CSE/ECE 343/543: Machine Learning
Assignment-2 Report

Programming Questions:

Q1. GridSearchCV and Support Vectors

1.Best parameters of SVM using the train set using GridSearchCV
kernel:'rbf'
c:1.0
gamma:0.01

Accuracy of train data set on best parameters:0.9914254818689677
Accuracy on Test data set on best parameters:0.5514814814814815

Time to train model:22757 sec(Approx 6.3 hrs)
PCA n_component=45

2. SVM train on Support Vectors:

Accuracy of train data set on best parameters:0.98484065460809
Accuracy on Test data set on best parameters:0.54361445783175
Time to train model:963 sec(Approx 16 min)
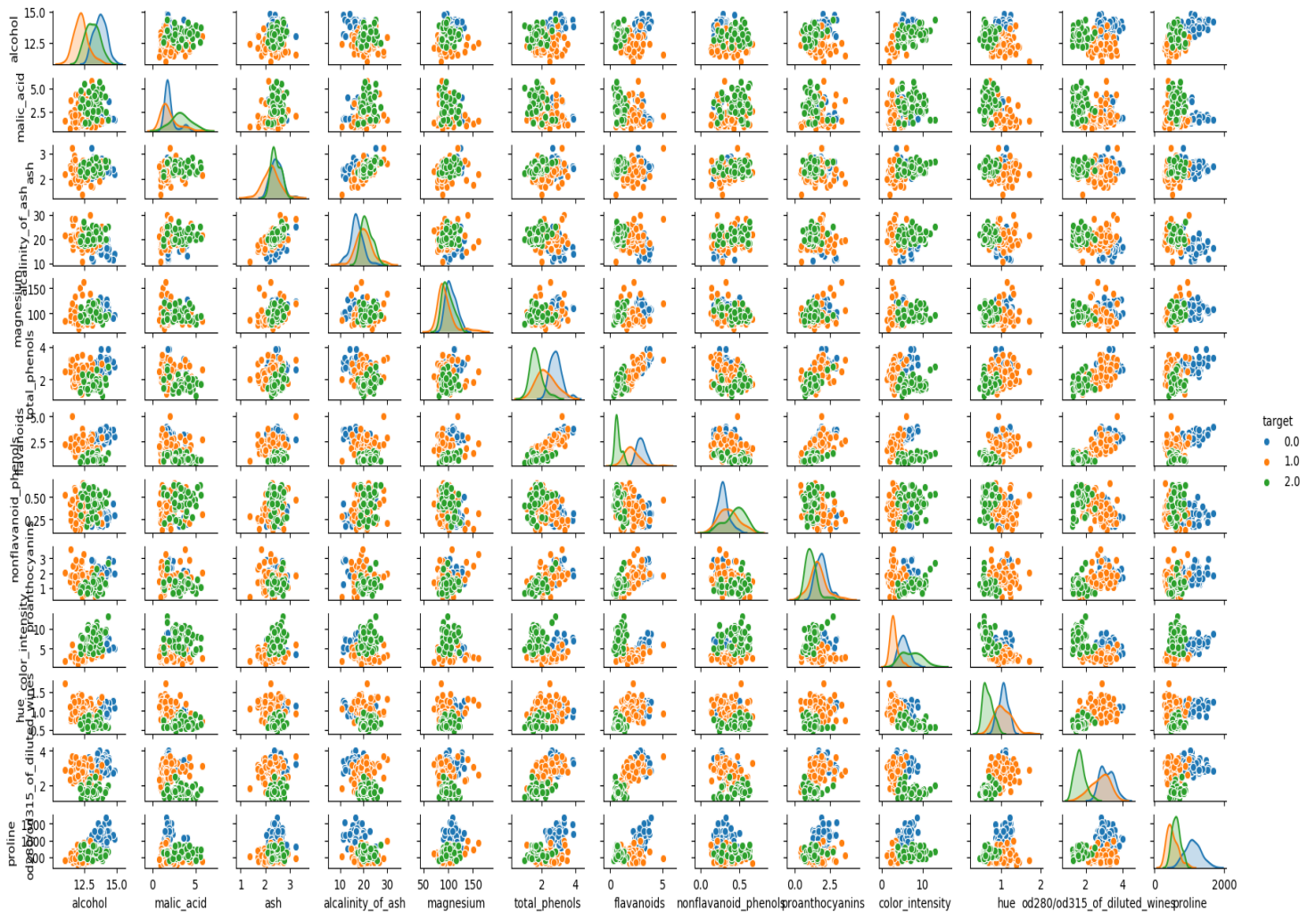PCA n_component=45

3.Compare the accuracy in 1 and 2:

Both accuracies are almost same because in 2 support vectors are being used to train the model which is subset of actual data.

Q2. SVM vs Naive Bayes vs Decision Trees

1.Pairwise relations in dataset plot using seaborn:

Inferences from this pairwise plot are following:

(i).All the at diagonal histograms shows variation of features of all the three classes.

(ii).All graphs are same graph at symmetric position after interchanging axis.

(iii).It shows clustering of data of different classes.

2.a.Evaluation scores for OneVsAll are following:

F-1 Score of oneVsAll using svm for train data: 0.9918020787586005
Accuracy of oneVsAll using svm for train data: 0.9919354838709677

Classwise accuracy for train data:
Train Accuracy of class 0   100.0
Train Accuracy of class 1   99.19354838709677
Train Accuracy of class 2   99.19354838709677

F-1 Score of oneVsAll using svm for test data: 0.983239053006495
Accuracy of oneVsAll using svm for test data: 0.9814814814814815

Classwise accuracy for test data:
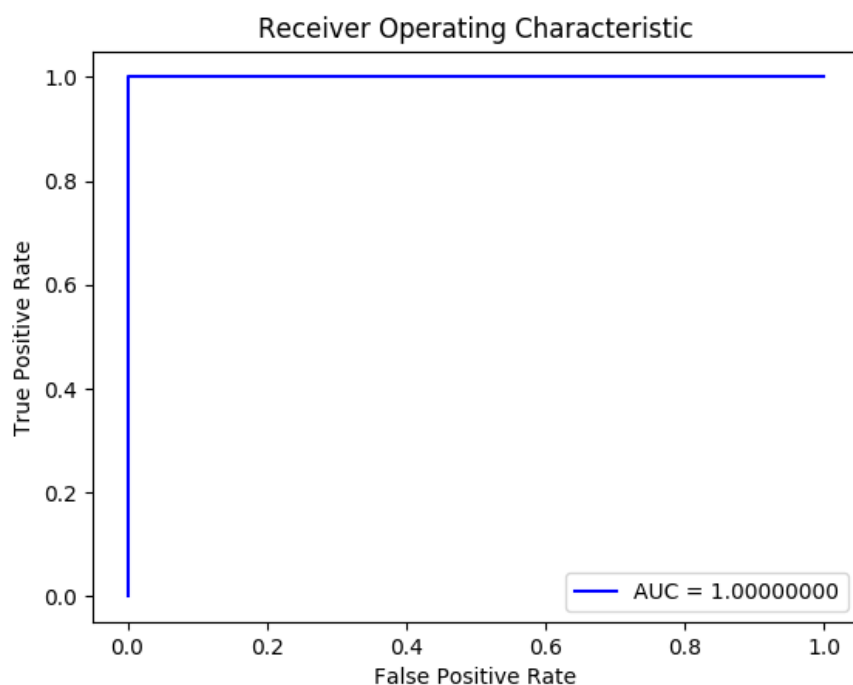Test Accuracy of class 0   98.14814814814815
Test Accuracy of class 1   98.14814814814815
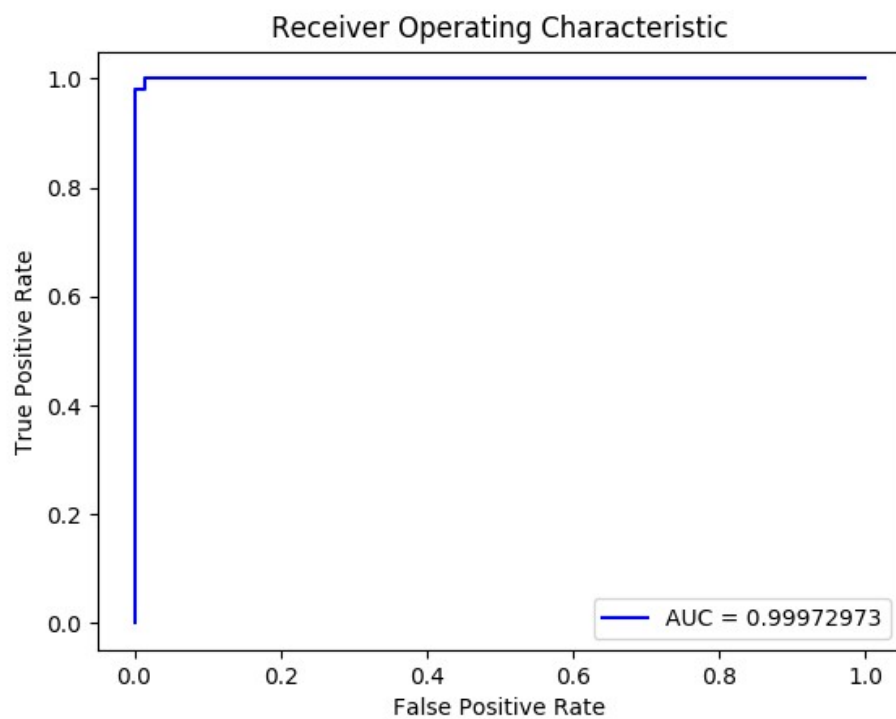Test Accuracy of class 2   100.0

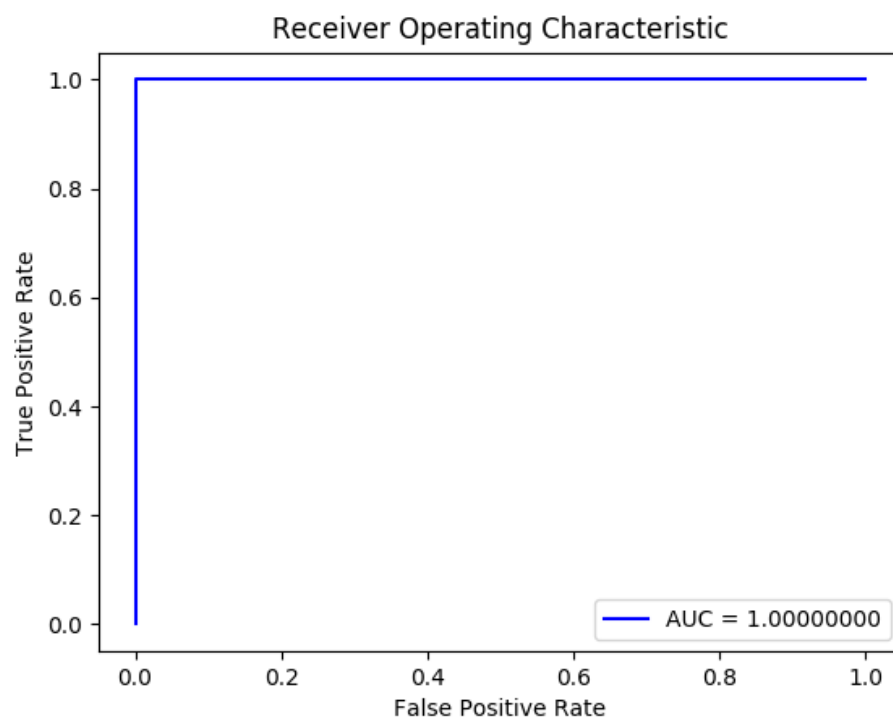### Roc Curve for OneVsAll:
For train Data:
Class0:

Class1:

Receiver Operating Characteristic

AUC = 0.99972973
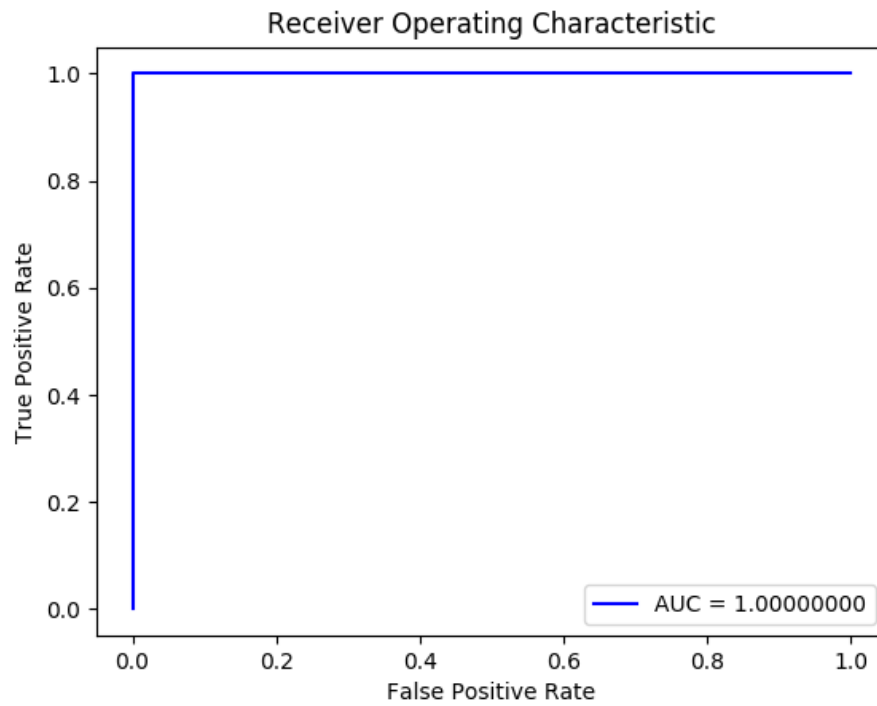
Class2:

Receiver Operating Characteristic

AUC = 1.00000000
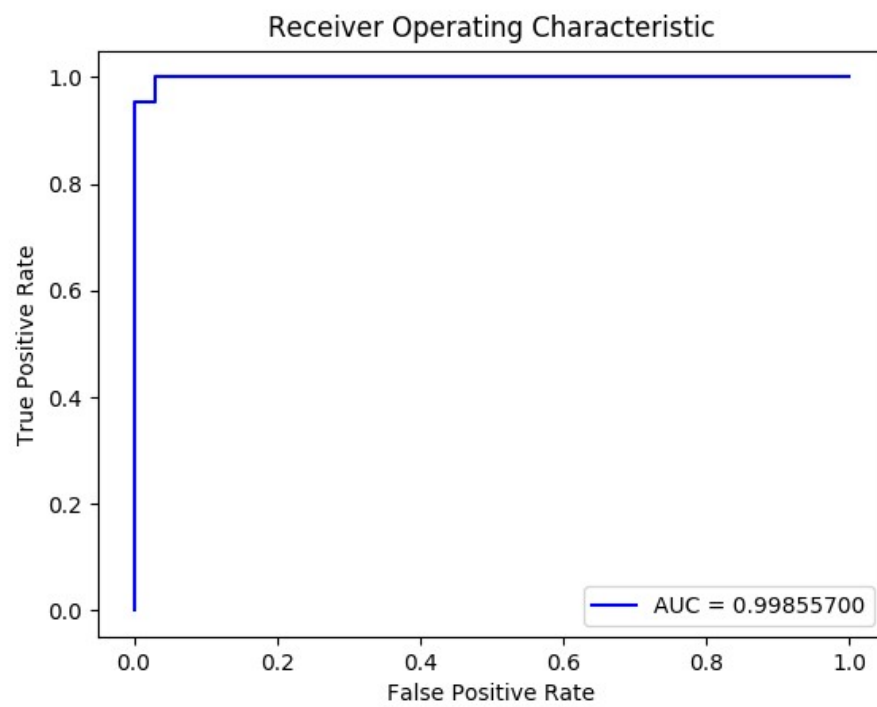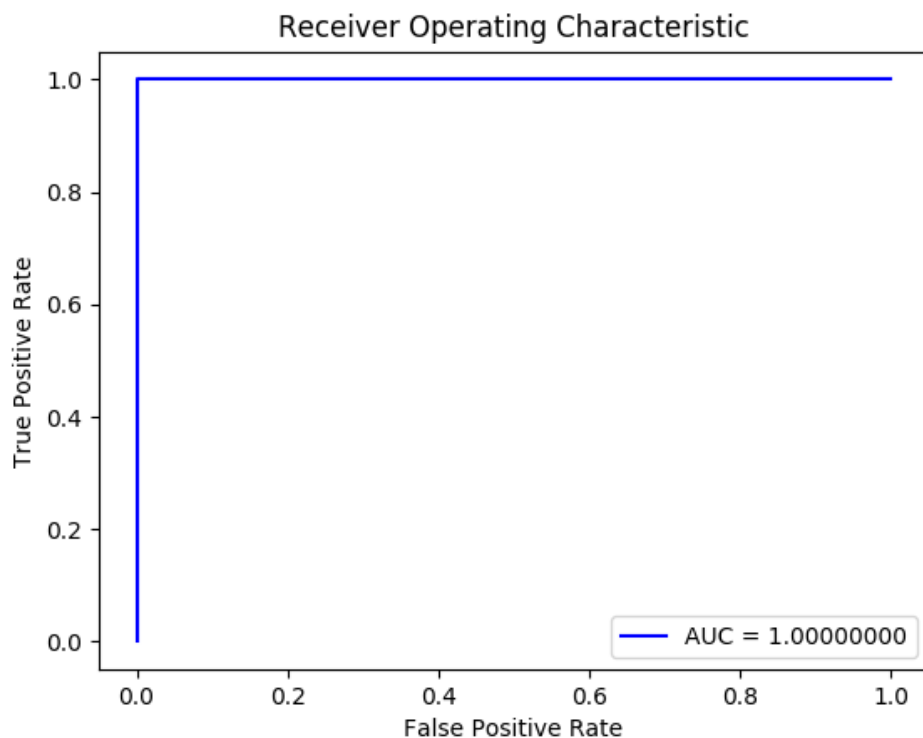
For test Data:
Class0:

Class1:

Class2:

b.Evaluation scores for OneVsOne are following

F-1 Score of OneVsOne using svm for train data: 0.9682259531719665
Accuracy of OneVsOne using svm for train data: 0.967741935483871
F-1 Score of OneVsOne using svm for test data: 0.9848406546080964
Accuracy of oneVsone using svm for test data: 0.9814814814814815

Classwise accuracy for train data:
Class 0 train data accuracy 0.9759036144578314
Class 1 train data accuracy 1.0

Class 2 train data accuracy 0.989010989010989

Classwise accuracy for test data:
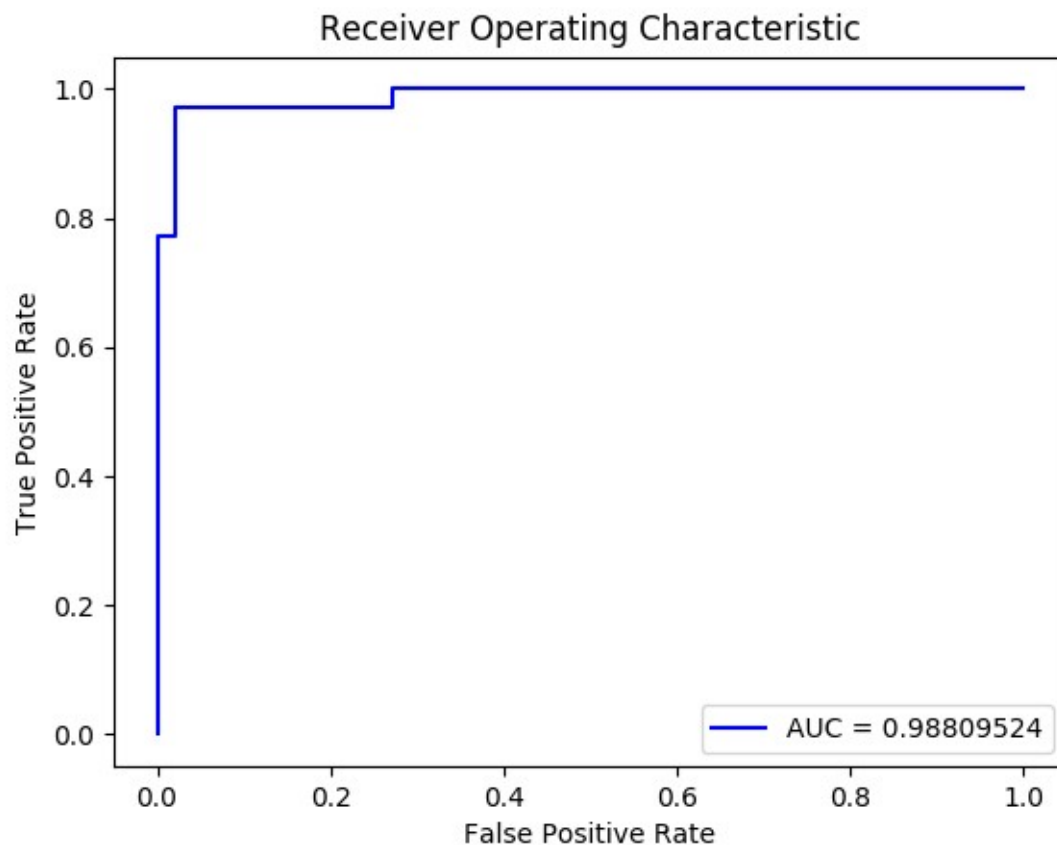Class 0 test data accuracy 0.9722222222222222
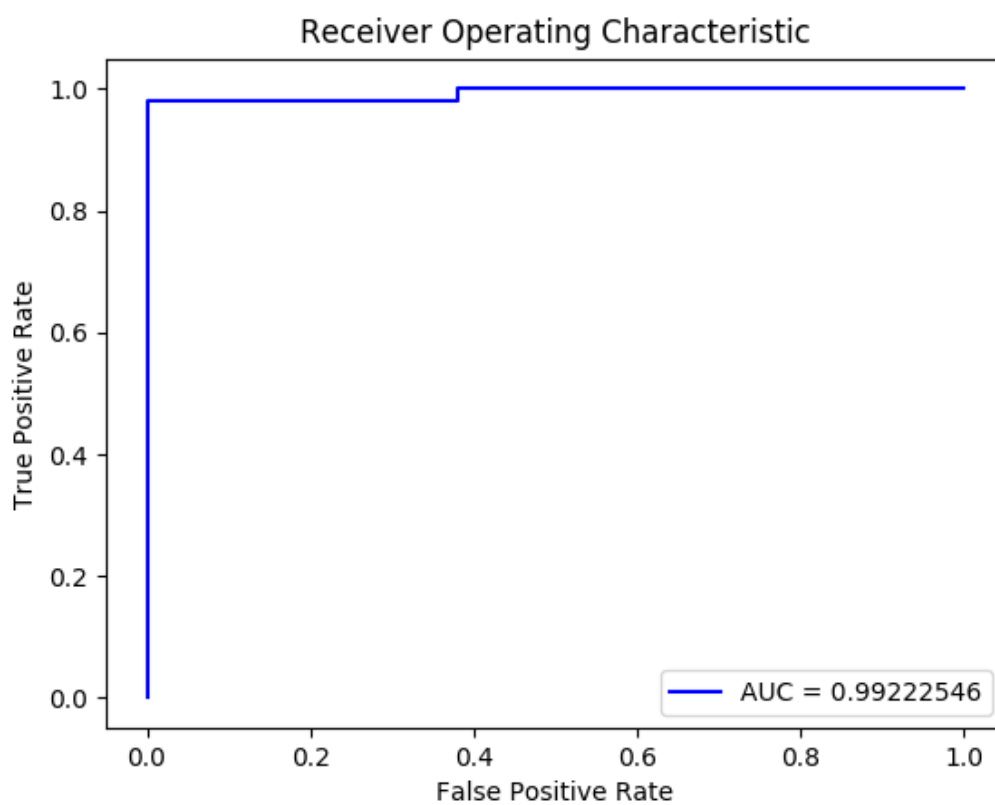Class 1 test data accuracy 1.0
Class 2 test data accuracy 1.0

Roc Curve for OneVsOne:

For train Data:
Class0:

Class1:

Receiver Operating Characteristic

True Positive Rate

False Positive Rate

AUC = 1.00000000

Receiver Operating Characteristic

True Positive Rate

False Positive Rate

AUC = 0.99222546
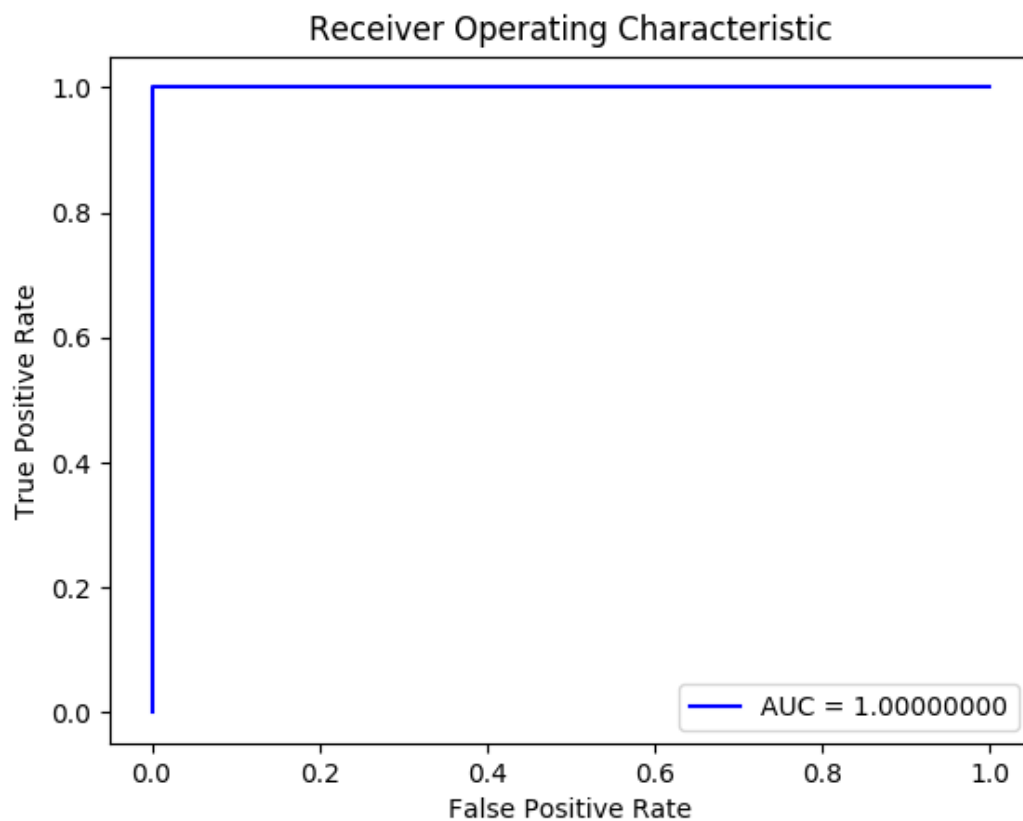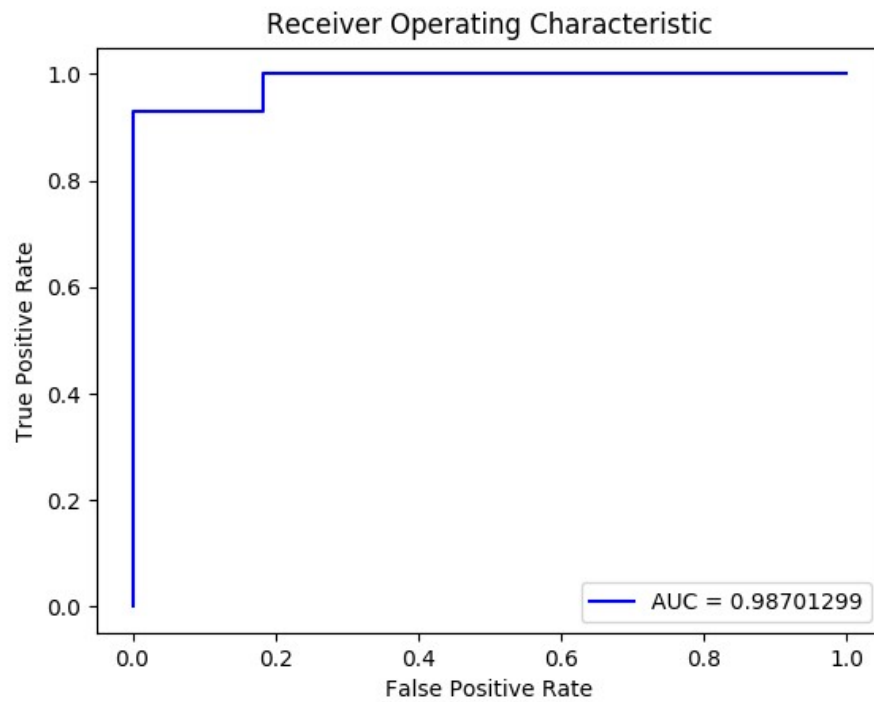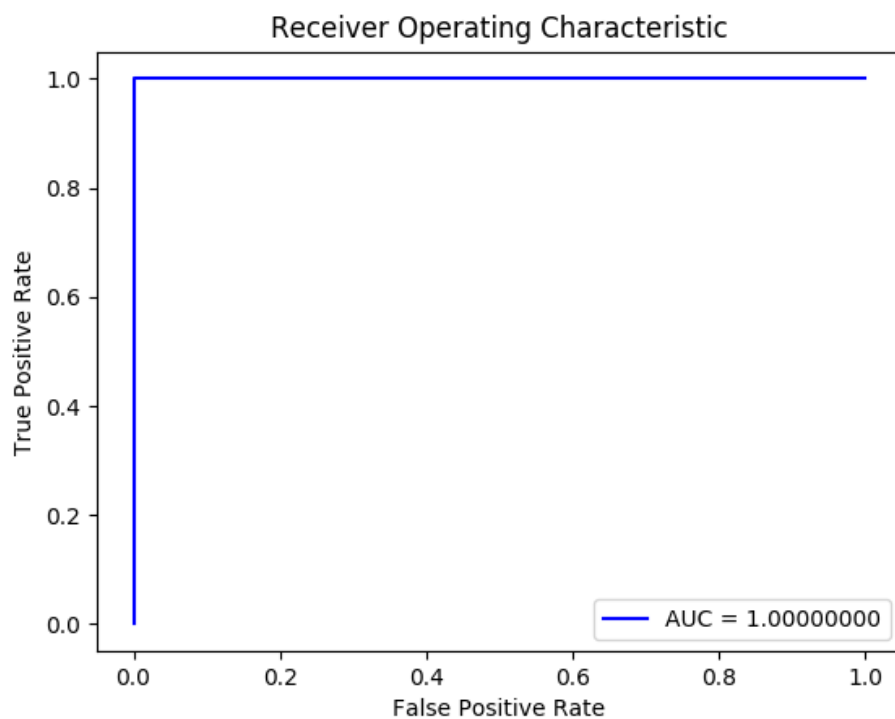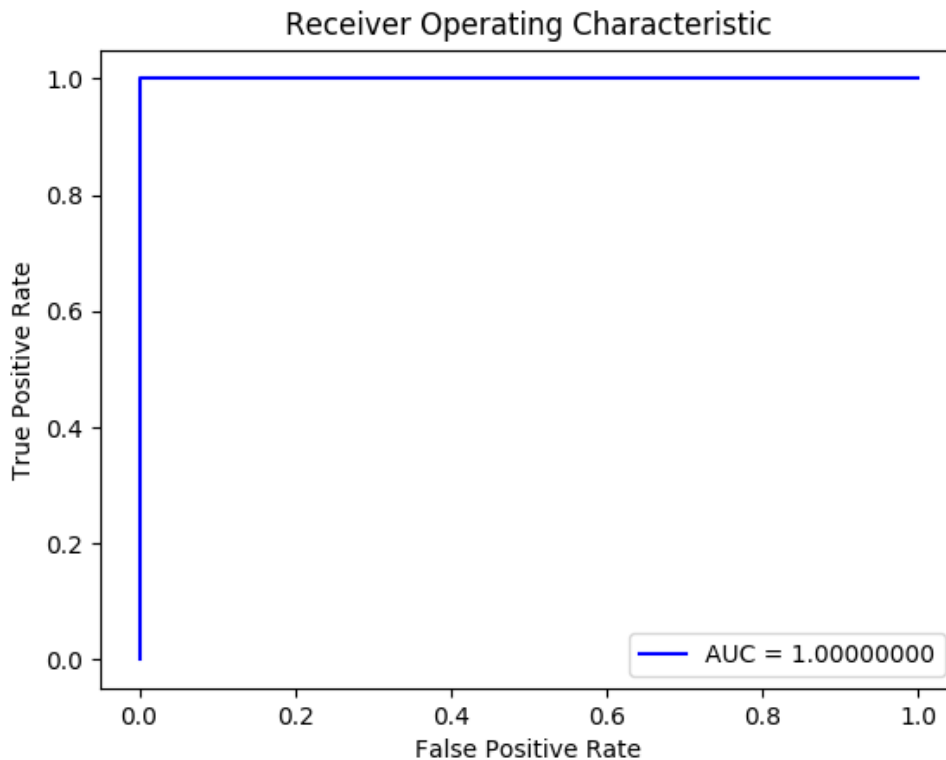
Class2:

For test Data:
Class0:

Class1:

Class2:

Observations:
(i)Time taken to fit SVM One-vs-Rest in seconds :1.997764
(ii)Time taken to fit SVM One-vs-One in seconds :0.280699
(iii)Model OneVsAll is giving more accuracy on train as well as test data as compare to OneVsOne so it is more prefered model to use.

3.Evaluation scores for Gaussian Naive Bayes are following:

Train data Accuracy of Gaussian Niave Bayes: 0.9758064516129032
Train data F-1 Score of Gaussian Niave Bayes: 0.9767210834795854
Test data Accuracy of Gaussian Niave Bayes: 1.0
Test data F-1 Score of Gaussian Niave Bayes: 1.0

Classwise accuracy for test data:
Train data Accuracy of Class 0 of Gaussian Niave Bayes: 0.95
Train data Accuracy of Class 1 of Gaussian Niave Bayes: 0.98
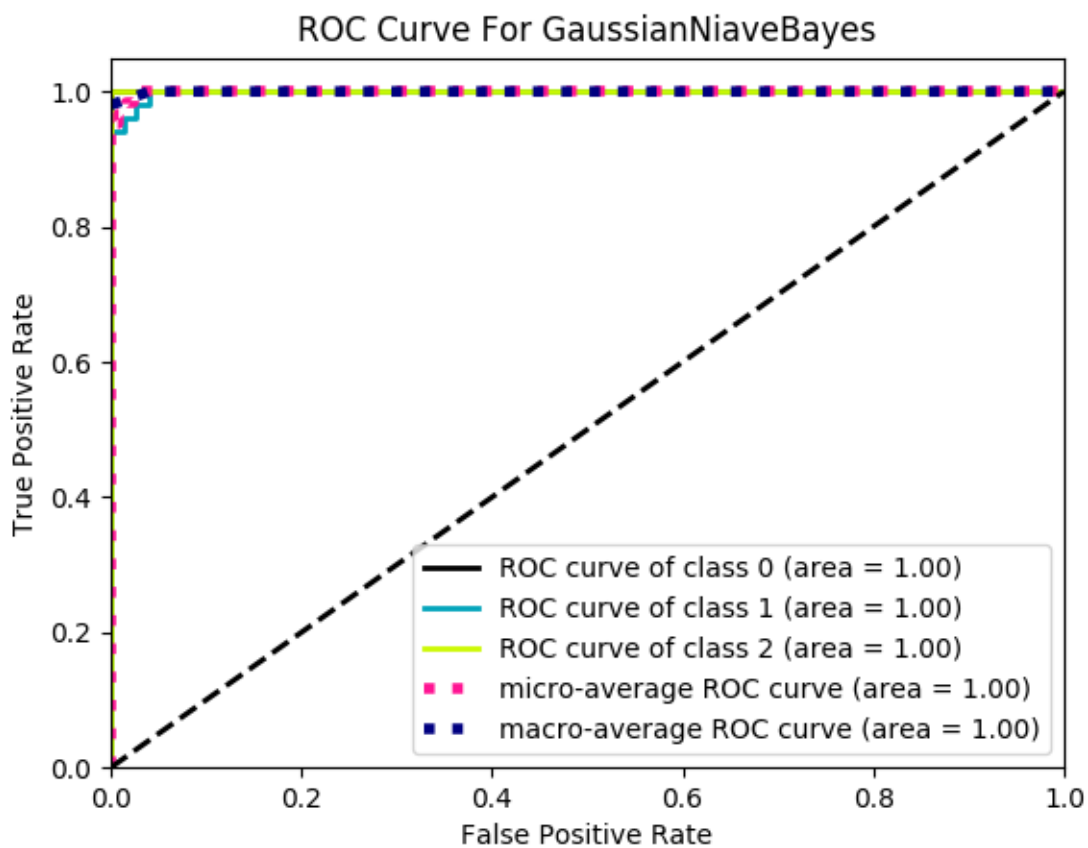Train data Accuracy of Class 2 of Gaussian Niave Bayes: 1.0

Classwise accuracy for test data:
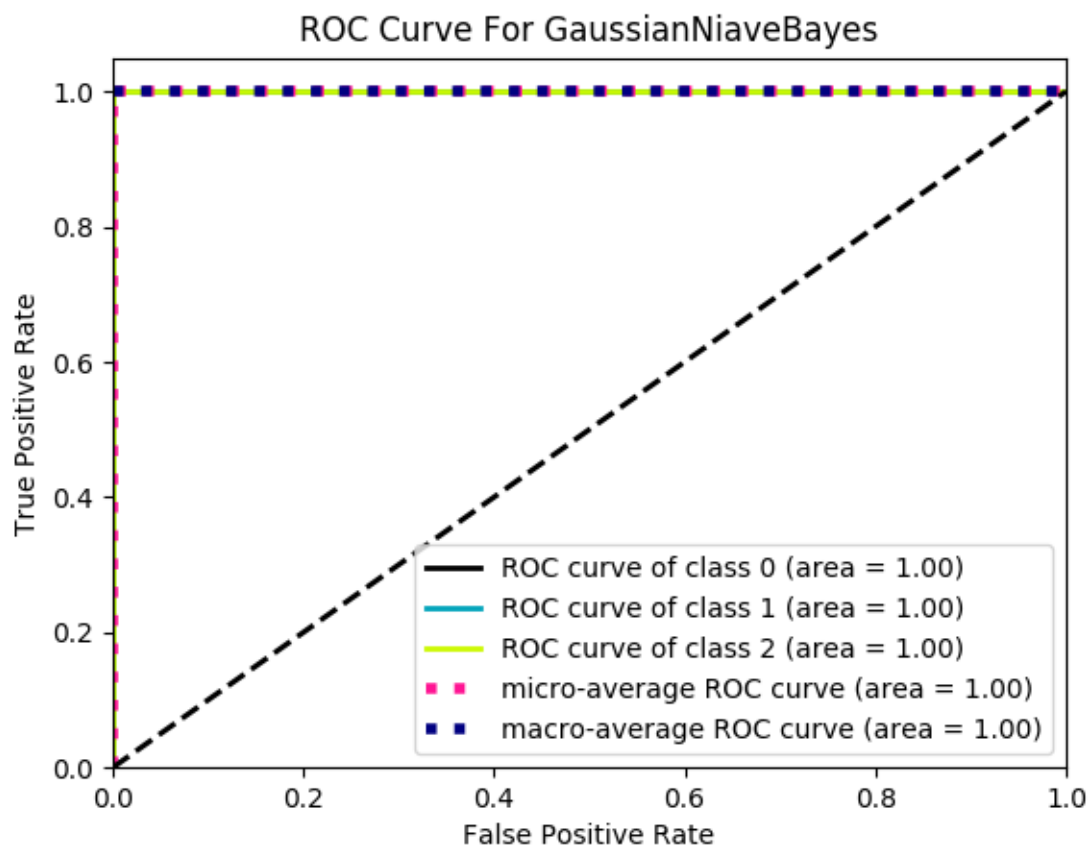Test data Accuracy of Class 0 of Gaussian Niave Bayes: 1.0
Test data Accuracy of Class 1 of Gaussian Niave Bayes: 1.0
Test data Accuracy of Class 2 of Gaussian Niave Bayes: 1.0
Roc Curve for Gaussian Niave Bayes on train data:

Roc Curve for Gaussian Niave Bayes on test data:

ROC Curve For GaussianNiaveBayes

Observations:

(i).Time taken to fit GaussianNiaveBayes model in seconds :0.000657
(ii).Accuracy of training and testing is high and also in test data it is predicting all the classes with accuracy.But Accuracy on test data is higher than train data it shows that our data should split in a magical way.We can eleminate it be random cross validation.

4.Evaluation scores for DecisionTree are following:

Hyperparameter Tuning using GridSearchCV:
Best Parameters:

'max_depth': 70
'max_features': 3
'min_samples_leaf': 2

'min_samples_split': 7

Accuracies:

Train data Accuracy of Decision Tree Classifier: 0.9274193548387096
Train data F-1 Score of Decision Tree Classifier: 0.9266331395044266
Classwise accuracy for train data:

Train data Accuracy of Class 0 of DecisionTree: 0.975
Train data Accuracy of Class 1 of DecisionTree: 0.94
Train data Accuracy of Class 2 of DecisionTree: 0.8529411764705882

Test data Accuracy of Decision Tree Classifier: 0.9074074074074074
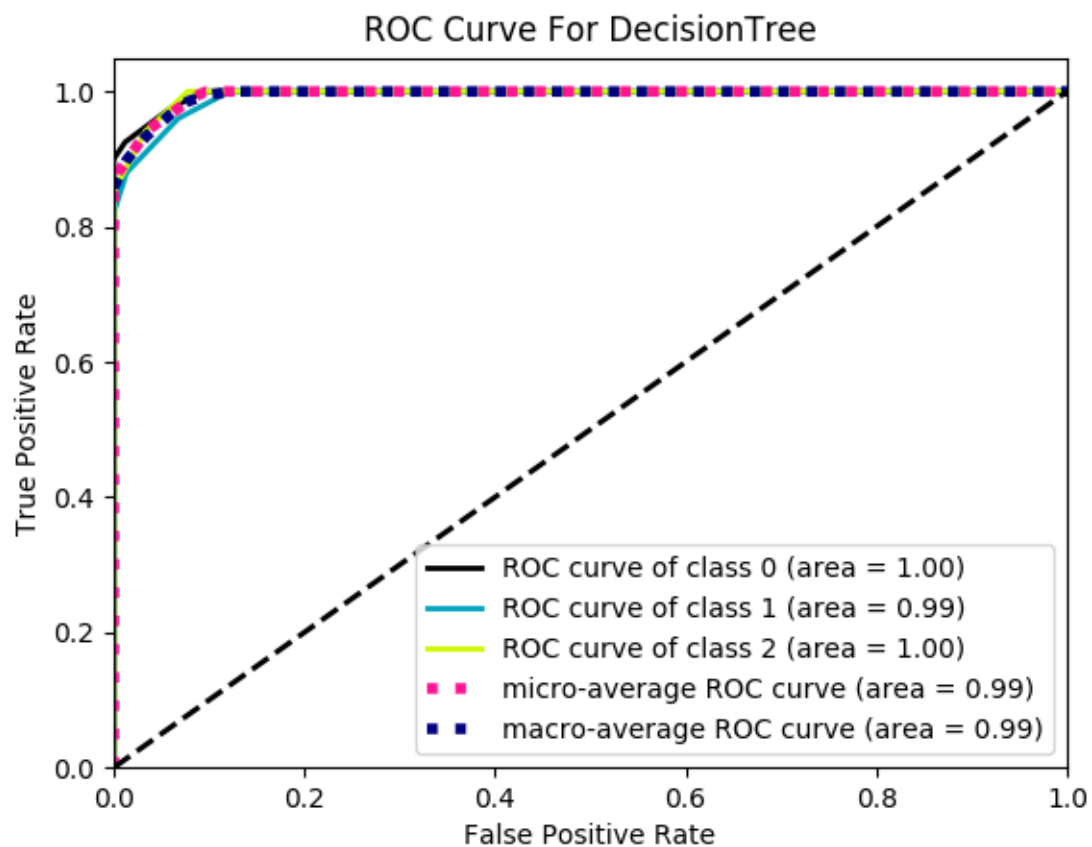Test data F-1 Score of Decision Tree Classifier: 0.9093183239524701
Classwise accuracy for test data:
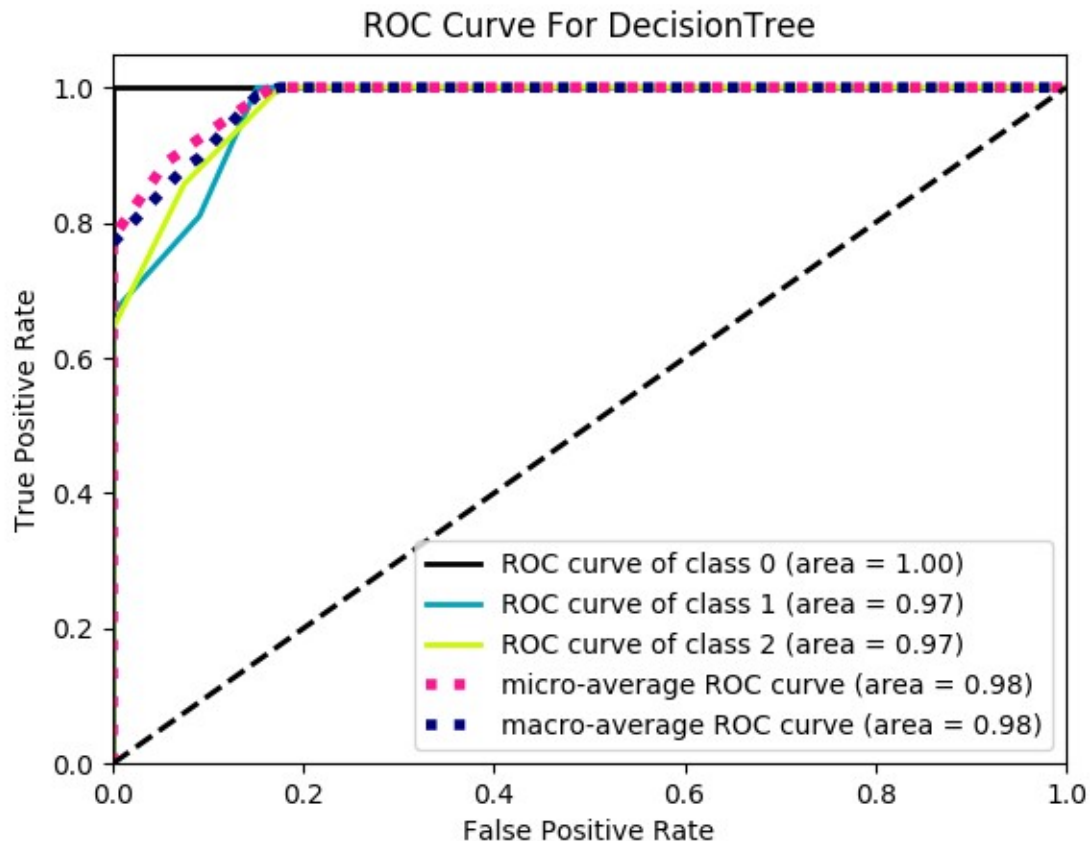Test data Accuracy of Class 0 of DecisionTree: 1.0
Test data Accuracy of Class 1 of DecisionTree: 0.9523809523809523
Test data Accuracy of Class 2 of DecisionTree: 1.0

Roc Curve for DecisionTree on train data:

ROC Curve For DecisionTree

ROC curve of class 0 (area = 1.00)	
ROC curve of class 1 (area = 0.99)	
ROC curve of class 2 (area = 1.00)	
micro-average ROC curve (area = 0.99)	
macro-average ROC curve (area = 0.99)	

Roc Curve for DecisionTree on test data:

ROC Curve For DecisionTree

Observations:

(i)Time taken to fit DecisionTree model in seconds :0.000024
(ii)Accuracy is significantly low than other models we can improve it by expension of parameters of GridSearchCv.

5.Comparision of above models:

(i)Overall Test accuracy sorted order of all models:

Gaussian Niave Bayes >SVM One-vs-Rest=>SVM One-vs-One>Decision Tree

(ii)Model fitting Time sorted order of all models:

Decision Tree<Gaussian Niave Bayes <SVM One-vs-One< SVM One-vs-Rest

(iii) After observing above data Gaussian Niave Bayes has high Test accuracy as well as significantly high Train accuracy so this is the best model for given dataset.
(iv) Gaussian Niave Bayes is taking more time than Decision Tree but less time than other models for best accuracy there we can do a trade off with time and still Gaussian Niave Bayes is best among all models.