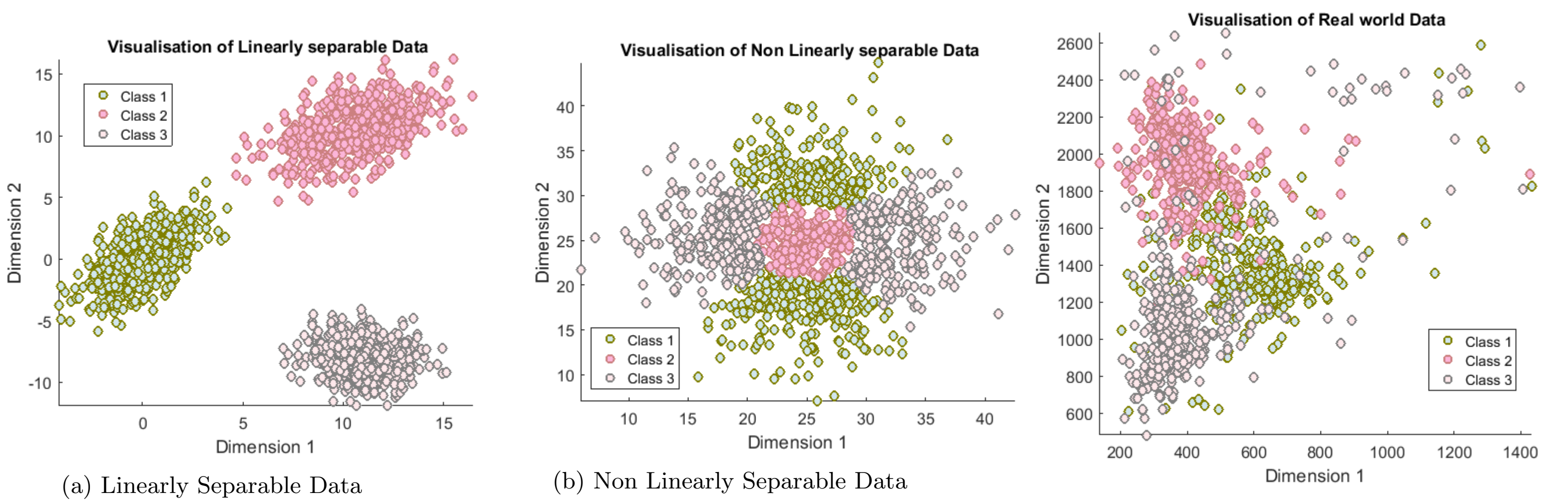
Pattern Recognition Assignment II

Muhammad Arsath K F(CS16S035) and Srihari Maruthachalam (CS16S024) Group 17

September 20, 2017

1 Visual Inference



(a) Linearly Separable Data

(b) Non Linearly Separable Data

(c) Real World Data

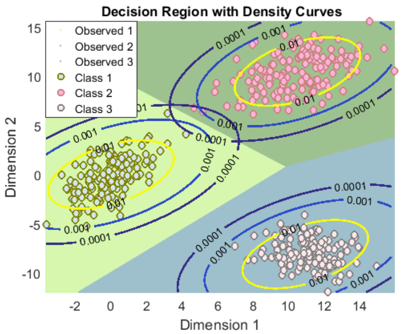
Figure 1: Visualization of Data

Intuition and Inference

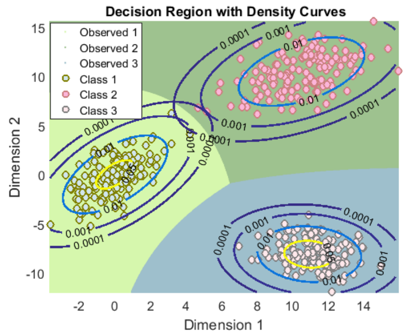
Given linearly separable data is well separated. On the other hand, non linearly separable is non convex data. Real world data is much over lapped across classes.

- For non convex data, we had a strategy to solve the classification problem. We made clusters within non convex class using KMeans and considering the clusters as a separate class(Discussed in Section 5.1).
- We can clearly see that there are few outliers in real world data. We had a strategy to detect the outliers using likelihood value(Discussed in Section 5.2).
- We attempted to solve the problem in the following five ways
  - Bayes with Covariance same for all classes(hereafter referred as Case 1)
  - Bayes with Covariance different for all classes(hereafter referred as Case 2)
  - Naive Bayes with C $= \sigma^2 * I$(hereafter referred as Case 3)
  - Naive Bayes with C same for all classes(hereafter referred as Case 4)
  - Naive Bayes with C different for all classes(hereafter referred as Case 5)

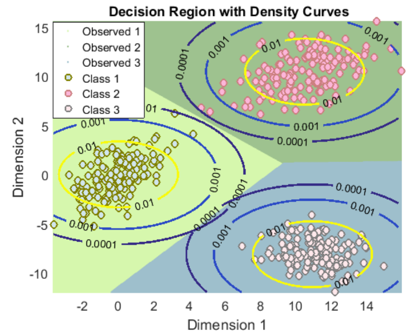
2 Linearly Separable Data



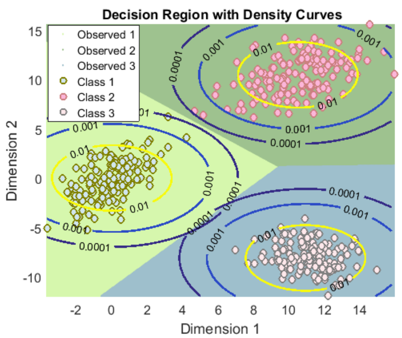
(a) Case 1



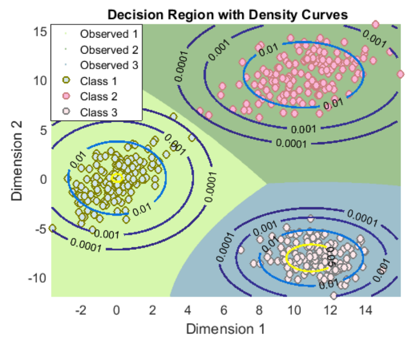
(b) Case 2



(c) Case 3



(d) Case 4

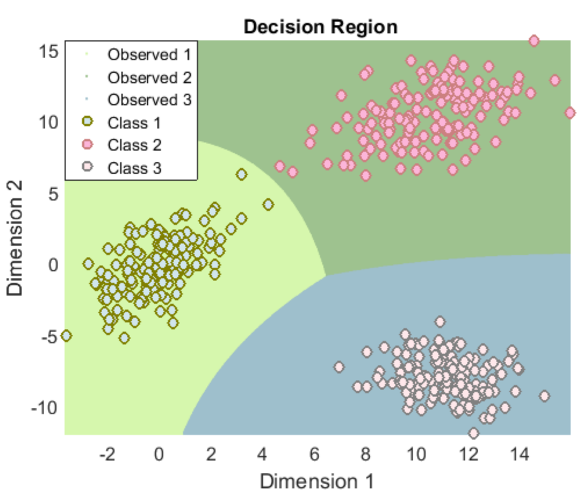


(e) Case 5

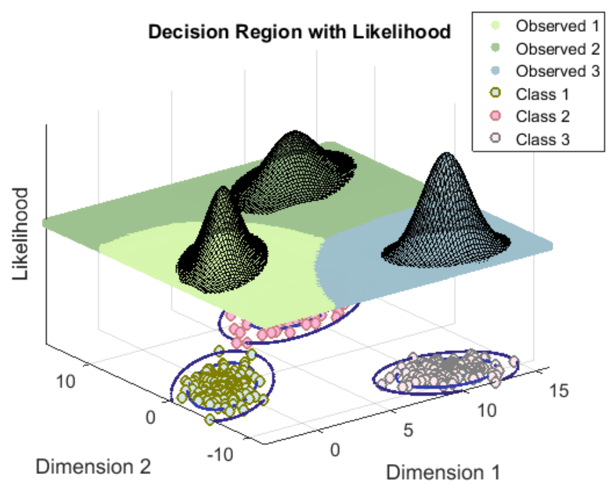
Figure 2: Density Curves for Linearly separable data

Inference

Decision boundaries are linear when covariance is same(case 1, 3 and 4) and Non linear when covariance is different(case 2 and 5). As we have well separated data points, we achieve best result for all the cases. Decision boundary and likelihood distribution for different covariance are shown below.



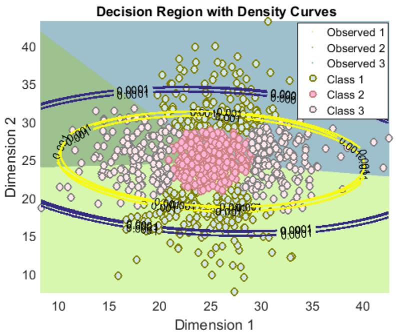
(a) Decision boundary for Linearly separable data



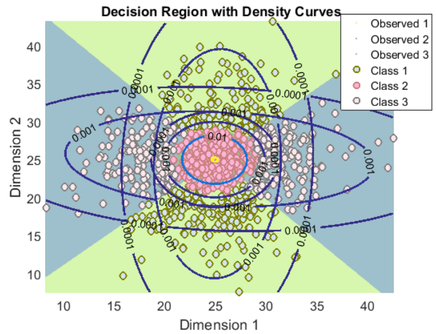
(b) Likelihood distribution for Linearly separable data

Figure 3: Linearly separable data

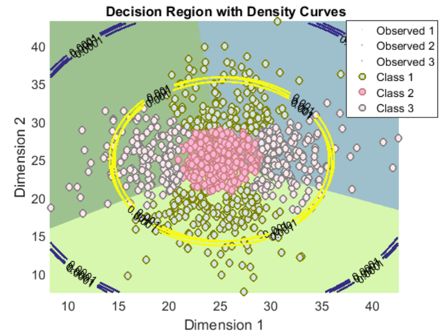
3 Non Linearly Separable Data



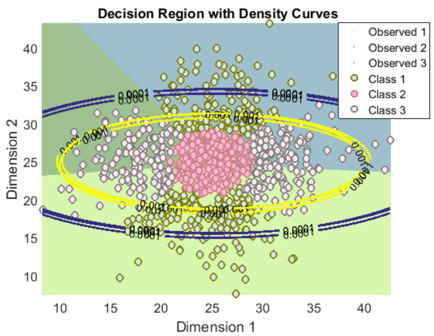
(a) Case 1



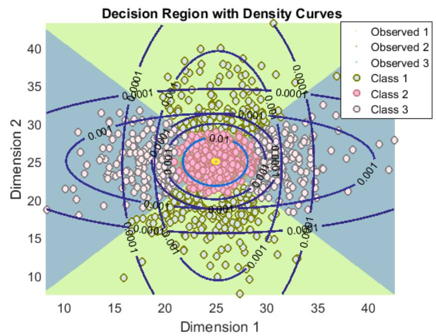
(b) Case 2



(c) Case 3



(d) Case 4

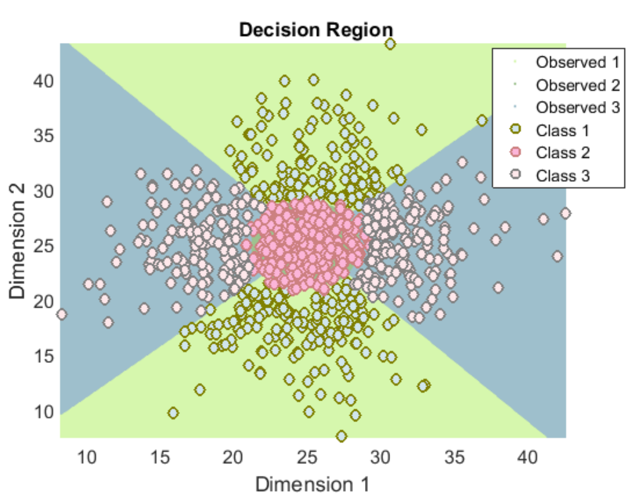


(e) Case 5

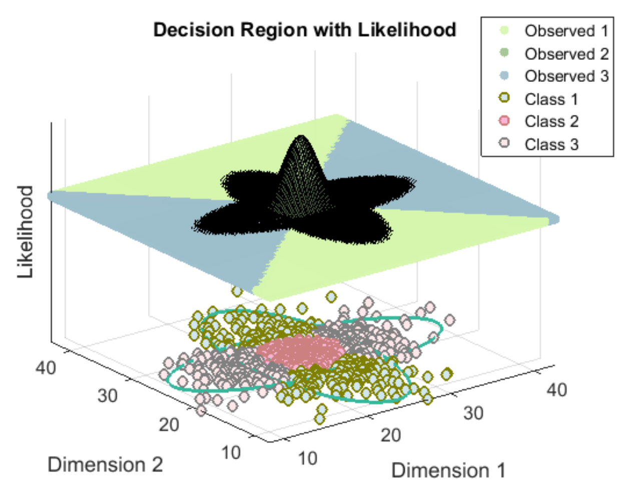
Figure 4: Density Curves for Non linearly separable data

Inference

Since mean of all classes are almost same, covariance decides the decision boundary. Here we can clearly see that same covariance for all classes does not perform well. Also, we observed that case 2 and 5 seems to have similar density curves because covariance across dimension 1 and dimension 2 is minimal for the given data. Decision boundary and likelihood distribution for different covariance are shown below.



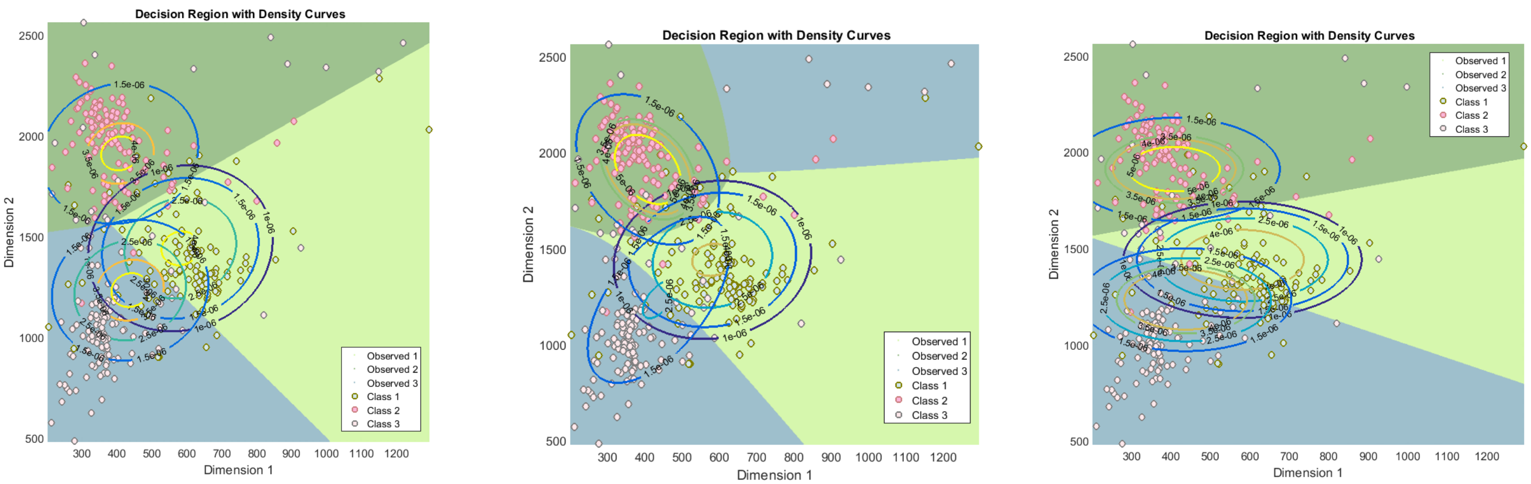
(a) Decision boundary for Non linearly separable data



(b) Likelihood distribution for Non linearly separable data

Figure 5: Non Linearly Separable Data

4 Real World Data



(a) Case 1 (b) Case 2 (c) Case 3

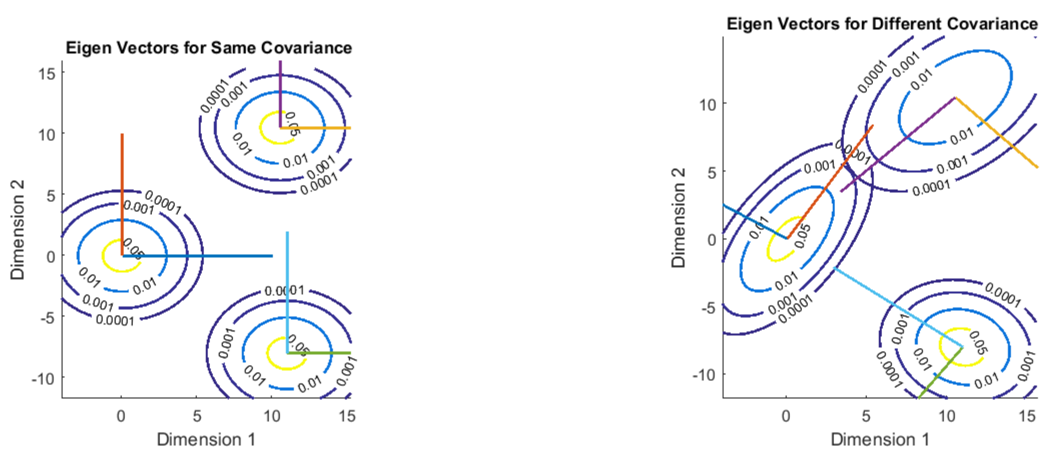
Figure 6: Density Curves for Real world data

Inference

From the likelihood contours, we can see that different covariance for each classes performs well. Also, we can see that Bayes classifiers are capable of classifying non convex data points as well. Another inference is that mean of class 3 is shifted because of the influence of outliers.

5 Eigen Vectors for Covariance

Here we shown Eigen vectors for covariance of linearly separable data.



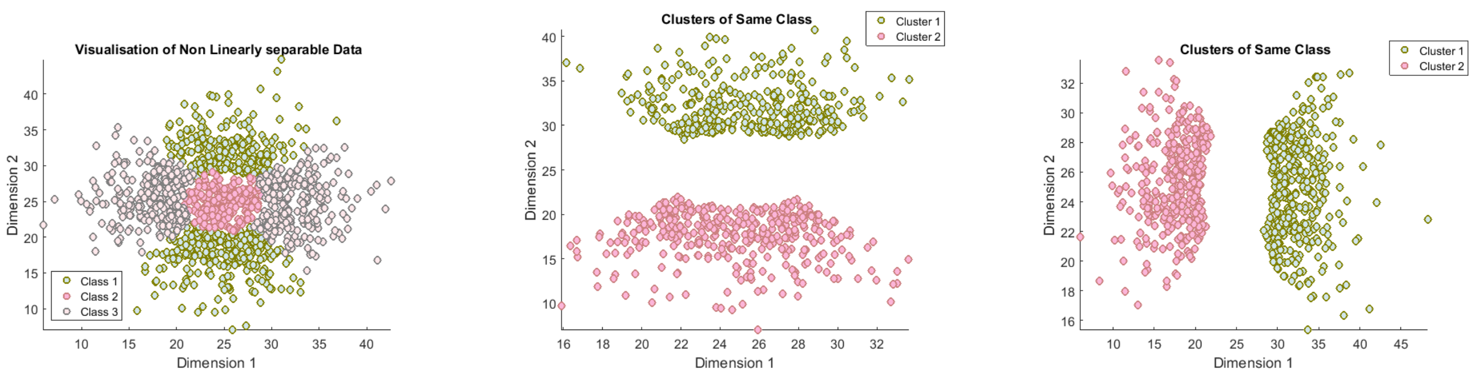
(a) Eigen vectors for Case 3 (b) Eigen vectors for Case 2

Figure 7: Eigen Vectors for Covariance

6 Intuition and Idea

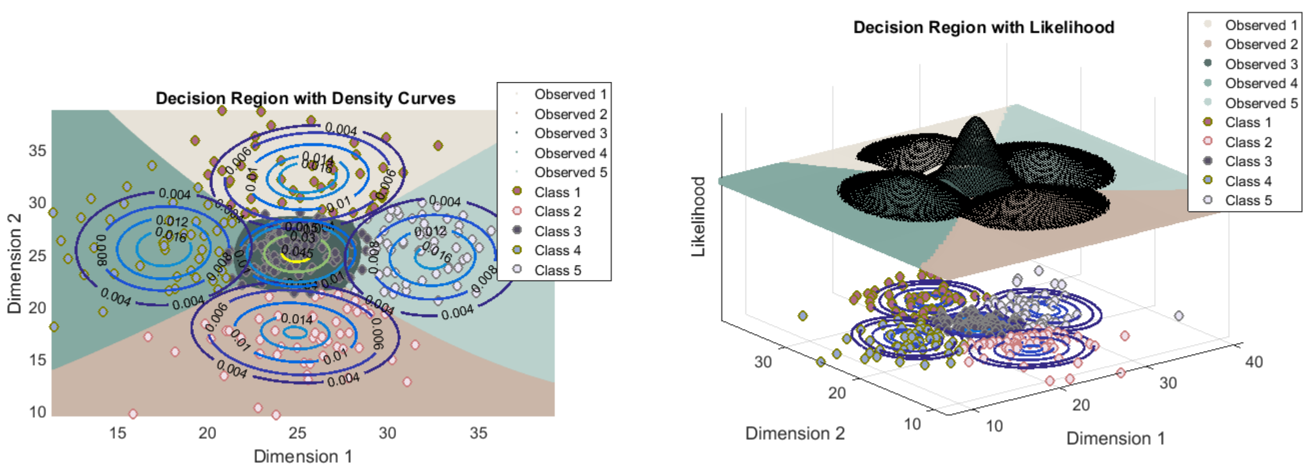
6.1 Idea for Solving Non Convex Data Set

- We preprocess the non convex classes into clusters using KMeans.
- Each cluster is treated as a separate class.
- We build statistical models for the distribution of classes and made given problem as convex problem.
- We postprocess by considering those models to be of same classes.
- While classifying new data point, likelihood from the clustered classes are assumed to likelihood of preprocessed non convex classes.



(a) Given Non Linearly Separable Data (b) Formed clusters from class 1 (c) Formed clusters from class 3

Figure 8: Clustering of non convex classes



(a) Density Curves for Convex data (b) Likelihood distribution for Convex data

Figure 9: Non convex data to convex data

6.2 Idea for Outlier Detection

Outliers are capable of shifting the mean of the classes. As a result, likelihood estimation becomes inefficient. We had an intuition of detecting outliers after estimating the parameters of the model. The idea is to find the data points of the classes which are having very low likelihood are assumed to be outliers. Outliers detected for the class 3 are highlighted in the below image plot.

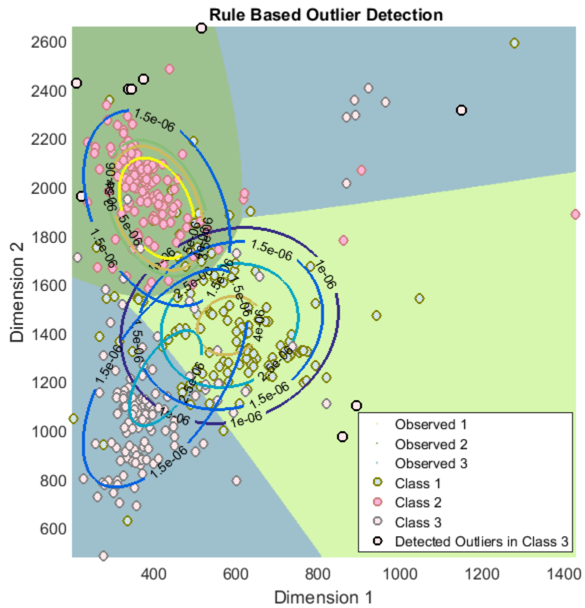
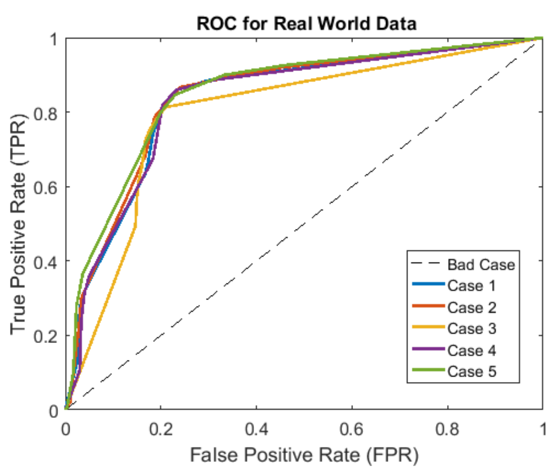


Figure 10: Detected Outliers from Likelihood

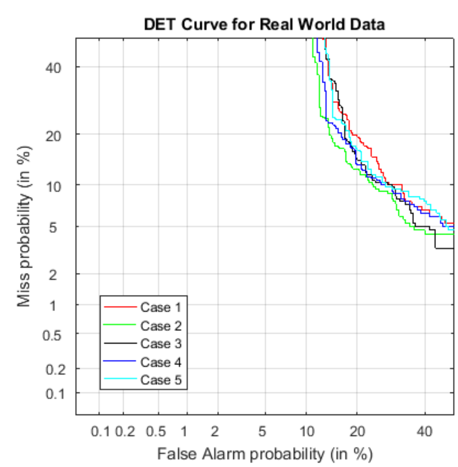
7 Comparison and Analysis

7.1 Receiver operating characteristic (ROC) Curves and Detection error trade-off (DET) Curves

Here we shown the ROC and DET curves for the given real world data.



(a) ROC curve for Real world data



(b) DET curve for Real world data

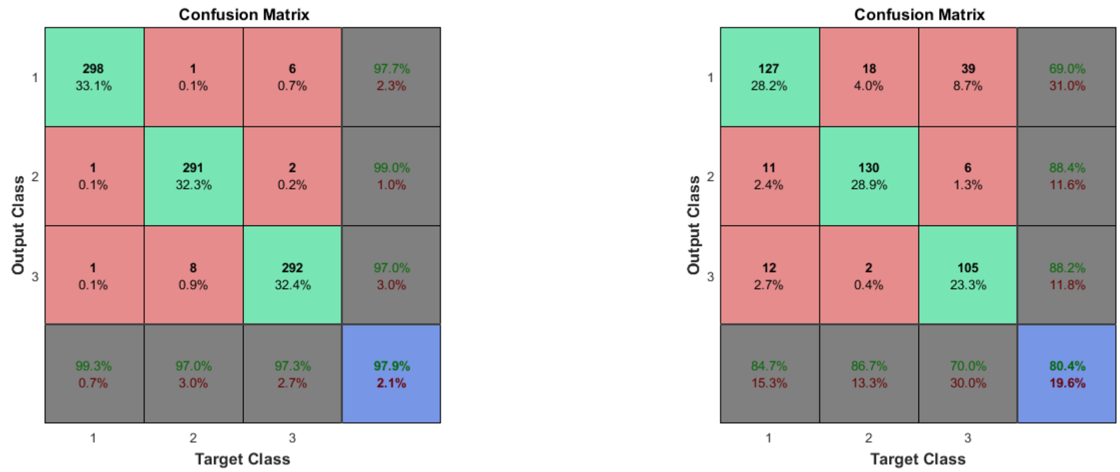
Figure 11: Comparison and Analysis of Real world data using ROC and DET curves

Inference

From the above ROC and DET curves, we can infer that different covariance for each classes(case 2) performs well.

7.2 Confusion Matrix

Confusion matrix for the non linearly separable and real world data are shown below.



(a) Confusion matrix for the non linearly separable data

(b) Confusion matrix for the real world data

Figure 12: Confusion matrices

Inference

From the confusion matrix, we can see that non linearly separable data performs well. We can also infer that most of the misclassifications in the real world data model are data points from class 2 and 3 classified as class 1.

7.3 Kullback–Leibler (KL) Divergence

Here we shown the KL Divergence across classes in linearly separable data and real world data.

$D_{KL}(P\|\|Q)$	P = 1	P = 2	P = 3
Q = 1	0	20.353	9.746
Q = 2	19.212	0	10.817
Q = 3	27.360	32.663	0

(a) KL Divergence for linearly separable data

$D_{KL}(P\|\|Q)$	P = 1	P = 2	P = 3
Q = 1	0	15.585e-4	5.683e-4
Q = 2	24.640e-4	0	21.783e-4
Q = 3	5.850e-4	20.315e-4	0

(b) KL Divergence for real world data

Table 1: Comparison using KL Divergence

Inference

As linearly separable data is well separated, KL Divergence across classes are in good shape. Since real world data is overlapping, we can see that KL Divergence is pretty low across classes.

8 References

- Bishop, Christopher M., "Pattern Recognition and Machine Learning", Information Science and Statistics, 2006.
- R.O.Duda, P.E.Hart and D.G.Stork, "Pattern Classification", John Wiley, 2001.
- Martin, A and R. Doddington, George and Kamm, Terri and Ordowski, Mark and Przybocki, Mark, "The DET curve in assessment of decision task performance", Fifth European Conference on Speech Communication and Technology, EUROSPEECH 1997.
- https://www.mathworks.com/
- Kullback–Leibler divergence
- NIST DETWare