CS6690 : PATTERN RECOGNITION

ASSIGNMENT 2

September 20, 2017

Group No : 12
Kanti Kumari and Sweta Kumari
Indian Institute of Technology Madras

0.1 Bayesian Classifier

It is a statistical classification model which classifies the data points based on the probability of the class which it belongs to using mean and covariance of the classes. This technique is based on Bayes' theorem and it calculates the likelihood of data points using Gaussian density function.
In the terms of difference between Bayes' and Naive Bayes classifiers, Naive Bayes' assumes independence among the feature vectors while Bayes' does not. Following are the cases which were used for classification of different types of provided datasets :

1. Bayesian classification model with same Covariance matrix for each class.
2. Bayesian classification model with different Covariance matrix for each class.
3. Naive Bayes classification model with covariance as σ^2*I.
4. Naive Bayes classification model with same covariance matrix for each class.
5. Naive Bayes classification model with different covariance matrix for each class.

0.2 Results

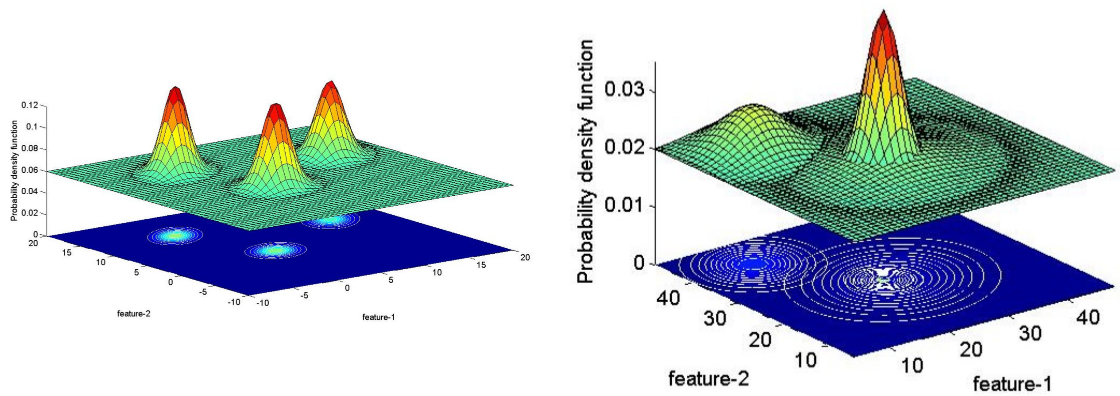

Figure 1: PDF (Gaussians) and their contours (constant density curves) for linearly (left) and non-linearly separable data(right).

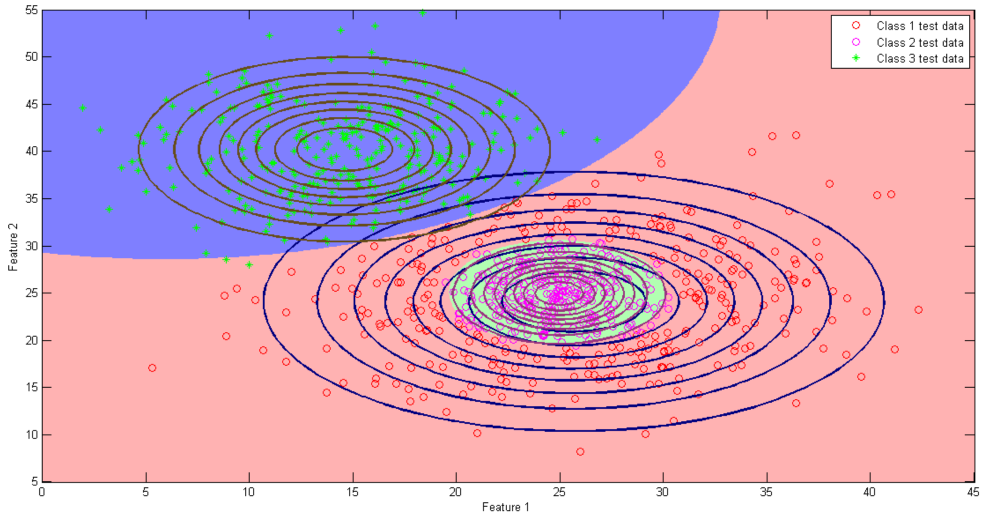

Figure 2: Descision boundary, decision surface and constant density curve for non linearly separable data.

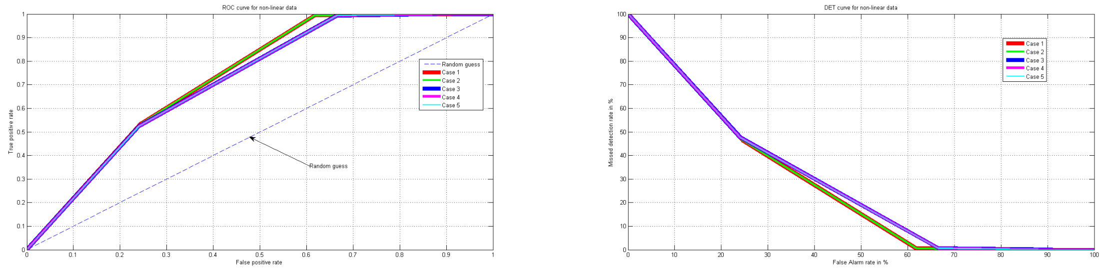

Figure 3: ROC and DET curve for non-linearly separable data.

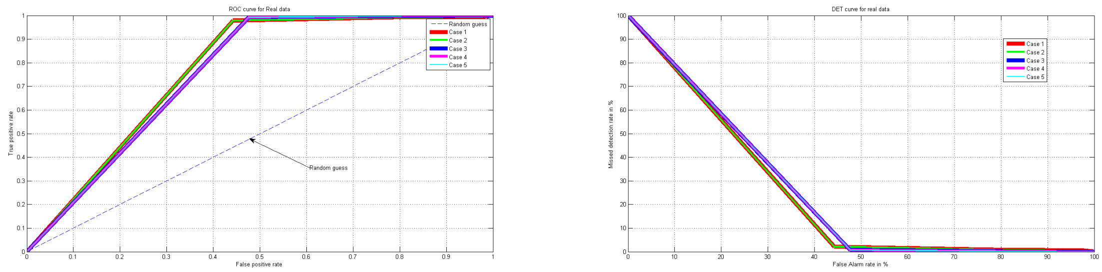

Figure 4: ROC and DET curve for real data.

0.3 Conclusion

After performing experiment with all of the given cases for different types of provided datasets, we observed accuracy as :

1. For linear dataset: Accuracy were same for all cases.
2. For non-linear dataset:

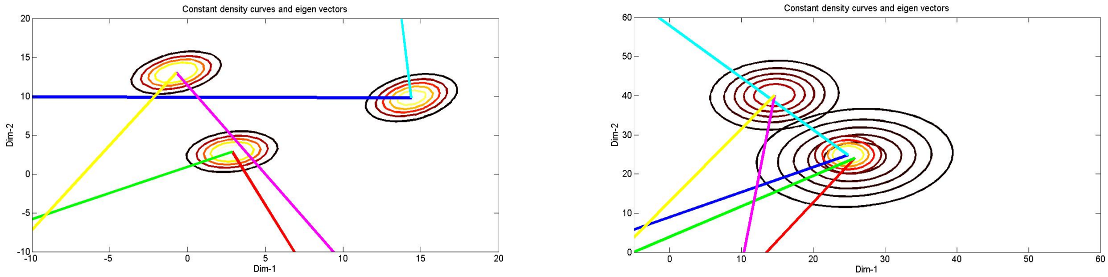

Figure 5: Constant density curves with eigenvectors for linear and non-linear data respectively.

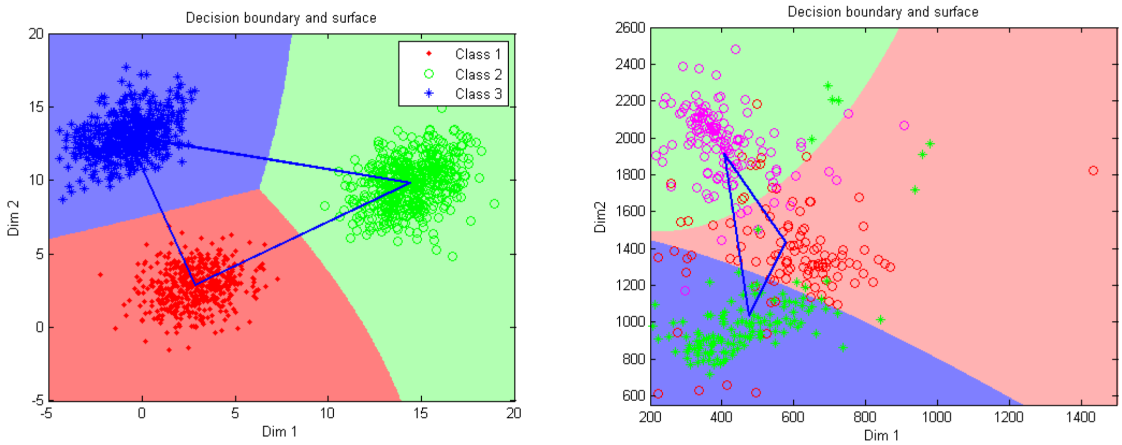

Figure 6: Decision Boundary and decision surface for linear and real data.

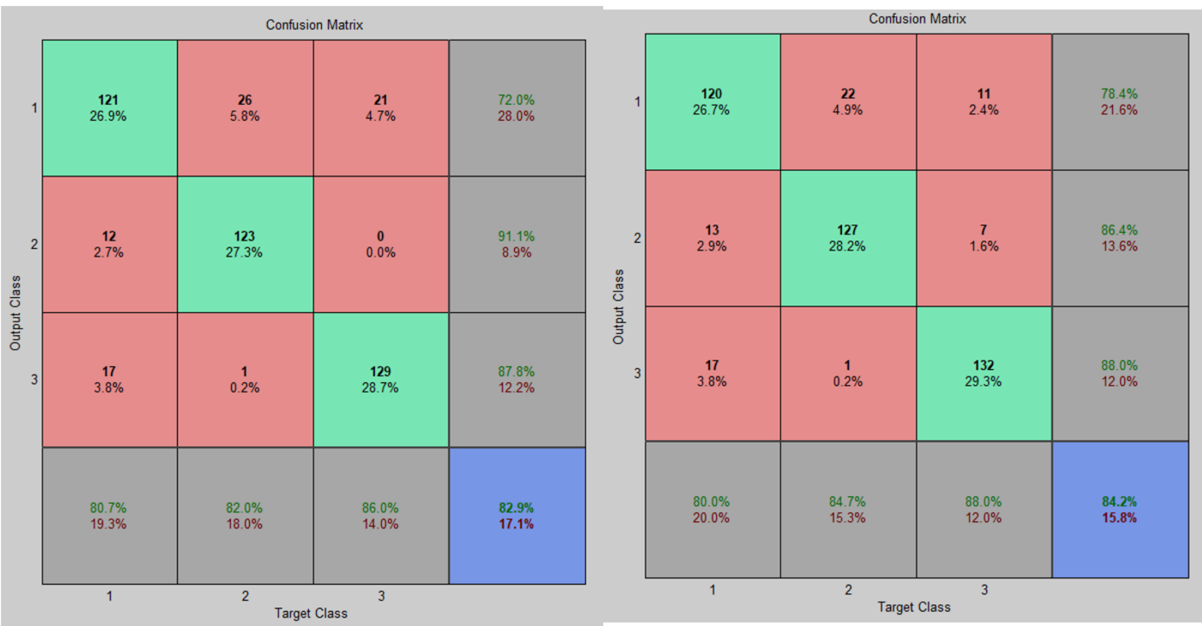

Figure 7: Confusion matrix for real data in Case 2 and 3 respectively.

Case number	Accuracy For Real Data (%)	Accuracy for Non-linear Data (%)
1	83.1	73.3
2	82.9	97.4
3	84.2	72.9
4	83.3	73.3
5	83.1	97.4

Figure 8: Case-wise Accuracy comparison table for real and non-linear data.

(a) Accuracy were same in case 2 and 5 (highest).

(b) Accuracy were same in case 1, 3 and 4 (lowest).

3. For real dataset:

   (a) Accuracy were same in case 1 and 5.

   (b) Accuracy were different in case 2, 3 and 4.