

Pattern Recognition

Assignment 2

Group No - 36

ED13D000 Bhargava Sai Ramu
CS17S008 Nitesh Methani

October 13, 2017

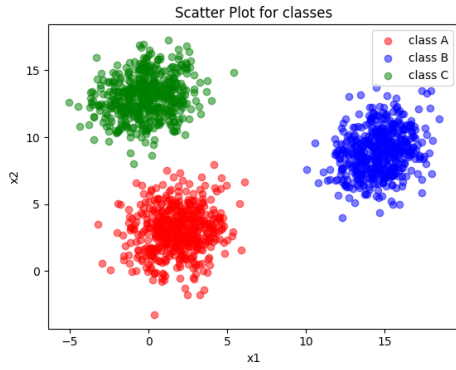
Contents

1	Introduction	1
2	Linearly Separable Data-set	2
3	Non-Linearly Separable Data-set	3
4	Linearly Non-Separable Data-set	4
5	Observations:	5

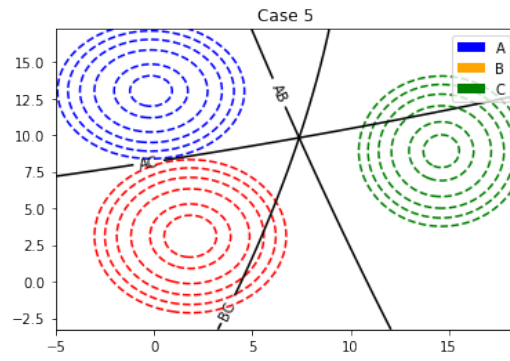
1 Introduction

We have three datasets to work with. One dataset is linearly separable that means all the classes can be separated using linear lines. Second dataset is Non-Linearly Separable dataset that means we should get non linear decision boundaries between each class. Third dataset is Linearly Non Separable dataset it means data cannot be separated using linear lines.

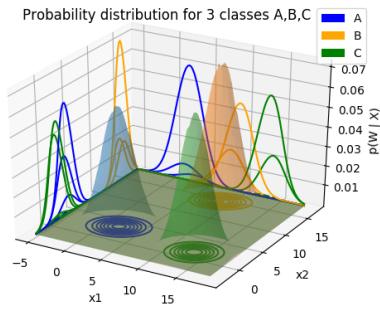
2 Linearly Separable Data-set



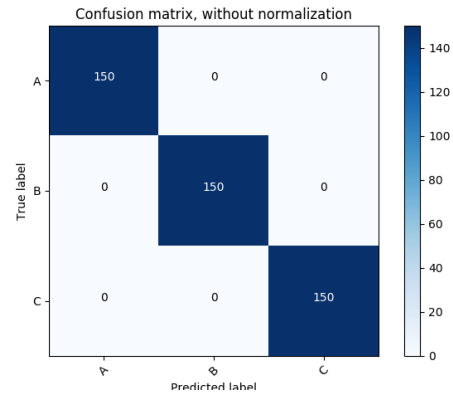
(a) Scatter Plot for each class



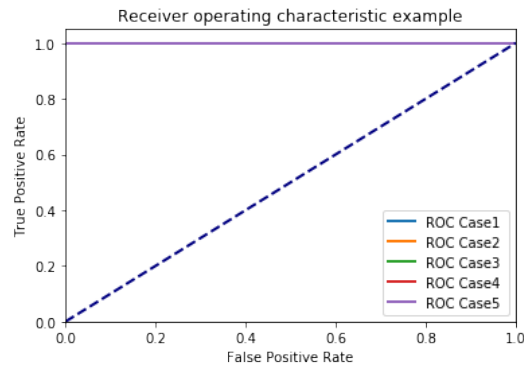
(b) Constant Density Curves and Decision Boundaries



(c) Gaussian Distributions for each class and feature



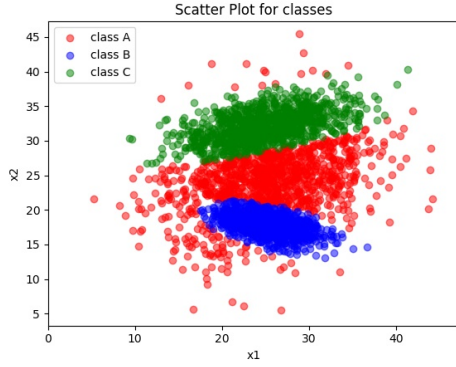
(d) Confusion Matrix



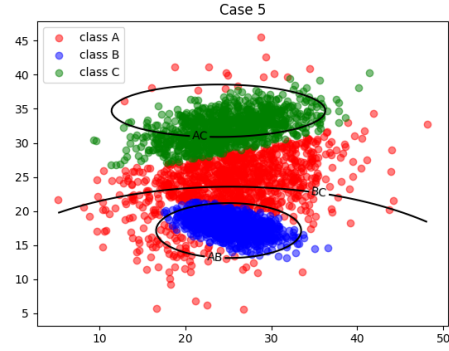
(e) ROC Curve

Figure 1: Summary of Linearly Separable Dataset

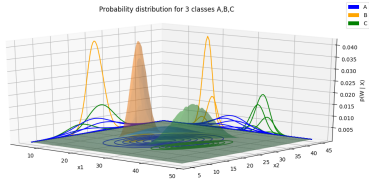
3 Non-Linearly Separable Data-set



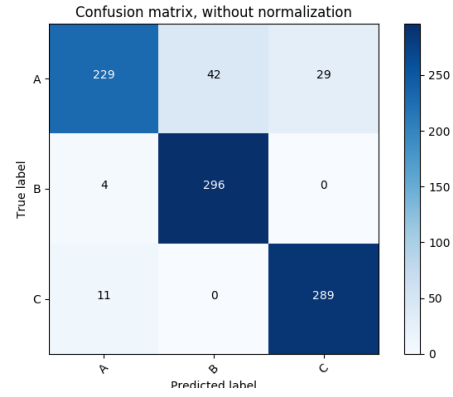
(a) Scatter Plot for each class



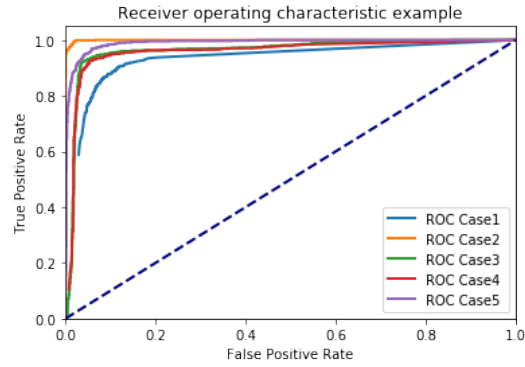
(b) Constant Density Curves and Decision Boundaries



(c) Gaussian Distributions for each class and feature



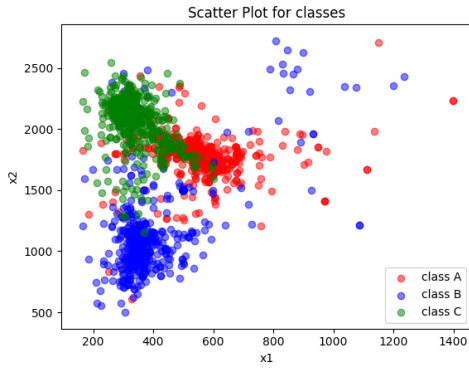
(d) Confusion Matrix



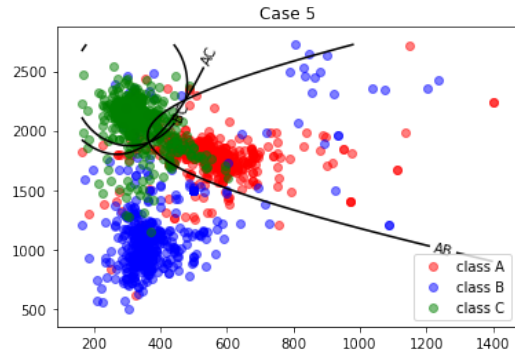
(e) ROC Curve

Figure 2: Summary of Linearly Non Separable Dataset

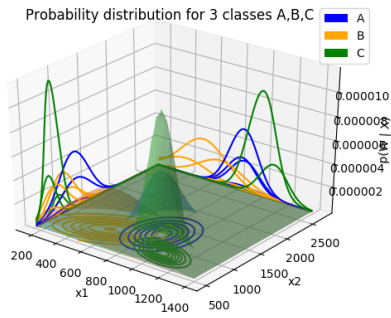
4 Linearly Non-Separable Data-set



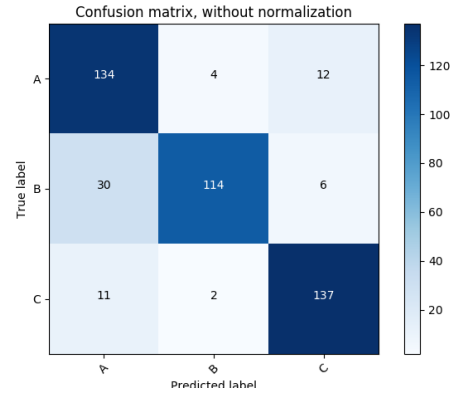
(a) Scatter Plot for each class



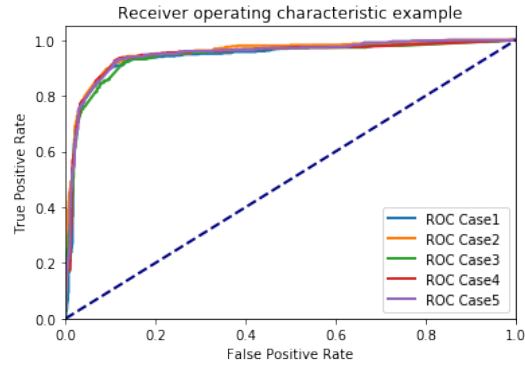
(b) Constant Density Curves and Decision Boundaries



(c) Gaussian Distributions for each class and feature



(d) Confusion Matrix



(e) ROC Curve

Figure 3: Summary of Non Linearly Separable Dataset

5 Observations:

- 1) If the constant density curves are circles that means the covariance matrix for all the classes is same which is nothing but our case 1 and case 4.
- 2) Eigen vectors are perpendicular to the lines joining the means of each distribution in case 1 and case 4.
- 3) Whereas for other cases when covariance matrix for all the classes is different, the constant density curves are ellipsoids.
- 4) Eigen vectors of these hyper-ellipsoids shows the distribution of the data points along the principal axes of the ellipsoid.
- 5) If I move along the major axis or minor axis density goes on reducing or increasing depending on the case.
- 6) The decision boundaries are not perpendicular in case 2 and case 5.
- 7) The decision boundaries are not perpendicular in case 1.
- 8) The nature of the ROC and DET curves that we got match with the theory we studied in class.
- 9) The confusion matrix also justify the metrics.
- 10) For real data, all the points are not properly classified since decision boundaries are not able to make distinction between all the three classes so accuracy and other metrics are not upto the mark in this case.
- 11) For linearly separable dataset and non-linearly separable dataset the accuracy is pretty high.