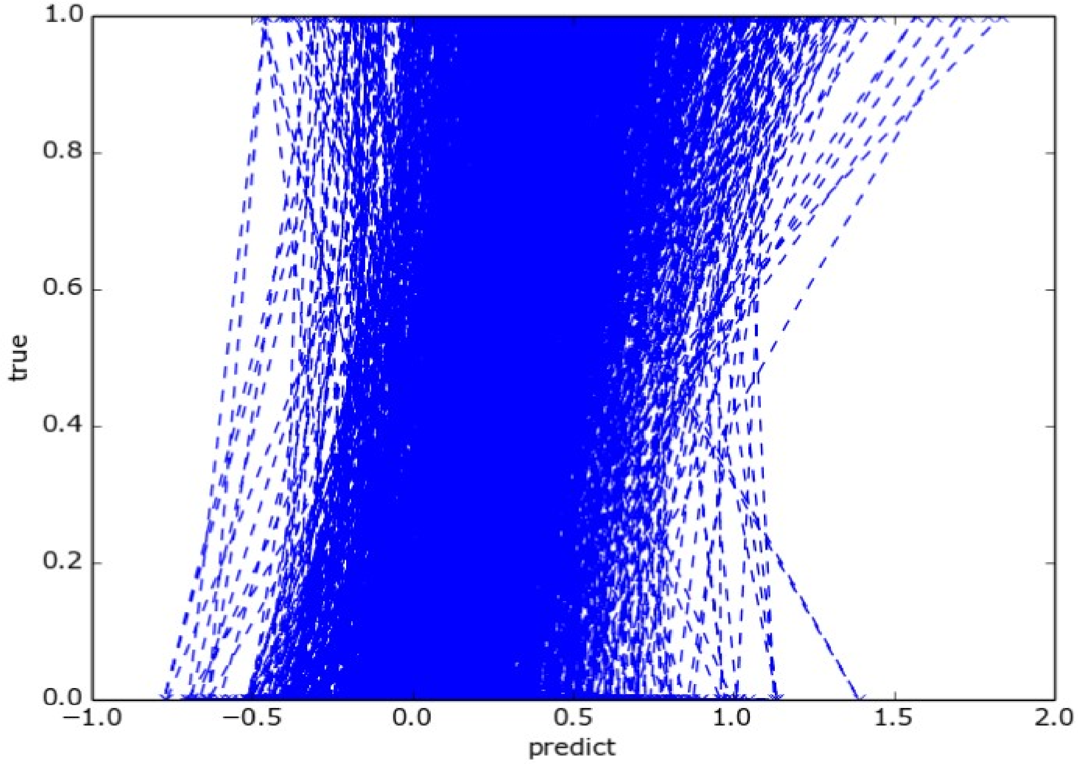
We observe that as partition size increases error decreases.



Lamda vs partition fraction graph



h) Train data
True vs predicted
We can see that the data is around the 45 degree axis
We can say the model is good



Test data: true vs predicted
We can see that the data is around the 45 degree axis
We can say the model is good even for test set