Q6 . Random Forests

Saurav Vara Prasad Channuri
ES16BTECH11007

a) *Accuracy in Scikit Learn random forests library =* ***1.0***

Accuracy in my Random Forests code = ***0.9196***
No. of random variables = 8
No. of decision trees = 5

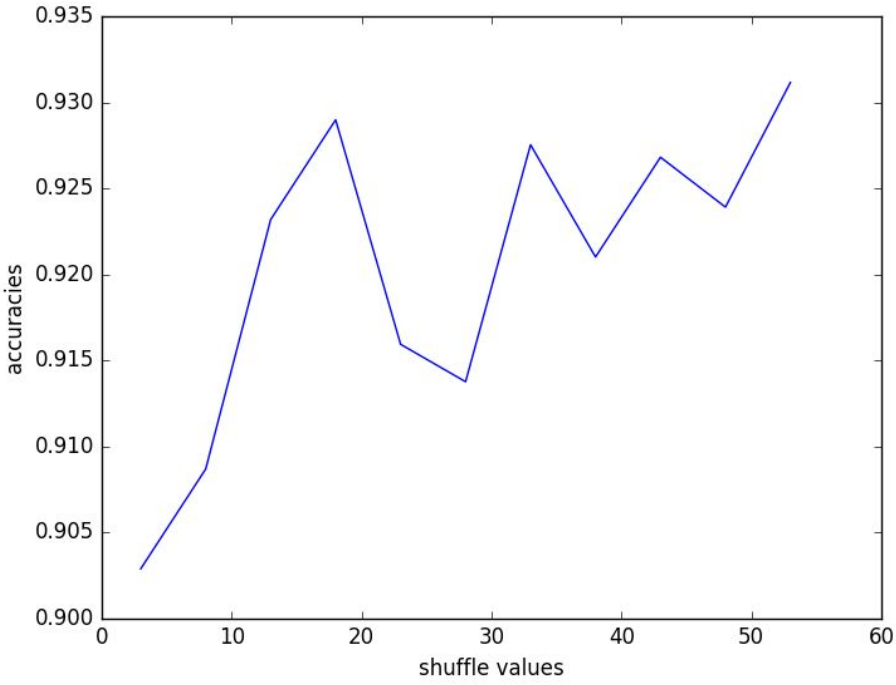
Accuracy in the Random forests code I wrote is a bit less than that of the scikit learn library. This maybe because of the number of random attributes I took for training each node.

The variation of accuracy as the number of random attributes changes is given in the graph of second question

b) *average of accuracies =* ***0.9203557312252966***
Random attributes size = 8

Accuracies were plotted taking certain number of random attributes each time for decision tree nodes.

As the number of attributes increases the accuracies are increasing, the accuracy is increasing. This maybe because it was able to capture best attribute that give highest information gain for each node which in previous cases was not a choice.



C) average of accuracies = **0.8763128176171655**

Random forests with Out of box error was implemented with taking 500 datapoints each time we train the tree with replacement along with the number of random attributes = 8

Datapoints size with replacement = 500
Random attributes size = 8

The accuracies are not consistent and also not as high as done in without OOB here. It maybe because of the number of datapoints that I took to train trees each time and also the