Report - Assignment 3

Foundations of Machine Learning

Prof. Vineeth N Balasubramanian

Naitik Malav | CS19BTECH11026

Deliverables:

1. I have submitted answers of question 1, 2 and 3 in a hand-written pdf format – Assign3_1_2_3.pdf
2. I have submitted two jupyter notebooks – Assign3_4.ipynb and Assign3_5.ipnyb as a solution of question 4^{th} and 5^{th}.
3. And Report.pdf

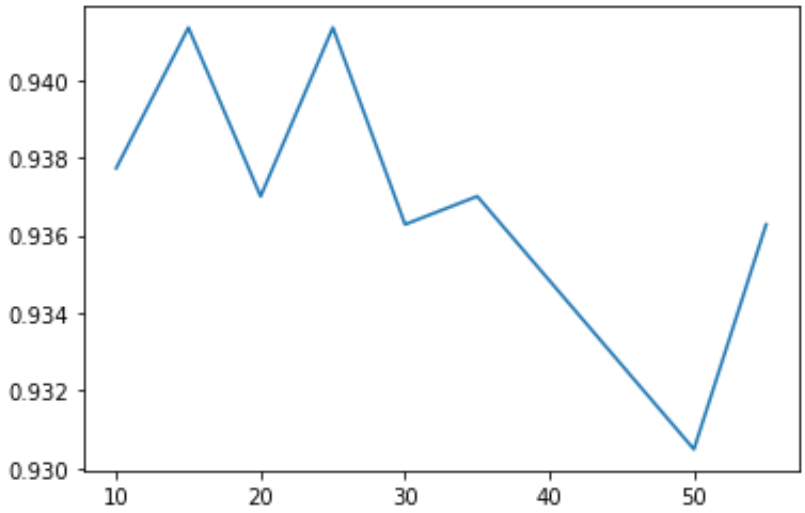
4^{th} –

a).
I have used the decision tree code from Assignment-1.

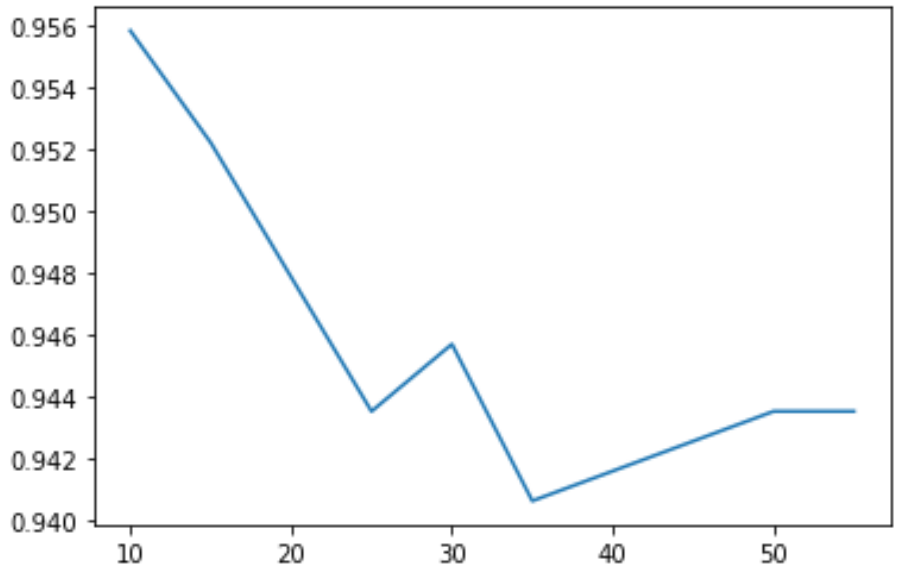
Accuracy in case of using built-in Sklearn library Random Forest Classifier is: 0.9371832005792905
while in case of my own Random forest classifier
average accuracy is: 0.891723400986602

b).
On increasing value of m we get improved prediction accuracy.
Time Taken for m= 10 is: 0.2570838928222656
Time Taken for m= 15 is: 0.4005300998687744
Time Taken for m= 20 is: 0.49711084365844727
Time Taken for m= 25 is: 0.6681728363037109
Time Taken for m= 30 is: 0.7091672420501709
Time Taken for m= 35 is: 0.8241963386535645
Time Taken for m= 50 is: 1.2231688499450684
Time Taken for m= 55 is: 1.3453032970428467

c).
OutOfBag Graph with variable m from 10 to 55:
This graph is varying when we run every time. This is because of the random va lues.



Test Error Graph with variable m from 10 to 55:



5th –

a).

Only first 47 columns have useful information for preparing dataset.
Dropping columns which don't contain any useful information. This can be don e using DataFrame.drop()

Converting Columns to useful columns using Binary Encoding, String Manipula tion, etc.

Also removing row entries where loan_status == 'Current'

b).
Best Accuracy for GradientBoostingClassifier :

Accuracy is: 0.9947817441763407
Precision is: 0.9956874
Recall score is: 0.994153

For n_estimators= 1
Accuracy is: 0.8491874474642757
Precision is: 0.8491874
Recall score is: 0.8491874

For n_estimators= 2
Accuracy is: 0.8491874474642757
Precision is: 0.8491874
Recall score is: 0.8491874

For n_estimators= 3
Accuracy is: 0.8491874474642757
Precision is: 0.8491874
Recall score is: 0.8491874

For n_estimators= 4
Accuracy is: 0.8491874474642757
Precision is: 0.8491874
Recall score is: 0.8491874

For n_estimators= 5
Accuracy is: 0.97219108994116
Precision is: 0.9721911
Recall score is: 0.9682907

For n_estimators= 6
Accuracy is: 0.97219108994116
Precision is: 0.9721911
Recall score is: 0.9682907

For n_estimators= 7
Accuracy is: 0.97219108994116
Precision is: 0.9721911
Recall score is: 0.9682907

For n_estimators= 8

Accuracy is: 0.97219108994116
Precision is: 0.9721911
Recall score is: 0.9682907

For n_estimators= 9
Accuracy is: 0.97219108994116
Precision is: 0.9721911
Recall score is: 0.9682907

For Decision Tree:

Accuracy is: 0.991874474642757
Precision is: 0.9918745
Recall score is: 0.9948075

On comparing GradientBoostingClassifier with Decision Tree, we have improv ement in accuracy and precision.