Name: Samarth Manjunath

UTA ID:1001522809

Subject: Data Mining

Assignment-2

Naïve Bayes Classifier and ROC curve

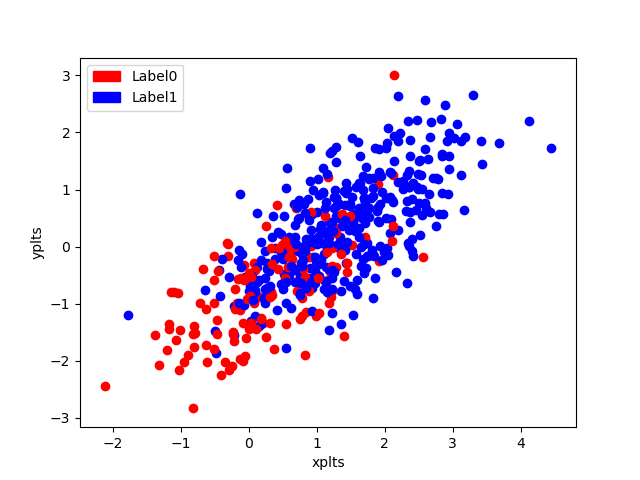
1. The program that I have designed takes in training data, trains the model based on the training data and finally predicts the labels to which the testing data belongs to.
2. The predicted data is this:



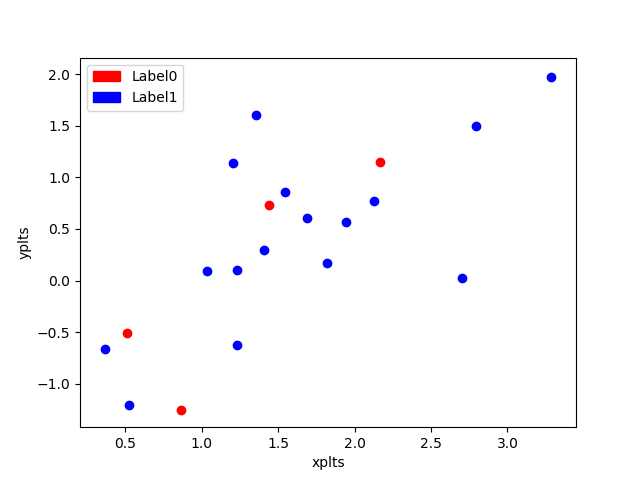
1. Parameters which had to be found is this:



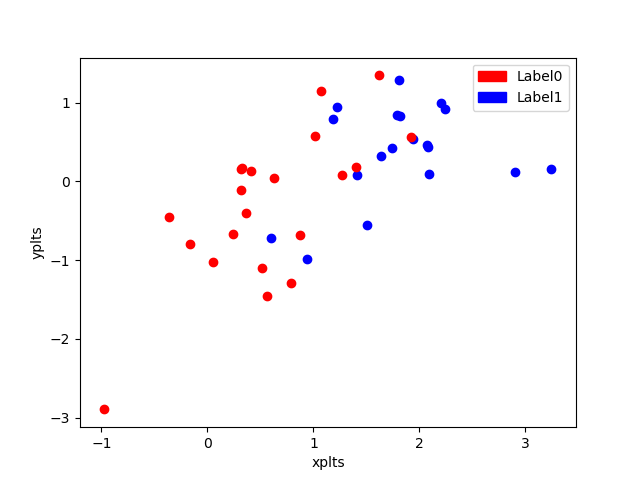
1. Scatter plot of train-set of 500 in each class is



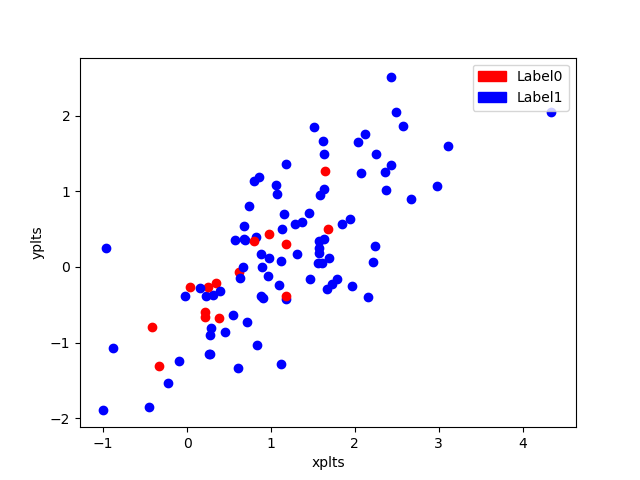
1. Scatter plot for train-set of 10 in each class is



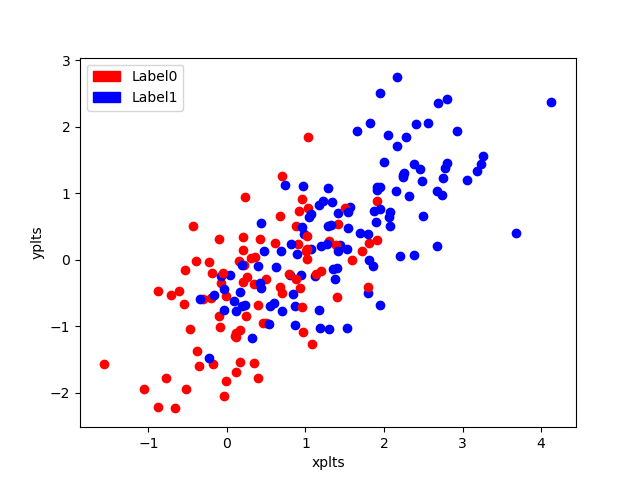
1. Scatter plot for train-set of 20 in each class is



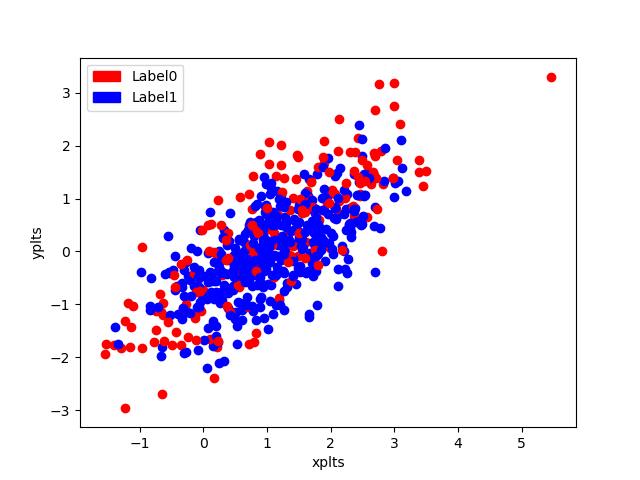
1. Scatter plot for train-set of 50 in each class is



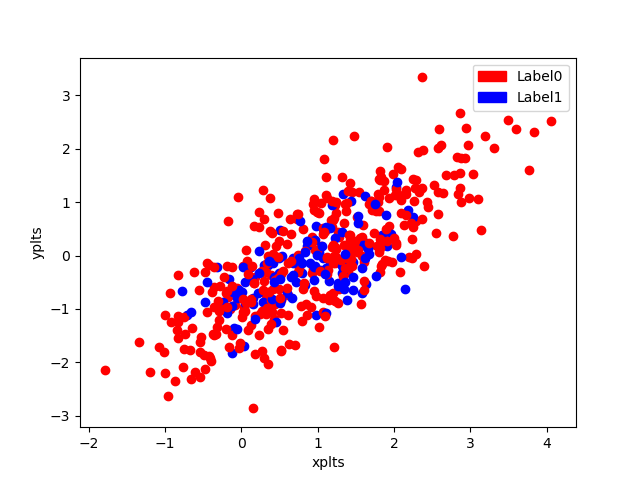
1. Scatter plot for train-set of 100 in each class is



1. Scatter plot for train-set of 600 in each class is

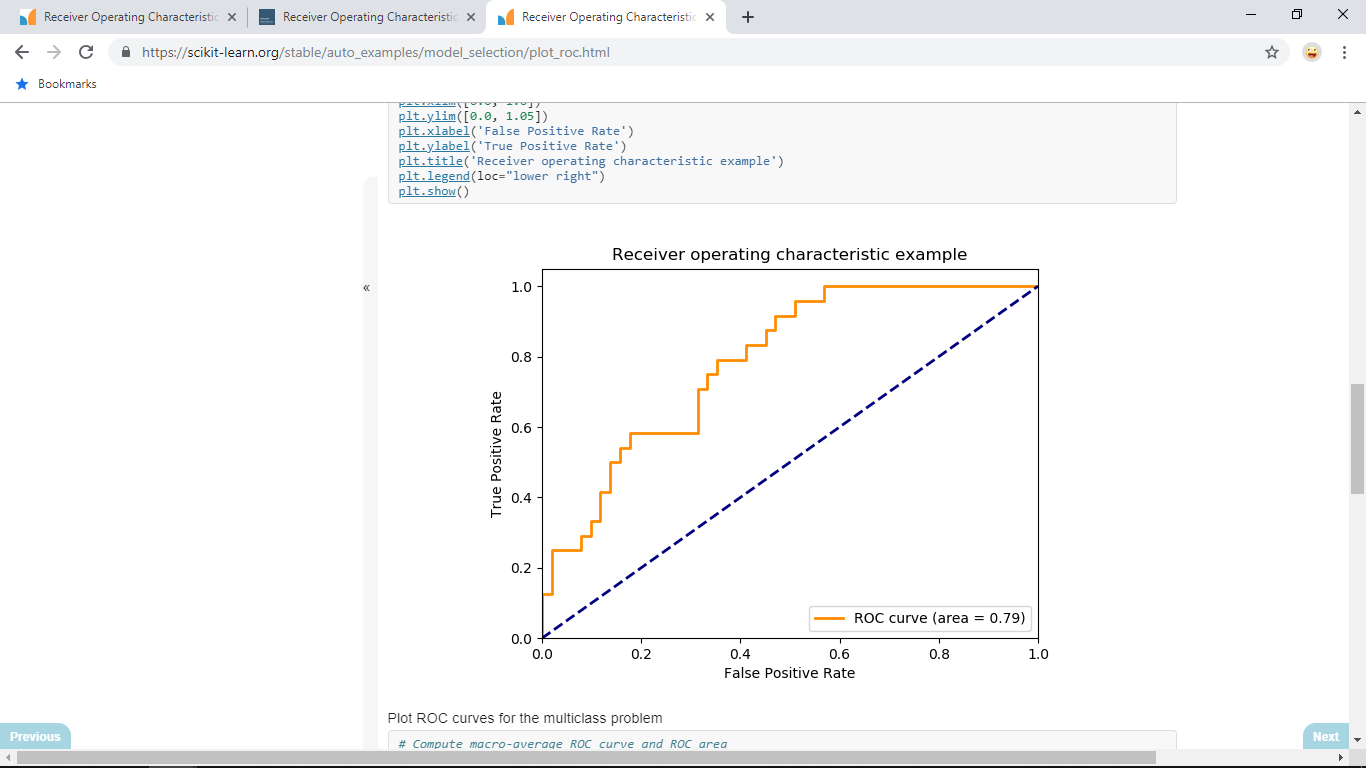


1. As an observation, the accuracy more or less decreases as the size of the training data set increases.
2. If suppose we change the train data set-0 as 700 and train-dataset-1 as 300, we get the following:



The accuracy decreases because there are more data points trained with respect to label 0 which is 700 than label 1 which is 300. The observations are plotted on scatter plot which is as above.

1. ROC curve



ROC curves typically feature true positive rate on the Y axis, and false positive rate on the X axis

ROC curves are typically used in binary classification to study the output of a classifier

The above shows an example of the same.