Train Accuracy	Validation Accuracy	Test Accuracy
1.0	1.0	1.0

Table 2: Performance of 1-NN Model on Dataset-1(a)

The confusion matrix for the test data for the method is as follows: $\begin{bmatrix} 30 & 0 & 0 & 0 \\ 0 & 30 & 0 & 0 \\ 0 & 0 & 29 & 0 \\ 0 & 0 & 0 & 29 \end{bmatrix}$

Here is the decision boundary for this method:

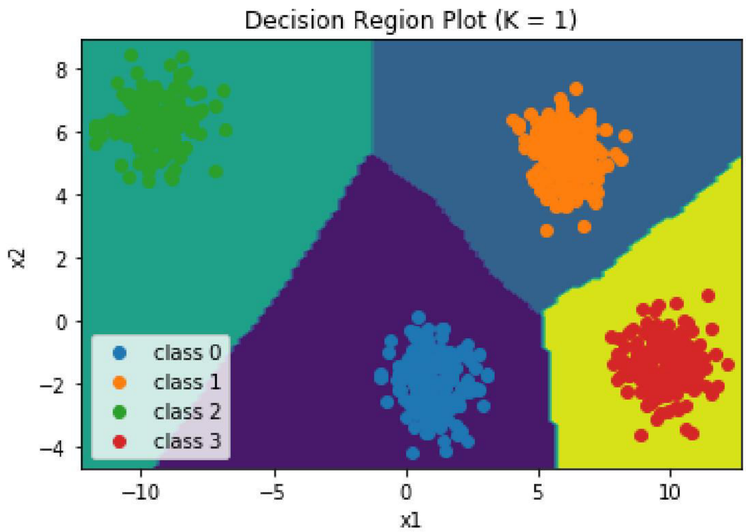

Figure 2: Decision Boundary Plot for K=1

For K=7:
The table below describes the train, validation and test accuracy for K=7.

Train Accuracy	Validation Accuracy	Test Accuracy
1.0	1.0	1.0

Table 3: Performance of 7-NN Model on Dataset-1(a)

The confusion matrix for the test data for the method is as follows: $\begin{bmatrix} 30 & 0 & 0 & 0 \\ 0 & 30 & 0 & 0 \\ 0 & 0 & 29 & 0 \\ 0 & 0 & 0 & 29 \end{bmatrix}$

Here is the decision boundary for this method:

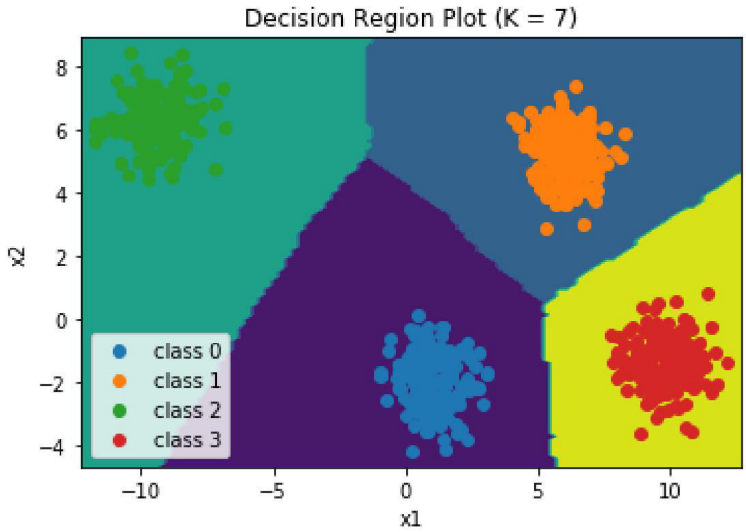

Figure 3: Decision Boundary Plot for K=7

For K=15:
The table below describes the train, validation and test accuracy for K=15.

Train Accuracy	Validation Accuracy	Test Accuracy
1.0	1.0	1.0

Table 4: Performance of 15-NN Model on Dataset-1(a)

The confusion matrix for the test data for the method is as follows: $\begin{bmatrix} 30 & 0 & 0 & 0 \\ 0 & 30 & 0 & 0 \\ 0 & 0 & 29 & 0 \\ 0 & 0 & 0 & 29 \end{bmatrix}$

Here is the decision boundary for this method:

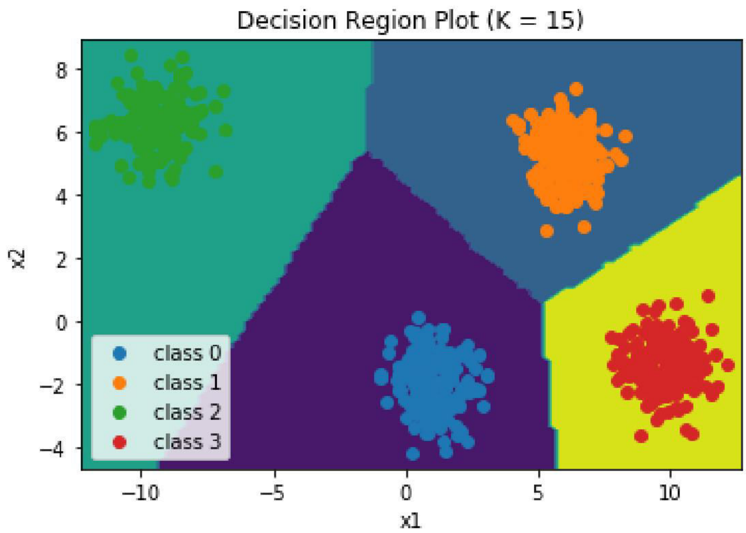

Figure 4: Decision Boundary Plot for K=15

1.1.1 Observations

- From figure[1], we can observe that the data is very well separated from each other classes(i.e linearly separable)

- From the table[1], we can observe that all the KNN models(K=1, K=7, K=15) are performing perfectly on Dataset 1(a), as they are linearly separable.

1.2 Naive Bayes Classifier with a Gaussian Distribution for Each Class

We considered three cases depending on the covariance matrices for the given dataset.

1.2.1 Covariance matrix for all the classes is the same and is $\sigma^2 I$

In this part we consider the case of a single variance value across all the classes and features. The table below describes the train, validation and test accuracy for this method.

Train Accuracy	Validation Accuracy	Test Accuracy
1.0	1.0	1.0

The confusion matrix for the test data for the method is as follows: $\begin{bmatrix} 30 & 0 & 0 & 0 \\ 0 & 30 & 0 & 0 \\ 0 & 0 & 29 & 0 \\ 0 & 0 & 0 & 29 \end{bmatrix}$

Here is the decision boundary for this method:

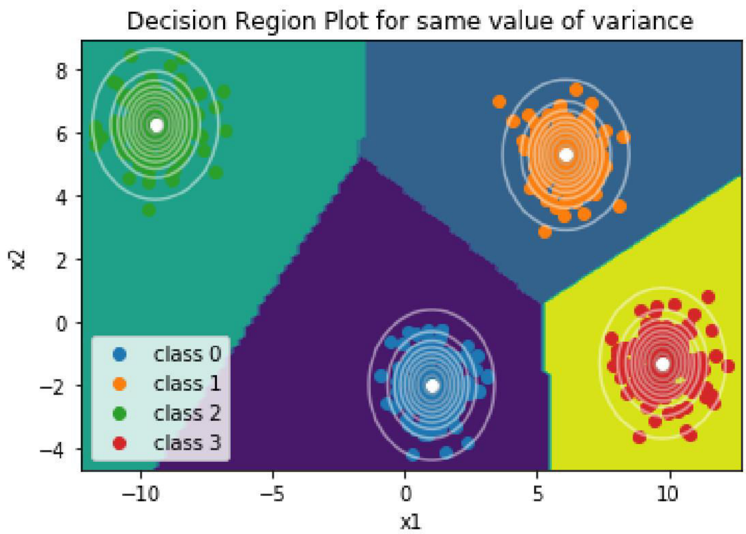

Figure 5: Decision Boundary Plot

1.2.2 Covariance matrix for all the classes is the same and is C

In this part we consider the case of a single variance value across all the classes and features. The table below describes the train, validation and test accuracy for this method.

Train Accuracy	Validation Accuracy	Test Accuracy
1.0	1.0	1.0

The confusion matrix for the test data for the method is as follows: $\begin{bmatrix} 30 & 0 & 0 & 0 \\ 0 & 30 & 0 & 0 \\ 0 & 0 & 29 & 0 \\ 0 & 0 & 0 & 29 \end{bmatrix}$

Here is the decision boundary for this method:

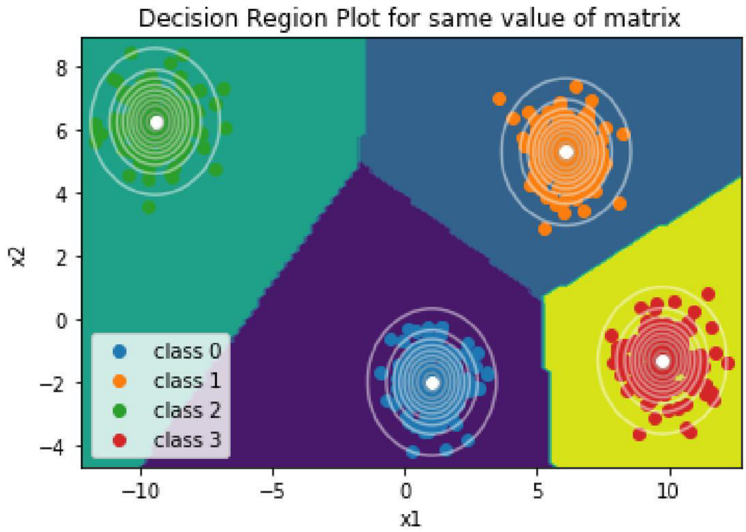

Figure 6: Decision Boundary Plot

1.2.3 Covariance matrix for all the classes is different

In this part we consider the case of a single variance value across all the classes and features. The table below describes the train, validation and test accuracy for this method.

Train Accuracy	Validation Accuracy	Test Accuracy
1.0	1.0	1.0

The confusion matrix for the test data for the method is as follows: $\begin{bmatrix} 30 & 0 & 0 & 0 \\ 0 & 30 & 0 & 0 \\ 0 & 0 & 29 & 0 \\ 0 & 0 & 0 & 29 \end{bmatrix}$

Here is the decision boundary for this method:

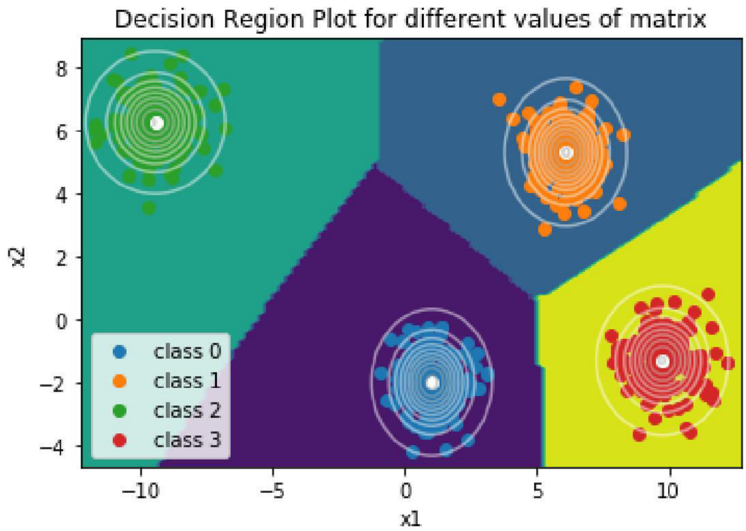

Figure 7: Decision Boundary Plot

1.2.4 Observations on Naive Bayes Classifier

Naive Bayes	Train Accuracy	Validation Accuracy	Test Accuracy
$C_i = C_j = \sigma^2 I$	1	1	1
$C_i = C_j$	1	1	1
$C_i \neq C_j$	1	1	1

- From figure[1], the data is linearly sepearable. And we knew that Naive Bayes classifier for all the above three classes, decision boundary is linear. So for dataset 1(a), it is perfectly classified for all the class labels.

2 Classification task on Dataset 1(b)

Given:

- **Train Dataset** - contains 600 samples. This is split to train and validation dataset in the ratio 80:20
- **Test Dataset** - consists of 300 samples

Following is the scatter plot of the dataset 1(b)

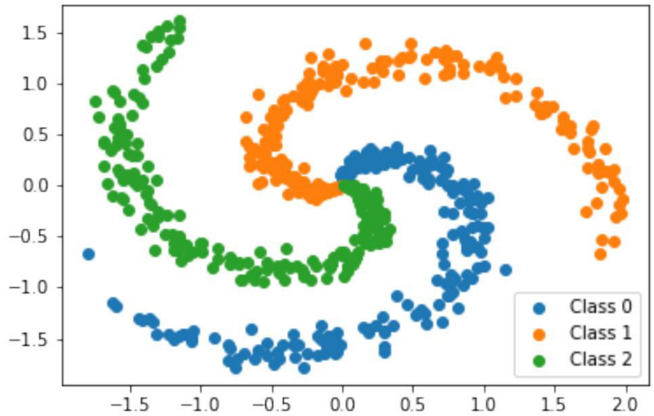

Figure 8: Scatter Plot of Dataset 1(b)

2.1 K Nearest Neighbours Classifier

In this sub section, K Nearest Neighbours (KNN) algortihm is used for classifying Dataset 1(b). The different values of K considered are K = 1, 7, 15.

KNN	**Train Accuracy**	**Validation Accuracy**
K=1	1	0.9917
K=7	0.9937	0.9917
K=15	0.9937	1

Table 5: Train and validation accuracies KNN

From Table 5, K=15 is the best model configuration, as it has the highest validation accuracy. The decision region plot for the best model configuration is given in Fig. 9. The accuracy and confusion matrix generated when this model is tested on the test dataset is given below:

- Test Accuracy: 1
- Test confusion matrix (K=15): $\begin{bmatrix} 29 & 0 & 0 \\ 0 & 30 & 0 \\ 0 & 0 & 30 \end{bmatrix}$

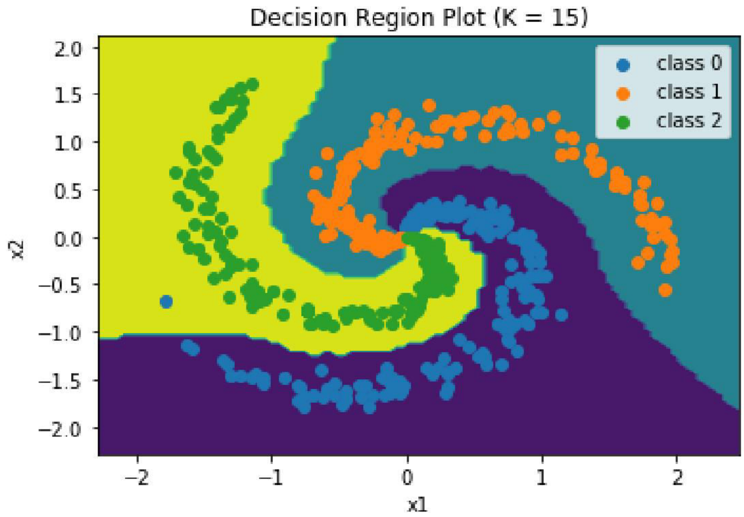

Figure 9: Decision Region Plot (K=15)

2.1.1 Observations

All the KNN models considered are performing remarkably well on Dataset 1(b), even though it is not linearly separable.

2.2 Bayes Classifier with GMM for each class (using full covariance matrices)

In this subsection, Gaussian Mixture Models (GMM) are used for performing the classification task on Dataset 1(b). Each class is represented by a mixture of Q Gaussians. (Q is the number of different gaussians used for defining the class conditional probability density function).

GMM	Train Accuracy	Validation Accuracy	# Parameters per class
Q=2	0.9541	0.925	12
Q=5	0.9979	0.9917	30
Q=10	0.9937	0.9917	60
Q=15	0.9853	0.9833	90
Q=20	0.9853	0.9833	120

Table 6: Train and validation accuracies GMM (full covariance matrix)

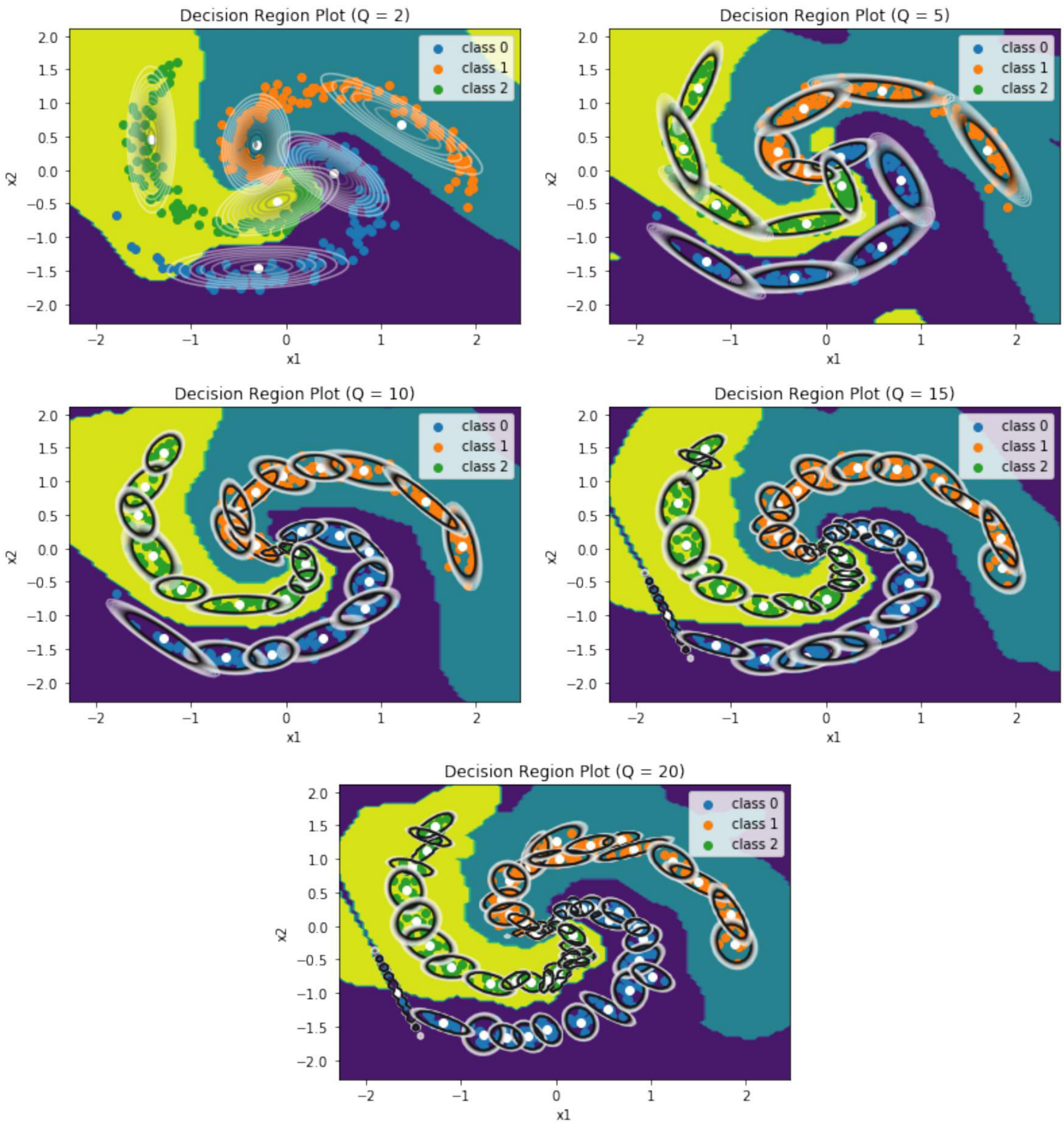

Figure 10: Decision Region Plots for GMM with full covariance matrix

From Table 6, Q=5 and Q=10 are the best model configurations, as they have the same and the highest validation accuracy. The decision region plot for all the model configuration is given in Fig. 10. The accuracy and confusion matrix generated by the best models on test dataset are given below:

For Q=5

- Test Accuracy: 0.9888
- Test confusion matrix: $\begin{bmatrix} 29 & 0 & 0 \\ 0 & 30 & 0 \\ 0 & 1 & 29 \end{bmatrix}$

For Q=10

- Test Accuracy: 1
- Test confusion matrix: $\begin{bmatrix} 29 & 0 & 0 \\ 0 & 30 & 0 \\ 0 & 0 & 30 \end{bmatrix}$

2.2.1 Observations

From Table 6 and Fig. 10, the following observations can be made:

- GMM with Q=2, is under fitting Dataset 1(b)
- With respect to validation accuracy, GMM with Q=5 and Q=10 appears to perform equally well on the classification task
- The small yellow patch near the centre in the decision region plot of GMM with Q=5, however, suggests that Q=10 has a lower generalization error and hence, is a slightly better model. This could also be the reason why the test accuracy of GMM with Q=5 is not 100%
- GMM with Q=15 and Q=20, are overfitting the data. This is evident in the decision region plots.

2.3 Bayes Classifier with GMM for each class (using diagonal covariance matrices)

In this subsection, Gaussian Mixture Models (GMM) with diagonal covariance matrices are used for performing the classification task on Dataset 1(b). Each class is represented by a mixture of Q Gaussians. (Q is the number of different gaussians used for defining the class conditional probability density function).

GMM	Train Accuracy	Validation Accuracy	# Parameters per class
Q=2	0.7223	0.7333	10
Q=5	0.9770	0.9750	25
Q=10	0.9896	1.0	50
Q=15	0.9916	0.9833	75
Q=20	0.9958	0.9917	100

Table 7: Train and validation accuracies GMM (diagonal covariance matrix)

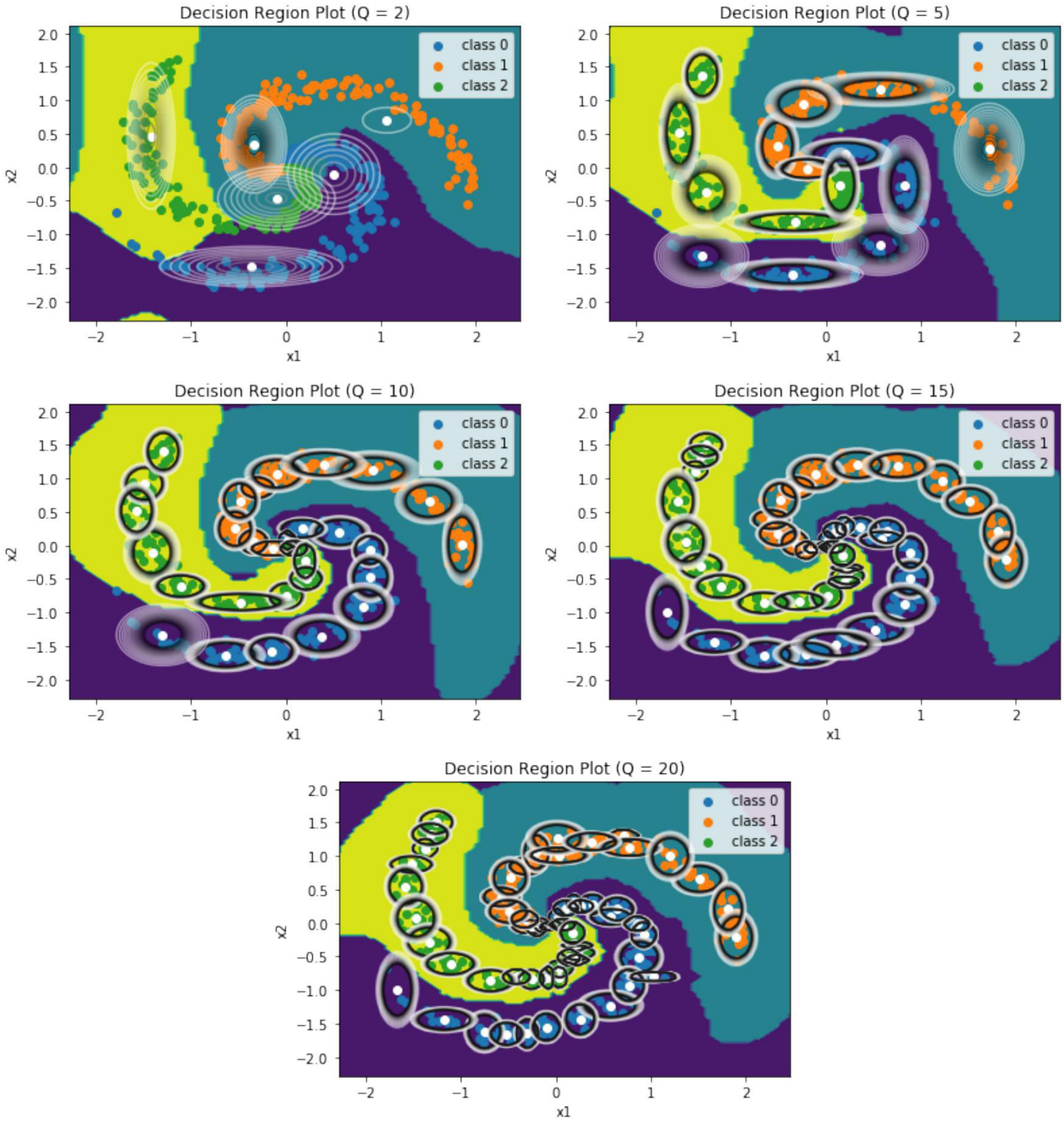

Figure 11: Decision Region Plots for GMM with diagonal covariance matrix

From Table 7, Q=10 is the best model configuration, as it has the highest validation accuracy. The decision region plot for all the model configuration is given in Fig. 11. The accuracy and confusion matrix generated by the best model on test dataset is given below:

- Test Accuracy: 0.9888
- Test confusion matrix: $\begin{bmatrix} 29 & 0 & 0 \\ 0 & 30 & 0 \\ 1 & 0 & 29 \end{bmatrix}$

2.3.1 Observations

From Table 7 and Fig. 11, the following observations can be made:

- GMM with Q=2 and Q=5, are under fitting Dataset 1(b)
- With respect to validation accuracy, Q=10 is the best model.

- From the decision plots, Q=15 and Q=20 are also performing well at the classification task without over fitting the data, unlike GMM with full covariance matrices. This could be because, the number of training parameters while considering only the diagonal elements of the covariance matrix is less compared to full covariance matrix.

2.4 Bayes Classifier with K Nearest Neighbours method for estimating class conditional probability density function

In this subsection, we use K-nearest Neighbours method for estimating class conditional probability density function for the classification of dataset 1(b), which is non-linearly separable.
Here, there are no model fitting. We have to have the train data all the time to calculate the distances from the given datapoint "x".

Following is the results for K = 10

- Train Accuracy = 0.99
- Test Accuracy = 0.989
- Train confusion matrix: $\begin{bmatrix} 197 & 0 & 3 \\ 0 & 197 & 3 \\ 0 & 0 & 200 \end{bmatrix}$
- Test confusion matrix: $\begin{bmatrix} 29 & 0 & 1 \\ 0 & 30 & 0 \\ 0 & 0 & 30 \end{bmatrix}$

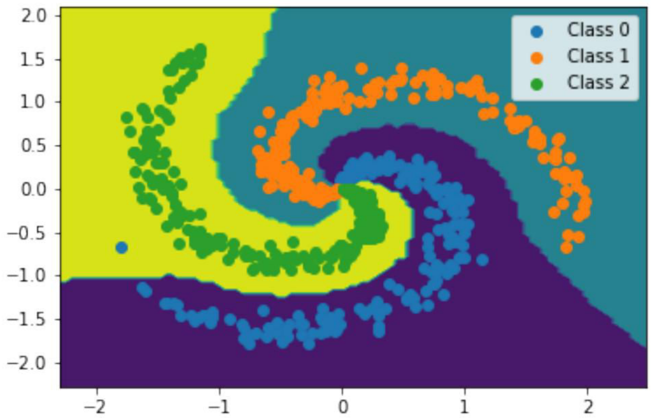

Figure 12: Decision Boundary for K = 10

Following is the result for K = 20

- Train Accuracy = 0.9833
- Test Accuracy = 0.978
- Train confusion matrix: $\begin{bmatrix} 194 & 0 & 6 \\ 1 & 196 & 3 \\ 0 & 0 & 200 \end{bmatrix}$
- Test confusion matrix: $\begin{bmatrix} 29 & 0 & 1 \\ 0 & 30 & 0 \\ 0 & 1 & 29 \end{bmatrix}$

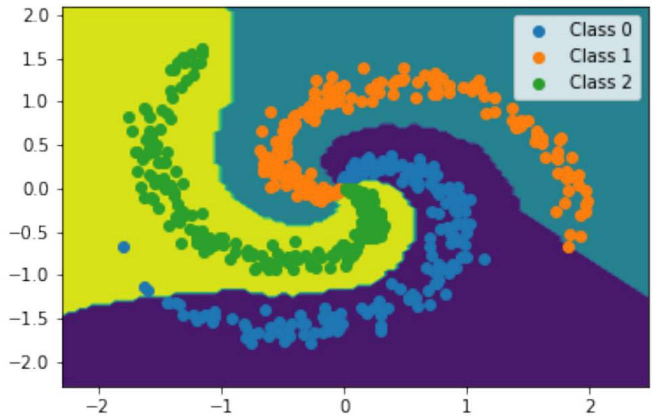

Figure 13: Decision Boundary for K = 20

2.4.1 Observations

- From the above results, Test Accuracy of K = 10(0.989) is greater than Test Accuracy of K = 20(0.978). Best model based on test accuracy is K = 10
- Theoretically, for very lesser values of K gives us over-fitting and higher values of K gives under-fitting.
- From the figure[12] and figure[13], we can clearly see that the figure[5] has smooth decision boundary, whereas in figure[6] did not have smooth decision boundary compared to figure[5].
- Time taken to run the model is pretty high when compared to training GMM with higher values of Q. This is because of for every value of x need to calculate distance from every training data point. Order of complexity of algorithm is O(N^2). But model is very simple to understand yet powerful enough to built the non-linear decision boundaries.

3 Classification task on Dataset 2(a)

Dataset 2(a): This is a real-world dataset, which has five classes namely coast, highway, forest, open-country, mountain. The **dimension** of each feature vector is **1x24**.

3.1 Bayes Classifier with GMM for each class (using full covariance matrices)

In this section, we fit a GMM model using **full covariance matrix** on the dataset 2(a) to perform multi-class classification. Each class is represented by a mixture of Q Gaussians(Q is the number of different gaussians used for defining the class conditional probability density function)

3.1.1 Hyperparameter Tuning of Q

Following table[8] is the results of Hyperparameter tuning of Q for choosing best model based on validation accuracy. Green marked is the best model.

Q	Train Accuracy	Validation Accuracy
1	0.6518	0.599
3	0.6311	0.574
5	0.6456	0.590
6	0.6208	0.583
8	0.6311	0.5619
10	0.6404	0.603
12	0.6508	0.595
14	0.6632	0.591
18	0.7004	0.574
20	0.7123	0.557

Table 8: Results for different values of Q in GMM with full covariance matrix

3.1.2 Best Model in Full Covariance Matrix

Best Model is selected based on validation accuracy of different models. The model with higher validation accuracy is considered as best model. From table[8], we can see that at **Q = 10 has the highest validation accuracy**. Accuracy and Confusion matrix of best model (i.e Q = 10)

- Q = 10
- Train Accuracy: 0.6404
- Validation Accuracy: 0.603
- Test Accuracy: 0.5057
- Train confusion matrix: $\begin{bmatrix} 96 & 21 & 2 & 66 & 21 \\ 4 & 138 & 0 & 9 & 31 \\ 11 & 2 & 98 & 10 & 14 \\ 22 & 23 & 0 & 146 & 28 \\ 9 & 40 & 0 & 35 & 142 \end{bmatrix}$
- Validation confusion matrix: $\begin{bmatrix} 18 & 6 & 0 & 18 & 3 \\ 0 & 36 & 0 & 1 & 10 \\ 7 & 3 & 27 & 2 & 8 \\ 5 & 4 & 1 & 25 & 7 \\ 4 & 6 & 2 & 9 & 40 \end{bmatrix}$
- Test confusion matrix: $\begin{bmatrix} 29 & 13 & 0 & 18 & 13 \\ 1 & 56 & 1 & 2 & 6 \\ 4 & 0 & 14 & 16 & 18 \\ 12 & 17 & 2 & 30 & 14 \\ 4 & 22 & 1 & 8 & 47 \end{bmatrix}$

3.2 Bayes Classifier with GMM for each class (using diagonal covariance matrices)

In this section, we fit a GMM model using **diagonal covariance matrix** on the dataset 2(a) to perform multi-class classification. Each class is represented by a mixture of Q Gaussians(Q is the number of different gaussians used for defining the class conditional probability density function)

3.2.1 Hyperparameter Tuning of Q

Following table[9] is the results of Hyperparameter tuning of Q for choosing best model based on validation accuracy. Green marked is the best model.

Q	Train Accuracy	Validation Accuracy
1	0.4969	0.5
3	0.5154	0.4752
5	0.5444	0.4958
6	0.5413	0.5165
8	0.5433	0.4917
10	0.5733	0.4876
12	0.5568	0.4958
14	0.5681	0.5206
15	0.5785	0.5289
16	0.5909	0.5619
18	0.6146	0.5702
20	0.5816	0.5371

Table 9: Results for different values of Q in GMM with diagonal covariance matrix

3.2.2 Best Model in Diagonal Covariance Matrix

Best Model is selected based on validation accuracy of different models. The model with higher validation accuracy is considered as best model. From table[9], we can see that at **Q = 18 has the highest validation accuracy**. Following are the Accuracy's and Confusion matrix's of best model (i.e Q = 18)

- Q = 18
- Train Accuracy: 0.6146
- Validation Accuracy: 0.5702
- Test Accuracy: 0.52011
- Train confusion matrix: $\begin{bmatrix} 96 & 16 & 3 & 73 & 18 \\ 5 & 130 & 0 & 13 & 34 \\ 12 & 4 & 93 & 17 & 9 \\ 32 & 15 & 0 & 145 & 27 \\ 22 & 5 & 5 & 39 & 131 \end{bmatrix}$
- Validation confusion matrix: $\begin{bmatrix} 18 & 8 & 0 & 13 & 6 \\ 2 & 35 & 0 & 3 & 7 \\ 6 & 2 & 26 & 8 & 5 \\ 6 & 2 & 0 & 30 & 4 \\ 7 & 5 & 5 & 15 & 29 \end{bmatrix}$
- Test confusion matrix: $\begin{bmatrix} 30 & 12 & 1 & 22 & 8 \\ 1 & 51 & 1 & 4 & 9 \\ 11 & 0 & 13 & 13 & 15 \\ 12 & 9 & 2 & 40 & 12 \\ 7 & 17 & 1 & 10 & 47 \end{bmatrix}$

3.3 Observations

- By plotting the information from table[8], we can observe that as value of Q increases, train accuracy, but the validation accuracy increases till Q = 10 and then decreases. This reflects the trade-off between underfitting(when value of Q is small) and overfitting(when value of Q is high)

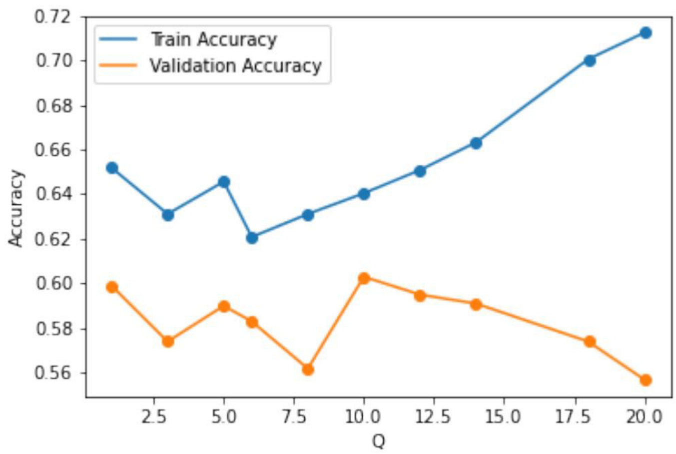

Figure 14: Plot of Training and Validation Accuracy for different values of Q for Full Covariance Matrix

- But in case of Diagonal Covariance matrix, training and validation accuracy increases till Q = 18 and then there is a drop in both training and validation accuracy. Refer figure[15]

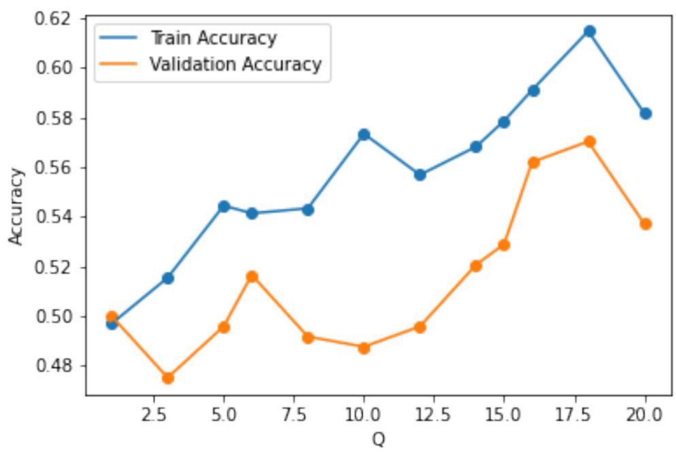

Figure 15: Plot of Training and Validation Accuracy for different values of Q for Diagonal Covariance Matrix

4 Classification task on Dataset 2(b)

Dataset 2(b): This is a real-world dataset, which has five classes namely coast, highway, forest, open-country, mountain. Set of local features are extracted from the original image and stored in a matrix of size (36x23). This is varying length dataset. In this section, we fit the model considering all local feature vector and apply GMM but when making a decision regarding class label, we collect the set of local feature vectors belonging to an image likelihoods, then argument maximum of product of those likelihoods will gives us the predicted label for that image.

4.1 Bayes Classifier with GMM for each class (using full covariance matrices)

In this section, we fit a GMM model using **full covariance matrix** on the dataset 2(b) to perform multi-class classification. Each class is represented by a mixture of Q Gaussians(Q is the number of different gaussians used for defining the class conditional probability density function)

4.1.1 Hyperparameter Tuning of Q

Following table[10] is the results of Hyperparameter tuning of Q for choosing best model based on test accuracy. Green marked is the best model.

Q	Training Accuracy	Test Accuracy
1	0.6537	0.5862
2	0.6082	0.5734
5	0.5363	0.5574
10	0.5066	0.49712
15	0.3619	0.3735
20	0.4801	0.49425
30	0.4421	0.4626
40	0.4115	0.4166
50	0.3008	0.3247
75	0.3396	0.3448
100	0.2677	0.3074

Table 10: Results for different values of Q in GMM with full covariance matrix

4.1.2 Best Model in Full Covariance Matrix

Best Model is selected based on validation accuracy of different models. The model with higher test accuracy is considered as best model. From table[10], we can see that at **Q = 1 has the highest test accuracy**. Accuracy and Confusion matrix of best model (i.e Q = 1)

- Q = 1
- Train Accuracy: 0.6537
- Test Accuracy: 0.5862
- Train confusion matrix: $\begin{bmatrix} 156 & 6 & 37 & 19 & 33 \\ 4 & 193 & 5 & 14 & 13 \\ 20 & 1 & 136 & 18 & 7 \\ 20 & 14 & 39 & 134 & 54 \\ 44 & 24 & 23 & 24 & 172 \end{bmatrix}$
- Test confusion matrix: $\begin{bmatrix} 46 & 2 & 14 & 1 & 10 \\ 0 & 53 & 2 & 8 & 3 \\ 4 & 0 & 27 & 14 & 7 \\ 9 & 8 & 13 & 31 & 14 \\ 10 & 10 & 6 & 9 & 47 \end{bmatrix}$

4.2 Bayes Classifier with GMM for each class (using diagonal covariance matrices)

In this section, we fit a GMM model using **diagonal covariance matrix** on the dataset 2(b) to perform multi-class classification. Each class is represented by a mixture of Q Gaussians(Q is the number of different gaussians used for defining the class conditional probability density function)

4.2.1 Hyperparameter Tuning of Q

Following table[11] is the results of Hyperparameter tuning of Q for choosing best model based on validation accuracy. Green marked is the best model.

Q	Training Accuracy	Test Accuracy
1	0.5636	0.5201
2	0.5355	0.4798
5	0.5438	0.5172
10	0.4669	0.4712
15	0.4768	0.4971
20	0.4710	0.4741
30	0.5	0.5
40	0.4859	0.4511
50	0.45950	0.4540

Table 11: Results for different values of Q in GMM with diagonal covariance matrix

4.2.2 Best Model in Diagonal Covariance Matrix

Best Model is selected based on validation accuracy of different models. The model with higher validation accuracy is considered as best model. From table[11], we can see that at **Q = 18 has the highest validation accuracy**. Following are the Accuracy's and Confusion matrix's of best model (i.e Q = 18)

- Q = 1
- Train Accuracy: 0.5636
- Test Accuracy: 0.5201
- Train confusion matrix: $\begin{bmatrix} 121 & 9 & 50 & 26 & 45 \\ 7 & 168 & 5 & 24 & 25 \\ 16 & 1 & 133 & 19 & 13 \\ 46 & 18 & 49 & 107 & 41 \\ 54 & 24 & 24 & 32 & 153 \end{bmatrix}$
- Test confusion matrix: $\begin{bmatrix} 32 & 1 & 13 & 10 & 17 \\ 0 & 45 & 3 & 10 & 8 \\ 4 & 0 & 25 & 11 & 12 \\ 5 & 5 & 14 & 37 & 14 \\ 13 & 9 & 8 & 10 & 42 \end{bmatrix}$

4.3 Observations

- From the table[10] and table[11], the trend of the training accuracy and validation accuracy is a decreasing trend (not strictly decreasing but trend is downwards.), this means that at Q = 1 we have the maximum training and test accuracy
- For the real-world datasets like Dataset(2b), there are some numerical challenges like overflow of exponent term in multivariate gaussian, singularity issues of covariance matrix. Singularity issue can be resolved by add some very small value to diagonal elements of the covariance matrix(motivated from Levenberg-Marquadt Method), so here there is some trade-off with accuracy.