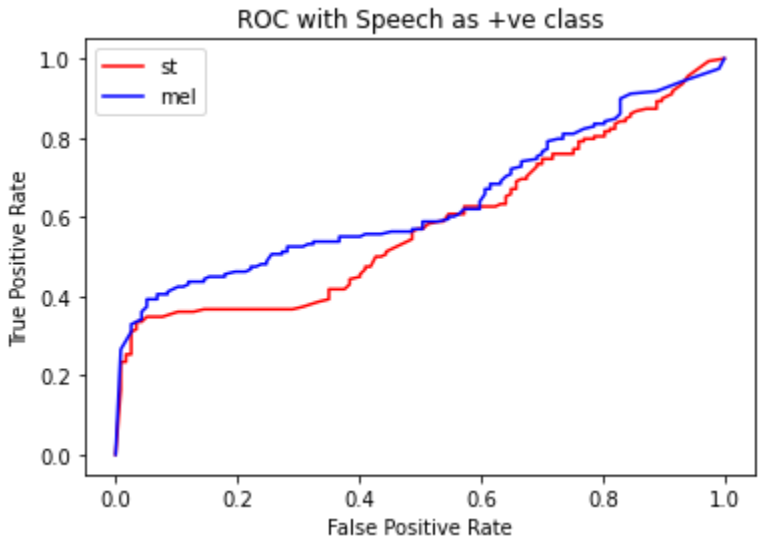
CS 669 / DS 403

Name - Mohit Verma
Roll No - B20215

Assignment 1

Google Colab Link -
https://colab.research.google.com/drive/194SIbYfY9aw5vlnG_K98muV-rxLJQ1xZ?usp=sharing

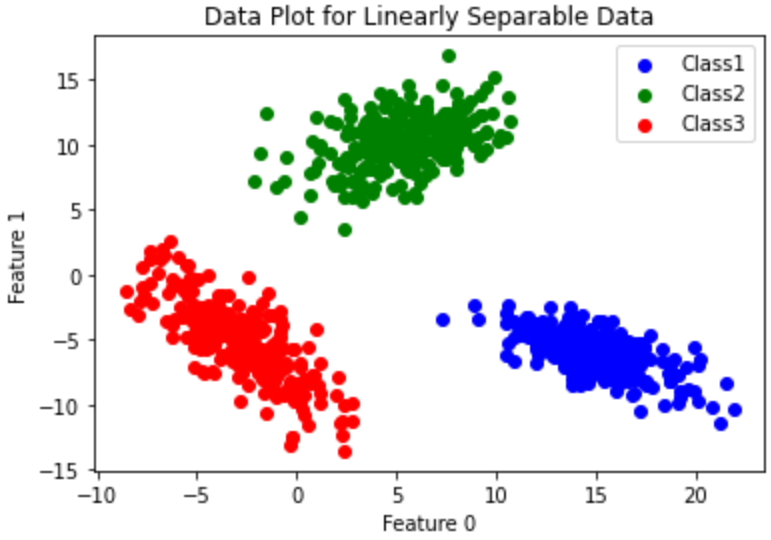
Q1)



Area of ROC mel-energy curve **0.635264.**
Area of ROC ST-energy curve **0.5813859.**

1. From the graph we can observe that the data doesn't follow the unimodal gaussian distribution. Hence, the accuracy we observe from the characteristic features is very poor.
2. We can also observe that the area of the ROC curve for MelEnergy is higher than the STEnergy curve. This means that the former is better than the latter for testing purposes.

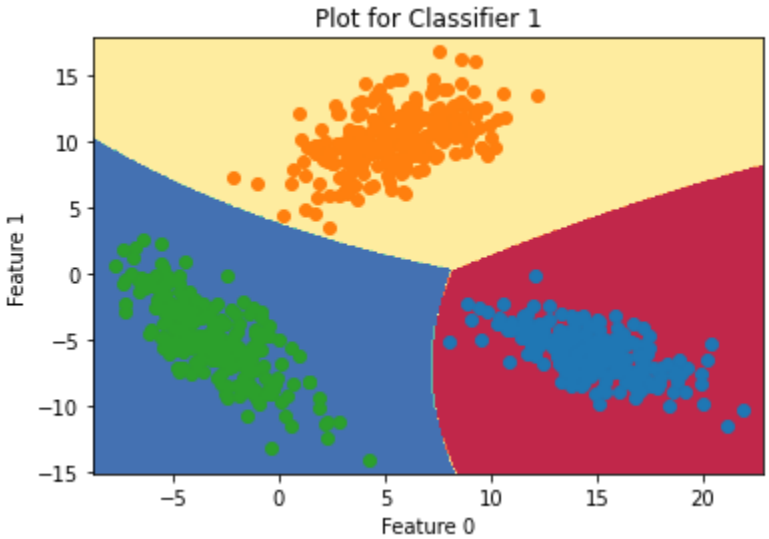
Q2)



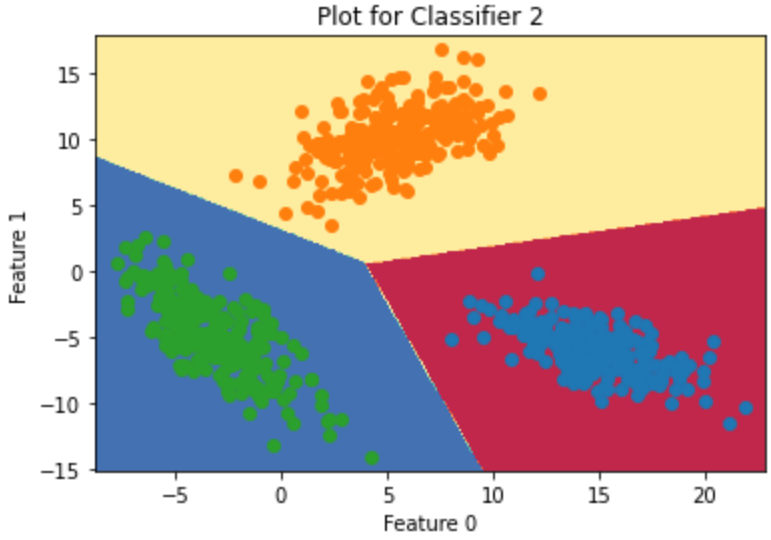
A Scatter Plot for the 3 Classes Data.

Metrics Table for the Linearly Separable Data

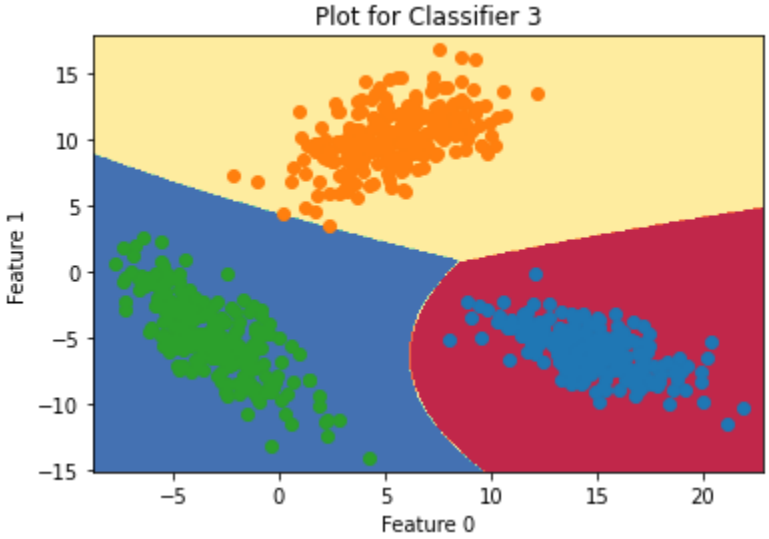
Classifier	Accuracy	Precision	Recall	F-Score
C1	1.0	1.0	1.0	1.0
C2	1.0	1.0	1.0	1.0
C3	0.9991	0.9986	0.9986	0.9986
C4	1.0	1.0	1.0	1.0



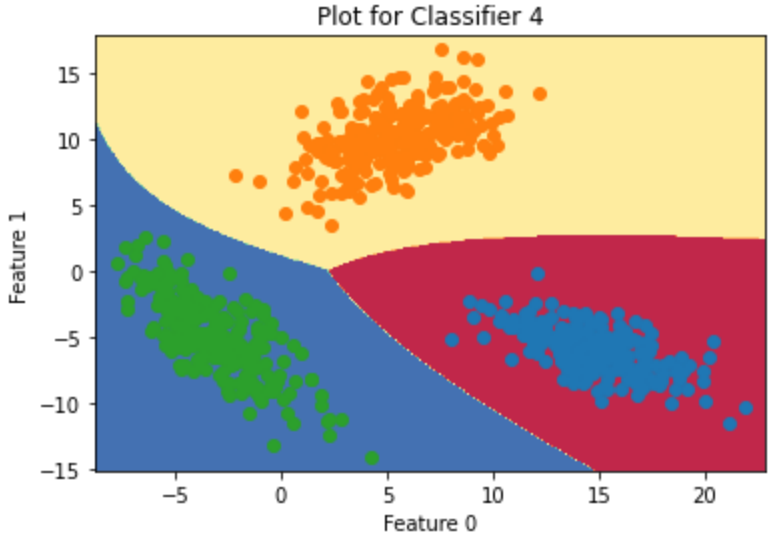
Decision Boundaries for Classifier 1



Decision Boundaries for Classifier 2



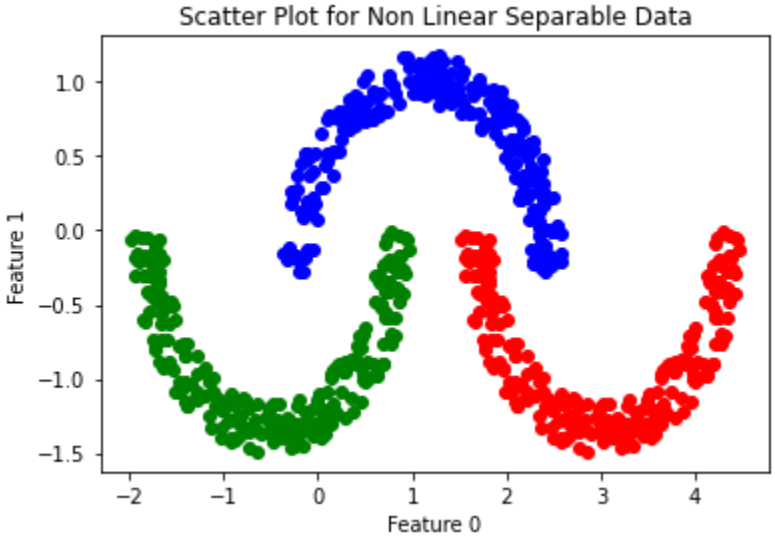
Decision Boundaries for Classifier 3



Decision Boundaries for Classifier 4

Inferences on Linearly Separable Data

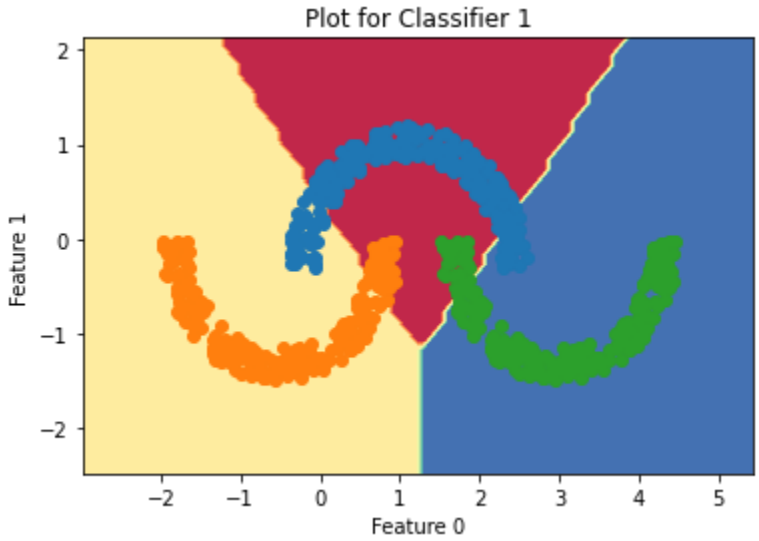
1. The Accuracy of all the classifiers is very close to 100%.
2. This high accuracy of all the classifiers suggests that the data for each of the classes is normally distributed.
3. Also, we can observe that the decision boundaries are able to perfectly separate the data points of any of the two classes.



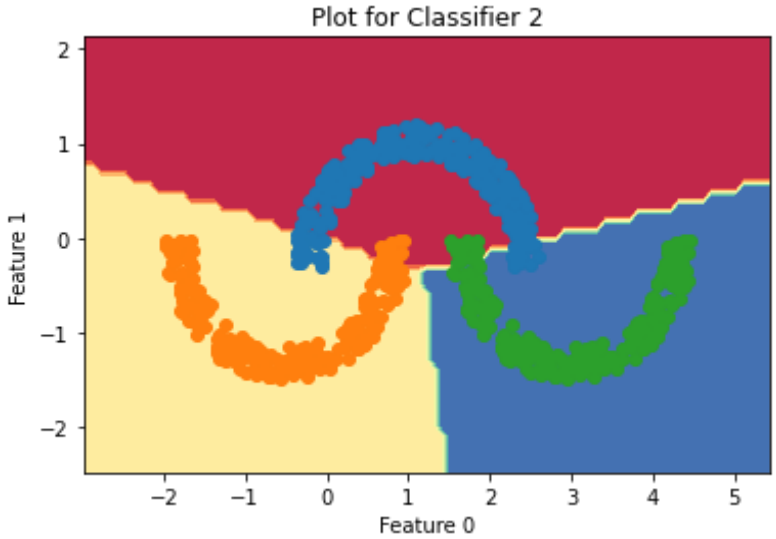
Scatter plot for Non Linearly Separable Data

Metrics Table for Non Linearly separable data

Classifier	Accuracy	Precision	Recall	F-Score
C1	0.872	0.808	0.808	0.808
C2	0.936	0.904	0.904	0.904
C3	0.938	0.908	0.908	0.908
C4	0.937	0.906	0.906	0.906

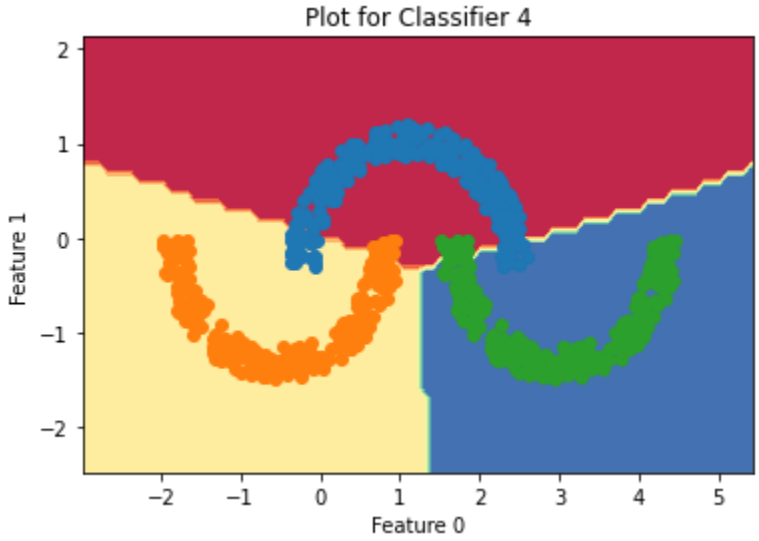


Decision Boundaries for Classifier 1



Decision Boundaries for Classifier 2

Decision Boundaries for Classifier 3



Decision Boundaries for Classifier 4

Inferences on Non Linearly Separable Data

1. We can observe that the accuracy of classifiers is around 90% which is still good.

2. This also suggests that the data is normally distributed as in the case of linearly separable data.
3. The decision boundaries are also not perfectly separating data of the two classes which is evident from the plots and the accuracy obtained.