Pattern Recognition

Assignment #3

Gaussian Mixture Model

Professor:
Hema A. Murthy

Nirav Bhavsar (CS17S016)
Sidharth Aggarwal (CS17S012)

1.1 Gaussian Mixture Model

In this assignment we are making a classifier which can classify any point that to which class it belongs to. For this we are designing our model with the training data. In this we are training the model with KMeans with convergence criteria to be the same means and inputed this coverged data input GMM and trained the model further with log-likelihood to be the convergence criteria. In this we will use the most common parametric form i.e, the Normal/Gaussian Probability function for the classification.

$$f(x) = 2\pi^{-\frac{d}{2}} \det \Sigma^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(x-\mu)^{\mathrm{T}}\Sigma^{-1}(x-\mu)\right)$$

In this assignment we are given two sets of data which includes Real data and Synthetic data. The Real data includes the feature vectors for the images. And on these data sets we have to apply gaussian mixture model to classify the data point to one the class. And further we have done the operations with different covariance matrices as below:-

1. GMM with Diagonal Covariance Matrices
2. GMM with Non-Diagonal Covariance Matrices

1.2 GMM on Synthetic Data

In this we were given artificial data which is of form circles and accordingly we have applied the kmean and GMM for making clusters. Before that we have tried to make the Elbow Graph to see for which K(Number of Clusters) it will work well. After getting the outputs from the GMM, we have tested our model by giving the test point and according to those results found out the accuracy with ROC and DET graphs.

As we can see from the below plots for the Real Data. In these plots we have shown scattered data with the contours of the clusters. Another plot is of the confusion matrix which tells the accuracy of the classifier for every classes. The ROC plot tells the accuracy of the classification done by plotting the graph between TPR(True Positive Rate) and FPR(False Positive Rate). Another graph which tell the accuracy of the model with by making a graph by the Miss Probability and the False Alarm Probability.

	Raw Dataset		Normalized Dataset	
K	Full CoV	Diagonal CoV	Full CoV	Diagonal CoV
4	64.33 %	60.5 %	67.16 %	62.8 %
6	73.4%	69.3%	75.16%	71.66%
10	89.2%	80.4%	91.6%	83.2 %
16	100 %	88.9 %	100 %	91.67 %
18	100 %	99.33 %	100 %	100 %

Figure 1.1: Comparision of different system accuracy for Synthetic Dataset

Diagonal Covariance for K = 6 with Accuracy 71.66%

		Actual Class	
		Class 1	Class 2
Predicted	Class 1	222	92
	Class 2	78	208

Full Covariance for K = 6 with Accuracy 75.16%

		Actual Class	
		Class 1	Class 2
Predicted	Class 1	210	59
	Class 2	90	241

Diagonal Covariance for K = 16 with Accuracy 91.67%

		Actual Class	
		Class 1	Class 2
Predicted	Class 1	275	27
	Class 2	25	273

Figure 1.2: Confusion Matrix of different system for K = 6 & 16

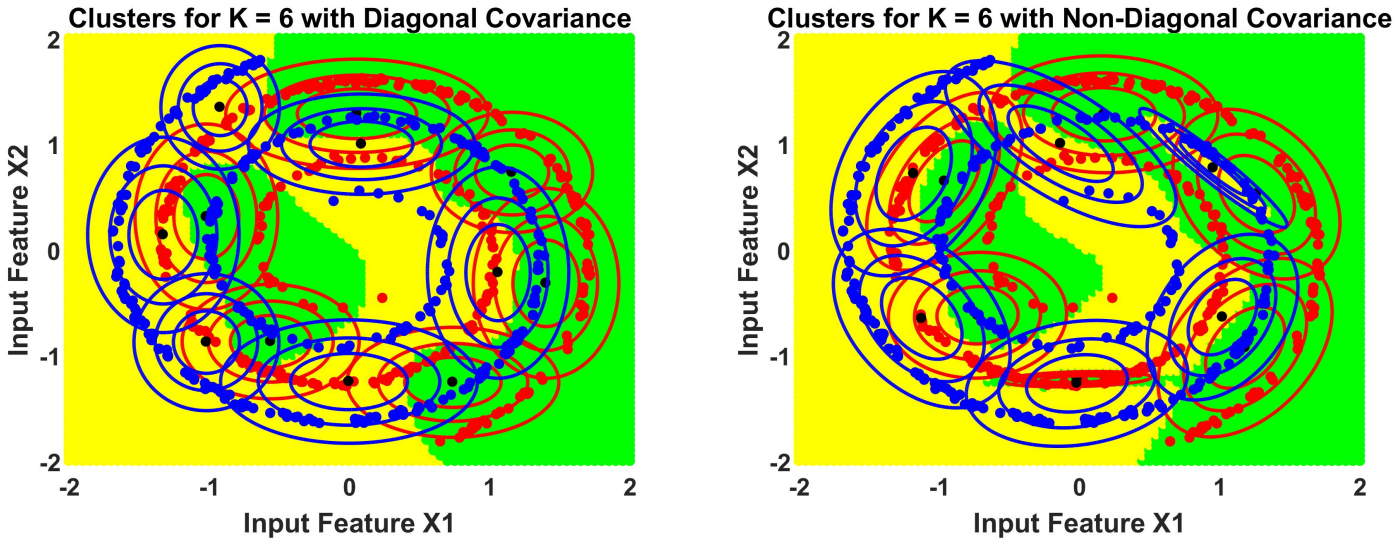


Figure 1.3: Contours & Decision boundary on training data for K = 6

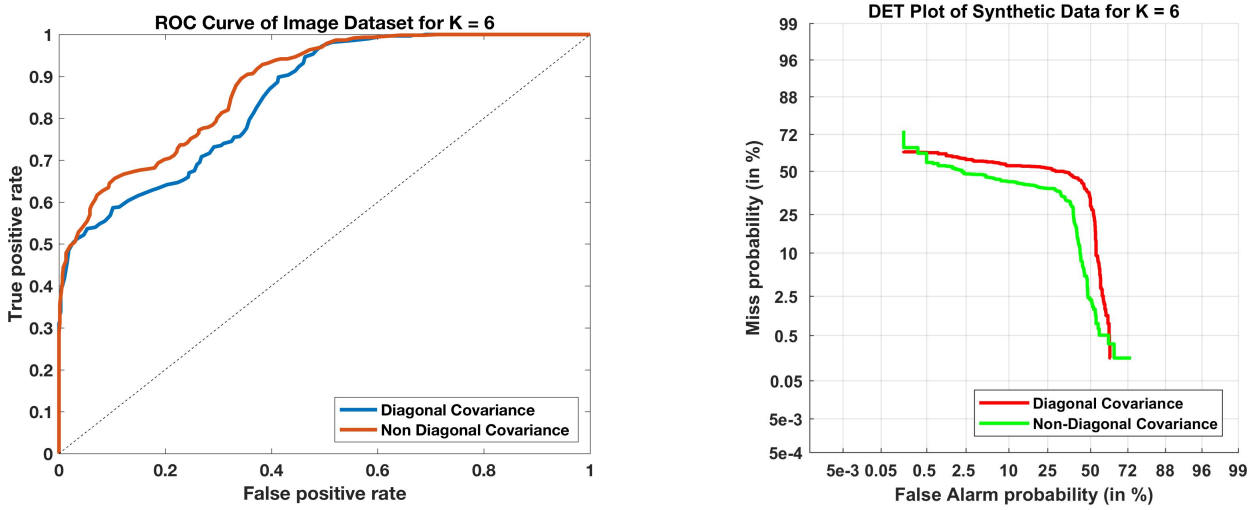


Figure 1.4: ROC & DET of Synthetic Dataset for K = 6

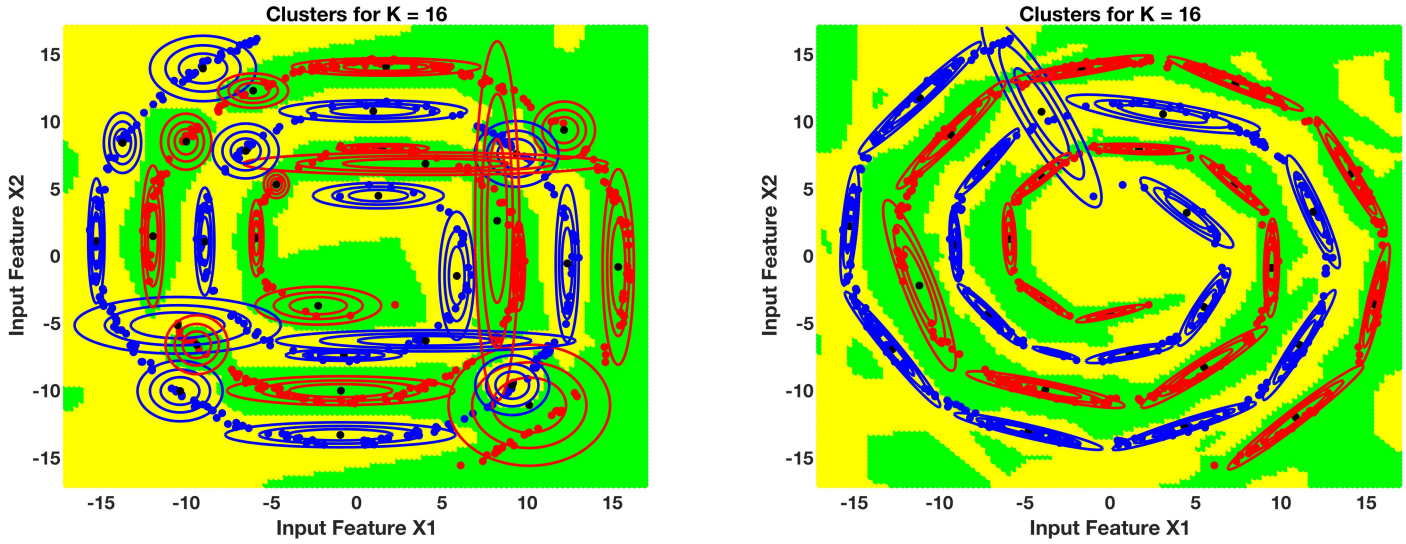


Figure 1.5: Contours & Decision boundary on training data for K = 16

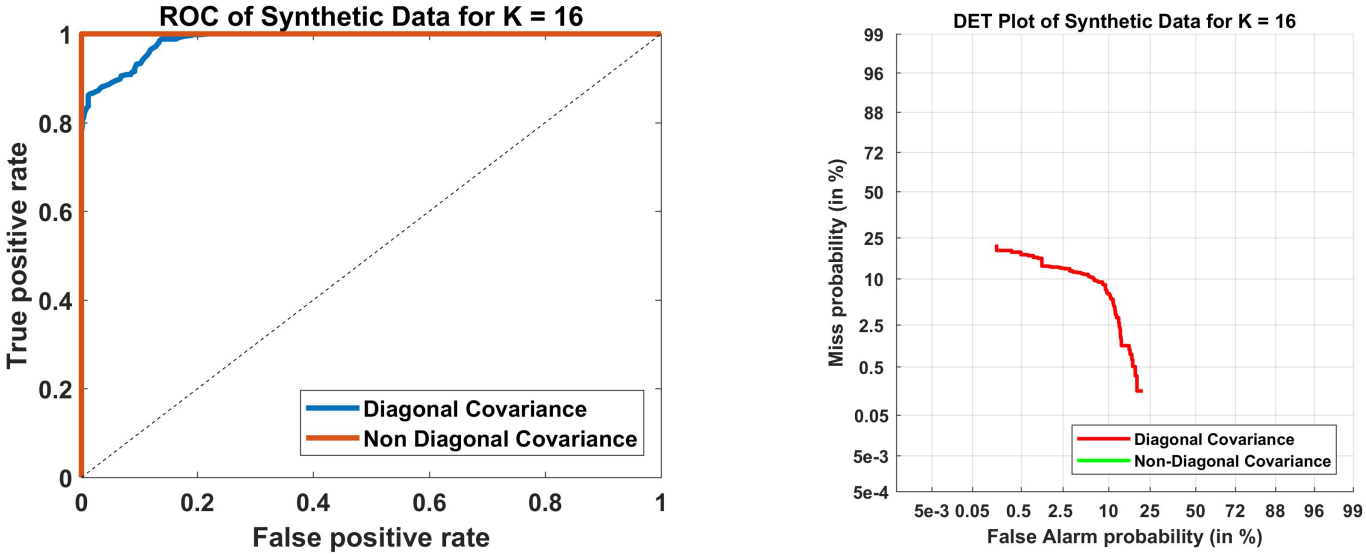
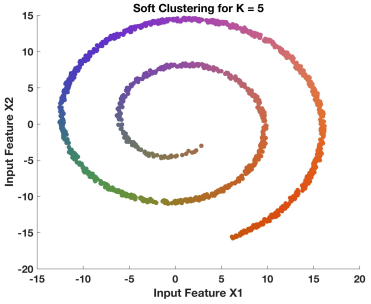


Figure 1.6: ROC & DET of Synthetic Dataset for K = 16



1.3 GMM on Real Data

In this we were given Real data which is the feature vectors of different images and accordingly we have applied the kmean and GMM for making clusters. Before that we have tried to make the Elbow Graph to see for which K(Number of Clusters) it will work well. After getting the outputs from the GMM, we have tested our model by giving the test point and according to those results found out the accuracy with ROC and DET graphs.

	Average Scoring		Voting Method		Normalized Dataset	
K	Full CoV	Diagonal CoV	Full CoV	Diagonal CoV	Full CoV	Diagonal CoV
2	74.16 %	72.5 %	72.5 %	72.5 %	100 %	84.16 %
3	82.5 %	84.16 %	79.16 %	82.5 %	100 %	85.83 %
4	83.33 %	85 %	82.5 %	81.67 %	100 %	95 %

Figure 1.7: Comparision of different system accuracy for Real Image Dataset

Diagonal Covariance – Normalized Dataset for K = 4 with Accuracy 95%

		Actual Class		
		Highway	Street	Buildings
Predicted	Highway	36	0	0
	Street	4	40	2
	Buildings	0	0	38

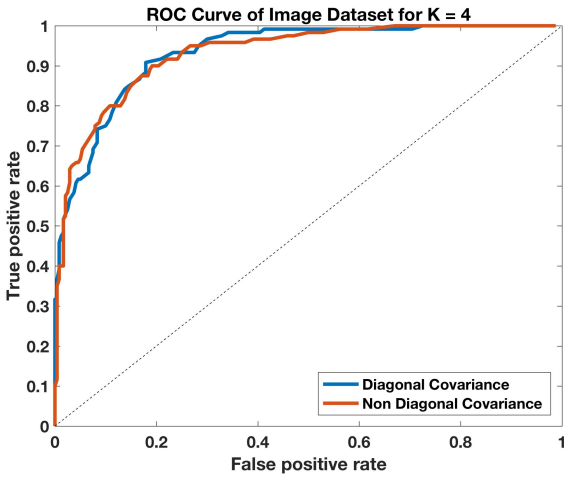
Full Covariance – Voting Method for K = 4 with Accuracy 82.5%

		Actual Class		
		Highway	Street	Buildings
Predicted	Highway	36	1	7
	Street	3	35	5
	Buildings	1	4	28

Diagonal Covariance for K = 4 with Accuracy 85%

		Actual Class		
		Highway	Street	Buildings
Predicted	Highway	35	0	5
	Street	5	37	5
	Buildings	0	3	30

Figure 1.8: Confusion Matrix of different system for K = 4



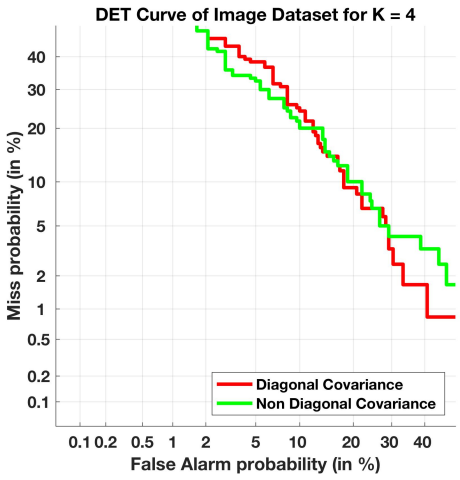


Figure 1.9: ROC & DET of Real Image Dataset for K = 4

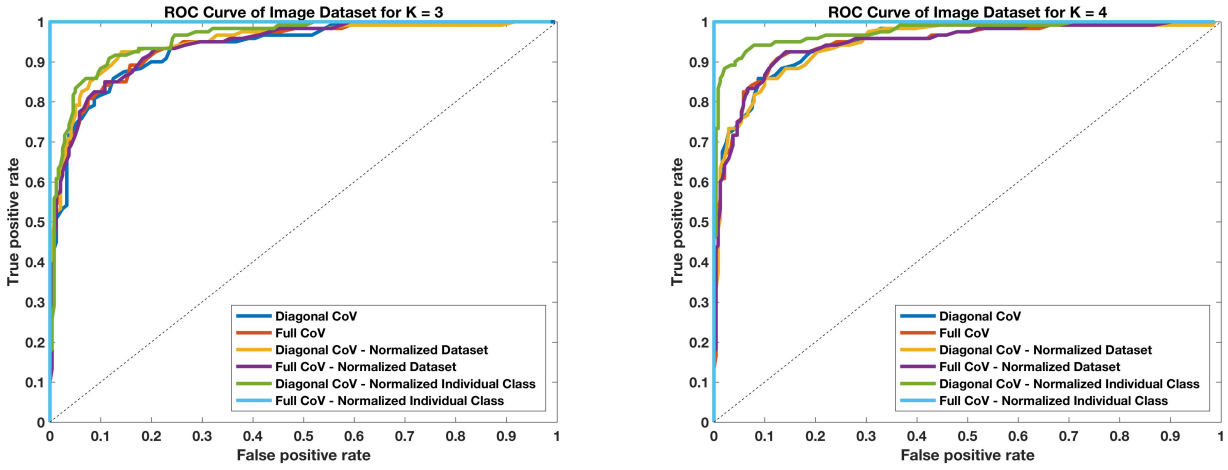


Figure 1.10: ROC comaparing different system of Real Image Dataset for K = 3 & 4

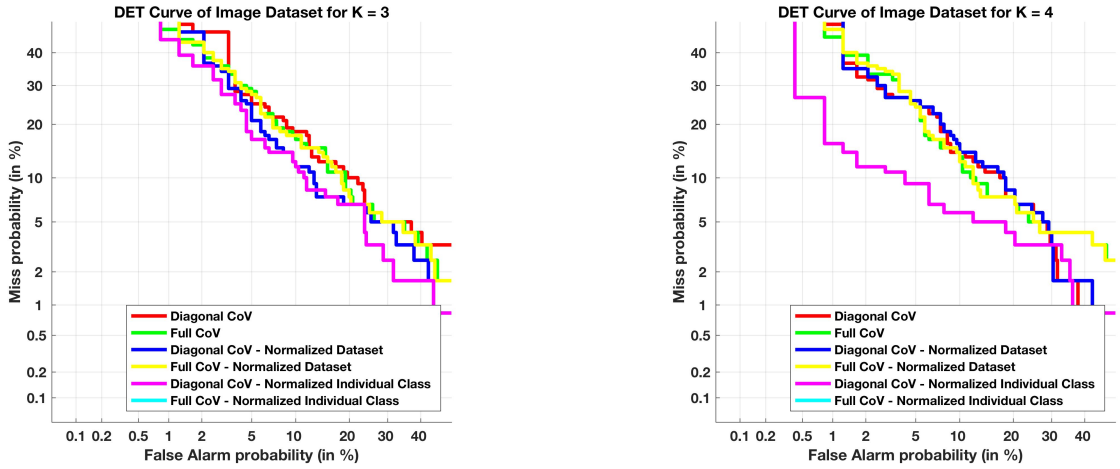


Figure 1.11: DET comaparing different system of Real Image Dataset for K = 3 & 4

Observations :-

From the plot of the contours as we can observe that after applying the GMM on the data set the system gets covergered and the gaussians are only on the data points. And further we observed that for the Diagonal Covariance the principal axis of the contours are parallel to the standard axis but in the case of the Non-Diagonal Covariance the principal axis of the coutours are may or may not be parallel to the standard axis. And we also observed that as we increase the number of clusters to certain extent the convergence is coming out to be more appropriate.

As from ROC plot we observed that for different values of the clusters the accuracy comes out to be different and in out case with 16 clusters the model is giving 100% accuracy for the classification of test points. And also further we observed that of non-diagonal covariance the accuracy is better than the accuracy in the diagonal covariance.