

Assignment 2

Question 6

Explanation :: Using subsample function, sample set for each tree is defined. Number of trees is limited to 5, to reduce the run time.(for quick display of results). This can be changed by changing number 5 to desired number of trees in line 204. For each sub sample, tree is constructed using decision tree.

6a.py :: Given above is basic structure of code. Accuracy is calculated by predicting the output for test data over all constructed trees(using bagging_and_predict function).

Using sklearn randomforest classifier, accuracy is calculated (by fit and score)

6b.py :: Code for b part

6c.py :: Code for c part

a)

N_features = 57

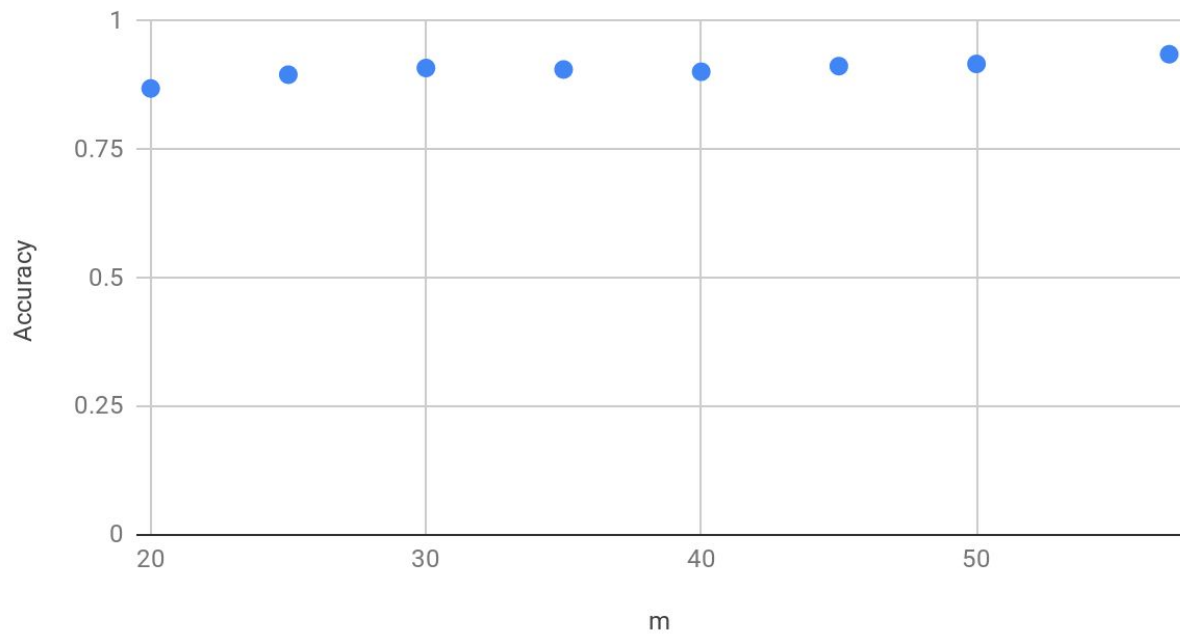
Accuracy without sklearn = 0.934105720492

Accuracy with sklearn = 0.948587979724837

(Written Random Forest classifier uses information gain for splitting and sklearn uses gini)

b)

Accuracy vs. m



c) $M=20,25,30,35,40$

OOB Error vs. m

