PATTERN RECOGNITION AND MACHINE LEARNING

Indian Institute of Technology Madras
M.Tech (Computer Science & Engineering)
SESSION 2018-2019

SUBMITTED BY:
OJAS MEHTA - CS18M038
PRANAV MURALI - CS18M041

SUBMITTED TO
Prof. C. CHANDRA SEKHAR
IIT Madras

GROUP NUMBER : 5

TASK 1 :K-Nearest Neighbour

DATASET 1 : 2-dimensional artificial data of 4 classes:

1. **METHOD**

   We have implemented the KNN model by following the below steps:

   a. Load the data
   b. Initialise the value of k
   c. For getting the predicted class, iterate from 1 to total number of training data points
   d. Calculate the distance between test data and each row of training data. Here we will use Euclidean distance as our distance metric since it's the most popular method. The other metrics that can be used are Chebyshev, cosine, etc.
   e. Sort the calculated distances in ascending order based on distance values
   f. Get top k rows from the sorted array
   g. Get the most frequent class of these rows

   $$K = \sum_{i=1}^{M} K_i$$

   Where 'M' is total number of classes and K_i is number of training examples among 'K' KNN's.

   $$i^* = argmax(K_i)$$

   So class label of input vector $\overline{x}$ is y_{i^*}

   h. Print the predicted class.

2. PLOTS

1) Linearly Separable Data:

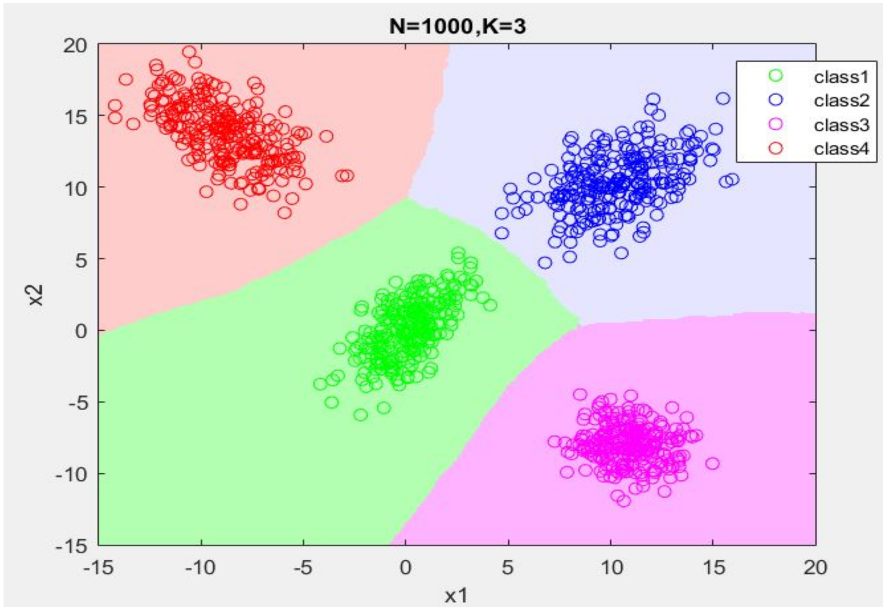

Figure 1.2.1:- Plot of decision region with training data superimposed.

2) Non-Linearly Separable Data:-

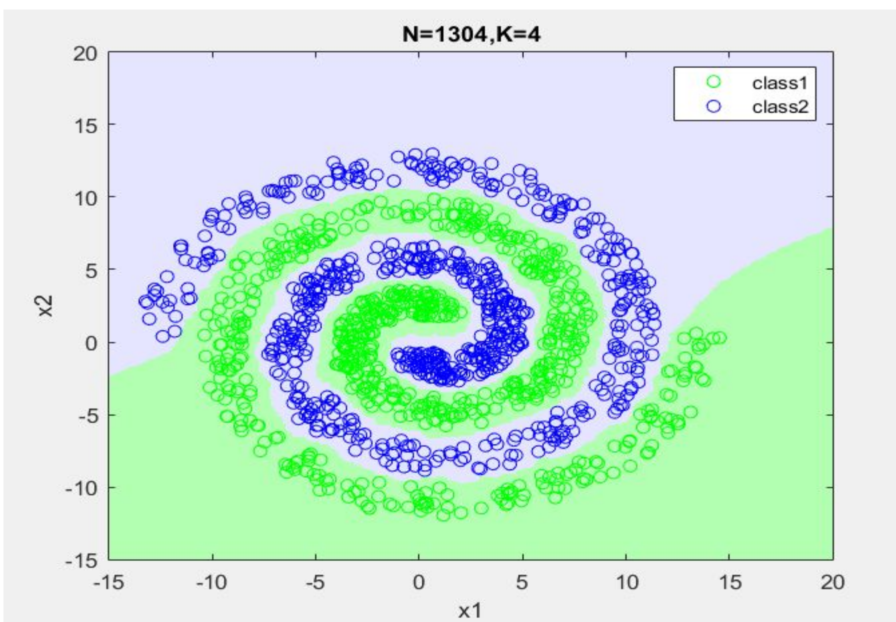

Figure 1.2.2:- Plot of decision region with training data superimposed.

3) **Overlapping Data:-**

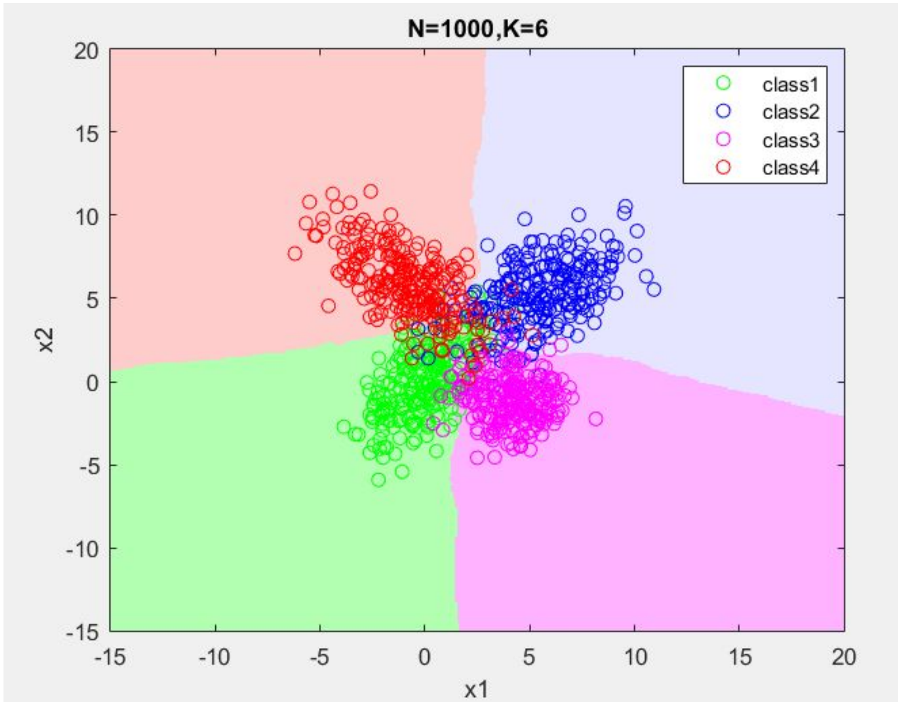

Figure 1.2.3:- Plot of decision region with training data superimposed.

3. RESULTS

1) **Linearly Separable Data:-**

Classification Accuracy on Training Data and Validation Data:-

Training Data	
K(Number of nearest neighbours)	**Classification Accuracy(In percentage)**
3	100
4	100
5	100
6	100

Fig 1.3.1:-Classification Accuracy on Training Data

Validation Data	
K(Number of nearest neighbours)	Classification Accuracy(In percentage)
2	100
3	100
5	100
6	100

Fig 1.3.2:-Classification Accuracy on Validation Data

<u>Classification Accuracy for Best Model on Test data</u>:- 100% (K=3)

<u>Confusion Matrix for Training Data:-</u>

Training Data(N=1000,K=3).				
	Class 1	Class 2	Class 3	Class 4
Class 1	250	0	0	0
Class 2	0	250	0	0
Class 3	0	0	250	0
Class 4	0	0	0	250

Fig 1.3.3:-Confusion Matrix for Training Data.

<u>Confusion Matrix for Test Data:-</u>

Test Data(N=400,K=3).				
	Class 1	Class 2	Class 3	Class 4
Class 1	100	0	0	0
Class 2	0	100	0	0
Class 3	0	0	100	0
Class 4	0	0	0	100

Fig 1.3.4:-Confusion Matrix for Test Data.

2) **Non-Linearly Separable Data:-**

Classification Accuracy on Training Data and Validation Data:-

Training Data	
K(Number of nearest neighbours)	**Classification Accuracy(In percentage)**
3	**100**
4	**100**
5	**100**
6	**100**

Fig 1.3.5:-Classification Accuracy on Training Data.

Validation Data	
K(Number of nearest neighbours)	**Classification Accuracy(In percentage)**
3	**100**
4	**100**
6	**100**
7	**100**

Fig 1.3.6:-Classification Accuracy on Validation Data.

<u>Classification Accuracy for Best Model on Test data</u>:- 100% (K=4)

<u>Confusion Matrix for Training Data:-</u>

Training Data(N=1304,K=3).		
	Class 1	**Class 2**
Class 1	**652**	**0**
Class 2	**0**	**652**

Fig 1.3.7:-Confusion Matrix for Training Data.

Confusion Matrix for Test Data:-

Test Data(N=520,K=4).		
	Class 1	**Class 2**
Class 1	260	0
Class 2	0	260

Fig 1.3.8:-Confusion Matrix for Test Data.

3) **Overlapping Data:-**

Classification Accuracy on Training Data and Validation Data:-

Training Data	
K(Number of nearest neighbours)	**Classification Accuracy(In percentage)**
2	93.30
3	92.9
4	91.8

Fig 1.3.9:-Classification Accuracy on Training Data.

Validation Data	
K(Number of nearest neighbours)	**Classification Accuracy(In percentage)**
1	86.23
2	88.6
5	88.67
6	91
7	90.5

Fig 1.3.10:-Classification Accuracy on Validation Data.

Classification Accuracy for Best Model on Test data:- 90% (K=6)

Confusion Matrix for Training Data:-

Training Data(N=1000,K=2).				
	Class 1	Class 2	Class 3	Class 4
Class 1	250	0	0	0
Class 2	13	237	0	0
Class 3	14	8	228	0
Class 4	18	12	2	218

Fig 1.3.11:-Confusion Matrix for Training Data.

Confusion Matrix for Testing Data:-

Testing Data(N=400,K=6).				
	Class 1	Class 2	Class 3	Class 4
Class 1	91	2	6	1
Class 2	4	91	2	3
Class 3	8	0	92	0
Class 4	8	6	0	86

Fig 1.3.12:-Confusion Matrix for Test Data.

4. INFERENCES

1. In case of linearly separable data, we observe that the accuracy remains the same for k=2,3,4,5. This is because the data points belonging to different classes are far apart. So increasing the number of nearest neighbours does not cause any significant change.
2. In case of non linearly separable data, we observe that decision region is non linear.
3. In case of overlapping data, from Fig. 1.3.10 we can observe that the accuracy increases as we keep on increasing k from 1 to 6. At k=6, we achieve maximum accuracy. We require more number of nearest neighbours because we would more number of points to make a good decision. With less k, classifier has a greater chance of making a wrong decision for the overlapping points.

TASK 2 :Naive-Bayes classifier with a Gaussian distribution

DATASET 1 : 2-dimensional artificial data of 4 classes:

1. **METHOD**

   Naive bayes classifier can be implemented in following steps:

   1. Calculate the covariance matrix for each of the classes.
      a. If each of the classes covariance matrix is assumed to be same then take the average of covariance matrix of all other classes.
      b. As naive bayes the data points are assumed to be independent, make the non-diagonal entries zero.
   2. Calculate the Gaussian probability for each class.
   3. Assign the class which has the maximum probability.
   4. Plot the decision region for each class, observe the decision surface.
   5. Posterior probability is calculated as follows:-

      $$p(y_i/\overline{x}) = \frac{p(\overline{x}/y_i)p(y_i)}{p(\overline{x})}$$

      $$i^* = argmax\ p(y_i/\overline{x})$$

      i) $p(y_i/\overline{x})$ is posterior probability for y_i given input vector $\overline{x}$

      ii) $p(\overline{x}/y_i)$ is likelihood for class y_i.

      iii) $p(y_i)$ is prior probability for class y_i.

   6. Class label for input vector $\overline{x}$ = y_{i^*} where $i^* = argmax(p(y_i)/\overline{x})$.

A. Covariance matrix of all classes are same which is $\sigma^2 I$

1) PLOTS

1. **Linearly Separable:-**

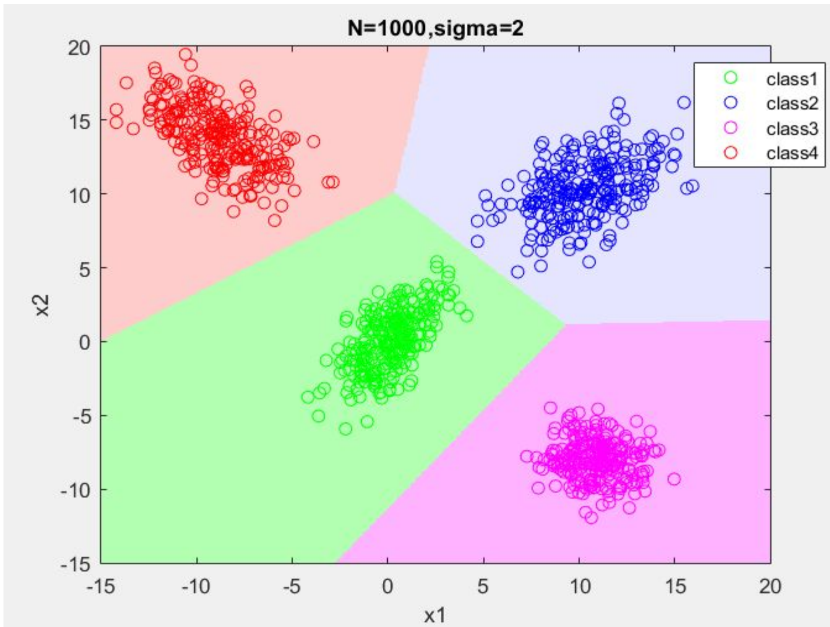

Fig A)2.1.1:- Plot of decision region with training data superimposed.

2. **Non-Linearly Separable:-**

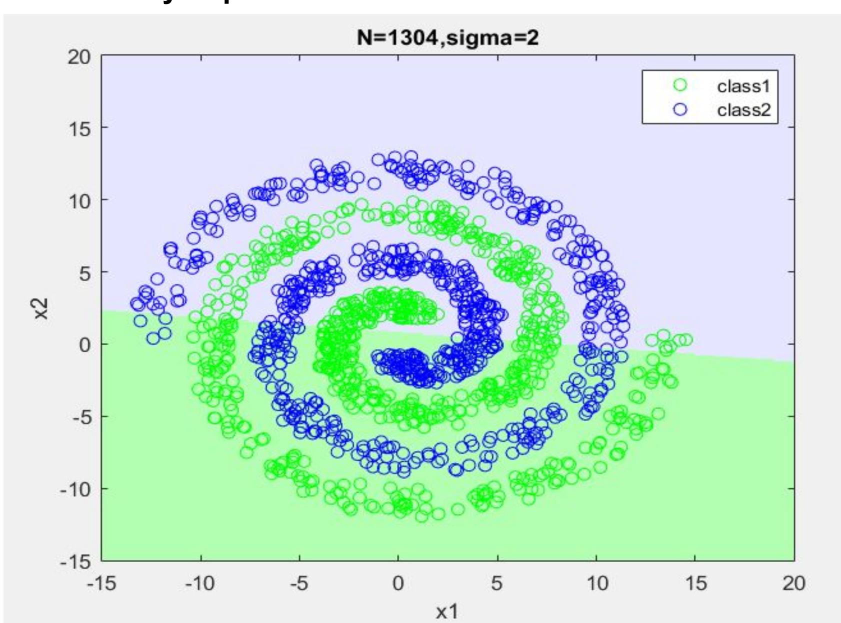

Fig A)2.1.2 :- Plot of decision region with training data superimposed.

3. **Overlapping:-**

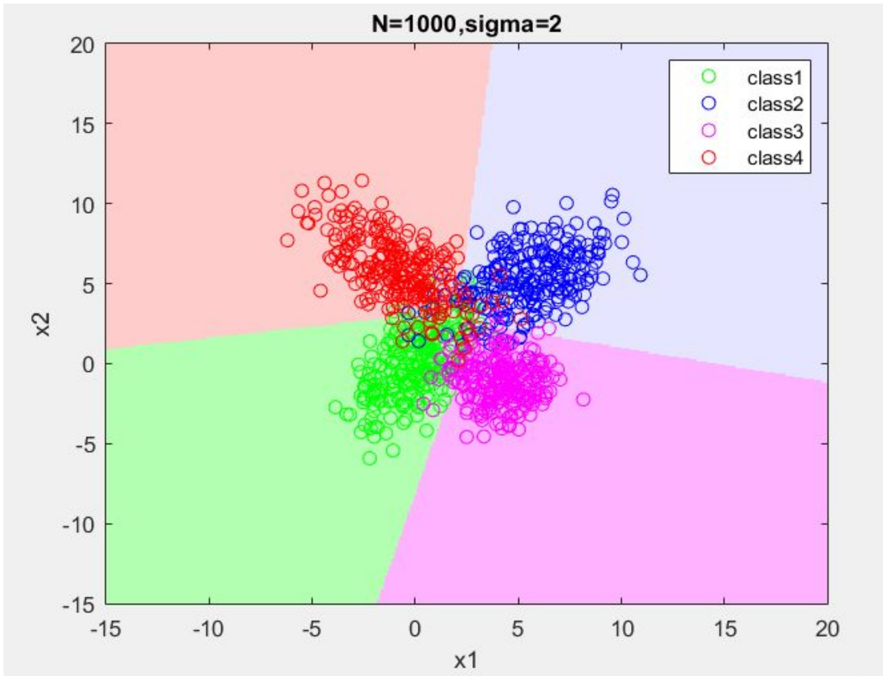

Fig A)2.1.3 :- Plot of decision region with training data superimposed.

2) RESULTS

1. **Linearly Separable:-**

Classification Accuracy on Training Data and Validation Data:-

Training Data	
σ (Sigma)	**Classification Accuracy(In percentage)**
2	**100**
3	**100**
4	**100**
5	**100**

Fig A) 2.2.1:-Classification Accuracy on Training Data

Validation Data	
σ (Sigma)	Classification Accuracy(In percentage)
2	100
3	100
4	100
5	100

Fig A)2.2.2:-Classification Accuracy on Validation Data

<u>Classification Accuracy for Best Model on Test data</u>:- 100% (σ = 2)

<u>Confusion Matrix for Training Data:-</u>

Training Data(N=1000,σ = 2).				
	Class 1	Class 2	Class 3	Class 4
Class 1	250	0	0	0
Class 2	0	250	0	0
Class 3	0	0	250	0
Class 4	0	0	0	250

Fig A)2.2.3:-Confusion Matrix for Training Data.

<u>Confusion Matrix for Test Data:-</u>

Test Data(N=400,σ = 2).				
	Class 1	Class 2	Class 3	Class 4
Class 1	100	0	0	0
Class 2	0	100	0	0
Class 3	0	0	100	0
Class 4	0	0	0	100

Fig A) 2.2.4:-Confusion Matrix for Test Data.

2. **Non-Linearly Separable:-**
   Classification Accuracy on Training Data and Validation Data:-

Training Data	
σ (Sigma)	**Classification Accuracy(In percentage)**
2	**54.91**
3	**54.91**
4	**54.91**

Fig A)2.2.5:-Classification Accuracy on Training Data.

Validation Data	
σ (Sigma)	**Classification Accuracy(In percentage)**
3	**57.16**
4	**57.16**
6	**57.16**

Fig A)2.2.6:-Classification Accuracy on Validation Data.

Classification Accuracy for Best Model on Test data:- 53.65% (σ=2)

Confusion Matrix for Training Data:-

Training Data(N=1304).		
	Class 1	**Class 2**
Class 1	**357**	**295**
Class 2	**293**	**359**

Fig A)2.2.7:-Confusion Matrix for Training Data.

Confusion Matrix for Test Data:-

Test Data(N=520).		
	Class 1	**Class 2**
Class 1	**152**	**108**
Class 2	**133**	**127**

Fig A)2.2.8:-Confusion Matrix for Test data.

3. **Overlapping:-**

Classification Accuracy on Training Data and Validation Data:-

Training Data	
σ (Sigma)	**Classification Accuracy(In percentage)**
2	**89.4**
3	**89.4**
4	**89.4**
10	**89.4**

Fig A) 2.2.9:-Classification Accuracy on Training Data

Validation Data	
σ (Sigma)	**Classification Accuracy(In percentage)**
2	**91.17**
3	**91.17**
4	**91.17**
10	**91.17**

Fig A)2.2.10:-Classification Accuracy on Validation Data

Classification Accuracy for Best Model on Test data:- 91.5% (σ = 2)

Confusion Matrix for Training Data:-

Training Data(N=1000,σ = 2).				
	Class 1	**Class 2**	**Class 3**	**Class 4**
Class 1	**226**	**12**	**2**	**10**
Class 2	**7**	**223**	**11**	**9**
Class 3	**21**	**4**	**225**	**0**
Class 4	**19**	**10**	**1**	**220**

Fig A)2.2.11:-Confusion Matrix for Training Data.

Confusion Matrix for Test Data:-

Test Data(N=400,σ = 2).				
	Class 1	Class 2	Class 3	Class 4
Class 1	94	0	5	1
Class 2	2	92	2	4
Class 3	10	0	90	0
Class 4	7	3	0	90

Fig A) 2.2.12:-Confusion Matrix for Test Data.

3) INFERENCES

In the above case, covariance matrix for all classes is same with diagonal entries $\sigma^2 I$. So the equation for the decision region would be linear and will be of form:-

$$\overline{w}^T \overline{x} + w_0 = 0$$

Where w_0 is a constant and $\overline{w}^T$ is (dx1) matrix where 'd' is the dimension of the input vector.

As the decision surface is linear, we get a good accuracy of 100% on linearly separable data, 91.17% accuracy on overlapping data and 54.91% accuracy on non-linearly separable data.

B. Covariance matrix of all classes are same which is C.

1) PLOTS

1. **Linearly Separable:-**

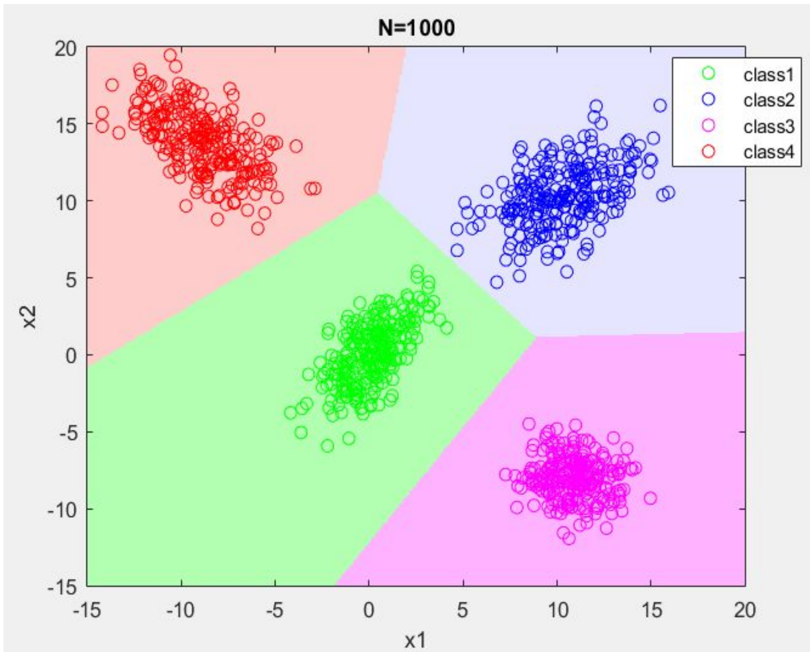

Fig B)2.1.1:-Plot of decision region with training data superimposed.

2. **Non-Linearly Separable:-**

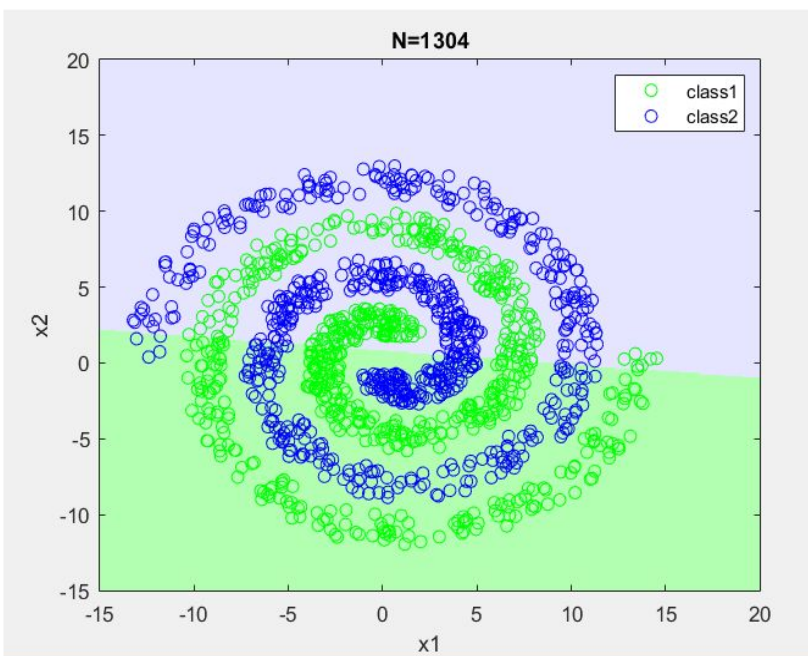

Fig B)2.1.2 :-Plot of decision region with training data superimposed.

3. **Overlapping:-**

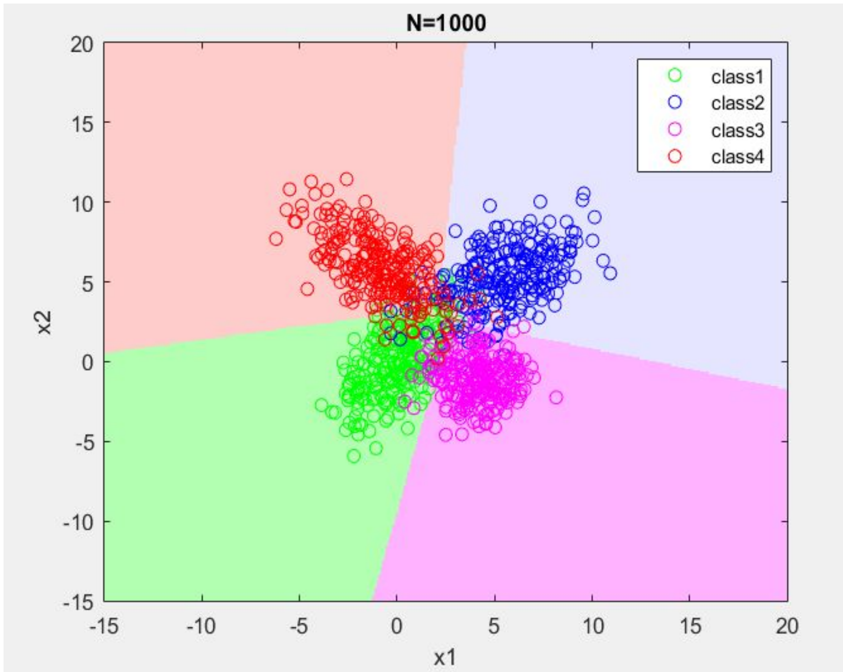

Fig B)2.1.3:-Plot of decision region with training data superimposed.

2) RESULTS

1. **Linearly Separable:-**

Classification Accuracy on Training Data and Validation Data:-

Training Data
Classification Accuracy(In percentage)
100

Fig B) 2.2.1:-Classification Accuracy on Training Data

Validation Data
Classification Accuracy(In percentage)
100

Fig B)2.2.2:-Classification Accuracy on Validation Data

<u>Classification Accuracy for Best Model on Test data</u>:- 100%

Confusion Matrix for Training Data:-

Training Data(N=1000).				
	Class 1	Class 2	Class 3	Class 4
Class 1	250	0	0	0
Class 2	0	250	0	0
Class 3	0	0	250	0
Class 4	0	0	0	250

Fig B)2.2.3:-Confusion Matrix for Training Data.

Confusion Matrix for Test Data:-

Test Data(N=400).				
	Class 1	Class 2	Class 3	Class 4
Class 1	100	0	0	0
Class 2	0	100	0	0
Class 3	0	0	100	0
Class 4	0	0	0	100

Fig B) 2.2.4:-Confusion Matrix for Test Data.

2. **Non-Linearly Separable:-**
   Classification Accuracy on Training Data and Validation Data:-

Training Data
Classification Accuracy(In percentage)
54.83

Fig B)2.2.5:-Classification Accuracy on Training Data.

Validation Data
Classification Accuracy(In percentage)
57.29

Fig B)2.2.6:-Classification Accuracy on Validation Data.

<u>Classification Accuracy for Best Model on Test data</u>:- 53.65%

<u>Confusion Matrix for Training Data:-</u>

Training Data(N=1304).		
	Class 1	Class 2
Class 1	357	295
Class 2	294	358

Fig B)2.2.7:-Confusion Matrix for Training Data.

<u>Confusion Matrix for Test Data:-</u>

Test Data(N=520).		
	Class 1	Class 2
Class 1	151	109
Class 2	132	128

Fig B)2.2.8:-Confusion Matrix for Test data.

3. **Overlapping:-**
   Classification Accuracy on Training Data and Validation Data:-

Training Data
Classification Accuracy(In percentage)
89.2

Fig B) 2.2.9:-Classification Accuracy on Training Data

Validation Data
Classification Accuracy(In percentage)
91.17

Fig B)2.2.10:-Classification Accuracy on Validation Data

<u>Classification Accuracy for Best Model on Test data</u>:- 91.7%.

<u>Confusion Matrix for Training Data:-</u>

Training Data(N=1000).				
	Class 1	**Class 2**	**Class 3**	**Class 4**
Class 1	224	10	5	11
Class 2	7	223	11	9
Class 3	21	4	225	0
Class 4	18	10	2	220

Fig B)2.2.11:-Confusion Matrix for Training Data.

<u>Confusion Matrix for Test Data:-</u>

Test Data(N=400).				
	Class 1	**Class 2**	**Class 3**	**Class 4**
Class 1	95	0	4	1
Class 2	2	92	2	4
Class 3	8	1	91	0
Class 4	7	3	0	90

Fig B) 2.2.12:-Confusion Matrix for Test Data.

3) INFERENCES

In the above case, covariance matrix for all classes is same which is C.
The decision surface is linear in this case.We get a good accuracy of 100% on linearly separable data, 91.17% accuracy on overlapping data and 54.91% accuracy on non-linearly separable data.

C. **Covariance matrix of each class is different.**

1) PLOTS

1. **Linearly Separable:-**

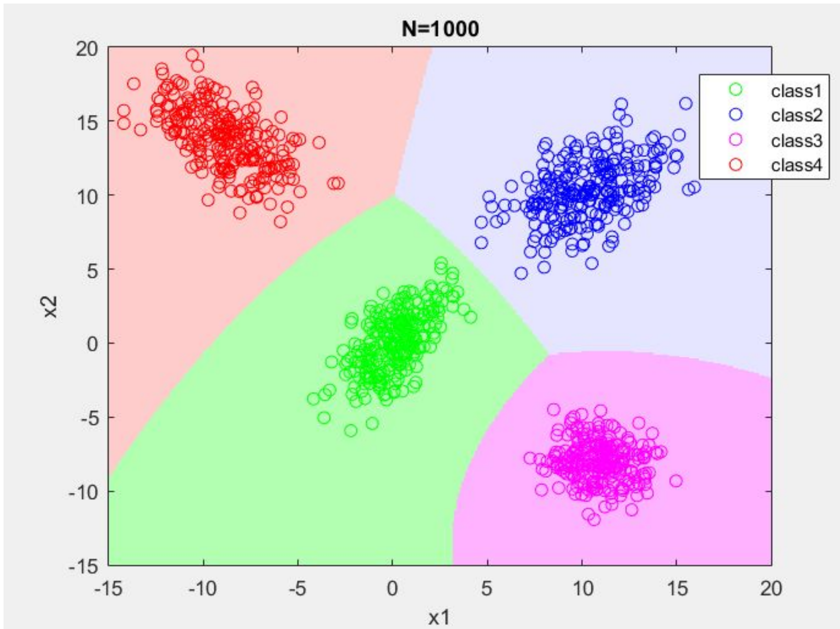

Fig C)2.1.1:-Plot of decision region with training data superimposed.

2. **Non-Linearly Separable:-**

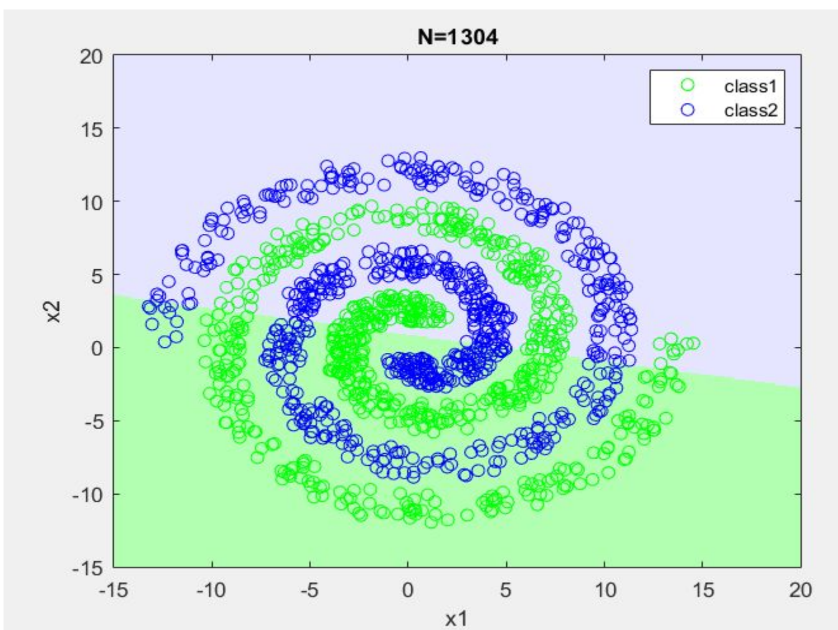

Fig C)2.1.2 :-Plot of decision region with training data superimposed.

3. **Overlapping:-**

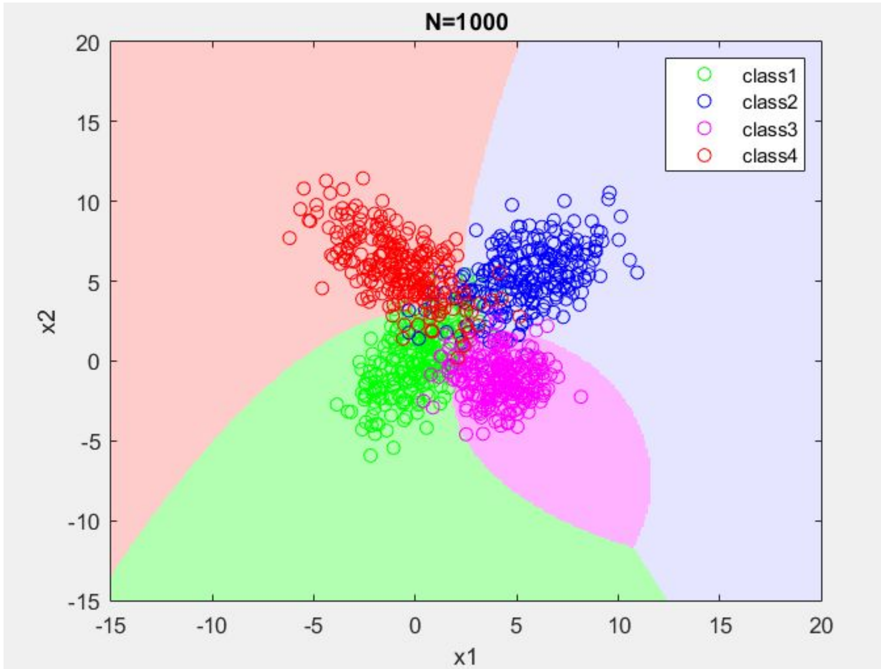

Fig C)2.1.3:-Plot of decision region with training data superimposed.

2) RESULTS

1. **Linearly Separable:-**

Classification Accuracy on Training Data and Validation Data:-

Training Data
Classification Accuracy(In percentage)
100

Fig C) 2.2.1:-Classification Accuracy on Training Data

Validation Data
Classification Accuracy(In percentage)
100

Fig C)2.2.2:-Classification Accuracy on Validation Data

Classification Accuracy for Best Model on Test data:- 100%

Confusion Matrix for Training Data:-

Training Data(N=1000).				
	Class 1	**Class 2**	**Class 3**	**Class 4**
Class 1	250	0	0	0
Class 2	0	250	0	0
Class 3	0	0	250	0
Class 4	0	0	0	250

Fig C)2.2.3:-Confusion Matrix for Training Data.

Confusion Matrix for Test Data:-

Test Data(N=400).				
	Class 1	**Class 2**	**Class 3**	**Class 4**
Class 1	100	0	0	0
Class 2	0	100	0	0
Class 3	0	0	100	0
Class 4	0	0	0	100

Fig C) 2.2.4:-Confusion Matrix for Test Data.

2. **Non-Linearly Separable:-**

Classification Accuracy on Training Data and Validation Data:-

Training Data
Classification Accuracy(In percentage)
56.83

Fig C)2.2.5:-Classification Accuracy on Training Data.

Validation Data
Classification Accuracy(In percentage)
58.91

Fig C)2.2.6:-Classification Accuracy on Validation Data.

Classification Accuracy for Best Model on Test data:- 54.85%

<u>Confusion Matrix for Training Data:-</u>

Training Data(N=1304).		
	Class 1	Class 2
Class 1	363	289
Class 2	292	360

Fig C)2.2.7:-Confusion Matrix for Training Data.

<u>Confusion Matrix for Test Data:-</u>

Test Data(N=520).		
	Class 1	Class 2
Class 1	154	105
Class 2	127	133

Fig C)2.2.8:-Confusion Matrix for Test data.

3. **Overlapping:-**

Classification Accuracy on Training Data and Validation Data:-

Training Data
Classification Accuracy(In percentage)
89.4

Fig C) 2.2.9:-Classification Accuracy on Training Data.

Validation Data
Classification Accuracy(In percentage)
91.33

Fig C)2.2.10:-Classification Accuracy on Validation Data

<u>Classification Accuracy for Best Model on Test data</u>:- 91.5%.

Confusion Matrix for Training Data:-

Training Data(N=1000).				
	Class 1	**Class 2**	**Class 3**	**Class 4**
Class 1	224	10	5	11
Class 2	7	223	11	9
Class 3	21	4	225	0
Class 4	18	10	2	220

Fig C)2.2.11:-Confusion Matrix for Training Data.

Confusion Matrix for Test Data:-

Test Data(N=400).				
	Class 1	**Class 2**	**Class 3**	**Class 4**
Class 1	95	0	4	1
Class 2	2	92	2	4
Class 3	8	1	91	0
Class 4	7	3	0	90

Fig C) 2.2.12:-Confusion Matrix for Test Data.

3) INFERENCES

In the above case, covariance matrix for all classes are different.
So the equation for the decision region would be non linear and will be of form:-

$$\overline{x}^T W \overline{x} + \overline{w}^T \overline{x} + w_0 = 0$$

Where w_0 is a constant, $\overline{x}^T W \overline{x}$ generates 2^{nd} degree monomials and $\overline{w}^T \overline{x}$ generates monomials of degree 1.

As the decision surface is non-linear, we get a good accuracy of 100% on linearly separable data, 91.5% accuracy on overlapping data and 54.85% accuracy on non-linearly separable data as non-linear surface can adapt itself to fit the data more properly than linear surface

TASK 3 :Bayes classifier with a Gaussian distribution

DATASET 1 : 2-dimensional artificial data of 4 classes:

1. **METHOD :**

   Bayes classifier can be implemented in following steps:

   1. Calculate the covariance matrix for each of the classes.
      a. If each of the classes covariance matrix is assumed to be same then take the average of covariance matrix of all other classes.

   2. Calculate the Gaussian probability for each class.
   3. Assign the class which has the maximum probability.
   4. Plot the decision region for each class, observe the decision surface.
   5. Posterior probability is calculated as follows:-

   $$p(y_i/\overline{x}) = \frac{p(\overline{x}/y_i)p(y_i)}{p(\overline{x})}$$

   $$i^* = argmax\ p(y_i/\overline{x})$$

   i) $p(y_i/\overline{x})$ is posterior probability for y_i given input vector $\overline{x}$

   ii) $p(\overline{x}/y_i)$ is likelihood for class y_i.

   iii) $p(y_i)$ is prior probability for class y_i.

   6. Class label for input vector $\overline{x}$ = y_{i^*} where $i^* = argmax(p(y_i)/\overline{x})$.

A. **Covariance matrix of all classes are same which is C.**

1) PLOTS

1. **Linearly Separable:-**

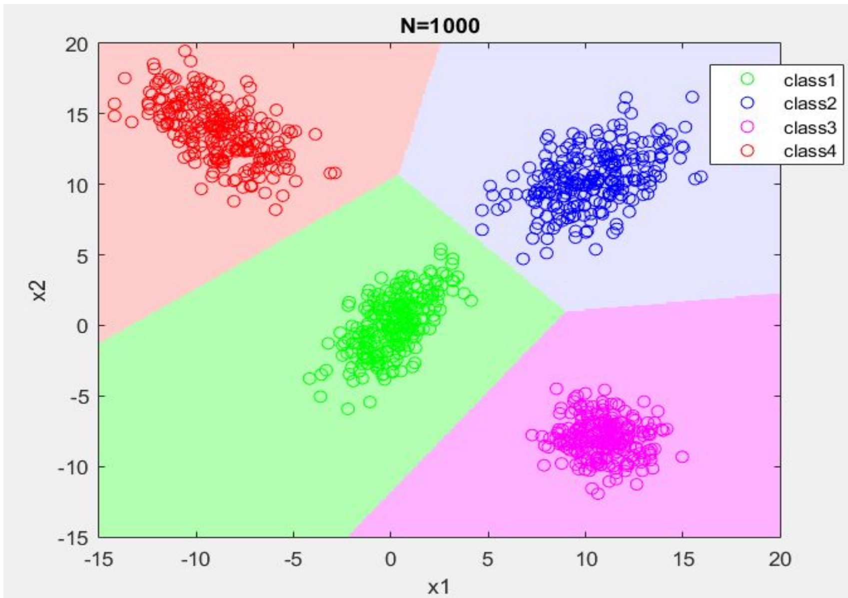

Fig A)3.1.1:-Plot of decision region with training data superimposed.

2. **Non-Linearly Separable:-**

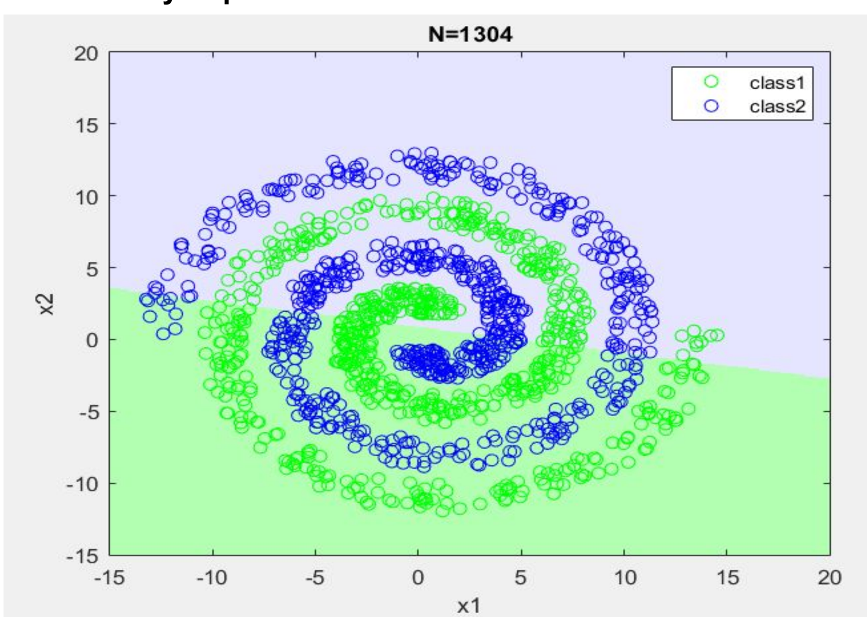

Fig A)3.1.2 :-Plot of decision region with training data superimposed.

3. **Overlapping:-**

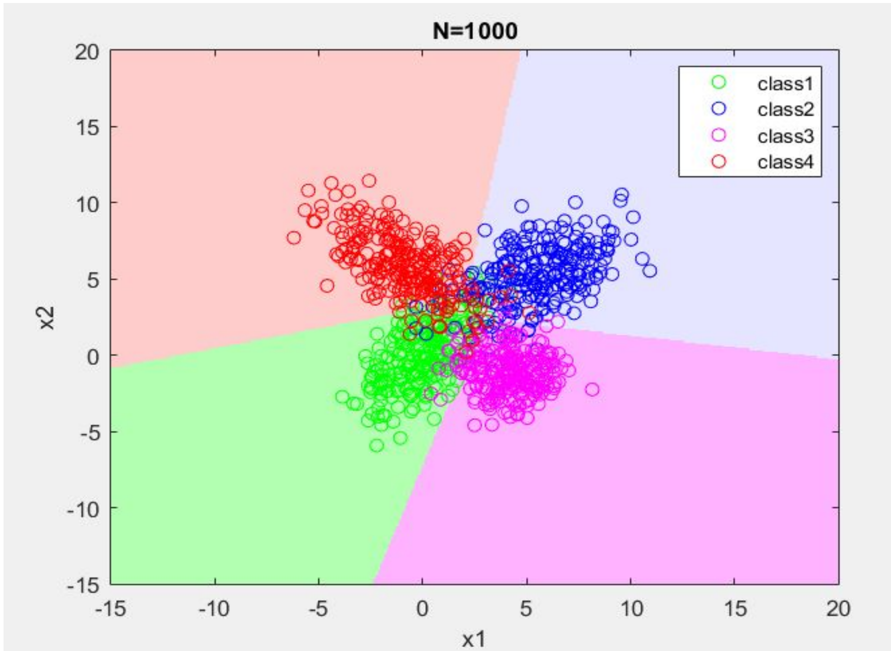

Fig A)3.1.3:-Plot of decision region with training data superimposed.

2) RESULTS

1. **Linearly Separable:-**

Classification Accuracy on Training Data and Validation Data:-

Training Data
Classification Accuracy(In percentage)
100

Fig A) 3.2.1:-Classification Accuracy on Training Data

Validation Data
Classification Accuracy(In percentage)
100

Fig A)3.2.2:-Classification Accuracy on Validation Data

<u>Classification Accuracy for Best Model on Test data</u>:- 100%

Confusion Matrix for Training Data:-

Training Data(N=1000).				
	Class 1	Class 2	Class 3	Class 4
Class 1	250	0	0	0
Class 2	0	250	0	0
Class 3	0	0	250	0
Class 4	0	0	0	250

Fig A)3.2.3:-Confusion Matrix for Training Data.

Confusion Matrix for Test Data:-

Test Data(N=400).				
	Class 1	Class 2	Class 3	Class 4
Class 1	100	0	0	0
Class 2	0	100	0	0
Class 3	0	0	100	0
Class 4	0	0	0	100

Fig A) 3.2.4:-Confusion Matrix for Test Data.

2. **Non-Linearly Separable:-**

Classification Accuracy on Training Data and Validation Data:-

Training Data
Classification Accuracy(In percentage)
55.29

Fig A)3.2.5:-Classification Accuracy on Training Data.

Validation Data
Classification Accuracy(In percentage)
57.16

Fig A)3.2.6:-Classification Accuracy on Validation Data.

Classification Accuracy for Best Model on Test data:- 53.85%

Confusion Matrix for Training Data:-

Training Data(N=1304).		
	Class 1	Class 2
Class 1	357	295
Class 2	288	364

Fig A)3.2.7:-Confusion Matrix for Training Data.

Confusion Matrix for Test Data:-

Test Data(N=520).		
	Class 1	Class 2
Class 1	151	109
Class 2	131	129

Fig A)3.2.8:-Confusion Matrix for Test data.

3. **Overlapping:-**

Classification Accuracy on Training Data and Validation Data:-

Training Data
Classification Accuracy(In percentage)
89.5

Fig A) 3.2.9:-Classification Accuracy on Training Data

Validation Data
Classification Accuracy(In percentage)
91.17

Fig A)3.2.10:-Classification Accuracy on Validation Data

Classification Accuracy for Best Model on Test data:- 91%

Confusion Matrix for Training Data:-

Training Data(N=1000).				
	Class 1	Class 2	Class 3	Class 4
Class 1	227	13	1	9
Class 2	8	223	11	8
Class 3	21	3	226	0
Class 4	19	11	1	219

Fig A)3.2.11:-Confusion Matrix for Training Data.

Confusion Matrix for Test Data:-

Test Data(N=400).				
	Class 1	Class 2	Class 3	Class 4
Class 1	95	1	4	0
Class 2	2	92	2	4
Class 3	10	0	90	0
Class 4	10	3	0	87

Fig A) 3.2.12:-Confusion Matrix for Test Data.

3) INFERENCES

In the above case, covariance matrix for all classes is same which is C.
The decision surface is linear in this case.We get a good accuracy of 100% on linearly separable data, 91% accuracy on overlapping data and 53.85% accuracy on non-linearly separable data.

B. Covariance matrix of each class is different.

1) PLOTS

1. **Linearly Separable:-**

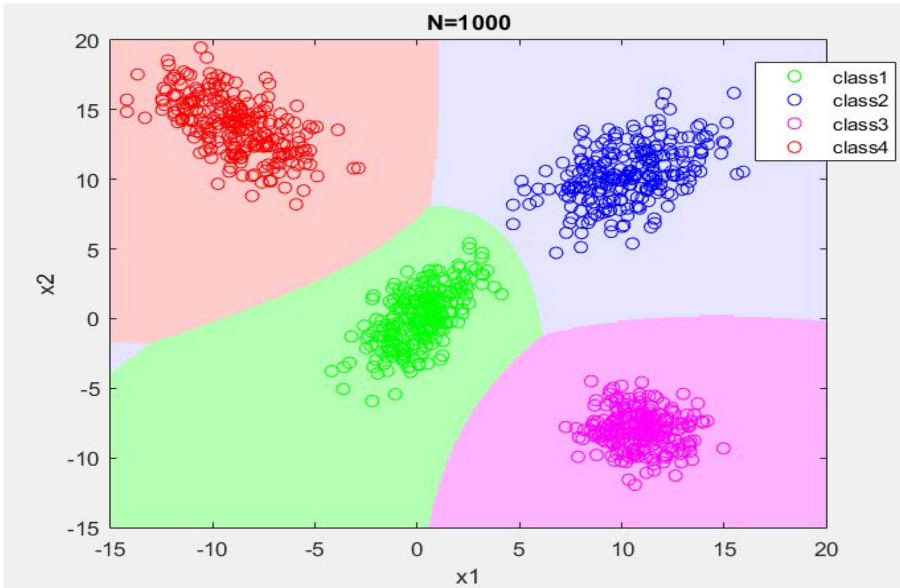

Fig B)3.1.1:-Plot of decision region with training data superimposed.

2. **Non-Linearly Separable:-**

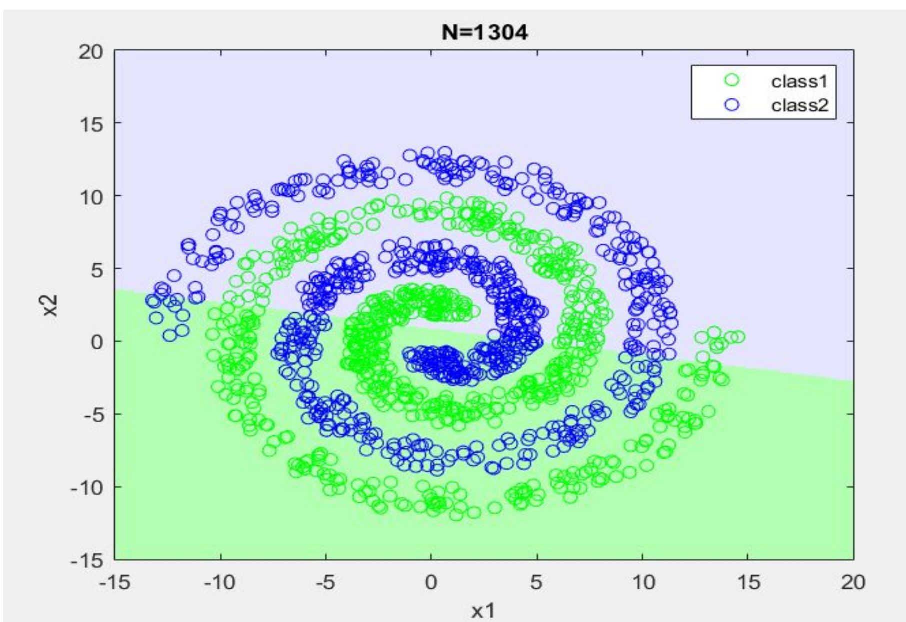

Fig B)3.1.2 :-Plot of decision region with training data superimposed.

3. **Overlapping:-**

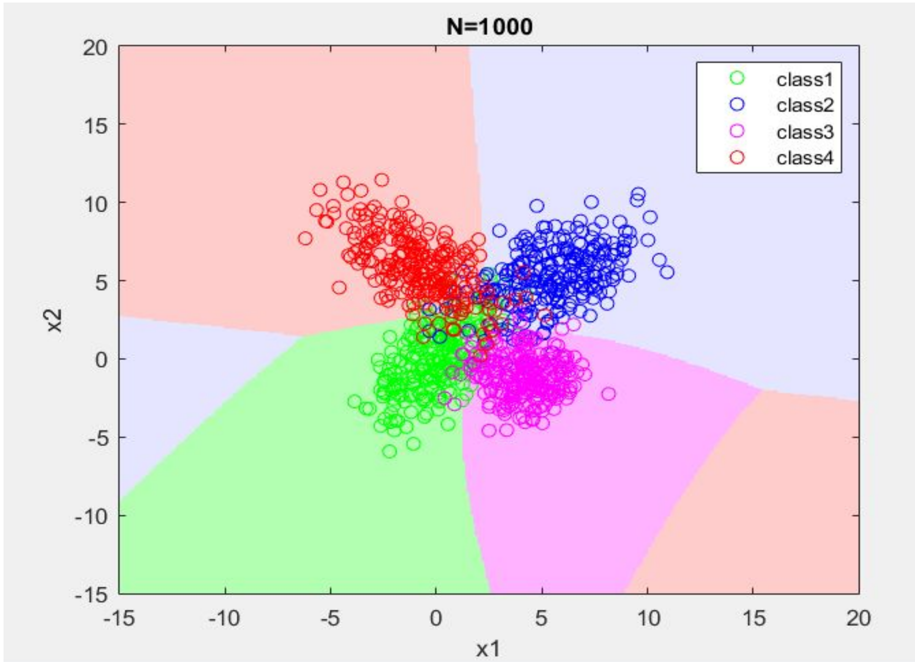

Fig B)3.1.3:-Plot of decision region with training data superimposed.

2) RESULTS

1. **Linearly Separable:-**

Classification Accuracy on Training Data and Validation Data:-

Training Data
Classification Accuracy(In percentage)
100

Fig B) 3.2.1:-Classification Accuracy on Training Data

Validation Data
Classification Accuracy(In percentage)
100

Fig B)3.2.2:-Classification Accuracy on Validation Data

<u>Classification Accuracy for Best Model on Test data</u>:- 100%

Confusion Matrix for Training Data:-

Training Data(N=1000).				
	Class 1	Class 2	Class 3	Class 4
Class 1	250	0	0	0
Class 2	0	250	0	0
Class 3	0	0	250	0
Class 4	0	0	0	250

Fig B)3.2.3:-Confusion Matrix for Training Data.

Confusion Matrix for Test Data:-

Test Data(N=400).				
	Class 1	Class 2	Class 3	Class 4
Class 1	100	0	0	0
Class 2	0	100	0	0
Class 3	0	0	100	0
Class 4	0	0	0	100

Fig B) 3.2.4:-Confusion Matrix for Test Data.

2. **Non-Linearly Separable:-**

Classification Accuracy on Training Data and Validation Data:-

Training Data
Classification Accuracy(In percentage)
55.14

Fig B)3.2.5:-Classification Accuracy on Training Data.

Validation Data
Classification Accuracy(In percentage)
57.42

Fig B)3.2.6:-Classification Accuracy on Validation Data.

<u>Classification Accuracy for Best Model on Test data</u>:- 53.46%

<u>Confusion Matrix for Training Data:-</u>

Training Data(N=1304).		
	Class 1	Class 2
Class 1	357	295
Class 2	290	362

Fig B)3.2.7:-Confusion Matrix for Training Data.

<u>Confusion Matrix for Test Data:-</u>

Test Data(N=520).		
	Class 1	Class 2
Class 1	150	110
Class 2	132	128

Fig B)3.2.8:-Confusion Matrix for Test data.

3. **Overlapping:-**

Classification Accuracy on Training Data and Validation Data:-

Training Data
Classification Accuracy(In percentage)
89.7

Fig B) 3.2.9:-Classification Accuracy on Training Data

Validation Data
Classification Accuracy(In percentage)
90.83

Fig B)3.2.10:-Classification Accuracy on Validation Data

<u>Classification Accuracy for Best Model on Test data</u>:- 91.75%

Confusion Matrix for Training Data:-

Training Data(N=1000).				
	Class 1	Class 2	Class 3	Class 4
Class 1	220	10	8	12
Class 2	7	225	8	10
Class 3	15	4	231	0
Class 4	16	11	2	221

Fig B)3.2.11:-Confusion Matrix for Training Data.

Confusion Matrix for Test Data:-

Test Data(N=400).				
	Class 1	Class 2	Class 3	Class 4
Class 1	93	1	5	1
Class 2	2	92	2	4
Class 3	7	1	92	0
Class 4	8	2	0	90

Fig B) 3.2.12:-Confusion Matrix for Test Data.

3) INFERENCES

In the above case, covariance matrix for all classes are different.
So the equation for the decision region would be non linear and will be of form:-

$$\overline{x}^T W \overline{x} + \overline{w}^T \overline{x} + w_0 = 0$$

Where w_0 is a constant, $\overline{x}^T W \overline{x}$ generates 2nd degree monomials and $\overline{w}^T \overline{x}$ generates monomials of degree 1.

As the decision surface is non-linear, we get a good accuracy of 100% on linearly separable data, 91.75% accuracy on overlapping data and 53.46% accuracy on non-linearly separable data.

TASK 4 :Naive-Bayes classifier with GMM for each class.

DATASET 1 : 2-dimensional artificial data of 4 classes:

1. **METHOD :**

   The gaussian mixture model is represented as

   $$p(\overline{x}) = \sum_{q=1}^{Q} w_q N\left(\overline{x}/\overline{\mu_q}, \overline{c_q}\right)$$

   Where, w_q = mixture coefficient of q^th^ component

   $\overline{\mu_q}$ = mean of all data points belonging to q^th^ component.

   $\overline{c_q}$ = covariance of all data points belonging to q^th^ component.

   We used Expectation-Maximization algorithm to find parameters of Gaussian Mixture models. The steps were as follows:

   a) Initialized the mixing coefficients, means and covariances using k-means clustering and calculated the initial likelihood and responsibilities using them as

      , **Repeat(steps b, c) until convergence(log likelihood doesn't change significantly):**

   b) Calculated the parameters θ = {ω , μ , Σ } using the current responsibilities (γ).

   $$w_q = \frac{N_q}{N}$$

   $$\overline{\mu_q} = \frac{1}{N_q} \sum_{n=1}^{N} \gamma_{nq} \overline{x_n}$$

   $$c_q = \frac{1}{N_q} \sum_{n=1}^{N} \gamma_{nq} (\overline{x_n} - \overline{\mu_q})(\overline{x_n} - \overline{\mu_q})^T$$

   c) Estimated the log likelihood using these parameters as follows

   $$ln\ p(X|w, \mu, \Sigma) = \sum_{n=1}^{N} ln(\sum_{k=1}^{K} w_k\ \mathcal{N}(x_n|\mu_k, \Sigma_k))$$

   d) Calculated the new gamma using these parameters as

   $$\gamma_{nq} = \frac{w_q N(\overline{x_n}/\overline{\mu_q}, c_q)}{\sum_{j=1}^{Q} p(z_j = 1) p(\overline{x}/z_j = 1)}$$

- The features in Naive Bayes theorem are assumed to be independent and hence the covariance matrix obtained is diagonal.

- This is particularly useful for higher number of dimensions as the parameters of covariance matrix to be calculated are reduced to d from d(d-1)/2.

- The naive Bayes assumption may result in a not so good representation of class conditional density.

2. PLOT

1) Linearly separable data:-

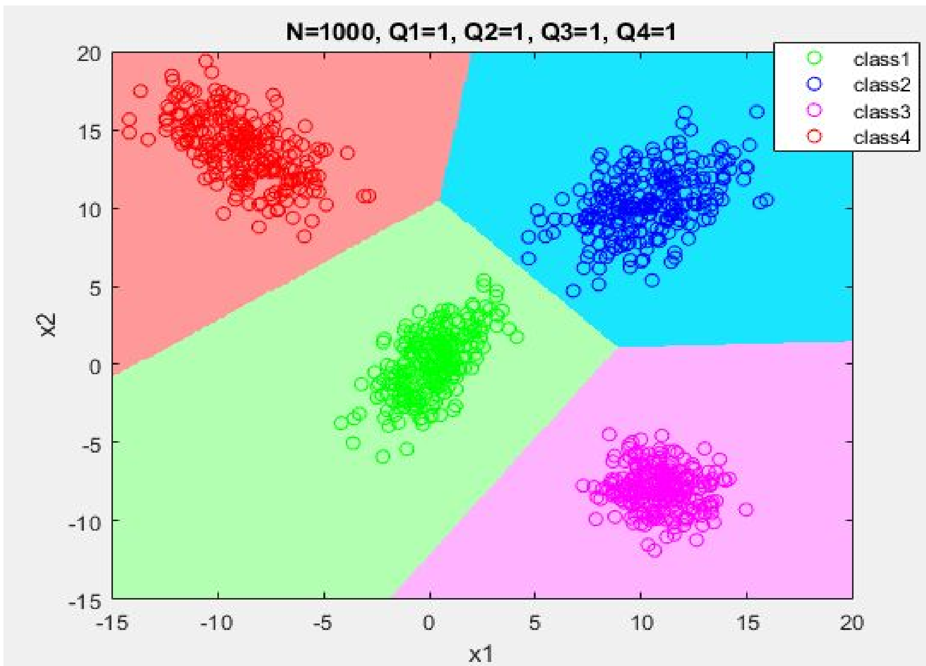

Fig 4.2.1:- Plot of decision region with training data superimposed.

2) **Non-Linearly Separable Data:-**

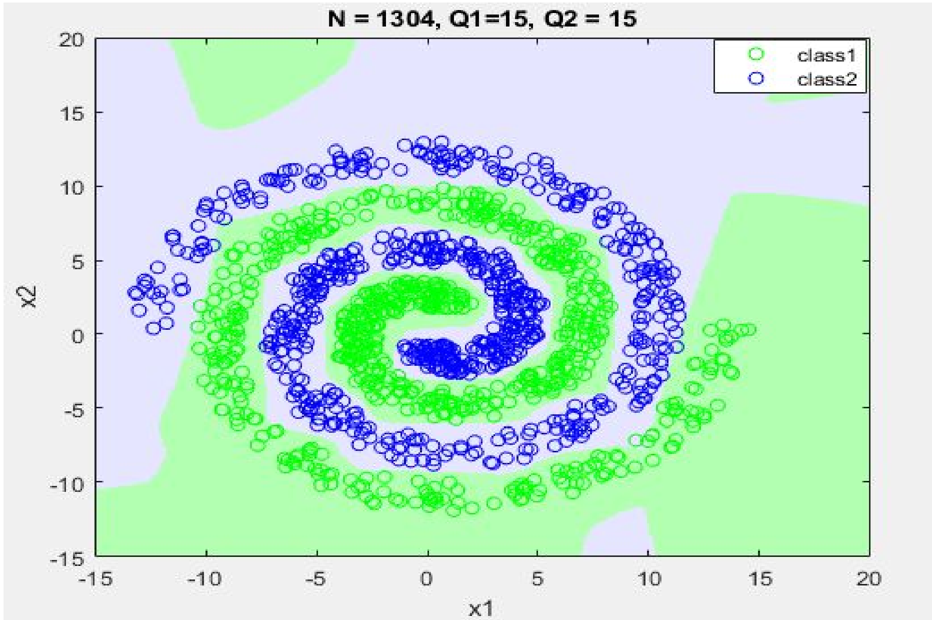

Fig 4.2.2:- Plot of decision region with training data superimposed.

3) **Overlapping Data:-**

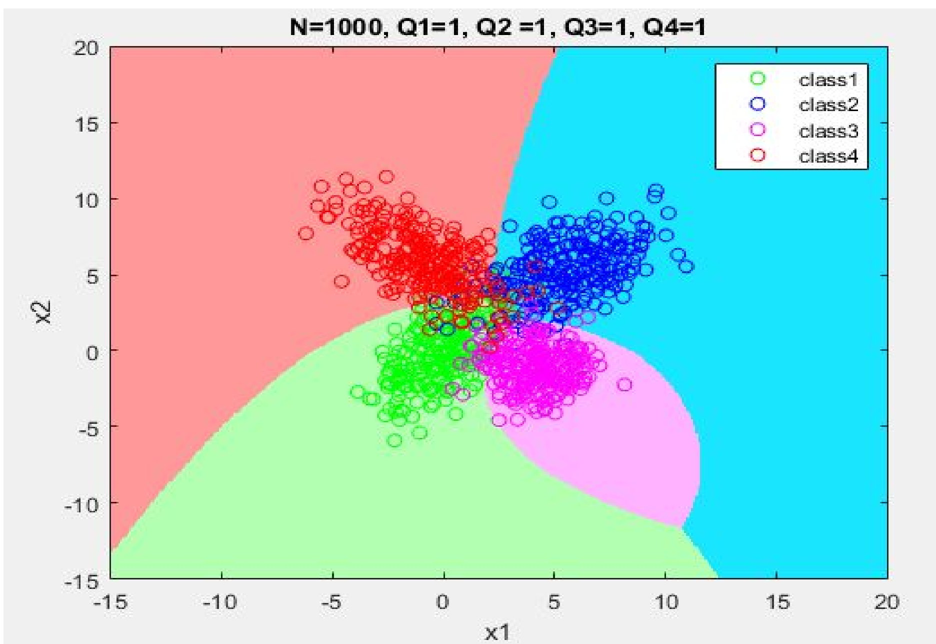

Fig 4.2.3:- Plot of decision region with training data superimposed.

3. RESULTS

1) Linearly Separable Data:-

Classification Accuracy on Training Data and Validation Data:-

Training Data				
Q1	Q2	Q3	Q4	Classification Accuracy(In percentage)
1	1	1	1	100
2	1	1	1	100
2	2	3	3	100
3	3	3	3	100

Fig 4.3.1:-Classification Accuracy on Training Data

Validation Data				
Q1	Q2	Q3	Q4	Classification Accuracy(In percentage)
1	1	1	1	100
2	1	1	1	100
2	2	3	3	100
3	3	3	3	100

Fig 4.3.2:-Classification Accuracy on Validation Data

Classification Accuracy for Best Model on Test data:- 100% (Q1=1, Q2=1, Q3=1, Q4 = 1)

Confusion Matrix for Training Data:-

Training Data(N=1000, Q1=1, Q2=1, Q3=1, Q4 = 1)				
	Class 1	Class 2	Class 3	Class 4
Class 1	250	0	0	0
Class 2	0	250	0	0
Class 3	0	0	250	0
Class 4	0	0	0	250

Fig 4.3.3:-Confusion Matrix for Training Data.

Confusion Matrix for Test Data:-

Test Data(N=400, Q1=1, Q2=1, Q3=1, Q4 = 1).				
	Class 1	Class 2	Class 3	Class 4
Class 1	100	0	0	0
Class 2	0	100	0	0
Class 3	0	0	100	0
Class 4	0	0	0	100

Fig 4.3.4:-Confusion Matrix for Test Data.

2) **Non-Linearly Separable Data:-**

Classification Accuracy on Training Data and Validation Data:-

Training Data		
Q1	Q2	Classification Accuracy(In percentage)
1	1	54.68
5	5	74.85
10	10	95.71
15	15	99.92
16	16	99.85

Fig 4.3.5:-Classification Accuracy on Training Data.

Validation Data		
Q1	Q2	Classification Accuracy(In percentage)
1	1	56.91
5	5	78.01
10	10	96.04
15	15	99.49
16	16	99.49

Fig 4.3.6:-Classification Accuracy on Validation Data.

Classification Accuracy for Best Model on Test data:- 99.49% (Q1 = 15, Q2 = 15)

Confusion Matrix for Training Data:-

Training Data(N=1304, Q1 = 15, Q2 = 15).		
	Class 1	Class 2
Class 1	651	1
Class 2	0	652

Fig 4.3.7:-Confusion Matrix for Training Data.

Confusion Matrix for Test Data:-

Test Data(N=1304,Q1 = 15, Q2 = 15).		
	Class 1	Class 2
Class 1	259	1
Class 2	4	256

Fig 4.3.8:-Confusion Matrix for Test Data.

3) **Overlapping Data:-**

Training Data				
Q1	**Q2**	**Q3**	**Q4**	**Classification Accuracy(In percentage)**
1	**1**	**1**	**1**	**89.40**
2	**2**	**1**	**1**	**89.20**
2	**3**	**1**	**1**	**89.00**
3	**3**	**1**	**1**	**90.17**

Fig 4.3.9:-Classification Accuracy on Training Data

Validation Data				
Q1	**Q2**	**Q3**	**Q4**	**Classification Accuracy(In percentage)**
1	**1**	**1**	**1**	**91.33**
2	**2**	**1**	**1**	**90.83**
2	**3**	**1**	**1**	**90.67**
3	**3**	**1**	**1**	**90.17**

Fig 4.3.10:-Classification Accuracy on Validation Data

<u>Classification Accuracy for Best Model on Test data</u>:- 90.50% (Q1 = 1, Q2=1, Q3=1, Q4=1)

Confusion Matrix for Training Data:-

Training Data(N=1000,Q1 = 1, Q2=1, Q3=1, Q4=1).				
	Class 1	Class 2	Class 3	Class 4
Class 1	223	13	4	10
Class 2	0	250	0	0
Class 3	0	0	250	0
Class 4	0	0	0	250

Fig 4.3.11:-Confusion Matrix for Training Data.

Confusion Matrix for Testing Data:-

Testing Data(N=400, Q1 = 1, Q2=1, Q3=1, Q4=1).				
	Class 1	Class 2	Class 3	Class 4
Class 1	94	1	4	1
Class 2	2	93	1	4
Class 3	10	2	88	0
Class 4	9	4	0	87

Fig 4.3.12:-Confusion Matrix for Test Data.

4. INFERENCES

1) The classification in Linearly Separable Data was a simpler problem and therefore 100% accuracy was achieved on training data (Fig 4.3.1) and test data(Fig 4.3.2) even after taking single gaussian component in each class i.e. Q1 = 1, Q2= 1, Q3 = 1, Q4 = 1.
   On using same covariance for each class the decision surfaces obtained were linear(Fig 4.2.1)
2) In the classification of Nonlinearly Separable Data since the different covariances matrices are used, the decision surface obtained is quadratic in nature and the fact is evident from the figure 4.2.2
   The best classification accuracy of 99.49% on test data was obtained for Q1 = 15, Q2=15.(Fig 4.3.6)
3) In the classification of overlapping data also the decision surface obtained is quadratic in nature (Fig 4.2.3)
   The best classification accuracy of 90.50% on test data was obtained for Q1 = 1, Q2=1, Q3 = 1, Q4 =1(Fig 4.3.10).

TASK 5 :Bayes classifier with GMM for each class.

DATASET 1 : 2-dimensional artificial data of 4 classes:

1. **METHOD :**

- The Expectation-Maximization as elaborated in Method of Task 4 was used to estimate the parameters. In which, the initial parameters were found using K-means algorithm.
  The total log likelihood is also calculated just like mentioned in Task 4.

- Since unlike naive bayes, here dependence between the feature is taken into consideration, the number of parameters for covariance matrices are d(d-1)/2 which increases the computation.

- The full covariance matrix as a result models the class conditional density better than the diagonal matrix obtained with Naive Bayes assumption.

2. **PLOT**

   1) **Linearly separable data:-**

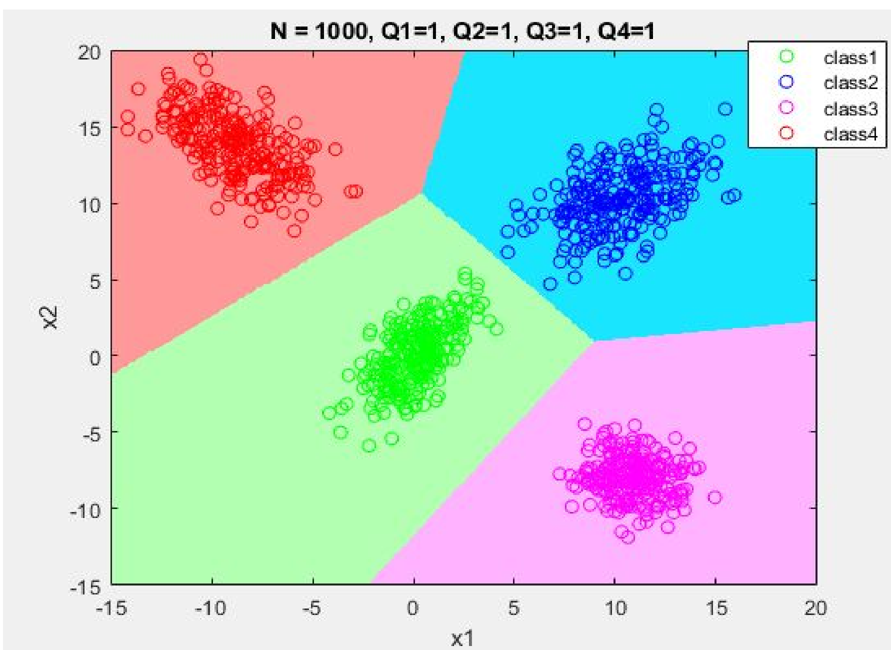

Fig 5.2.1:- Plot of decision region with training data superimposed.

2) **Non-Linearly Separable Data:-**

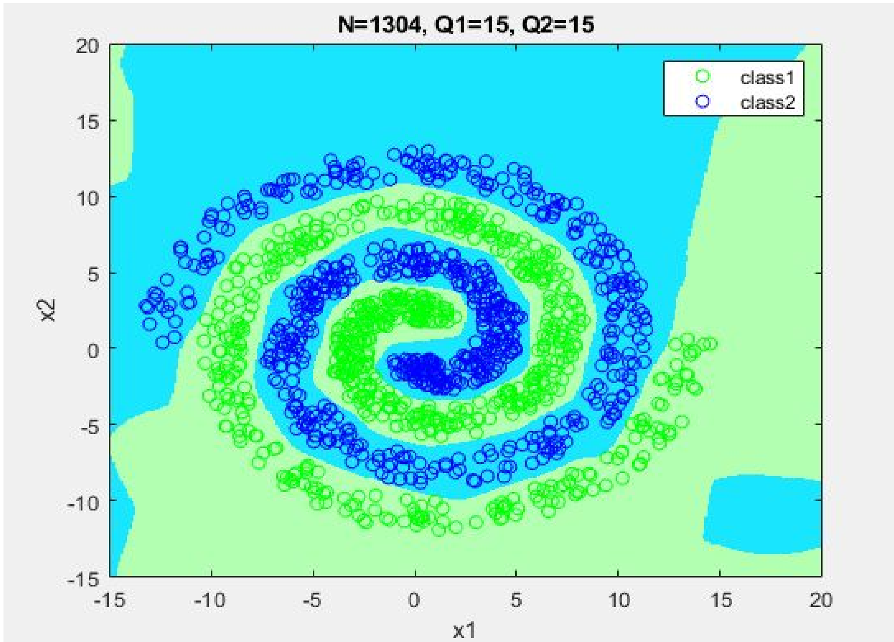

Fig 5.2.2:- Plot of decision region with training data superimposed.

3) **Overlapping Data:-**

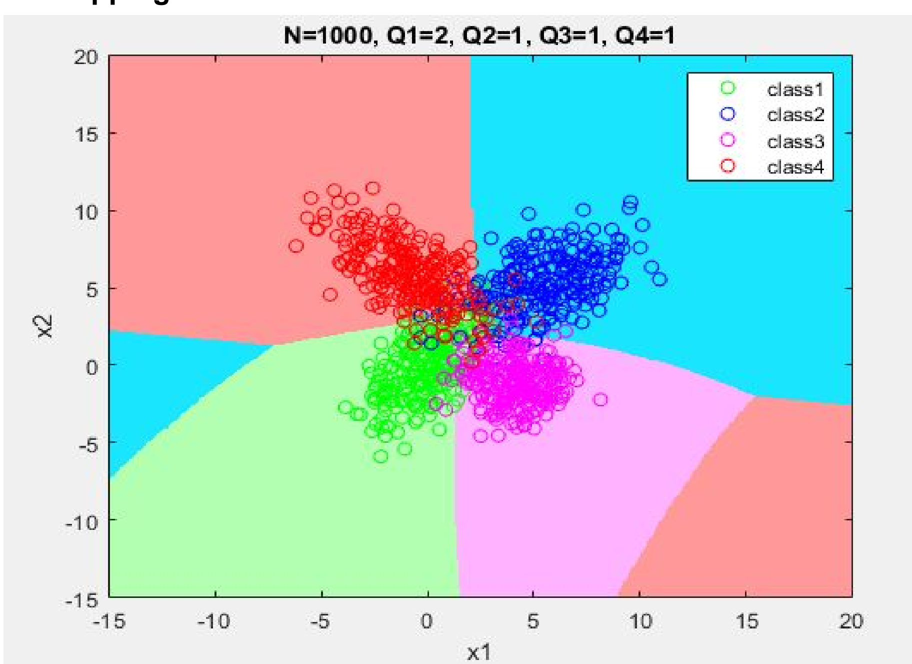

Fig 5.2.3:- Plot of decision region with training data superimposed.

3. RESULTS

1) Linearly Separable Data:-

Classification Accuracy on Training Data and Validation Data:-

Training Data				
Q1	**Q2**	**Q3**	**Q4**	**Classification Accuracy(In percentage)**
1	1	1	1	100
2	1	1	1	100
2	2	3	3	100
3	3	3	3	100

Fig 5.3.1:-Classification Accuracy on Training Data

Validation Data				
Q1	**Q2**	**Q3**	**Q4**	**Classification Accuracy(In percentage)**
1	1	1	1	100
2	1	1	1	100
2	2	3	3	100
3	3	3	3	100

Fig 5.3.2:-Classification Accuracy on Validation Data

<u>Classification Accuracy for Best Model on Test data</u>:- 100% (Q1=1, Q2=1, Q3=1, Q4 = 1)

Confusion Matrix for Training Data:-

Training Data(N=1000,Q1=1, Q2=1, Q3=1, Q4 = 1).				
	Class 1	Class 2	Class 3	Class 4
Class 1	250	0	0	0
Class 2	0	250	0	0
Class 3	0	0	250	0
Class 4	0	0	0	250

Fig 5.3.3:-Confusion Matrix for Training Data.

Confusion Matrix for Test Data:-

Test Data(N=400, Q1=1, Q2=1, Q3=1, Q4 = 1).				
	Class 1	Class 2	Class 3	Class 4
Class 1	100	0	0	0
Class 2	0	100	0	0
Class 3	0	0	100	0
Class 4	0	0	0	100

Fig 5.3.4:-Confusion Matrix for Test Data.

2) Non-Linearly Separable Data:-

Classification Accuracy on Training Data and Validation Data:-

Training Data		
Q1	Q2	Classification Accuracy(In percentage)
1	1	55.14
5	5	85.58
10	10	99.77
15	15	100

Fig 5.3.5:-Classification Accuracy on Training Data.

Validation Data		
Q1	Q2	Classification Accuracy(In percentage)
1	1	0.5754
5	5	0.8325
10	10	0.9962
15	15	0.9987

Fig 5.3.6:-Classification Accuracy on Validation Data.

<u>Classification Accuracy for Best Model on Test data</u>:- 100% (Q1 = 15, Q2 = 15)

<u>Confusion Matrix for Training Data:-</u>

Training Data(N=1304,Q1 = 15, Q2 = 15).		
	Class 1	Class 2
Class 1	652	0
Class 2	0	652

Fig 5.3.7:-Confusion Matrix for Training Data.

<u>Confusion Matrix for Test Data:-</u>

Test Data(N=520,Q1 = 15, Q2 = 15).		
	Class 1	Class 2
Class 1	260	0
Class 2	0	260

Fig 5.3.8:-Confusion Matrix for Test Data.

3) Overlapping Data:-

Training Data				
Q1	Q2	Q3	Q4	Classification Accuracy
1	1	1	1	89.70
2	2	1	1	89.40
3	2	1	1	90.10
2	3	1	1	89.30

Fig 5.3.9:-Classification Accuracy on Training Data

Validation Data				
Q1	Q2	Q3	Q4	Classification Accuracy(In percentage)
1	1	1	1	90.43
2	2	1	1	90.57
3	2	1	1	90.57
2	3	1	1	90.57

Fig 5.3.10:-Classification Accuracy on Validation Data

Classification Accuracy for Best Model on Test data:- 91.75% (Q1 = 2, Q2=2, Q3=1, Q4=1)

Confusion Matrix for Training Data:-

Training Data (N=1000,Q=2, Q2=2, Q3=1, Q4=1).				
	Class 1	Class 2	Class 3	Class 4
Class 1	220	10	8	12
Class 2	7	225	8	10
Class 3	15	4	231	0
Class 4	16	11	2	221

Fig 5.3.11:-Confusion Matrix for Training Data.

Confusion Matrix for Testing Data:-

Testing Data(N=400,Q1=2, Q2=2, Q3=1, Q4=1).				
	Class 1	**Class 2**	**Class 3**	**Class 4**
Class 1	93	1	5	1
Class 2	2	92	2	4
Class 3	7	1	92	0
Class 4	8	2	0	90

Fig 5.3.12:-Confusion Matrix for Test Data.

4) INFERENCES

1) The classification in Linearly Separable Data was a simpler problem and therefore 100% accuracy was achieved on training data (Fig 5.3.1) and test data(Fig 5.3.2) even after taking single gaussian component in each class i.e. Q1 = 1, Q2= 1, Q3 = 1, Q4 = 1.
On using same covariance for each class the decision surfaces obtained were linear (Fig 5.2.1)

2) In the classification of Nonlinearly Separable Data since the different covariances matrices are used, the decision surface obtained is quadratic in nature and the fact is evident from the figure:(Fig 5.2.2)

   The best classification accuracy of 99.49% on test data was obtained for Q1 = 15, Q2=15.

3) In the classification of overlapping data also the decision surface obtained is quadratic in nature figure:(Fig 5.2.3)
The best classification accuracy of 91.75% on test data was obtained for Q1 = 1, Q2=1, Q3 = 1, Q4 =1(Fig 5.3.10)

TASK 6 :Bayes classifier with K-nearest neighbours method for estimation of class-conditional probability density function.

DATASET 1 : 2-dimensional artificial data of 4 classes:

1. **METHOD :**

    1. To estimate probability density function using KNN, we use fixed 'K' which is number of nearest neighbours and determine V_i which is volume of hypersphere of class 'i' after taking 'K' nearest neighbours.
    2. Steps to determine class of input $\overline{x}$:-
        - Identity K nearest neighbours of $\overline{x}$ from class 'i'.
        - For each class find the maximum distance of $\overline{x}$ by compute the distance to each point belonging to class 'i'.
        - Use this maximum distance as the radius of the hypersphere.
        - Probability of a input belonging to a particular class is given by $p(\overline{x}/y_i) * p(y_i)$.
        - Class of input $\overline{x}$ is given by:-
          $$i^* = argmax\frac{K_i}{N_i * V_i}$$
          $$i^* = argmax\frac{1}{r_i}$$

2. **PLOTS**
    1) **Linearly Separable Data:-**

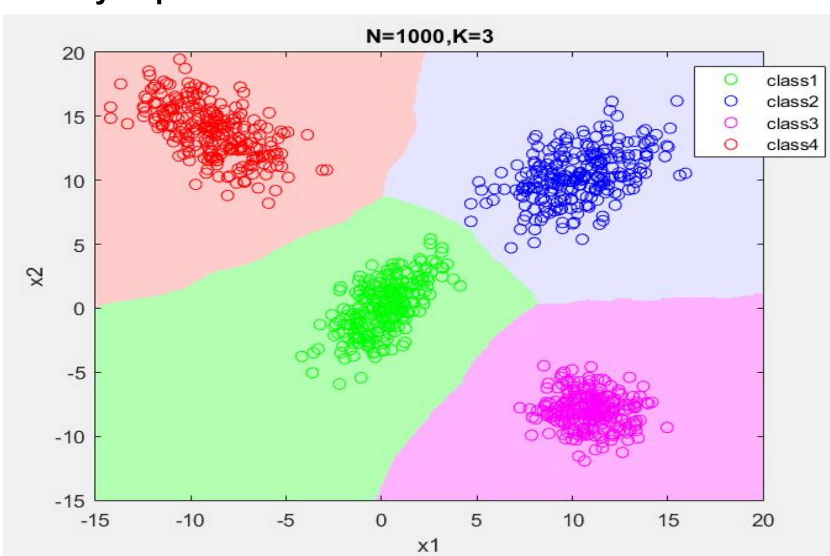

Fig 6.2.1:- Plot of training decision region with training data superimposed.

2) **Non-Linearly Separable Data:-**

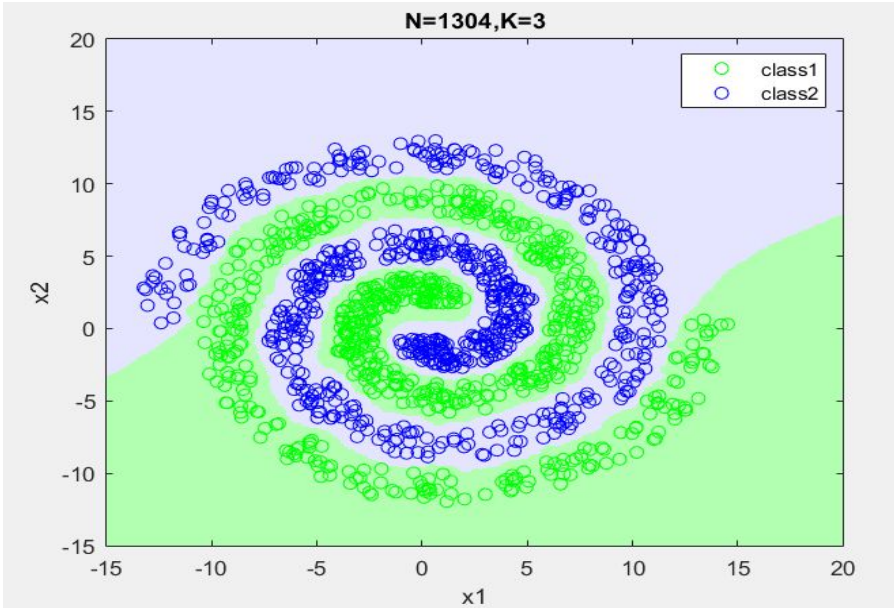

Fig 6.2.2:- Plot of training decision region with training data superimposed.

3) **Overlapping Data:-**

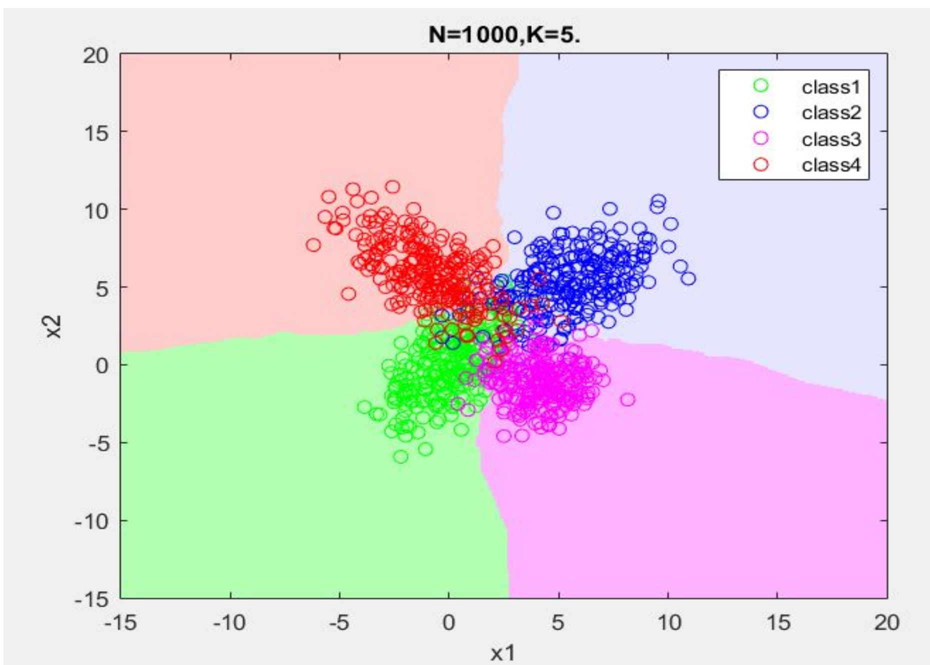

Fig 6.2.3:- Plot of training decision region with training data superimposed.

3. RESULTS

1) Linearly Separable Data:-

Classification Accuracy on Training Data and Validation Data:-

Training Data	
K(Number of nearest neighbours)	**Classification Accuracy(In percentage)**
3	100
4	100
5	100

Fig 6.3.1:-Classification Accuracy on Training Data

Validation Data	
K(Number of nearest neighbours)	**Classification Accuracy(In percentage)**
3	100
4	100
5	100

Fig 6.3.2:-Classification Accuracy on Validation Data

Classification Accuracy for Best Model on Test data:- 100% (K=3)

Confusion Matrix for Training Data:-

Training Data(N=1000,K=3).				
	Class 1	Class 2	Class 3	Class 4
Class 1	250	0	0	0
Class 2	0	250	0	0
Class 3	0	0	250	0
Class 4	0	0	0	250

Fig 6.3.3:-Confusion Matrix for Training Data.

<u>Confusion Matrix for Test Data:-</u>

Test Data(N=400,K=3).				
	Class 1	Class 2	Class 3	Class 4
Class 1	100	0	0	0
Class 2	0	100	0	0
Class 3	0	0	100	0
Class 4	0	0	0	100

Fig 6.3.4:-Confusion Matrix for Test Data.

2) **Non-Linearly Separable Data:-**

Classification Accuracy on Training Data and Validation Data:-

Training Data	
K(Number of nearest neighbours)	Classification Accuracy(In percentage)
3	100
4	100
5	100

Fig 6.3.5:-Classification Accuracy on Training Data.

Validation Data	
K(Number of nearest neighbours)	Classification Accuracy(In percentage)
3	100
4	100
6	100

Fig 6.3.6:-Classification Accuracy on Validation Data.

<u>Classification Accuracy for Best Model on Test data</u>:- 100% (K=3)

Confusion Matrix for Training Data:-

Training Data(N=1304,K=3).		
	Class 1	**Class 2**
Class 1	652	0
Class 2	0	652

Fig 6.3.7:-Confusion Matrix for Training Data.

Confusion Matrix for Test Data:-

Test Data(N=520,K=4).		
	Class 1	**Class 2**
Class 1	260	0
Class 2	0	260

Fig 6.3.8:-Confusion Matrix for Test Data.

3) **Overlapping Data:-**

Classification Accuracy on Training Data and Validation Data:-

Training Data	
K(Number of nearest neighbours)	**Classification Accuracy(In percentage)**
3	91.4
4	90.70
5	91
6	91

Fig 6.3.9:-Classification Accuracy on Training Data.

Validation Data	
K(Number of nearest neighbours)	Classification Accuracy(In percentage)
3	89.67
4	90.33
5	90.67
6	90.73

Fig 6.3.10:-Classification Accuracy on Validation Data.

<u>Classification Accuracy for Best Model on Test data</u>:- 91% (K=6)

<u>Confusion Matrix for Training Data:-</u>

Training Data(N=1000,K=6).				
	Class 1	Class 2	Class 3	Class 4
Class 1	223	8	10	9
Class 2	8	229	8	5
Class 3	15	4	231	0
Class 4	18	9	1	222

Fig 6.3.11:-Confusion Matrix for Training Data.

<u>Confusion Matrix for Testing Data:-</u>

Testing Data(N=400,K=6).				
	Class 1	Class 2	Class 3	Class 4
Class 1	93	2	4	1
Class 2	4	91	1	4
Class 3	6	1	93	0
Class 4	8	4	1	87

Fig 6.3.12:-Confusion Matrix for Test Data.

4. INFERENCES

1) In case of linearly separable data, we observe that the accuracy remains the same for k=2,3,4,5. This is because the data points belonging to different classes are far apart. So increasing the number of nearest neighbours does not cause any significant change.

2) In case of non linearly separable data, we observe that decision surface is non linear.

3) In case of overlapping data, from Fig. 6.3.10 we can observe that the accuracy increases as we keep on increasing k from 3 to 6. At k=6, we achieve maximum accuracy. We require more number of nearest neighbours because we would more number of points to make a good decision about density estimation. With less k, classifier has a greater chance of making a wrong decision for the overlapping points.

TASK 7 :Naive Bayes classifier with GMM for each class

DATASET 2 : A)Data set for static pattern classification
B)Image data set for varying length pattern (Set of local feature vectors representation) classification

1. **METHOD for Naive Bayes and Bayes classifier with GMM for each class :**

- The Expectation-Maximization algorithm, as elaborated in Method of Task 4 was used to estimate the parameters. In which, the initial parameters were found using K-means algorithm.
  The total log likelihood is also calculated just like mentioned in Task 4.
  The covariances are as explained in Task 5.

- All local features are assumed to be independent and hence Class of input $\overline{x}$ is given by:-

$$p(\overline{x}/y_i) = \prod_{t=1}^{T} p(\overline{x_t}/y_i)$$

  Where 'T' is total number of segments present in input $\overline{x}$

2. RESULTS

A) Data set for static pattern classification

Classification Accuracy on Training Data and Validation Data:-

Training Data			
Q1 (No of components for class 1)	Q2 (No of components for class 2)	Q3 (No of components for class 3)	Classification Accuracy(In percentage)
2	1	2	93.33
2	1	1	92.38
1	1	1	92.38

Fig A)7.2.1:-Classification Accuracy on Training Data

Validation Data			
Q1	Q2	Q3	Classification Accuracy(In percentage)
2	1	2	100
2	1	1	100
1	1	1	100

Fig A)7.2.3:-Classification Accuracy on Validation Data

Classification Accuracy for Best Model on Test data:- 100% (Q1=2, Q2=1, Q3=1

Confusion Matrix for Training Data:-

Training Data(N=105,Q1=2, Q2=1, Q3=2).			
	Class 1	Class 2	Class 3
Class 1	35	0	0
Class 2	1	33	3
Class 3	0	3	32

Fig A)7.2.3:-Confusion Matrix for Training Data.

Confusion Matrix for Test Data:-

Test Data(N=30,Q1=2,Q2=1,Q3=2).			
	Class 1	Class 2	Class 3
Class 1	10	0	0
Class 2	0	10	0
Class 3	0	0	10

Fig A)7.2.4:-Confusion Matrix for Test Data.

B) Image data set for varying length pattern (Set of local feature vectors representation) classification.

Classification Accuracy on Training Data and Validation Data:-

Training Data					
Q1	Q2	Q3	Q4	Q5	Classification Accuracy
2	2	1	1	1	60.23
2	2	1	2	1	63.89

Fig B)7.2.1:-Classification Accuracy on Training Data

Validation Data					
Q1	Q2	Q3	Q4	Q5	Classification Accuracy
2	2	1	1	1	58.69
2	2	1	2	1	61.31

Fig B)7.2.2:-Classification Accuracy on Validation Data

Classification Accuracy for Best Model on Test data:- 62.66% (Q1=2,Q2=2,Q3=1,Q4=2,Q5=1).

Confusion Matrix for Training Data:-

Training Data(N1=49,N2=61,N3=46,N4=56,N5=101) (Q1=2, Q2=2, Q3=1,Q4=2,Q5=1).					
	Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Class 1	40	0	0	5	0
Class 2	0	23	29	9	0
Class 3	0	7	21	17	1
Class 4	0	29	2	24	1
Class 5	7	1	1	0	92

Fig B)7.2.3:-Confusion Matrix for Training Data.

Confusion Matrix for Test Data:-

Test Data(N1=51,N2=55,N3=46,N4=39,N5=42), (Q1=2, Q2=2, Q3=1, Q4=2, Q5=1)					
	Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Class 1	45	2	0	4	0
Class 2	16	20	11	5	3
Class 3	7	2	21	11	5
Class 4	0	19	0	20	0
Class 5	2	0	0	0	40

Fig B)7.2.3:-Confusion Matrix for Training Data.

3. INFERENCES

- The mixture model was able to classify images with 62.66% when the gaussian components in each class were Q1 = 2, Q2 = 2,Q3 =1, Q4 =2 Q5 =1 respectively. Though images were divided in feature vectors of variable length model was able to classify by considering the probability of each feature vector and taking their product and keeping naive bayes assumption of feature independence.

TASK 8 : Bayes classifier with GMM for each class

DATASET 2 : A)Data set for static pattern classification
B)Image data set for varying length pattern (Set of local feature vectors representation) classification

1. METHOD:-

- The Expectation-Maximization algorithm, as elaborated in Method of Task 4 was used to estimate the parameters. In which, the initial parameters were found using K-means algorithm.
  The total log likelihood is also calculated just like mentioned in Task 4.
  The covariances are as explained in Task 5.

- All local features are assumed to be independent and hence Class of input $\overline{x}$ is given by:-

$$p(\overline{x}/y_i) = \prod_{t=1}^{T} p(\overline{x_t}/y_i)$$

  Where 'T' is total number of segments present in input $\overline{x}$

2. RESULTS

A) Data set for static pattern classification

Classification Accuracy on Training Data and Validation Data:-

Training Data			
Q1	**Q2**	**Q3**	**Classification Accuracy**
2	1	2	96.19
2	1	1	95.24
1	1	1	95.24

Fig A)8.2.1:-Classification Accuracy on Training Data

Validation Data			
Q1	Q2	Q3	Classification Accuracy
2	1	2	100
2	1	1	100
1	1	1	100

Fig A)8.2.3:-Classification Accuracy on Training Data

<u>Classification Accuracy for Best Model on Test data</u>:- 100% (Q1=1, Q2=1, Q3=1)

<u>Confusion Matrix for Training Data:-</u>

Training Data(N=105,Q1=2, Q2=1, Q3=2).			
	Class 1	Class 2	Class 3
Class 1	35	0	0
Class 2	1	33	1
Class 3	0	2	31

Fig A)8.2.3:-Confusion Matrix for Training Data.

<u>Confusion Matrix for Test Data:-</u>

Test Data(N=30,Q1=1,Q2=1,Q3=1).			
	Class 1	Class 2	Class 3
Class 1	10	0	0
Class 2	0	10	0
Class 3	0	0	10

Fig A)8.2.4:-Confusion Matrix for Test Data.

B) Image data set for varying length pattern (Set of local feature vectors representation) classification.

Classification Accuracy on Training Data and Validation Data:-

Training Data					
Q1	Q2	Q3	Q4	Q5	Classification Accuracy
2	2	1	1	1	63.26
2	2	1	2	1	64.53

Fig B)8.2.1:-Classification Accuracy on Training Data

Validation Data					
Q1	Q2	Q3	Q4	Q5	Classification Accuracy
2	2	1	1	1	59.10
2	2	1	2	1	61.77

Fig B)8.2.2:-Classification Accuracy on Validation Data

Classification Accuracy for Best Model on Test data:- 63.52% (Q1=1,Q2=2,Q3=1,Q4=2,Q5=1).

Confusion Matrix for Training Data:-

Training Data(N1=49,N2=61,N3=46,N4=56,N5=101) (Q1=2, Q2=2, Q3=1,Q4=2,Q5=1).					
	Class 1	Class 2	Class 3	Class 4	Class 5
Class 1	45	1	0	3	0
Class 2	0	17	12	31	1
Class 3	0	0	21	5	20
Class 4	17	4	9	26	0
Class 5	7	1	0	0	93

Fig B)8.2.3:-Confusion Matrix for Training Data.

Confusion Matrix for Test Data:-

Test Data(N1=51,N2=55,N3=46,N4=39,N5=42), (Q1=2, Q2=2, Q3=1, Q4=2, Q5=1).					
	Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Class 1	48	2	0	1	0
Class 2	4	22	10	8	11
Class 3	2	17	27	0	0
Class 4	0	1	18	16	4
Class 5	1	0	0	6	35

Fig B)8.2.4:-Confusion Matrix for Test Data.

3. INFERENCES

- The mixture model was able to classify images with 63.52% when the gaussian components in each class were Q1 = 2, Q2 = 2,Q3 =1, Q4 =2 Q5 =1 respectively. Though images were divided in feature vectors of variable length model was able to classify by considering the probability of each feature vector.
- The accuracy here is better than the one with naive bayes assumption(62.66%) because features aren’t assumed independent anymore and full covariance matrix is considered.

<u>TASK 9 : Bayes classifier with K-nearest neighbours method for estimation of class-conditional probability density function</u>

DATASET 2 : A)Data set for static pattern classification
B)Image data set for varying length pattern (Set of local feature vectors representation) classification

1. **METHOD :**

    1. To estimate probability density function using KNN, we use fixed ‘K’ which is number of nearest neighbours and determine V_i which is volume of hypersphere of class ‘i’ after taking ‘K’ nearest neighbours.
    2. Steps to determine class of input $\overline{x}$:-
        - Identity K nearest neighbours of segment $\overline{x_t}$ of $\overline{x}$ from class ‘i’ where ‘t’ is t^{th} segment of input $\overline{x}$.
        - For each class find the maximum distance of segment of $\overline{x}$ by computing the distance to each point belonging to class ‘i’.
        - Use this maximum distance as the radius of the hypersphere.
        - Probability of a input segment $\overline{x_t}$ of input $\overline{x}$ belonging to a particular class is given by:-
          $p(\overline{x_t}/y_i) * p(y_i)$
        - Class of input $\overline{x}$ is given by:-
          $$p(\overline{x}/y_i) = \prod_{t=1}^{T} p(\overline{x_t}/y_i)$$
          Where ‘T’ is total number of segments present in input $\overline{x}$

2. RESULTS

A) Data set for static pattern classification

Classification Accuracy on Training Data and Validation Data:-

Training Data	
K(No of Nearest neighbours)	**Classification Accuracy(In percentage)**
1	100
2	90
3	90

Fig A)9.2.1:-Classification Accuracy on Training Data

Validation Data	
K(No of Nearest neighbours)	**Classification Accuracy(In percentage)**
1	80
2	100
3	100

Fig A)9.2.2:-Classification Accuracy on Validation Data

Classification Accuracy for Best Model on Test data:- 100%(K=2)

Confusion Matrix for Training Data:-

Training Data(N=105,K=1).			
	Class 1	**Class 2**	**Class 3**
Class 1	35	0	0
Class 2	0	35	0
Class 3	0	0	35

Fig A)9.2.3:-Confusion Matrix for Training Data.

<u>Confusion Matrix for Test Data:-</u>

Test Data(N=30,K=2).			
	Class 1	Class 2	Class 3
Class 1	10	0	0
Class 2	0	10	0
Class 3	0	0	10

Fig A)9.2.4:-Confusion Matrix for Test Data.

B) Image data set for varying length pattern (Set of local feature vectors representation) classification.

Training Data	
K(No of Nearest neighbours)	Classification Accuracy(In percentage)
5	59.79
7	70.73
8	68.12

Fig B)9.2.1:-Classification Accuracy on Training Data

Validation Data	
K(No of Nearest neighbours)	Classification Accuracy(In percentage)
5	63.92
7	67.23
8	65.47

Fig B)9.2.2:-Classification Accuracy on Validation Data

<u>Classification Accuracy for Best Model on Test data</u>:- 66.78%

<u>Confusion Matrix for Training Data:-</u>

Training Data(N1=140,N2=100,N3=130,N4=150,N5=160,K=7).					
	Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Class 1	139	1	0	0	0
Class 2	48	50	0	2	0
Class 3	30	0	72	20	8
Class 4	26	0	30	88	6
Class 5	1	26	1	0	132

Fig B)9.2.3:-Confusion Matrix for Training Data.

<u>Confusion Matrix for Test Data:-</u>

Test Data(N1=42,N2=56,N3=40,N4=47,N5=86,N=271,K=7).					
	Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Class 1	39	0	1	2	0
Class 2	16	15	4	21	0
Class 3	21	1	18	0	0
Class 4	12	5	1	29	0
Class 5	1	0	0	5	80

Fig B)9.2.3:-Confusion Matrix for Testing Data.

3. INFERENCES

For real world data varying length data, we get an accuracy of 66.78% using KNN method for probability density estimation which is better than naive bayes GMM. This is because KNN is non-parametric, i.e. it makes no assumption about the data distribution whereas naive bayes classifier assumes that attributes are conditionally independent to each other given the class, and are normally distributed (for real-valued attributes). As a result, Naive Bayes can only have linear, elliptic, or parabolic decision boundaries, which makes the flexibility of KNN's decision boundary a huge advantage.