CS5691 Pattern Recognition and Machine Learning

Jan – May 2021

Assignment - 2

Group 18:

BS17B033 – Shreya Nema

ME17B065 – Sahil Ansari

ME17B158 – Omkar Nath

Contents

1 Dataset 1A – Linearly Separable 2D Data

1.1 Problem Statement

We are given a 2-dimensional artificial data, which is linearly separable for static pattern classification. For this, the dataset specifically allotted to Group 18 is used.

Based upon this, the following classifiers need to be built.

1. K-nearest neighbours' classifier, for K=1, K=7 and K=15
2. Naive-Bayes classifier with a Gaussian distribution for each class
    a. Covariance matrix for all the classes is the same and is $\sigma^2 I$
    b. Covariance matrix for all the classes is the same and is C
    c. Covariance matrix for each class is different

1.2 Classification Accuracies Obtained

Upon implementing the various classifiers, the results obtained are as follows:

1.2.1 K-nearest neighbours Classifier

Table 1: KNN Classifier Accuracies

K	Training Accuracy	Validation Accuracy
1	100%	100%
7	100%	100%
15	100%	100%

1.2.2 Naïve Bayes Classifier

Table 2: Naive Bayes Classifier Accuracies

Case	Training Accuracy	Validation Accuracy
A	100%	100%
B	100%	100%
C	100%	100%

1.3 Best Model, and Test accuracy

The best model amongst the models available, as well as its performance on the test dataset is as given below.

1.3.1 K-nearest neighbours Classifier

The best model for this case is **K = 7**.

This gives a test accuracy of **100%.**

1.3.2 Naïve Bayes Classifier

The best model for Naïve Bayes is **Case B** i.e., the Covariance matrix for all the classes is the same and is C.

This gives a test accuracy of **100%.**

1.4 Confusion Matrix for the best model

The confusion matrix that is obtained for the best model, as determined by the performance on the validation dataset, is given below for each of the individual cases.

1.4.1 K-nearest neighbors Classifier

The confusion matrices for the best model, i.e., K=7 is given below.

Table 3: KNN Classifier Confusion Matrix for Training Data

Training Data				
Class	0	1	2	3
0	200	0	0	0
1	0	200	0	0
2	0	0	199	0
3	0	0	0	200

Table 4: KNN Classifier Confusion Matrix for Validation Data

Validation Data				
Class	0	1	2	3
0	29	0	0	0
1	0	30	0	0
2	0	0	30	0
3	0	0	0	30

1.4.2 Naïve Bayes Classifier

The confusion matrices for the best model, i.e., case B (Covariance matrix for all the classes is the same and is C) are given below.

Table 5: Naive Bayes Classifier Confusion Matrix for Training Data

Training Data				
Class	0	1	2	3
0	200	0	0	0
1	0	200	0	0
2	0	0	199	0
3	0	0	0	200

Table 6: Naive Bayes Classifier Confusion Matrix for Validation Data

Validation Data				
Class	0	1	2	3
0	29	0	0	0
1	0	30	0	0
2	0	0	30	0
3	0	0	0	30

1.5 Decision Region Plots

Various plots to visualize the data and the classifiers are displayed here.

1.5.1 Training Data

The plot of the training data is as shown below.

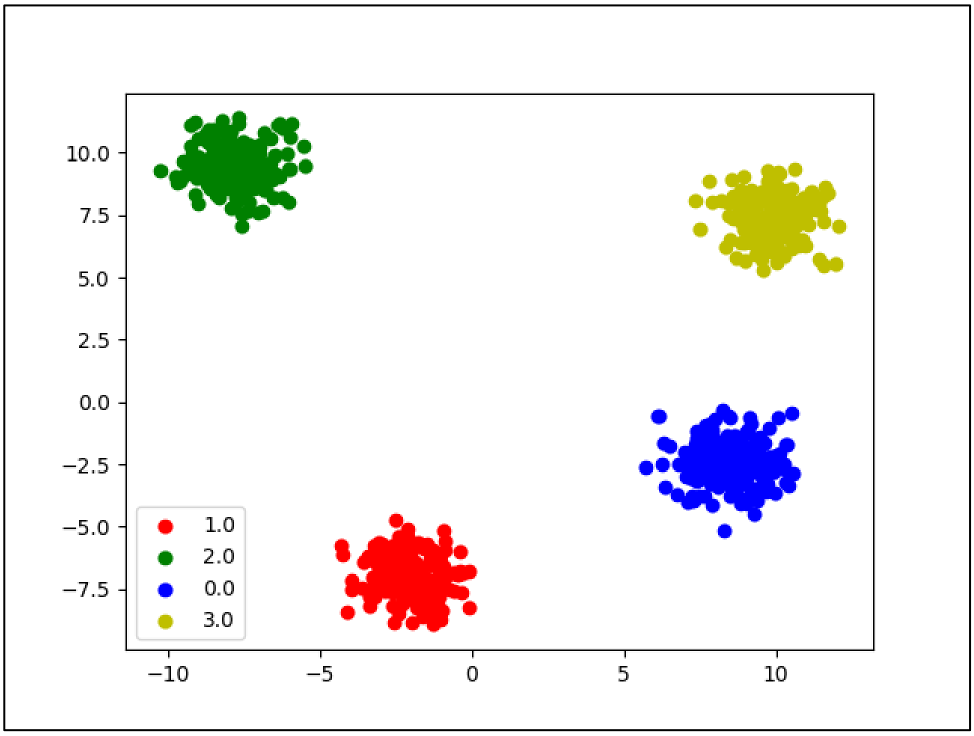

Figure 1: Training Data for Dataset 1A

1.5.2 K-nearest neighbours Classifier

The plot for the classification boundaries, for the KNN Classifier is as shown in the plot below.

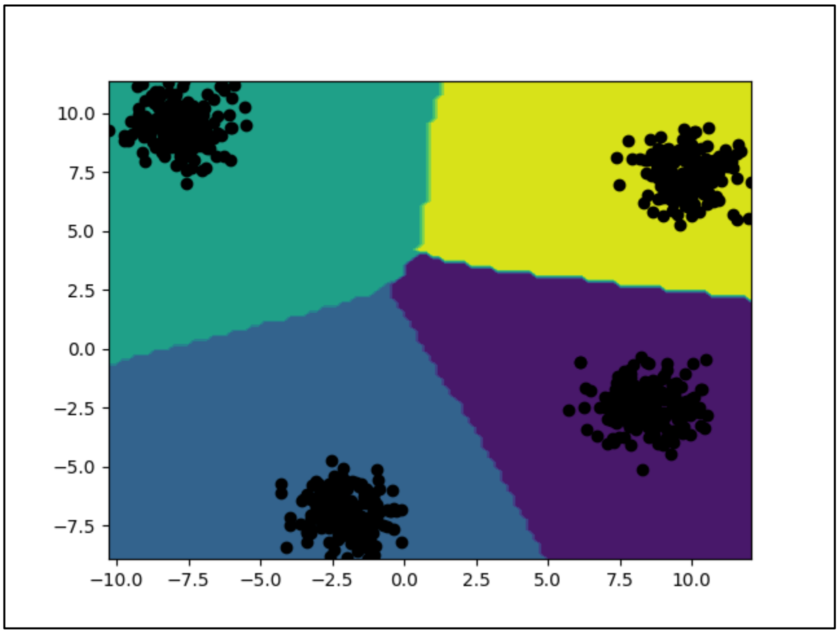

Figure 2: Classification Boundaries for KNN Classifier

1.5.3 Naïve Bayes Classifier

The plot for the classification boundaries, for the Naïve Bayes Classifier is as shown in the plot below.

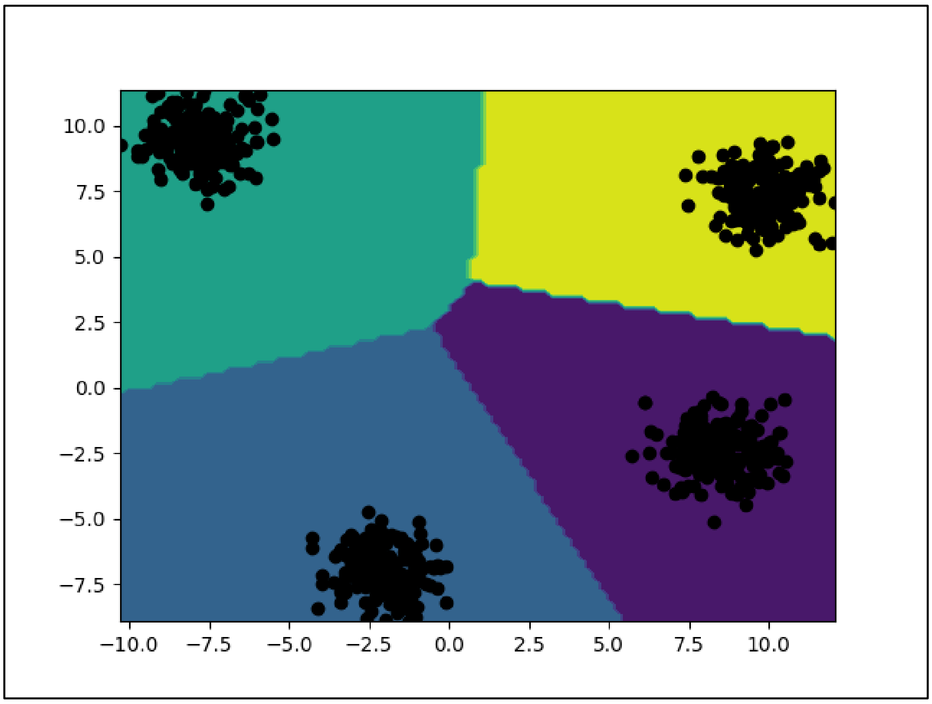

Figure 3: Classification Boundaries for Naive Bayes Classifier

Below is the plot showing the level curves for the Naïve Bayes Classifier

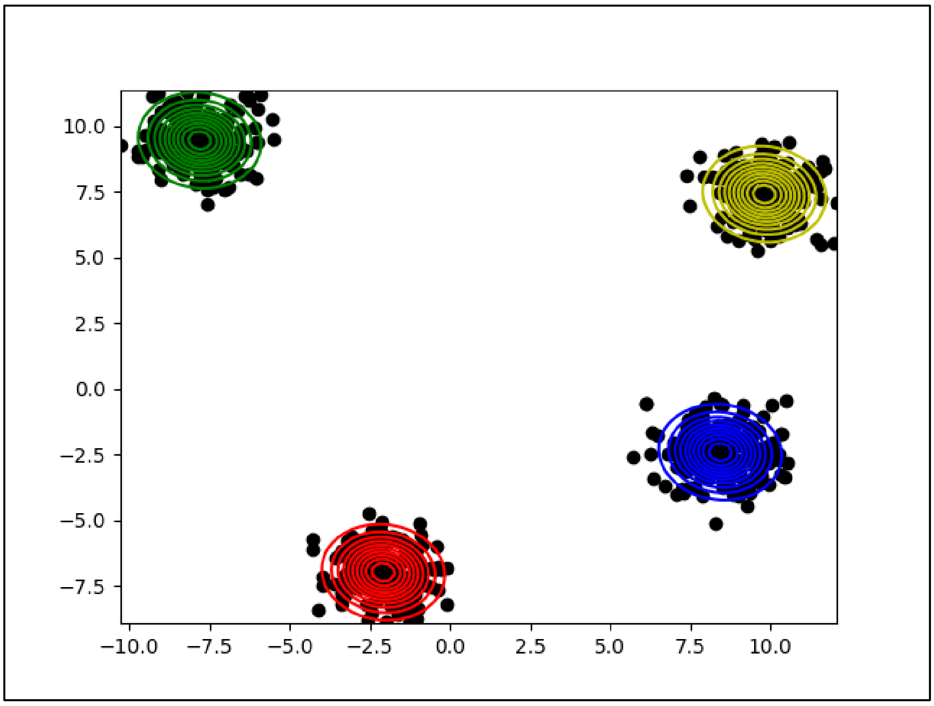

Figure 4: Contour Plots for Naive Bayes Classifier

1.6 Observations

Some key observations after looking at the results are as follows:

- The data is already well separated. The data from the four classes occupy the four corners as it can be seen in the graph.
- Hence, when we use a KNN classifier, we are able to achieve an accuracy of 100%.
- The same explanation holds for the naive bayes classifier as well. The gaussian classifier is able to perfectly model the data distribution as can be seen in the contour plots resulting in an accuracy of 100%

2 Dataset 1B – Nonlinearly Separable 2D Data

2.1 Problem Statement

We are given a 2-dimensional artificial data, which is nonlinearly separable for static pattern classification. For this, the dataset specifically allotted to Group 18 is used.

Based upon this, the following classifiers need to be built.

1. K-nearest neighbours' classifier, for K=1, K=7 and K=15
2. Bayes classifier with a GMM for each class, using full covariance matrices
3. Bayes classifier with a GMM for each class, using diagonal covariance matrices
4. Bayes classifier with K-nearest neighbours' method for estimation of class-conditional probability density function, for K=10 and K=20

2.2 Classification Accuracies Obtained

Upon implementing the various classifiers, the accuracies obtained are as follows.

2.2.1 K-nearest neighbours Classifier

Training and validation accuracies for KNN Classifier are as below.

Table 7: KNN Classifier Classification Accuracies

K	Training Accuracy	Validation Accuracy
1	98.99 %	100 %
7	98.99 %	98.88 %
15	98.99 %	97.75 %

2.2.2 GMM with Full Covariance Matrix

Training and validation accuracies for GMM using Full Covariance Matrices are as below.

Table 8: GMM with Full Covariance Matrix Classification Accuracies

Q	Training Accuracy	Validation Accuracy
4	99.33 %	97.75 %
5	98.99 %	98.88 %
6	99.33 %	99.22 %
7	98.83 %	98.55 %

2.2.3 GMM with Diagonal Covariance matrix

Training and validation accuracies for GMM using Diagonal Covariance Matrices are as below.

Table 9: GMM with Diagonal Covariance matrix Classification Accuracies

Q	Training Accuracy	Validation Accuracy
4	99.16 %	97.75 %
5	98.33 %	98.87 %
6	98.33 %	98.88 %
7	98.99 %	98.88 %

2.2.4 Bayes Classifier with KNN

Training and validation accuracies for Bayes Classifier with K-Nearest Neighbours are as below.

Table 10: Bayes Classifier with KNN Classification Accuracies

K	Training Accuracy	Validation Accuracy
10	98.83 %	97.74%
20	97.16 %	92.11 %

2.3 Best Model, and Test accuracy

The best model amongst the models available, as well as its performance on the test dataset is as given below.

2.3.1 K-nearest neighbours Classifier

The best model for KNN Classifier is **K = 7**.

This gives a test accuracy of **98.88%.**

2.3.2 GMM with Full Covariance Matrix

The best model for GMM using Full Covariance Matrices is **Q = 6**.

This gives a test accuracy of **99.22%.**

2.3.3 GMM with Diagonal Covariance matrix

The best model for GMM using Diagonal Covariance Matrices is **Q = 7**.

This gives a test accuracy of **98.88%**.

2.3.4 Bayes Classifier with KNN

The best model for Bayes Classifier with K-Nearest Neighbours is **K = 10**.

This gives a test accuracy of **97.74%.**

2.4 Confusion Matrix for the best model

The confusion matrix that is obtained for the best model, as determined by the performance on the validation dataset, is given below for each of the individual cases.

2.4.1 K-nearest neighbours Classifier

The confusion matrices for the best model, i.e., K=7 is given below.

Table 11: KNN Classifier Confusion Matrix for Training Data

Training Data			
Class	0	1	2
0	196	0	3
1	1	198	1
2	1	0	199

Table 12: KNN Classifier Confusion Matrix for Validation Data

Validation Data			
Class	0	1	2
0	29	0	0
1	1	29	0
2	0	0	30

2.4.2 GMM with Full Covariance Matrix

The confusion matrices for the best model, i.e., Q=6 is given below.

Table 13: GMM with Full Covariance Matrix Classifier Confusion Matrix for Training Data

Training Data			
Class	0	1	2
0	199	0	0
1	2	198	0
2	2	0	198

Table 14: GMM with Full Covariance Matrix Classifier Confusion Matrix for Validation Data

Validation Data			
Class	0	1	2
0	28	1	0
1	0	30	0
2	0	0	30

2.4.3 GMM with Diagonal Covariance matrix

The confusion matrices for the best model, i.e., Q=7 is given below.

Table 15: GMM with Diagonal Covariance Matrix Classifier Confusion Matrix for Training Data

Training Data			
Class	0	1	2
0	199	0	0
1	2	197	1
2	3	0	197

Table 16: GMM with Diagonal Covariance Matrix Classifier Confusion Matrix for Validation Data

Validation Data			
Class	0	1	2
0	29	0	0
1	0	29	1
2	0	0	30

2.4.4 Bayes Classifier with KNN

The confusion matrices for the best model, i.e., K=10 is given below.

Table 17: Bayes Classifier with KNN Confusion Matrix for Training Data

Training Data			
Class	0	1	2
0	196	0	3
1	0	197	3
2	0	1	199

Table 18: Bayes Classifier with KNN Confusion Matrix for Validation Data

Validation Data			
Class	0	1	2
0	28	0	1
1	1	29	0
2	0	0	30

2.5 Decision Region Plots

Various plots to visualize the data and the classifiers are displayed here.

2.5.1 The Dataset

The plot of the training data is as shown below.

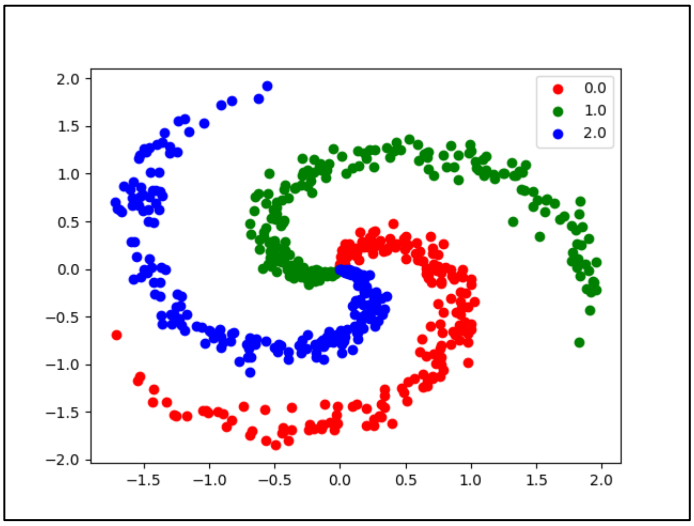

Figure 5: Plot of Dataset 1B

2.5.2 K-nearest neighbours Classifier

The plot for the classification boundaries, for the KNN Classifier is as shown in the plot below.

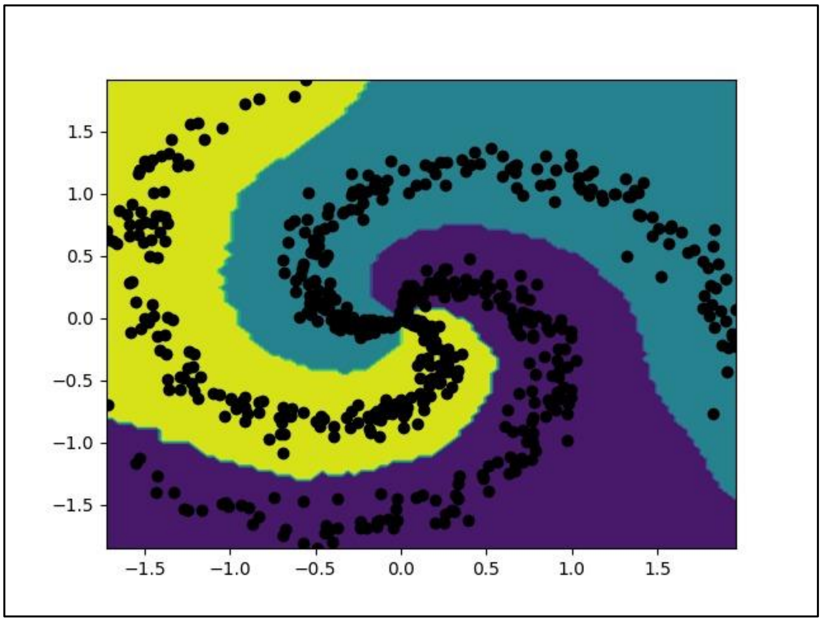

Figure 6: Boundaries for KNN Classifier

2.5.3 GMM with Full Covariance Matrix

The plot for the classification boundaries, for the GMM with Full Covariance Matrix Classifier is as shown in the plot below.

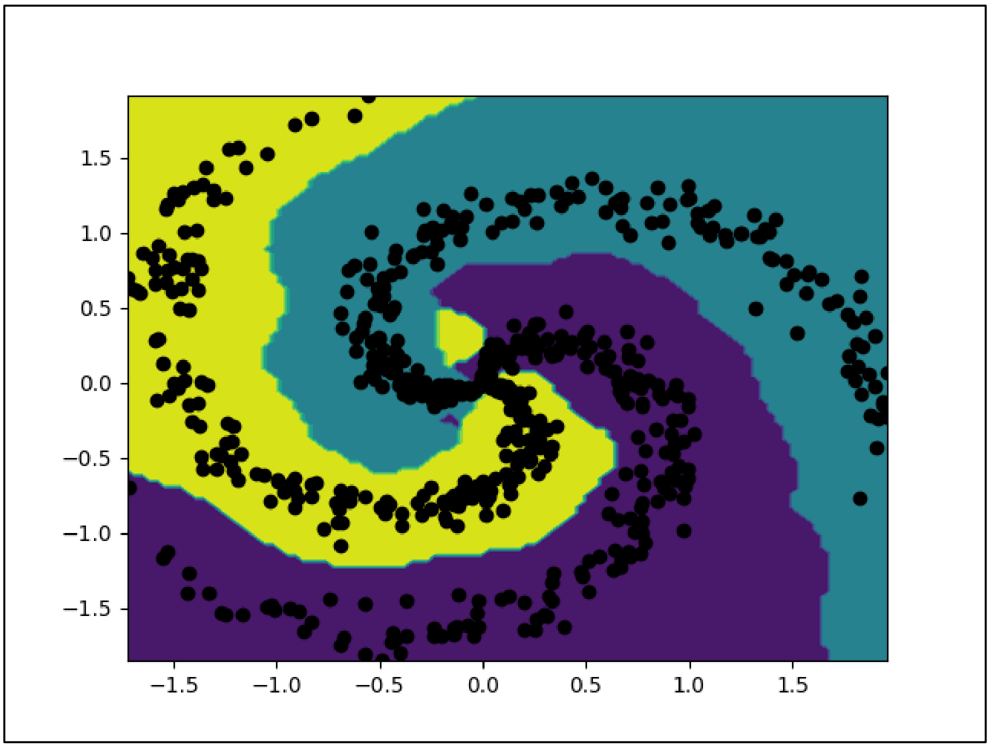

Figure 7: Boundaries for GMM with Full Covariance Matrix Classifier

Below is the plot showing the level curves for the GMM with Full Covariance Matrix Classifier

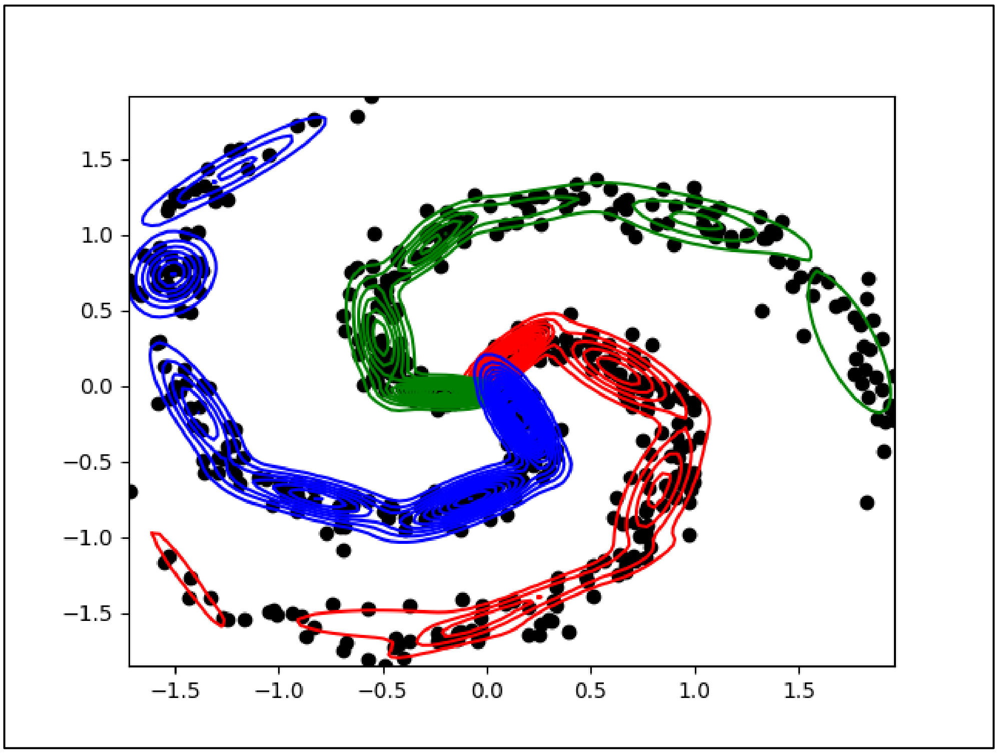

Figure 8: Contour Plots for GMM with Full Covariance Matrix Classifier

2.5.4 GMM with Diagonal Covariance matrix

The plot for the classification boundaries, for the GMM with Diagonal Covariance Matrix Classifier is as shown in the plot below.

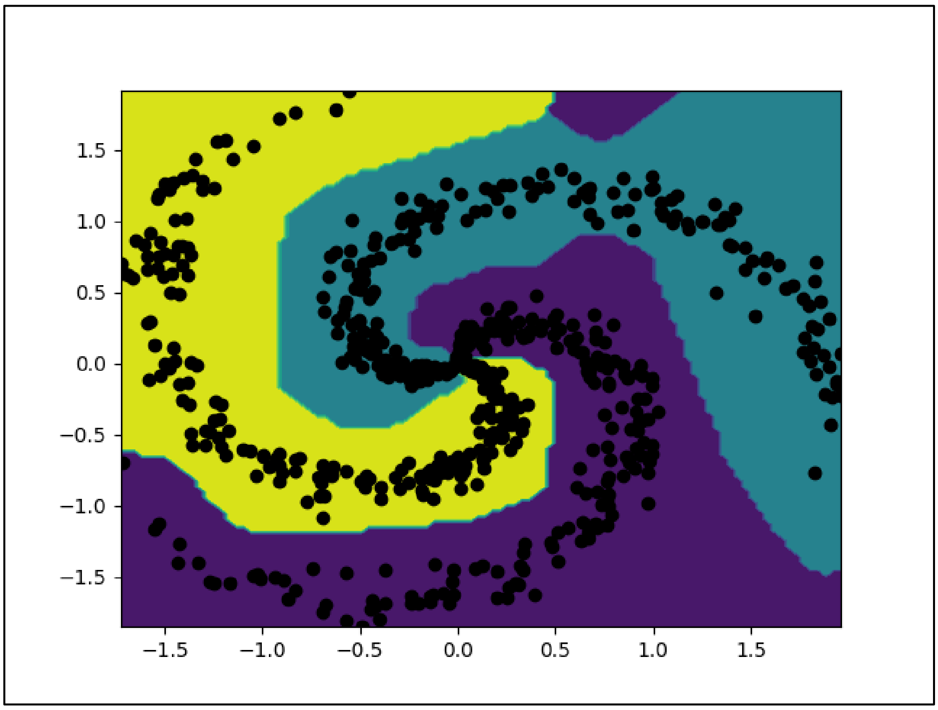

Figure 9: Boundaries for GMM with Diagonal Covariance Matrix Classifier

Below is the plot showing the level curves for the GMM with Diagonal Covariance Matrix Classifier

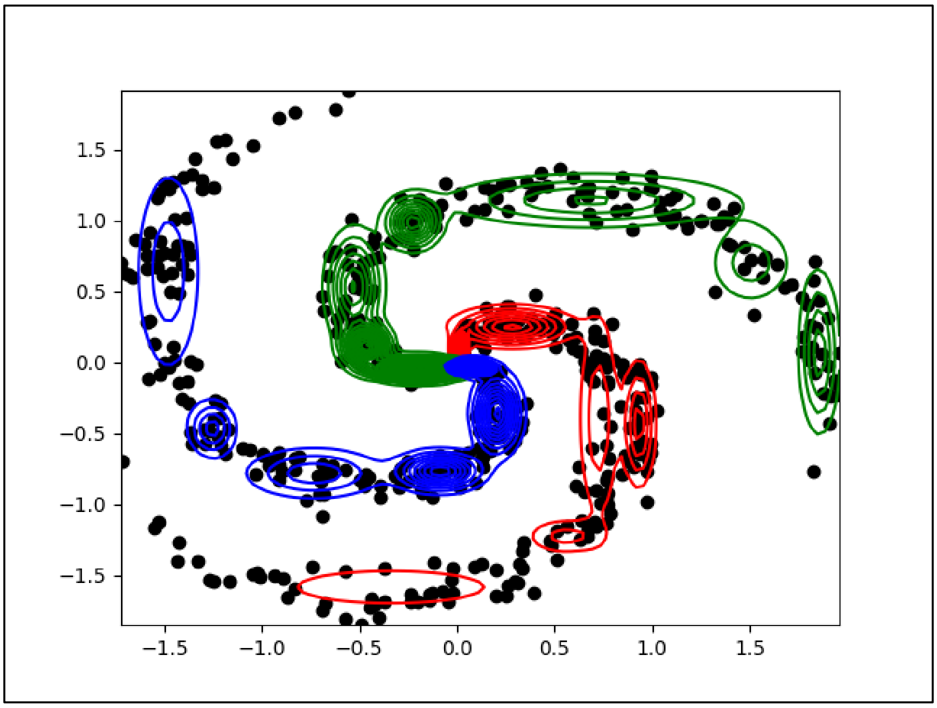

Figure 10: Contour Plot for GMM with Diagonal Covariance Matrix Classifier

2.5.5 Bayes Classifier with KNN

The plot for the classification boundaries, for the Bayes Classifier with KNN is as shown in the plot below.

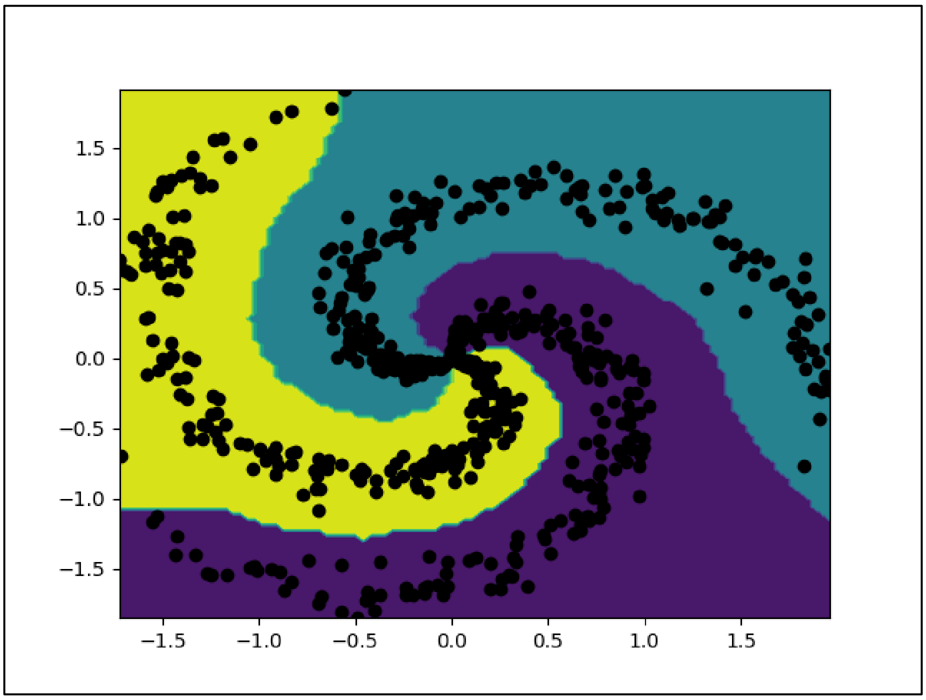

Figure 11: Boundaries for Bayes Classifier with KNN

2.6 Observations

Some key observations after looking at the results are as follows:

- In this case, the data is not that well separated as compared to 1a. The data is distributed in a spiral fashion. But data points belonging to the same class are close to each other.
- Hence, a KNN classifier is again able to perform well, as it can be seen in the contour plot, giving a best-case accuracy of 98.88 %.
- A GMM classifier is also able to give a performance at par with the KNN classifier giving a best-case accuracy of 99.22 % when using 6 modes for full covariance matrix and a best-case accuracy of 98.88 % for diagonal covariance matrix.
- Similarly, a Bayes classifier with K=10 is able to give a good performance with an accuracy of 97.74 %.
- We are getting such good accuracies due to the distribution of the data points.

3 Dataset 2A – Image Data Set for Static Classification

3.1 Problem Statement

We are given a Real-world data set i.e., an Image data set for static pattern classification.

For this data, as per the mapping given to us (Group 18), we work on the classes: **coast, forest, highway, mountain, and tall building**.

Based upon this, the following classifiers need to be built.

1. Bayes classifier with a GMM for each class, using full covariance matrices
2. Bayes classifier with a GMM for each class, using diagonal covariance matrices

3.2 Classification Accuracies Obtained

Upon implementing the various classifiers, the accuracies obtained are as follows.

3.2.1 GMM using Full Covariance Matrix

Training and validation accuracies for Bayes Classified with a GMM, using Full Covariance Matrices are as below.

Table 19: GMM using Full Covariance Matrix Classification Accuracies

Q value	Multiplier	Training Accuracy	Validation Accuracy
6	0.001	25.76	39.17
6	0.01	31.65	31.45
6	0.1	42.92	35.31
12	0.001	24.91	40.36
12	0.01	34.47	31.45
12	0.1	42.92	35.31
20	0.001	25.34	40.36
20	0.01	33.87	32.64
20	0.1	42.92	35.31
50	0.001	25.17	40.36
50	0.01	34.47	31.75
50	0.1	42.92	35.31

3.2.2 GMM using Diagonal Covariance Matrix

Training and validation accuracies for Bayes Classified with a GMM, using Diagonal Covariance Matrices are as below.

Table 20: GMM using Diagonal Covariance Matrix Classification Accuracies

Q value	Multiplier	Training Accuracy	Validation Accuracy
6	0.0	49.66	37.68
6	0.01	31.14	30.86
6	0.1	41.81	35.31
12	0.0	38.82	33.53
12	0.01	31.43	30.86
12	0.1	41.81	35.31
20	0.0	34.90	30.27
20	0.01	31.14	30.86
20	0.1	41.81	35.31
50	0.01	31.14	30.86
50	0.1	41.81	35.31

3.3 Best Model, and Test accuracy

The best model amongst the models available, as well as its performance on the test dataset is as given below.

3.3.1 GMM using Full Covariance Matrix

The best model for Bayes Classified with a GMM, using Full Covariance Matrices is **Q=6, and multiplier = 0.1.**

This gives a test accuracy of **30.86%**

3.3.2 GMM using Diagonal Covariance Matrix

The best model for Bayes Classified with a GMM, using Diagonal Covariance Matrices is **Q=6, and multiplier = 0.1.**

This gives a test accuracy of **35.31%**

3.4 Confusion Matrix for the best model

The confusion matrix that is obtained for the best model, is given below for each of the individual cases.

3.4.1 GMM using Full Covariance Matrix

The confusion matrices for the best model, i.e., Q=6 and multiplier=0.1 is given below.

Table 21: GMM with Full Covariance Matrix Classifier Confusion Matrix for Training Data

Training Data					
Class	Coast	Forest	Highway	Mountain	Tall Building
Coast	0	50	44	16	141
Forest	0	189	14	0	26
Highway	0	14	110	3	55
Mountain	0	52	11	27	171
Tall Building	0	44	20	8	177

Table 22: GMM with Full Covariance Matrix Classifier Confusion Matrix for Validation Data

Validation Data					
Class	Coast	Forest	Highway	Mountain	Tall Building
Coast	0	20	12	3	38
Forest	0	59	2	0	5
Highway	0	1	15	1	35
Mountain	0	23	8	2	42
Tall Building	0	25	3	0	43

3.4.2 GMM using Diagonal Covariance Matrix

The confusion matrices for the best model, i.e., Q=6 and multiplier=0.1 is given below.

Table 23: GMM with Diagonal Covariance Matrix Classifier Confusion Matrix for Training Data

Training Data					
Class	Coast	Forest	Highway	Mountain	Tall Building
Coast	142	0	18	39	52
Forest	78	0	30	25	96
Highway	12	0	154	8	8
Mountain	56	0	22	109	74
Tall Building	24	0	15	33	177

Table 24: GMM with Diagonal Covariance Matrix Classifier Confusion Matrix for Validation Data

Validation Data					
Class	Coast	Forest	Highway	Mountain	Tall Building
Coast	33	0	10	15	15
Forest	20	0	10	8	28
Highway	5	0	27	6	14
Mountain	18	0	8	31	18
Tall Building	17	0	4	14	36

3.5 Observations

Some key observations after looking at the results are as follows:

- For full covariance matrices often the diagonal elements for covariance matrices tend to zero. In such cases the pseudo inverse of the matrix using Standard Value Decomposition (SVD) cannot be computed. We added a small fraction to the diagonal elements of such matrices provided the determinant of the matrix is less than 0.0001.
- For the diagonal covariance matrix, the highest accuracy is obtained using 6 components for each class with no multiplier addition to the covariance matrix.
- For full covariance matrix, as expected the higher accuracy is obtained with no multiplier addition and number of components for each class equal to 6.

4 Dataset 2B – Image Data Set for Dynamic Classification

4.1 Problem Statement

We are given a Real-world data set i.e., an Image data set for varying length pattern classification.

For this data, as per the mapping given to us (Group 18), we work on the classes: **coast, forest, highway, mountain, and tall building**.

Based upon this, the following classifiers need to be built.

1. Bayes classifier with a GMM for each class, using full covariance matrices
2. Bayes classifier with a GMM for each class, using diagonal covariance matrices

4.2 Classification Accuracies Obtained

Upon implementing the various classifiers, the accuracies obtained are as follows.

4.2.1 GMM using Full Covariance Matrix

Training and validation accuracies for Bayes Classified with a GMM, using Full Covariance Matrices are as below.

Table 25: GMM using Full Covariance Matrix Classification Accuracies

Q Value	Iterations	Multiplier	Training Accuracy	Validation Accuracy
6	20	0.001	52.82	45.70
6	20	0.01	45.56	40.06
12	10	0.01	59.13	56.38
12	20	0.01	53.41	47.18
12	10	0.001	52.30	45.70
20	10	0.001	52.64	45.70

4.2.2 GMM using Diagonal Covariance Matrix

Training and validation accuracies for Bayes Classified with a GMM, using Diagonal Covariance Matrices are as below.

Table 26: GMM using Diagonal Covariance Matrix Classification Accuracies

Q Value	Iterations	Multiplier	Training Accuracy	Validation Accuracy
4	100	0.001	51.54	43.62
6	1	0.01	45.56	40.06
6	10	0.01	51.96	46.88
6	100	0.01	57.33	52.52
12	10	0.01	58.61	51.04
12	100	0.01	57.85	54.30
20	10	0.01	51.19	44.21

4.3 Best Model, and Test accuracy

The best model amongst the models available, as well as its performance on the test dataset is as given below.

4.3.1 GMM using Full Covariance Matrix

The best model for Bayes Classified with a GMM, using Full Covariance Matrices is **Q=12, and multiplier = 0.01.**

This gives a test accuracy of **56.38%**

4.3.2 GMM using Diagonal Covariance Matrix

The best model for Bayes Classified with a GMM, using Diagonal Covariance Matrices is **Q=12, and multiplier = 0.01.**

This gives a test accuracy of **54.30%**

4.4 Confusion Matrix for the best model

The confusion matrix that is obtained for the best model, is given below for each of the individual cases.

4.4.1 GMM using Full Covariance Matrix

The confusion matrices for the best model, i.e., Q=12 and multiplier=0.01 is given below.

Table 27: GMM with Full Covariance Matrix Classifier Confusion Matrix for Training Data

Training Data					
Class	Coast	Forest	Highway	Mountain	Tall Building
Coast	100	96	62	33	47
Forest	1	158	1	49	20
Highway	13	0	138	4	27
Mountain	11	6	42	103	99
Tall Building	4	1	35	14	194

Table 28: GMM with Full Covariance Matrix Classifier Confusion Matrix for Validation Data

Validation Data					
Class	Coast	Forest	Highway	Mountain	Tall Building
Coast	31	1	17	8	16
Forest	0	46	0	14	6
Highway	4	0	22	4	22
Mountain	2	1	7	38	27
Tall Building	0	0	12	6	53

4.4.2 GMM using Diagonal Covariance Matrix

The confusion matrices for the best model, i.e., Q=6 and multiplier=0.1 is given below.

Table 29: GMM with Diagonal Covariance Matrix Classifier Confusion Matrix for Training Data

Training Data					
Class	Coast	Forest	Highway	Mountain	Tall Building
Coast	152	25	31	15	28
Forest	3	185	2	14	25
Highway	55	1	93	3	29
Mountain	63	24	22	58	93
Tall Building	28	8	15	6	190

Table 30: GMM with Diagonal Covariance Matrix Classifier Confusion Matrix for Validation Data

Validation Data					
Class	Coast	Forest	Highway	Mountain	Tall Building
Coast	48	8	10	3	4
Forest	3	52	0	7	4
Highway	22	0	11	0	19
Mountain	15	3	5	20	31
Tall Building	9	4	2	3	52

4.5 Observations

Some key observations after looking at the results are as follows:

- As the data set is real world data having large size, the time taken by each model to train for full covariance matrix was found to be double the time taken by diagonal covariance matrix. Because of which higher iteration values for the full matrix couldn’t be estimated.
- As the number of iterations increases the accuracy of the model tends to increase.
- The optimal component count is found to be the same for both diagonal and full covariance matrices i.e., 12 components.