

# Pattern Recognition

## Assignment 3

### Group No - 36

ED13D000 Bhargava Sai Ramu  
CS17S008 Nitesh Methani

October 13, 2017

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Observations</b>	<b>1</b>
<b>3</b>	<b>Synthetic Dataset</b>	<b>1</b>
<b>4</b>	<b>Image Dataset</b>	<b>4</b>
<b>5</b>	<b>Conclusion</b>	<b>5</b>

## 1 Introduction

We have two datasets to work with. One dataset is synthetic data that contains two classes. Second dataset consists of three classes, each class contains 36x23 images.

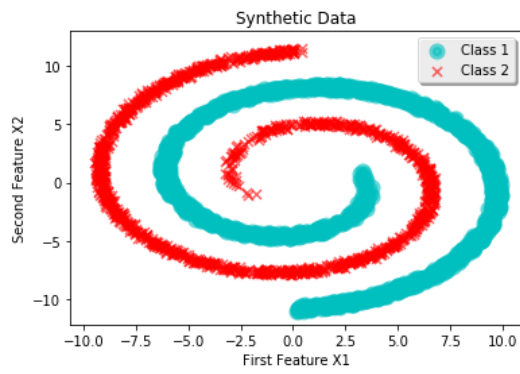
## 2 Observations

Attributes	Full Covariance	Diagonal Covariance	Image Dataset
For class 1, K	7	5	10
For class 2, K	10	4	10
For class 3, K	-	-	10
EM for class 1	9	15	35
EM for class 2	24	21	30
EM for class 3	-	-	29
Accuracy	100%	94.5%	55%

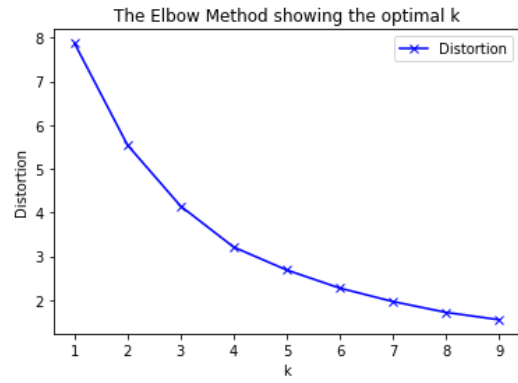
Figure : Summary of the assignment

## 3 Synthetic Dataset

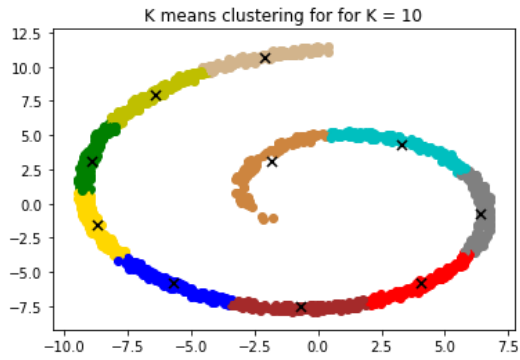
1) Full Covariance Matrix



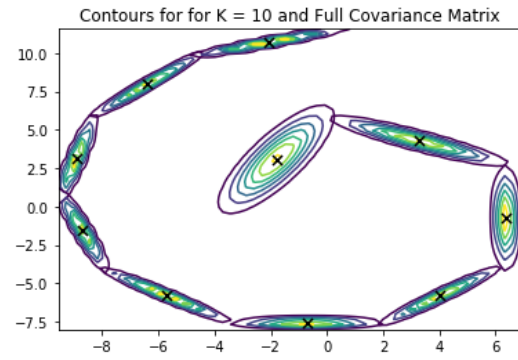
(a) Synthetic Data



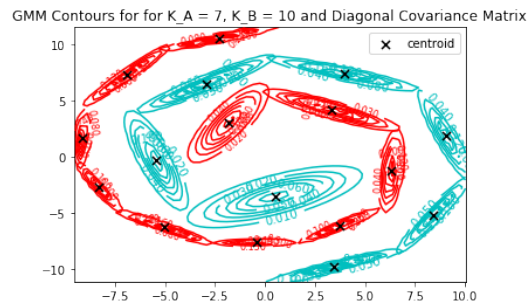
(b) Ellbow Curve



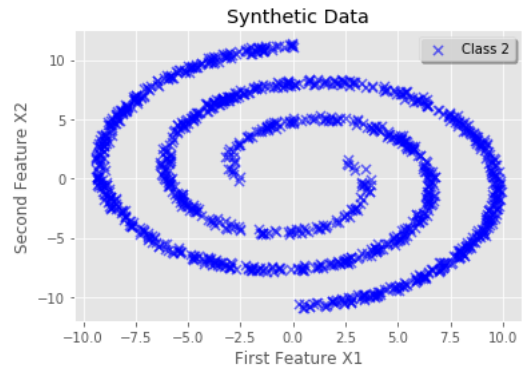
(c) K Means Clustering for class A



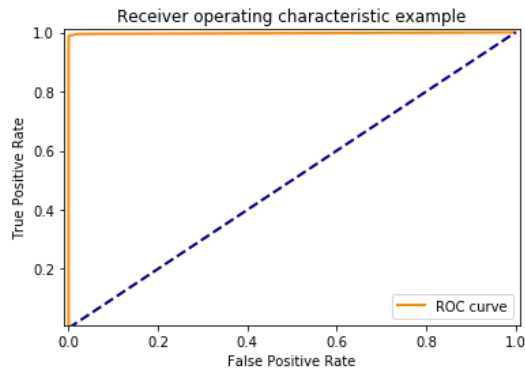
(d) K Means Contours for class A



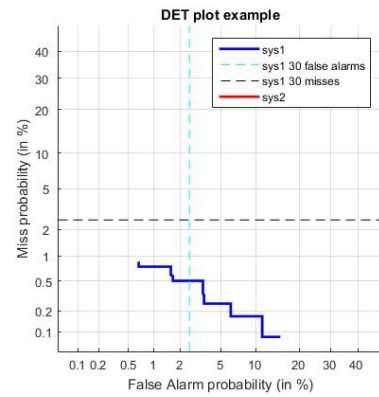
(e) GMM Contours



(f) Test Data



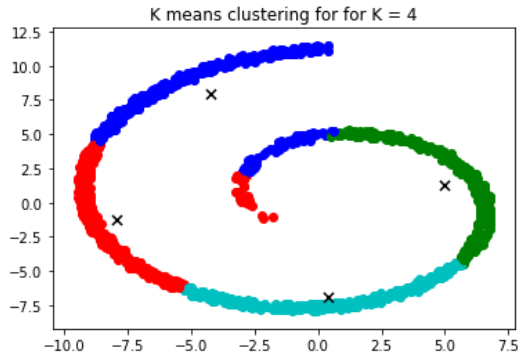
(e) ROC Curve



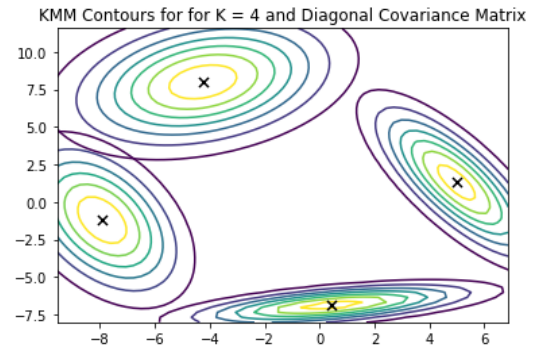
(f) DET Curve

Figure 1: Synthetic Data with Full Covariance Matrix

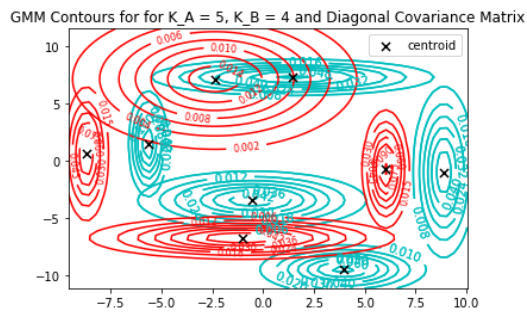
## 2) Diagonal Covariance Matrix



(a) K Means Clustering



(b) K Means Contours

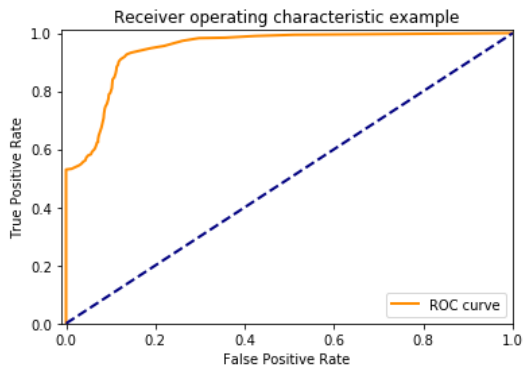


(c) GMM Contours

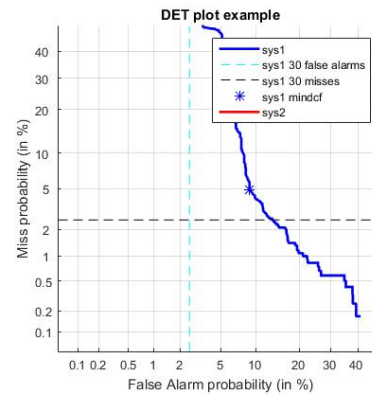
---

Confusion matrix, without normalization  
 $\begin{bmatrix} 540 & 60 \\ 4 & 596 \end{bmatrix}$   
 Normalized confusion matrix  
 $\begin{bmatrix} 0.9 & 0.1 \\ 0.01 & 0.99 \end{bmatrix}$

(d) Confusion Matrix



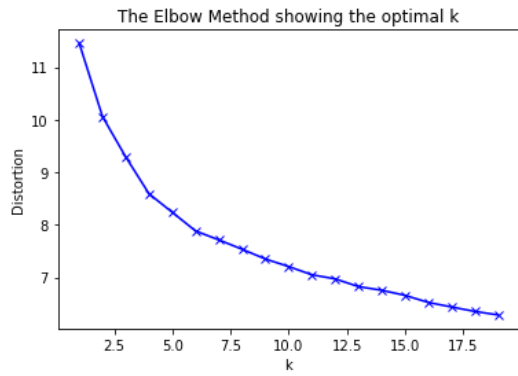
(e) RC Curve



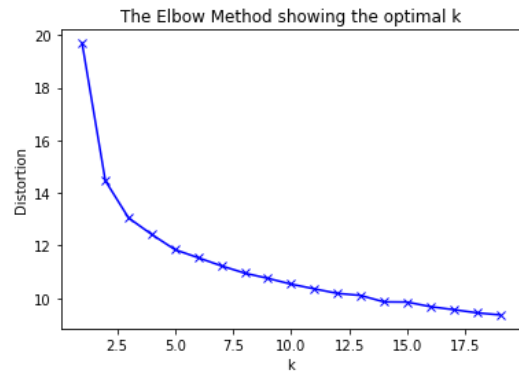
(f) DET Curve

Figure 2: Synthetic Data with Diagonal Covariance Matrix

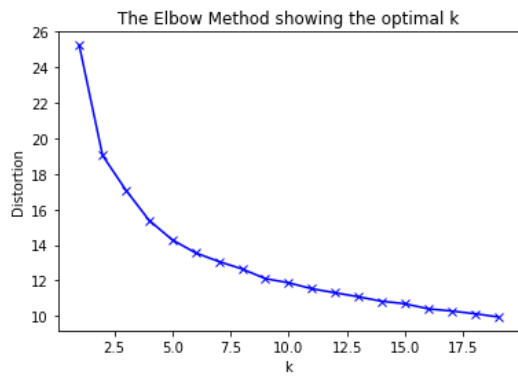
## 4 Image Dataset



(a) Ellbow Curve for Forest



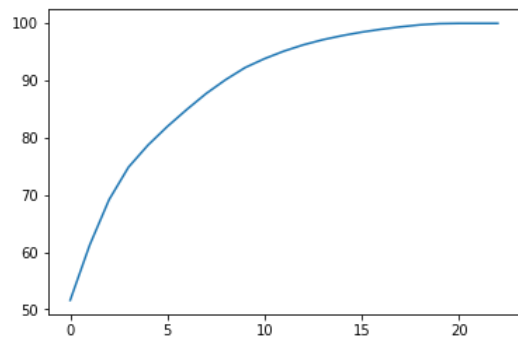
(b) Ellbow Curve for Tall Buildings



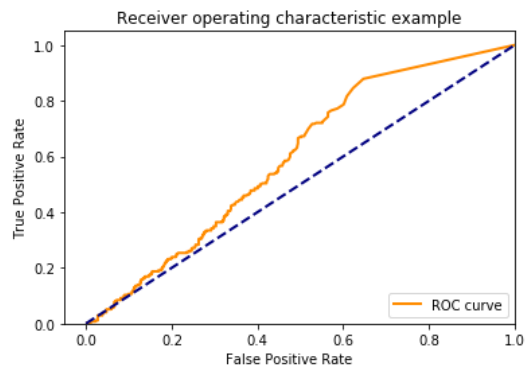
(c) Ellbow Curve for Highways

51	32	15
14	49	35
4	22	50

(d) Confusion Matrix



(c) PCA for class A



(d) ROC Curve

Figure 3: Plots for Image Dataset

## 5 Conclusion

- 1) As we have 23 features for each image, we ran an inbuilt PCA function to identify dominant features.
- 2) We transformed  $n \times 23$  to  $n \times 6$  dimension where 6 features are able to explain 80% of the data to avoid singularity
- 3) The model's accuracy was 55% on test data and 85% for higher cluster values there might not be any point clustered and GMM cannot be constructed
- 5) For Synthetic Data, for reasonable  $K$  (i.e.  $> 5$ ) we are able to fit both diagonal and complete covariance GMMs