Assignment 1
Report-CS669

BY AYUSH DWIVEDI
V21093

MSc Applied Mathematics

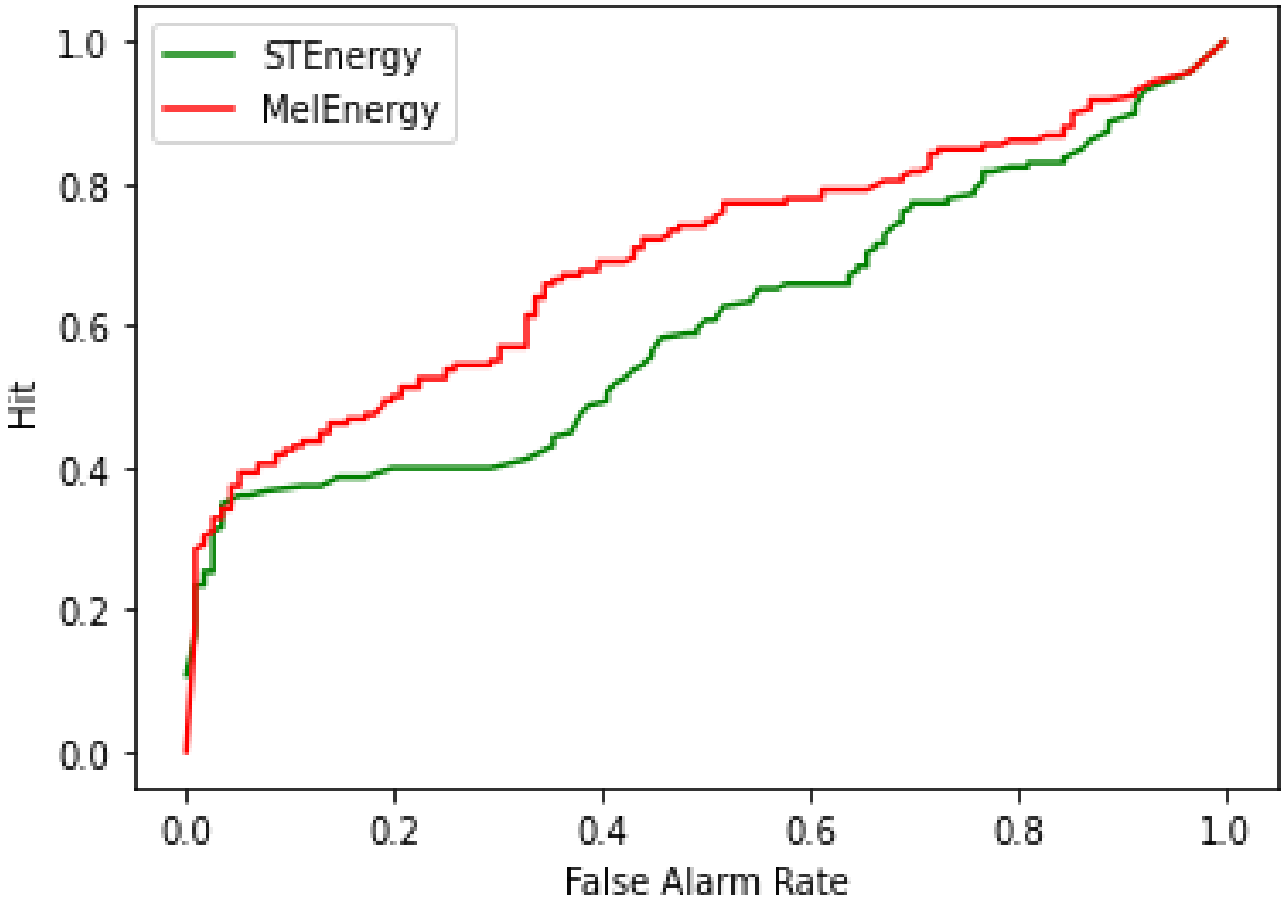
Question 1:

Signal Audio Detector:

We have been given with an audio file and two features for that audio file named STEnergy and MelEnergy and we also have training and testing data in segment2 and segment3 with class labels in Ground Truth file for training and testing data.

We have binary classes so we estimate the mean and variances for calculating the likelihood probabilities for Bayesian classifier with segment2 training data for the both the classes speech and non-speech. And then choose the 1000 threshold values for checking all the posterior probabilities and correspondingly calculate the False alarm rate and hit from the confusion matrix for chosen threshold .

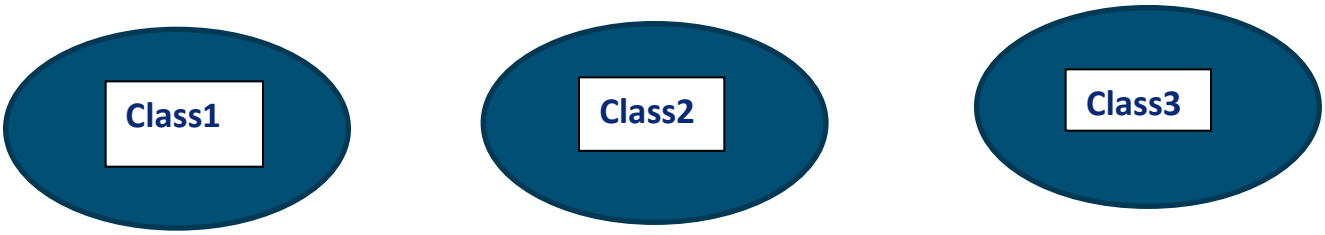
After that plot the ROC curve for both the features of 1000 points corresponding to each threshold values. Looking at the ROC Curve we get the the MelEnergy as the better feature than STEnengy due to high Hit values than corresponding False Alarm Rate.



Calculate: $P\left(\frac{s}{x}\right) = \frac{P\left(\frac{x}{s}\right) * P(s)}{P\left(\frac{x}{s}\right) * P(s) + P\left(\frac{x}{s'}\right) * P(s')}$ *for all the test data and test it with threshold values and then compare the predicted values with actual values . Form the confusion matrix and compute the TPR and FPR for corresponding threshold and plot it.*

Question:2

We have to fit the Bayesian Classifier for three classes and form four classifier as conditions mentioned in the question itself.



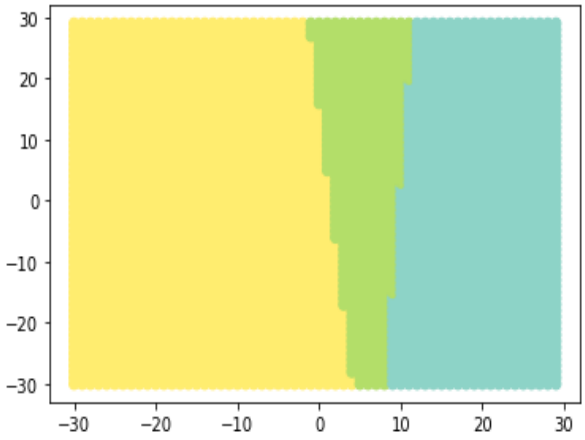
All the three classes contain the classes contain 2 dimensional data and divide your data first 50%-50% into training and testing data and calculate the mean and variance from the training data and add all your testing data into one . and pass through the discriminant functions of the classes and check whose probability is greater and assign it to that class.

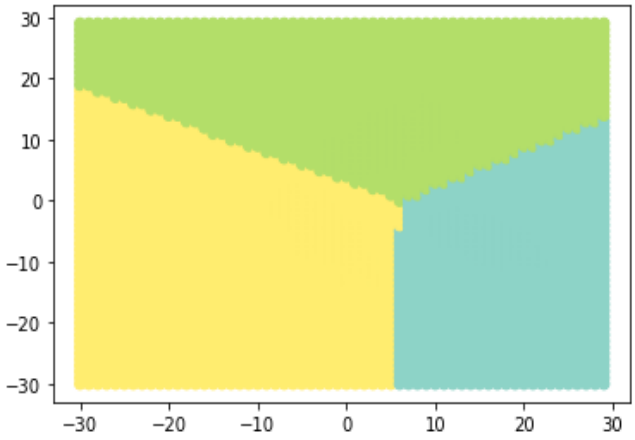
After assigning compare your predicted results with the actual values which you had already and compute all the values for the table by using the confusion matrix corresponding to each class and take their average the corresponding classifier.

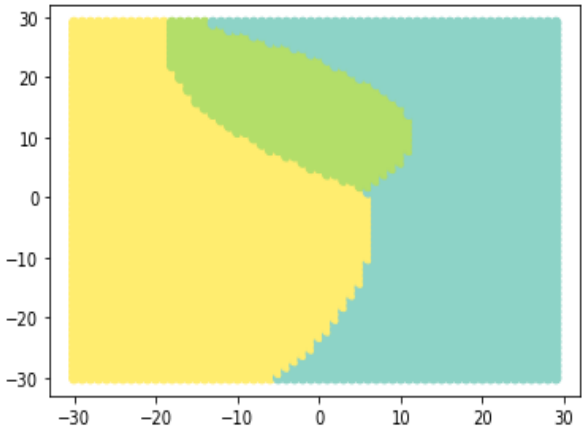
I am attaching my results for the linearly separable data for different classifiers and their decision boundaries.

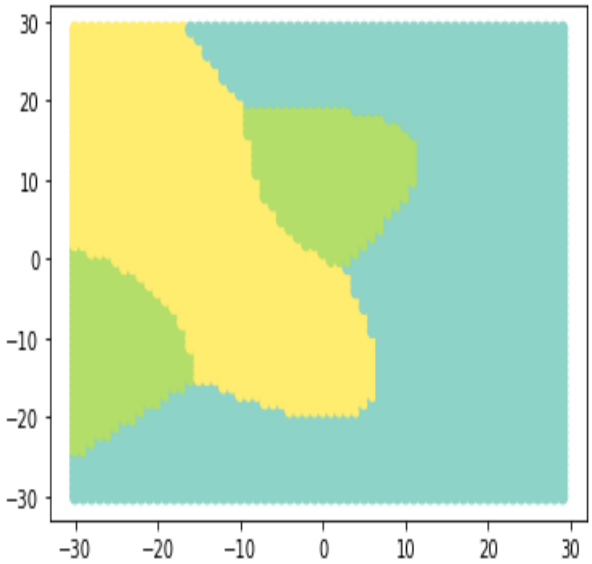
TABLE FOR LINEAR SEPARABLE DATA

Classifier	Accuracy	Precision	Recall	F-Score
C1	100	1	1	1
C2	99	0.98	0.98	0.98
C3	99.73	1	1	1
C4	99.87	1	1	1


30
20
10
0
−10
−20
−30
−30 −20 −10 0 10 20 30


30
20
10
0
−10
−20
−30
−30 −20 −10 0 10 20 30

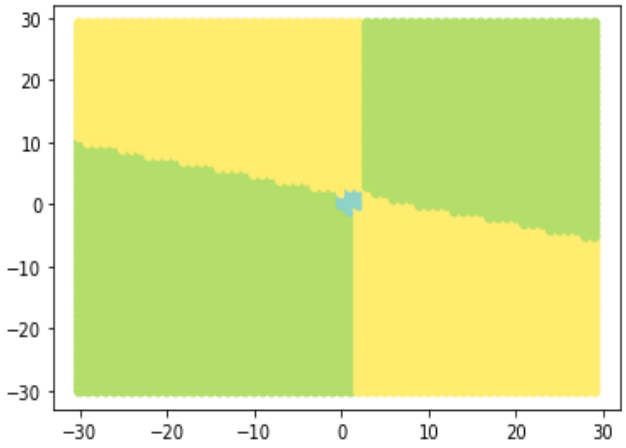

30
20
10
0
−10
−20
−30
−30 −20 −10 0 10 20 30

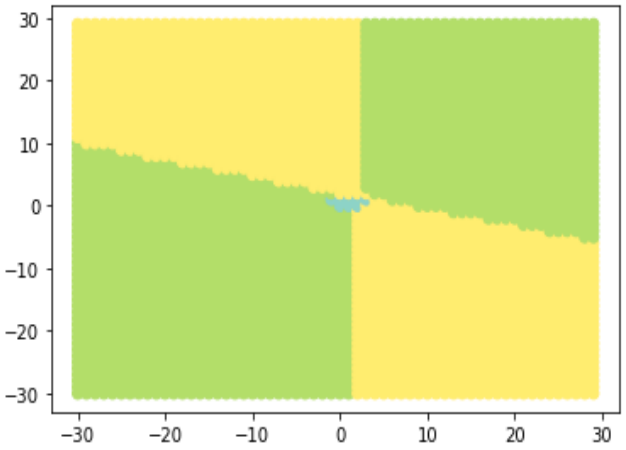

30
20
10
0
−10
−20
−30
−30 −20 −10 0 10 20 30

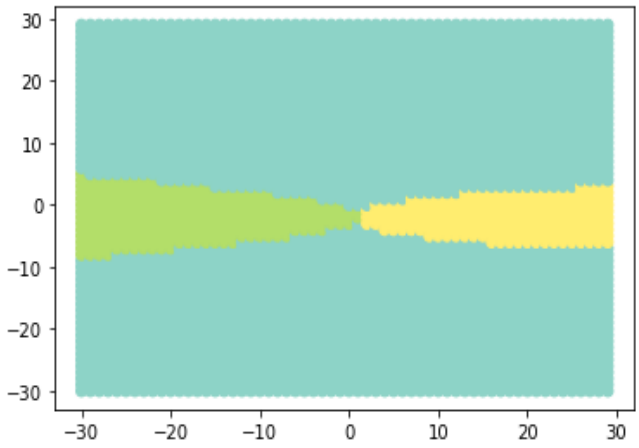
Similar steps you will have to repeat for the non-linearly separable data. Their corresponding table:

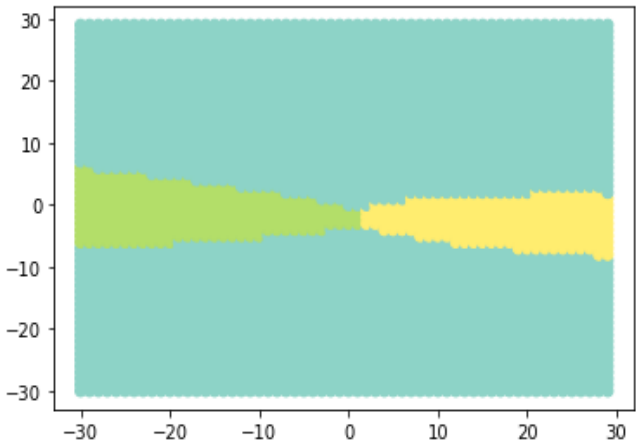
TABLE FOR NON_LINEAR DATA

Classifier	Accuracy	Precision	Recall	F-Score
C1	81.2	0.82	0.81	0.81
C2	90	0.9	0.9	0.9
C3	89.6	0.9	0.9	0.9
C4	89.47	0.9	0.89	0.9









Please Find the Attached Code on Google colab.

CLICK HERE