DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING, IIT MADRAS

CS5691: Assignment 2
Team 1

Sheth Dev Yashpal
CS17B106

Harshit Kedia
CS17B103

1 PART 2A

1.1 BAYESIAN CLASSIFICATION

1.1.1 LINEARLY SEPARABLE DATA

Figure 1.1: The first figure is for Case 2: Bayes with different covariance matrices. The second figure corresponds to Case 3: Naive Bayes with $C = \sigma^2$. The third figure is the ROC for Linearly Separable Case.

- For the Linearly Separable Data-set we observe that we can easily classify the data even with the worst possible classifier i.e. the Univariate case. We are able to separate all the points with 100% accuracy and no False Positive or Negative even with the worst case.

1.1.2 NON-LINEARLY SEPARABLE DATA

Figure 1.2: The first figure is for Case 1: Bayes with same covariance matrices. The second figure corresponds to Case 5: Naive Bayes with C different for all classes. The third figure is the ROC for Linearly Separable Case.

- For Non-Linearly Separable data we are able to get good classification in both the cases of Bayes Classification with different Covariance and Naive Bayes Classification with different Covariance. Also shown here, we can see that the best case Linear Classifier also doesn't work in the case of Non-Linear data which is as expected.

- We plot the ROC curves for all our models and observe that the model for Case 2 (Bayes with covariance different for all classes) performs almost at par with Case 5 (Naive Bayes with C different for all classes) as can be observed from the decision surface plots.

1.1.3 Real Data

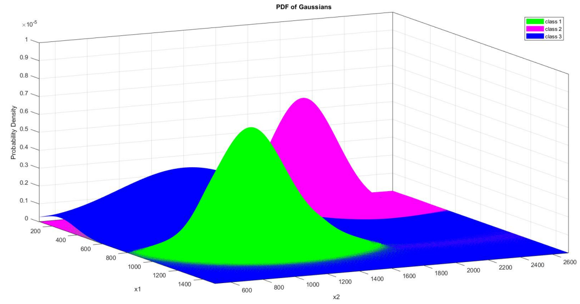


Figure 1.3: The plot for 2 Dimensional Gaussians for Real Dataset Provided.

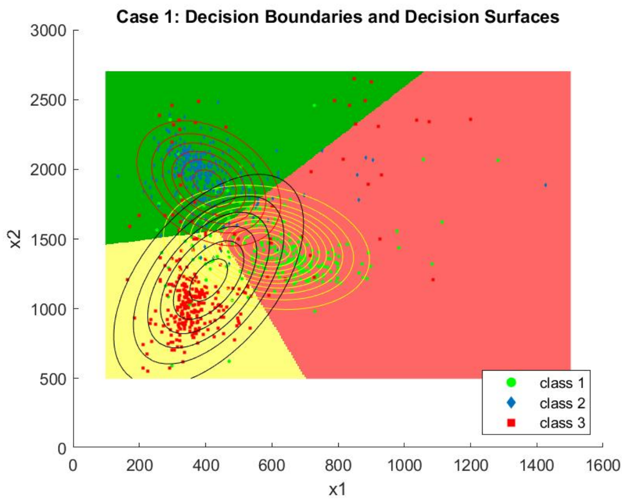


Figure 1.4: Decision Boundary and Decision Surface of Real Data for Case 1: Bayes with same covariance matrices.

- Here in the first plot we can observe the 2D Gaussian for Real Data.

- In the second figure we are classifying the data according to Case 1 (Bayes with Covariance same for all classes) and we observe that it is quite prone to error as expected as the data is not expected to be Linearly Separable.

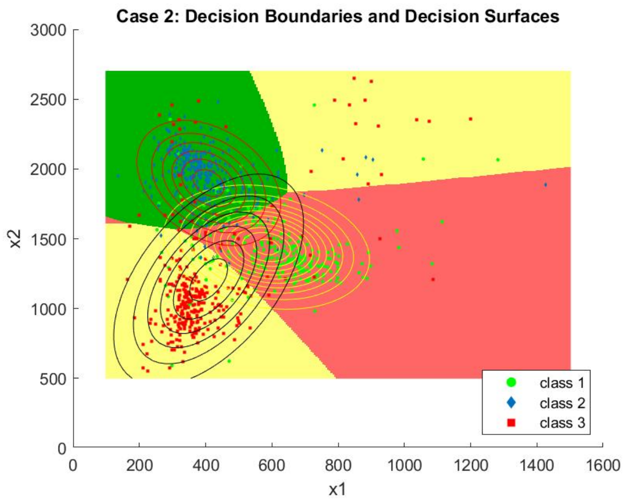


Figure 1.5: Decision Boundary and Decision Surface of Real Data for Case 2: Bayes with different covariance matrices.

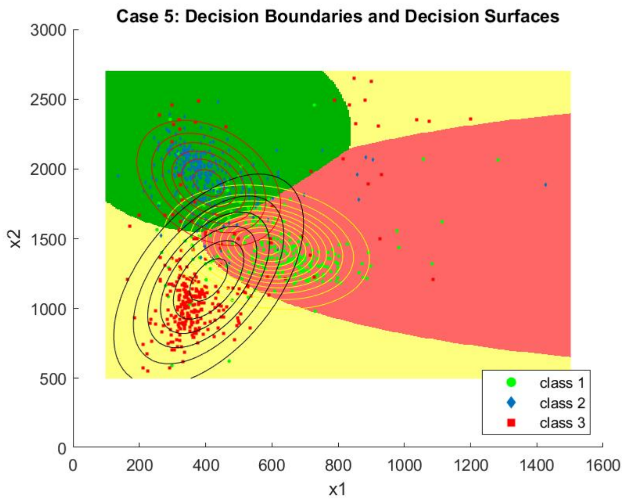


Figure 1.6: Decision Boundary and Decision Surface of Real Data for Case 5: Naive Bayes with C different for all classes.

- The first figure shows the classification of data according to the model of Case 2 (Bayes with Covariance different for all classes) which is arguably the best performing model for the real data set.

- In the second figure we are classifying the data according to Case 5 (Naive Bayes with C different for all classes.) and we observe that it performs very close to the best model though the decision boundaries are very different from the Bayes case.

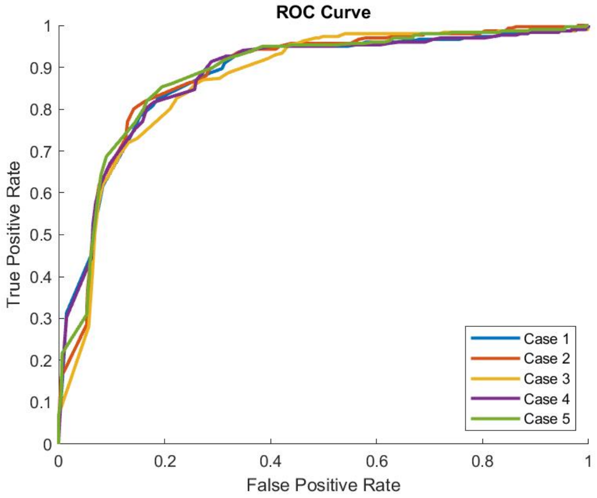


Figure 1.7: ROC Curve for Real Data-set for all the five Cases of classification.

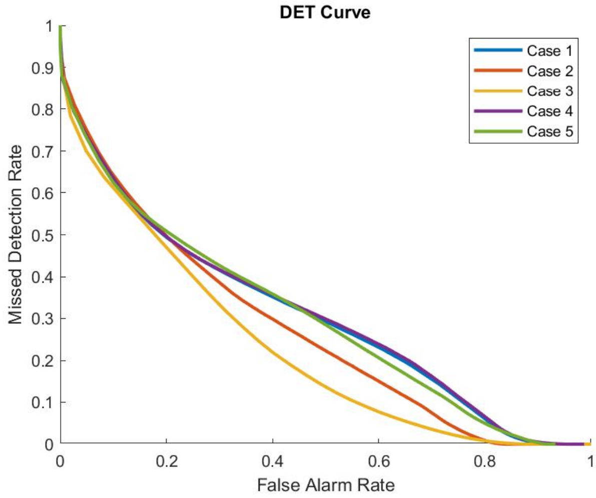


Figure 1.8: DET Curve for Real Data-set for all the five Cases of classification.

- In the first figure above, as expected the ROC curve is best for the models in Case 2 and Case 5 which we observed also with the decision boundary plots.

- In the second figure the DET curve for Case 3 is surprisingly better than the the curve for Case 2 or 5. We think thet the method of our calculation for DET curves for this result which is not accurate and therefore we are getting counter intutive results.

Output Class \ Target Class	1	2	3	Total
1	81	10	24	115
2	9	89	3	101
3	10	1	73	84
Total	100	100	100	300

Table 1.1: Confusion Matrix.

- Here we give the confusion matrix for our data and observe that our model is able to correctly classify 81% of all the images.

2 Part 2B

2.1 Gaussian Mixture Models

2.1.1 Synthetic Data

- The figure 2.1 and 2.2 we have each class being represented as a cluster of only 1 Gaussian distribution. Here we see the contour of the Gaussian fitting the entire model from approximately the centre.

- For figure 2.1 and 2.2s we clearly see that the decision boundary is almost a straight line.

- The figure 2.3 and 2.4 we have each class being represented as a cluster of 4 Gaussian distributions. Here we see the contour of the Gaussian trying fit each branch / spike in the entire model but due insufficient number of Gaussians, failing to fit properly.

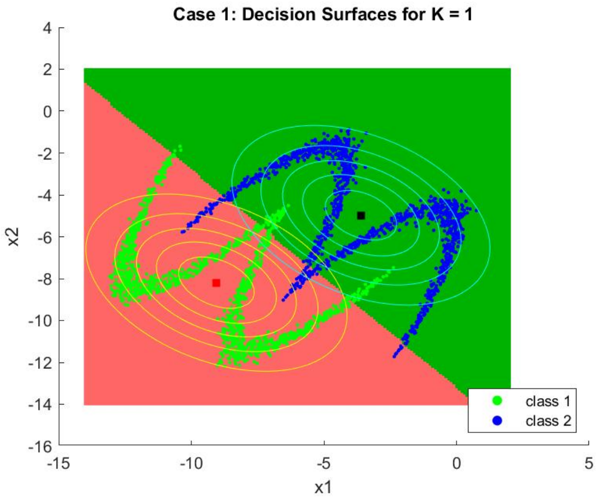


Figure 2.1: GMMs for Synthetic Data-sets with $K = 1$ clusters for each class in the case of Non-Diagonal Covariance Matrices.

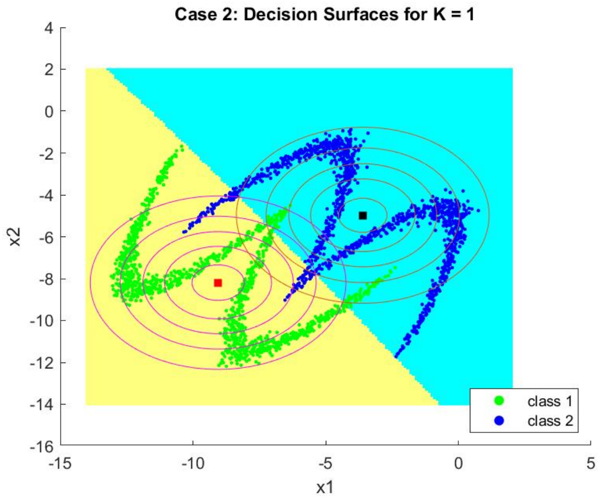


Figure 2.2: GMMs for Synthetic Data-sets with $K = 1$ clusters for each class in the case of Diagonal Covariance Matrices.

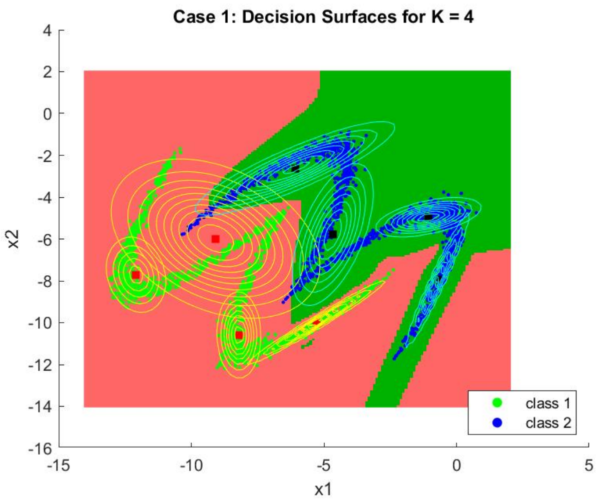


Figure 2.3: GMMs for Synthetic Data-sets with $K = 4$ clusters for each class in the case of Non-Diagonal Covariance Matrices.

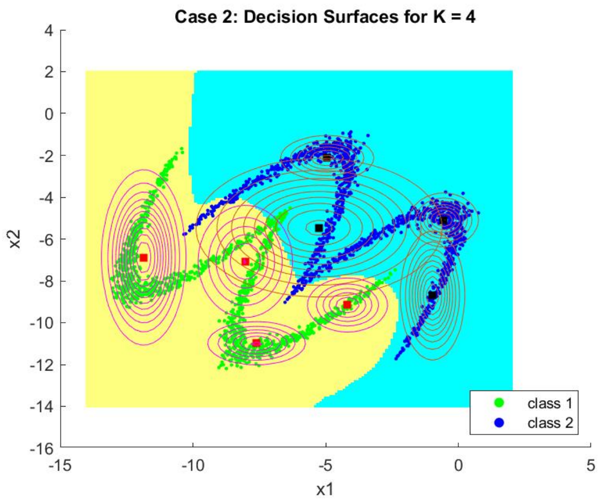
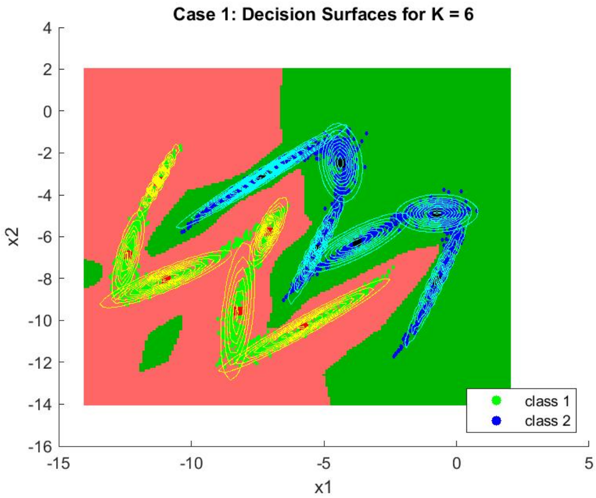


Figure 2.4: GMMs for Synthetic Data-sets with $K = 4$ clusters for each class in the case of Diagonal Covariance Matrices.

- For figure 2.3 and 2.4 we see that the decision surface is trying to take the "m" or "w" shape. The contours of the spikes are aligned with the the spikes direction due to presence of skewed covariance matrix in first case.

- The figure 2.5 and 2.6 we have each class being represented as a cluster of 6 Gaussian distributions. Here we see the contour of the Gaussian fitting each branch / spike in the entire model from approximately the centre of respective spike.
- For figure 2.5 and 2.6 we also see that the decision surface is almost of the "m" or "w" shape. The contours of the spikes are aligned with the the branch's direction and each branch is fit brilliantly with help of the non-diagonality but the later case fails because of diagonal covariance matrix.

- The figure 2.7 and 2.8 we have each class being represented as a cluster of 8 Gaussian distributions. Here we see the contour of the Gaussian fitting each branch / spike in the entire model from very precise points of respective spike. Also it seems like the branches are being over-fitted.
- For figure 2.7 and 2.8 we also see that the decision surface does very exact classification. Unlike for K = 6, due to over-fitting of the clusters, the later also shows a significant inprovement in the decision surface.



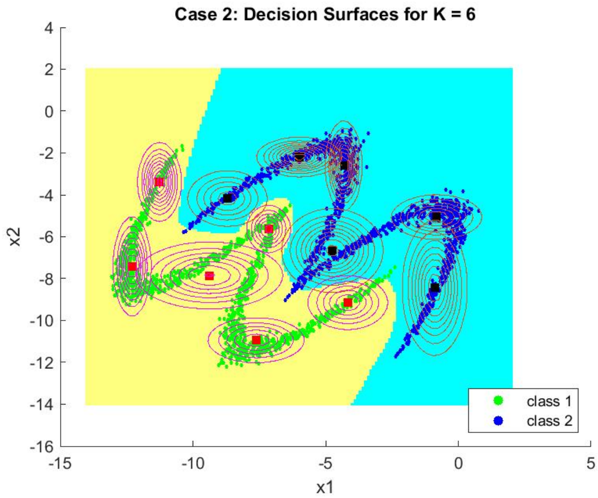
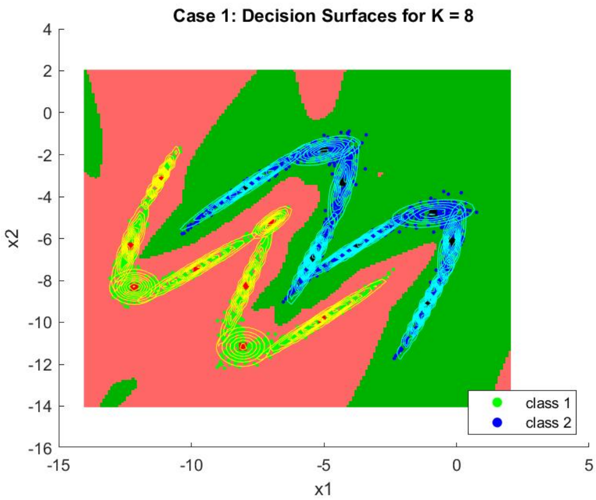


Figure 2.5: GMMs for Synthetic Data-sets with $K = 6$ clusters for each class in the case of Non-Diagonal Covariance Matrices.

Figure 2.6: GMMs for Synthetic Data-sets with $K = 6$ clusters for each class in the case of Diagonal Covariance Matrices.



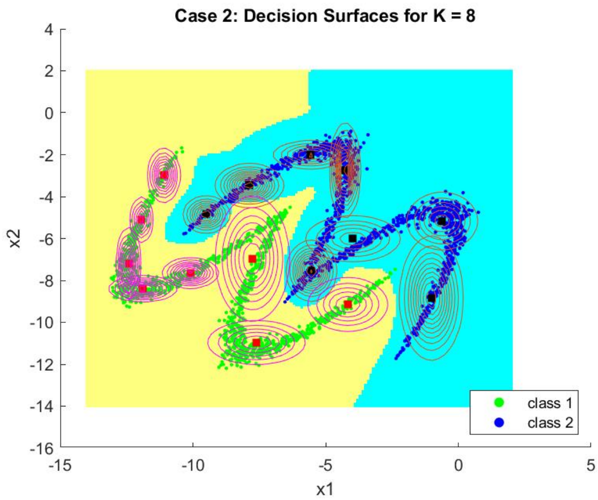


Figure 2.7: GMMs for Synthetic Data-sets with $K = 8$ clusters for each class in the case of Non-Diagonal Covariance Matrices.

Figure 2.8: GMMs for Synthetic Data-sets with $K = 8$ clusters for each class in the case of Diagonal Covariance Matrices.

2.1.2 Image Data

- Table 2.1 is the Confusion Matrix for Image Data in the Non-Diagonal case and we are able to classify 49.6% images correctly.

- Table 2.2 is the Confusion Matrix for Image Data in the Non-Diagonal case and we are able to classify 53.2% images correctly.

- Surprisingly the results obtained with diagonal case are slightly better than that for non-diagonal case.

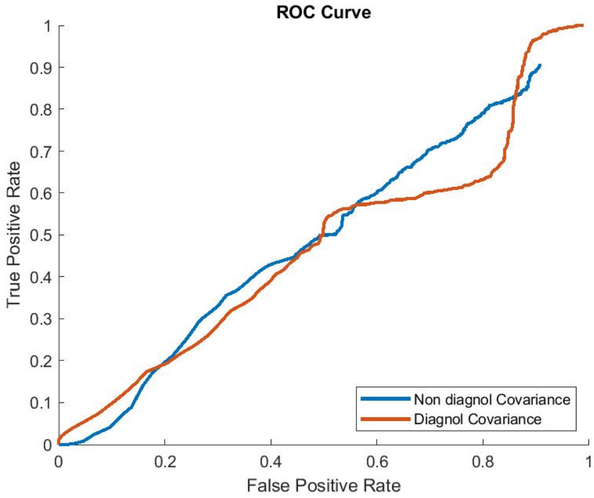


Figure 2.9: ROC Curve for the GMMs for Synthetic Data.

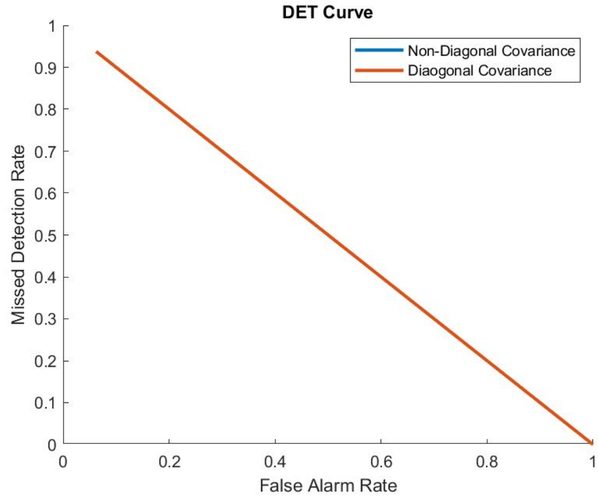


Figure 2.10: DET Curve for the GMMs for Synthetic Data.

Output Class \ Target Class	1	2	3	4	5	Total
1	16	2	1	0	0	19
2	17	29	13	8	4	71
3	7	9	30	26	9	81
4	10	4	3	16	4	37
5	0	6	3	0	33	42
Total	50	50	50	50	50	250

Table 2.1: Confusion Matrix for K = 4 and 10 iterations each for K-Means and GMM for Non-Diagonal Case.

Output Class \ Target Class	1	2	3	4	5	Total
1	19	8	2	8	1	38
2	19	26	20	4	6	75
3	0	7	26	1	3 2	39
4	12	7	1	24	3	47
5	0	2	1	1	38	42
Total	50	50	50	50	50	250

Table 2.2: Confusion Matrix for K = 4 and 10 iterations each for K-Means and GMM for Diagonal Case.