**Implementation Assignment – 3**

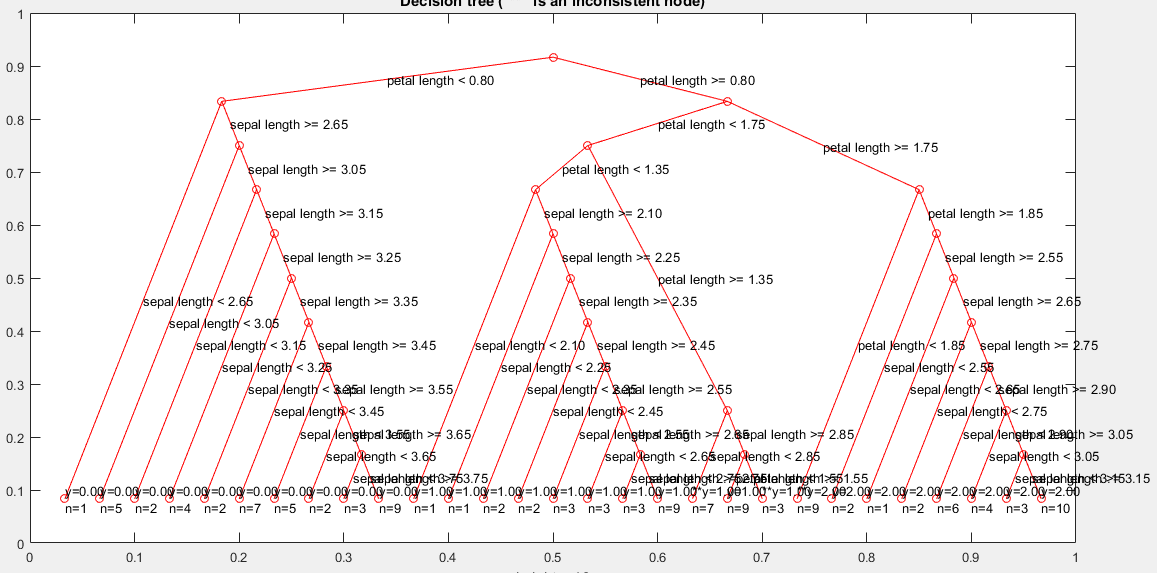
**Group Members:**

Purbasha Chatterjee – 931707158 (37%, completed part-2 implementation)

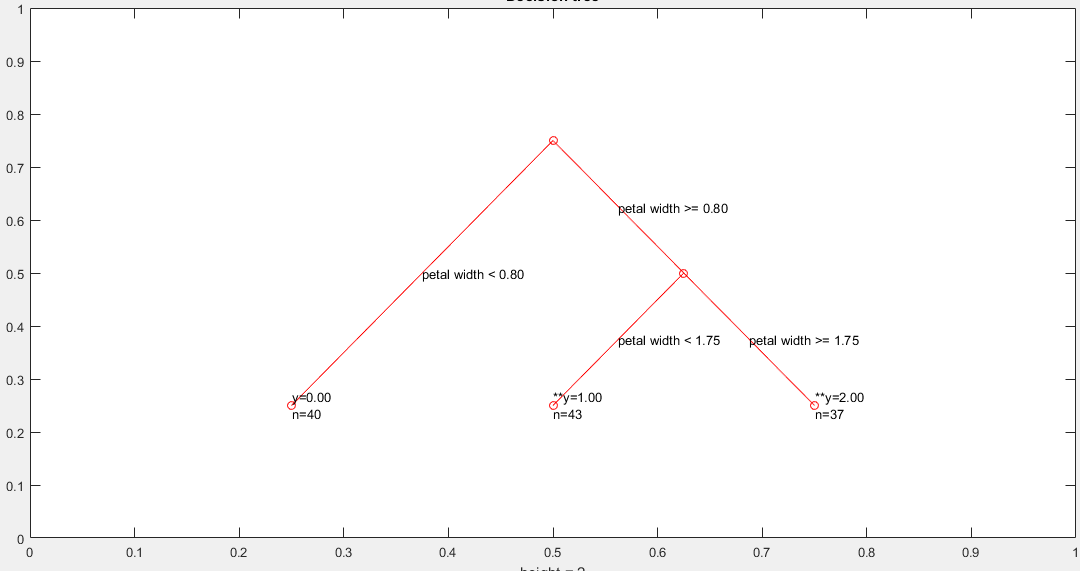
Meghamala Sinha – 932944466 (38%, completed part-1 implementation)

Alex Way Curtis – 932191830 (25%, partial contribution in part-1 and part-2)

2) a) For feature bagging we tested around different values of k, in random we recorded the tree for sepal length and petal length decision tree.

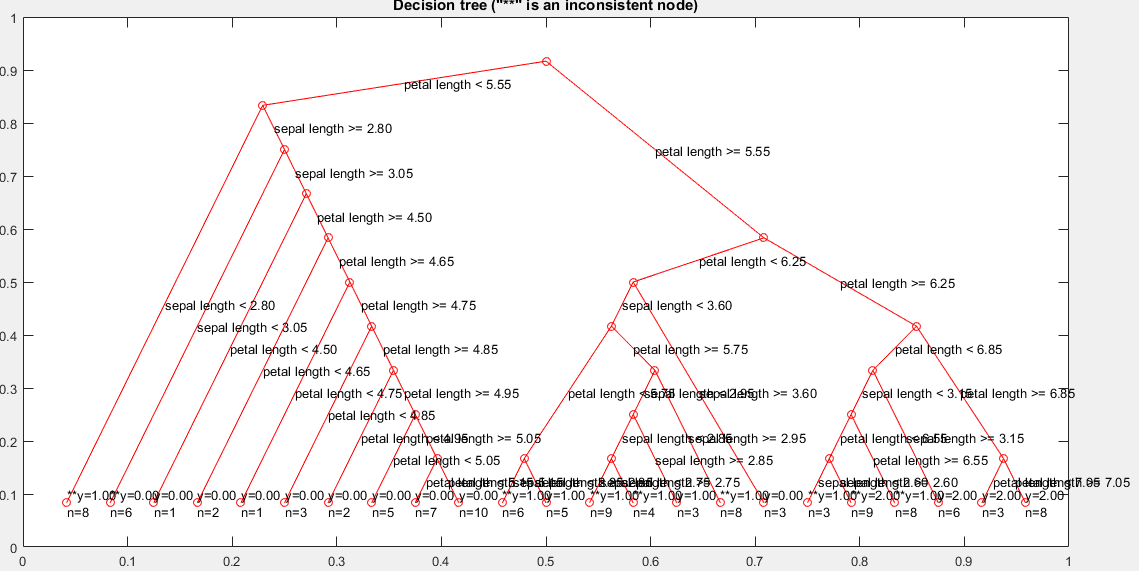


For k=10, decision tree.

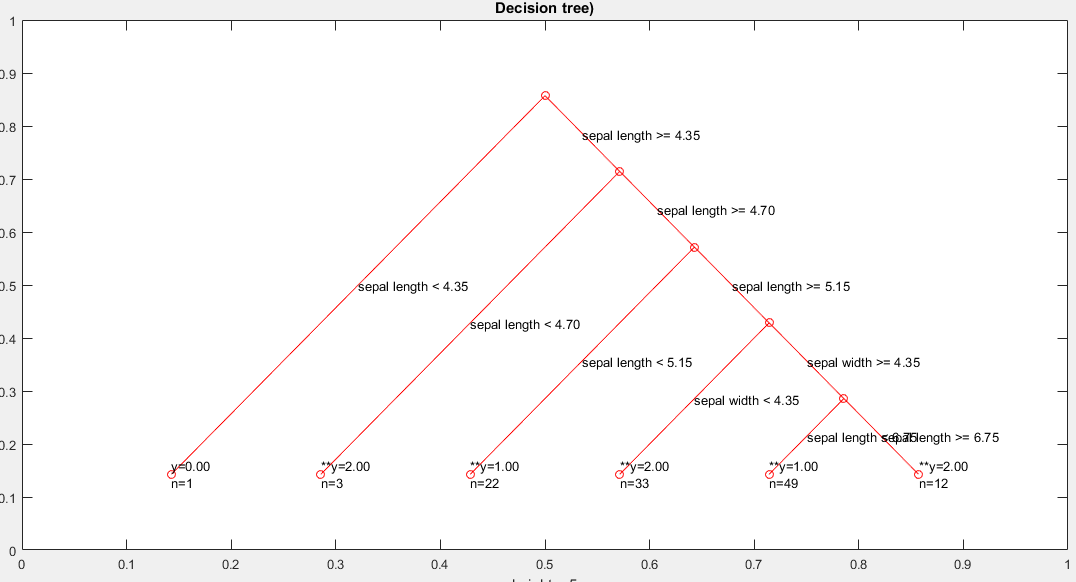


For k=50, decision tree

b) Next, we perform bagging or bootstrap aggregating, we again record the random training samples for the feature as below.



For k=10, decision tree



For k=50, decision tree

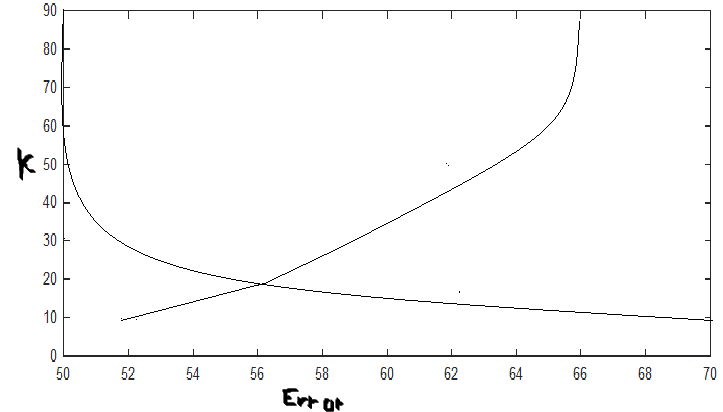
c) Now we record the tabular chart for the majority vote of the class by all the L values ranging from L=5 to 30 with an interval of 5 for two random feature that is sepal length and petal length.

|  |  |
| --- | --- |
| **Class Vote** | **L** |
| 1 | 5 |
| 2 | 10 |
| 2 | 15 |
| 0 | 20 |
| 2 | 25 |
| 0 | 30 |

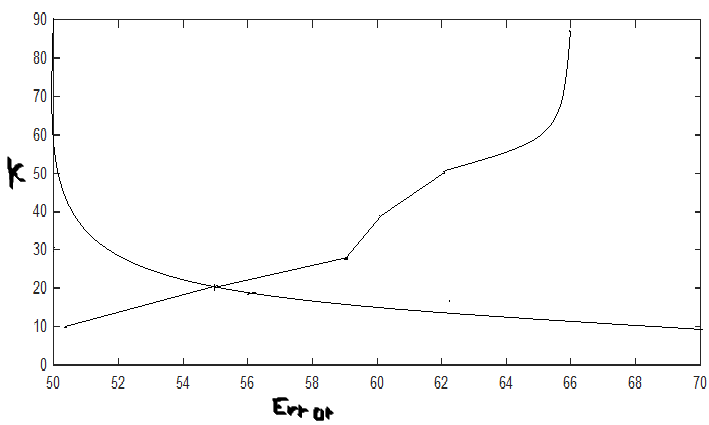
Hence , we get class 2 that is Iris Virginica as the majority vote here.

d)

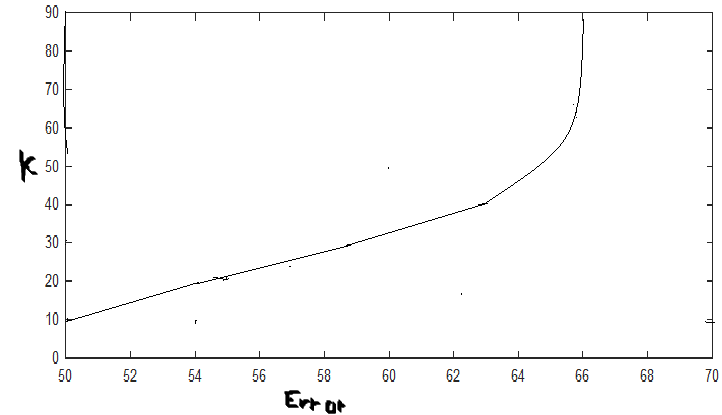
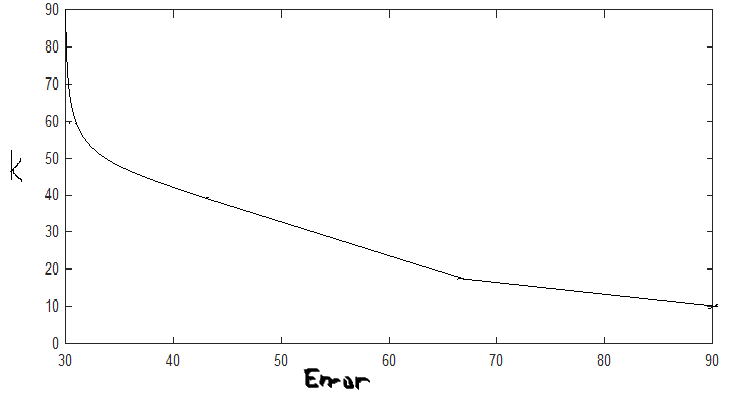
For L=5;



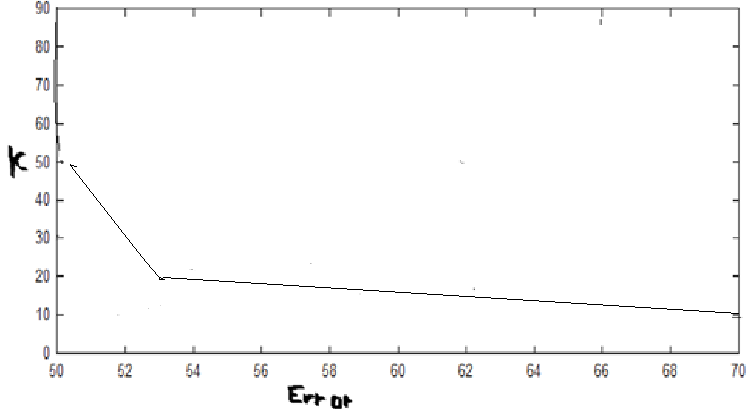
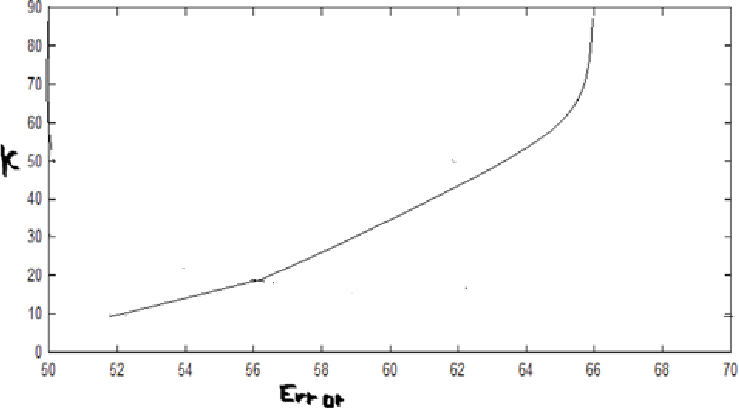
For L=10;



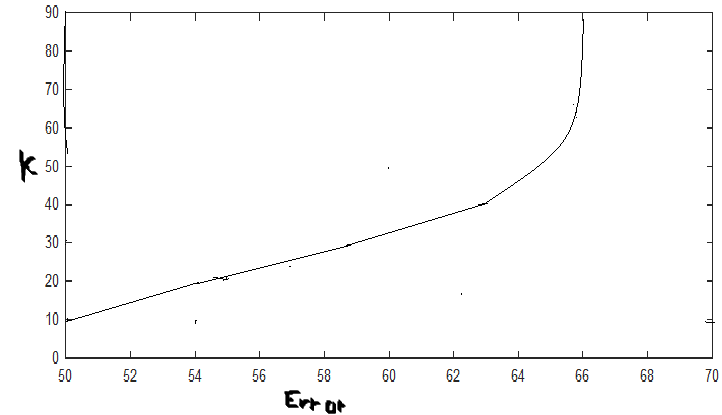
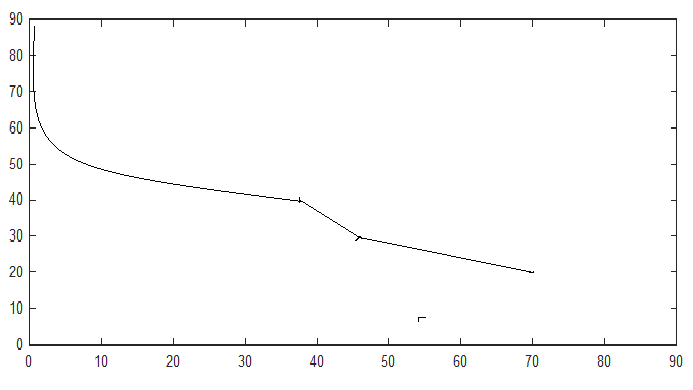
For L=15;

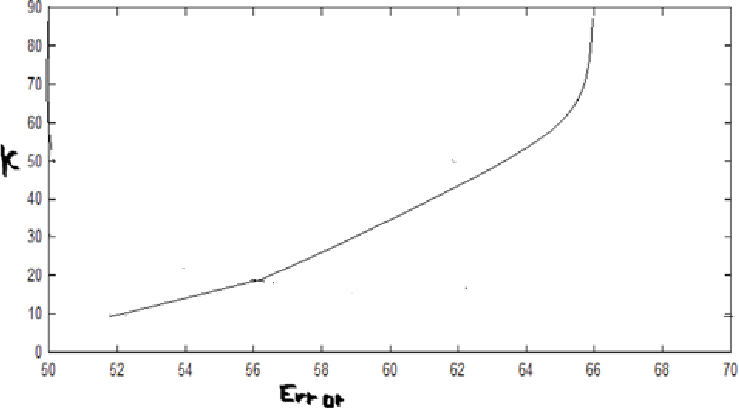
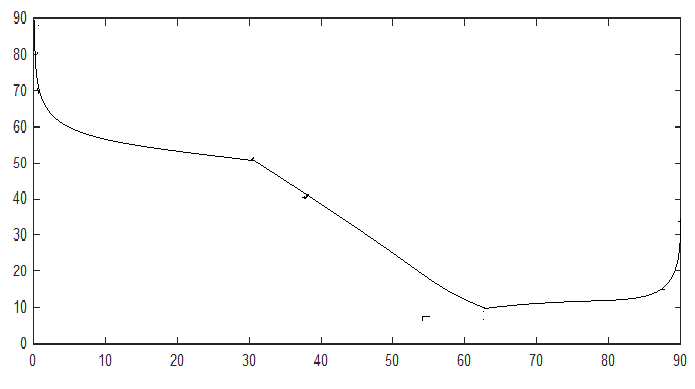
For L=20:



For L=25:

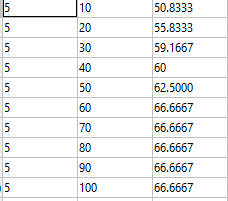
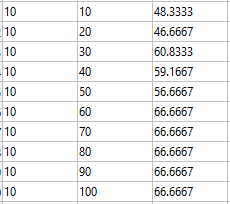
 

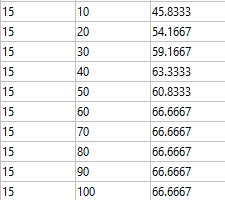
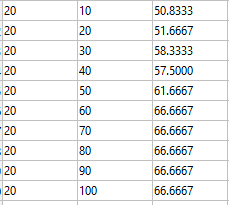
For L=30:

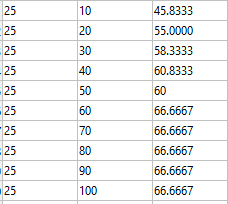
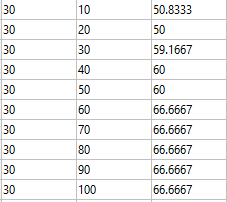
 

For training data :

L k error L k error

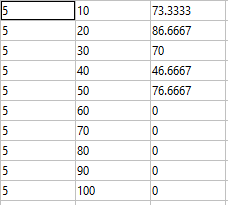
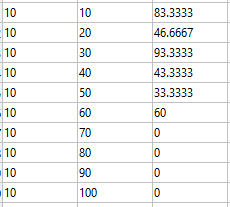
 

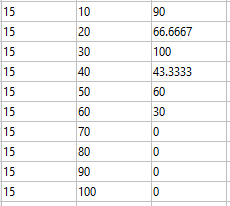
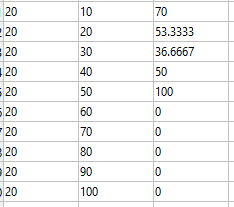
 

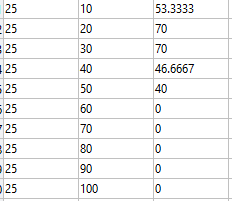
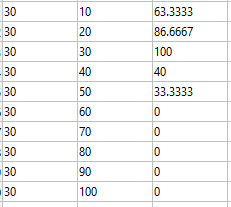
 

For testing data :

L k error L k error

e) With the increase in number of trees we see that the error decreases and hence we can predict more accurately for training data whereas for the testing data it increases for a certain point and then start decreasing. With the use of random feature selection, we see that the accuracy improves as now the training time reduces due to less features and also it reduces the overfitting. With the increase in k we see that the error increases because for more the value of k we have less depth tree and hence difficult to predict accurately for training and in case of testing we see that error increases at certain point and then decreases.