ASSIGNMENT 2

CS5691 Pattern Recognition and Machine Learning

CS5691 Assignment 2

Team Members:

CS24M039	Sagar Biswas
CS24M043	Saurabh Kumar Sahu

Indian Institute of Technology, Madras

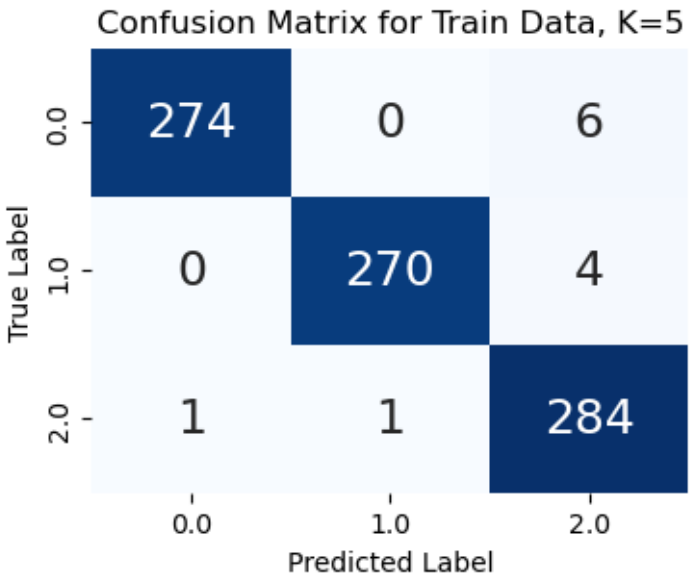
Dataset-1:

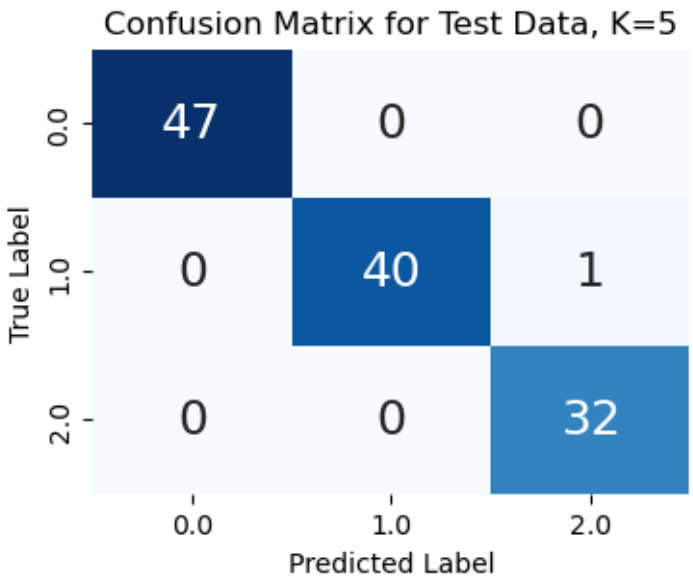
1. K-Nearest Neighbour Classifier for K=1, K=5 and K=9

Table of Classification Accuracies of the Model for Training data, Validation data, and Test data:

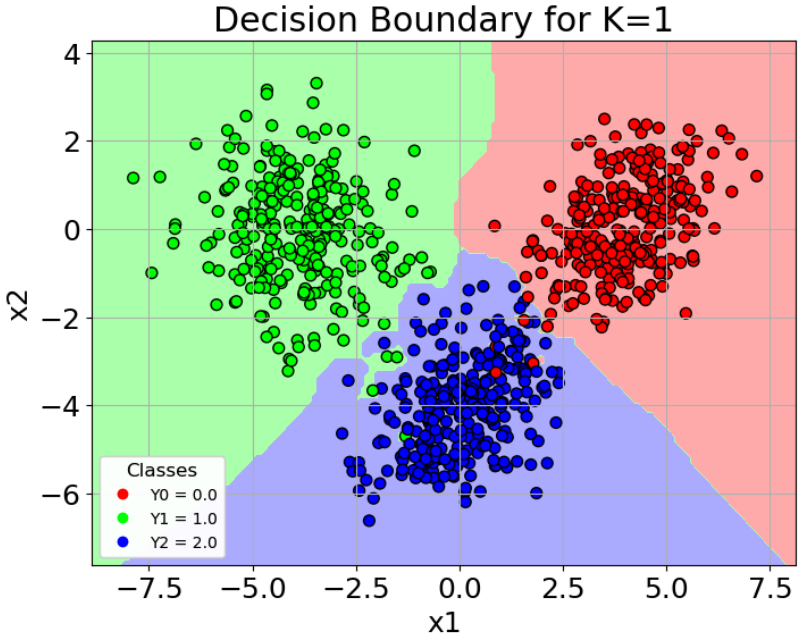
k-value	Accuracy train	Accuracy Val	Accuracy test
1	1.000000	0.9875	0.991667
5	0.985714	0.9875	0.991667
9	0.986905	0.9875	0.983333

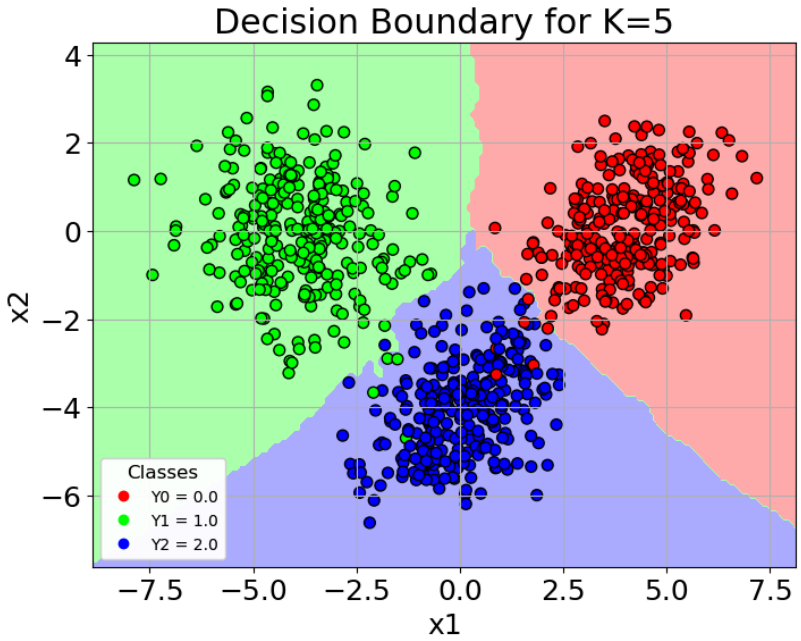
Confusion Matrix for the best configuration of the model, on training data and test data:

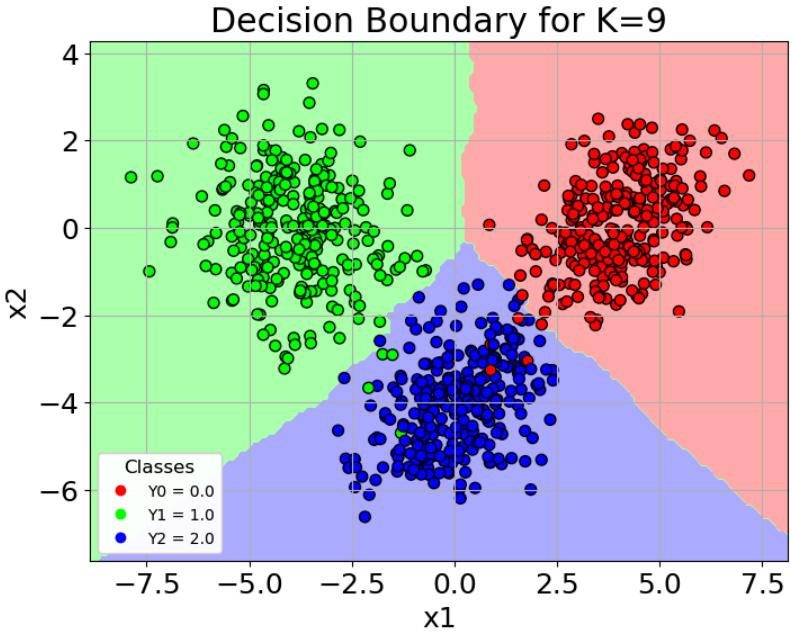




Decision Region Plots:







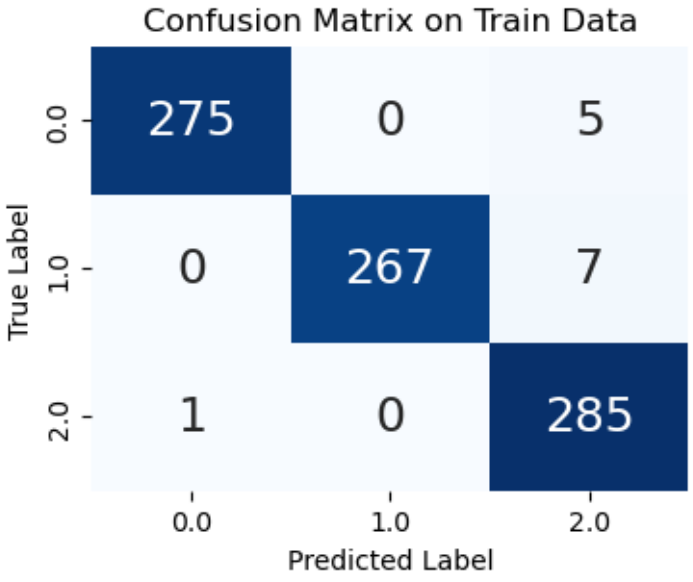
2. Bayes Classifier with a Gaussian Distribution for every class

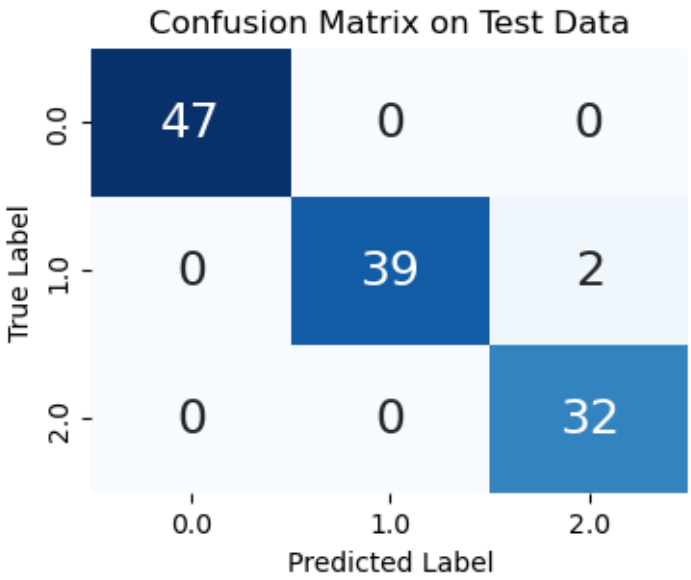
a. Covariance matrices for all the classes are the same

Table of Classification Accuracies of the Model for Training data, Validation data, and Test data:

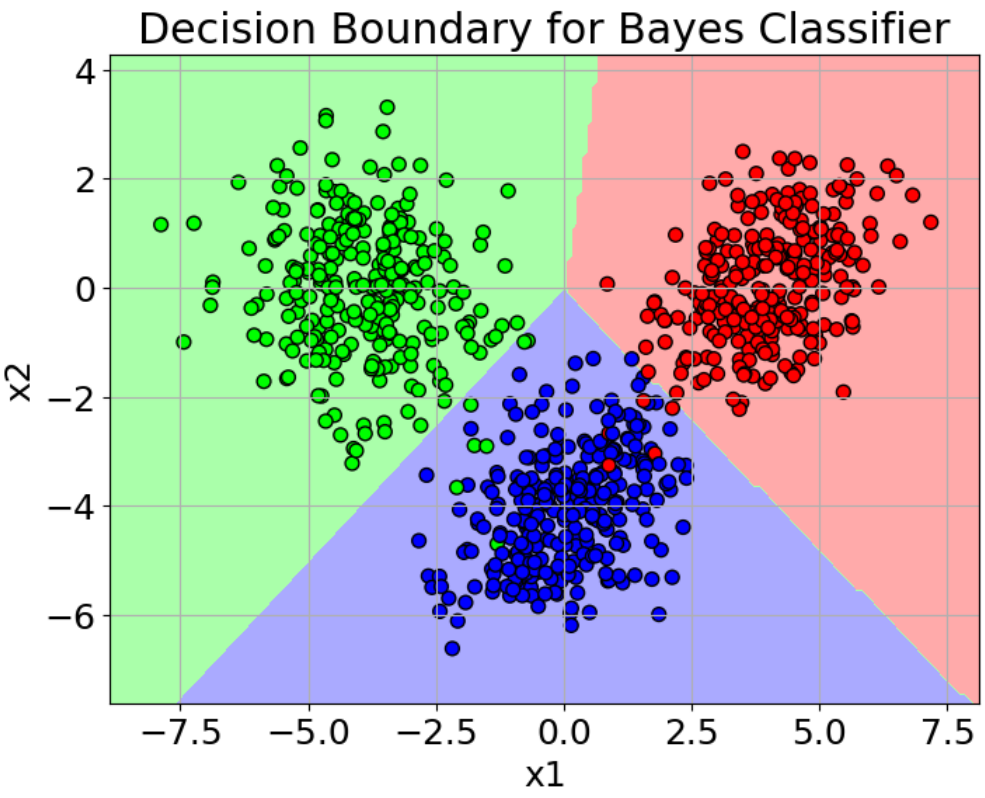
Dataset	Size	Accuracy
Train	840	0.984524
Validation	240	0.983333
Test	120	0.983333

Confusion Matrix for the best configuration of the model, on training data and test data:

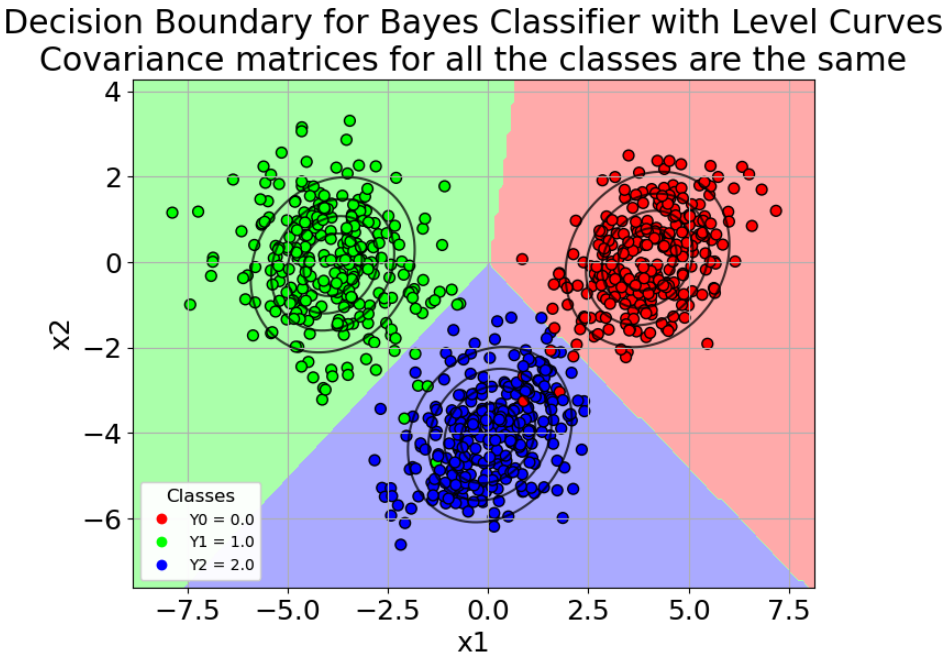




Decision Region Plot:



Decision Region Plot with Level curves on Training Data:

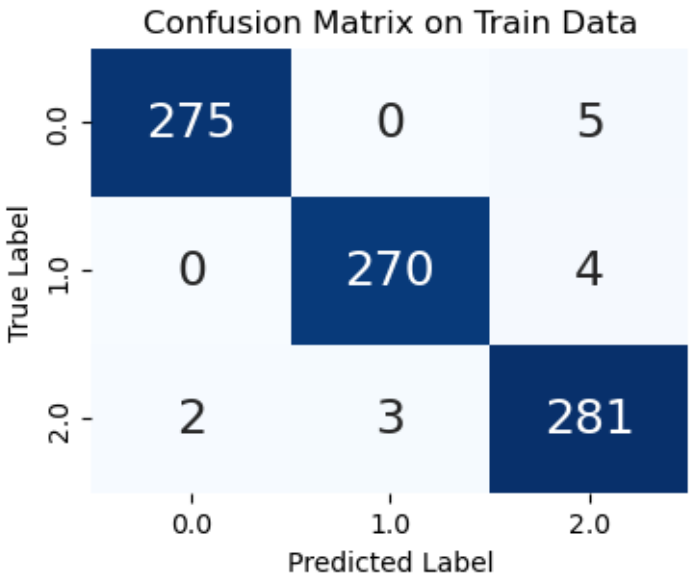


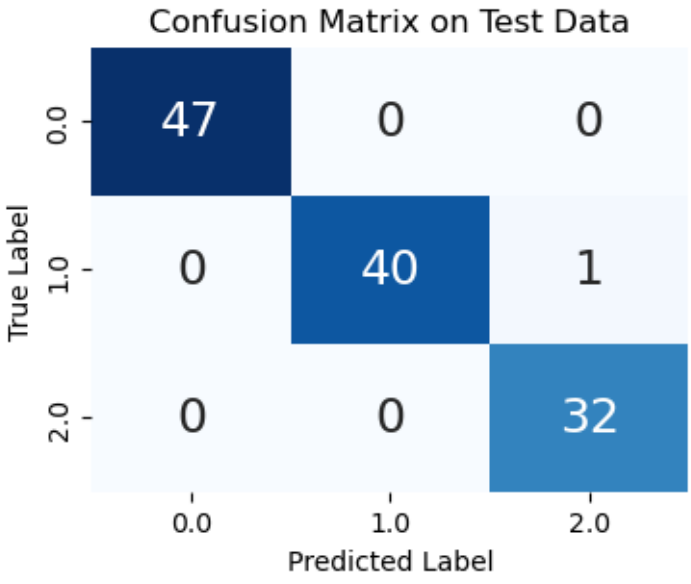
b. Covariance matrices are different

Table of Classification Accuracies of the Model for Training data, Validation data, and Test data:

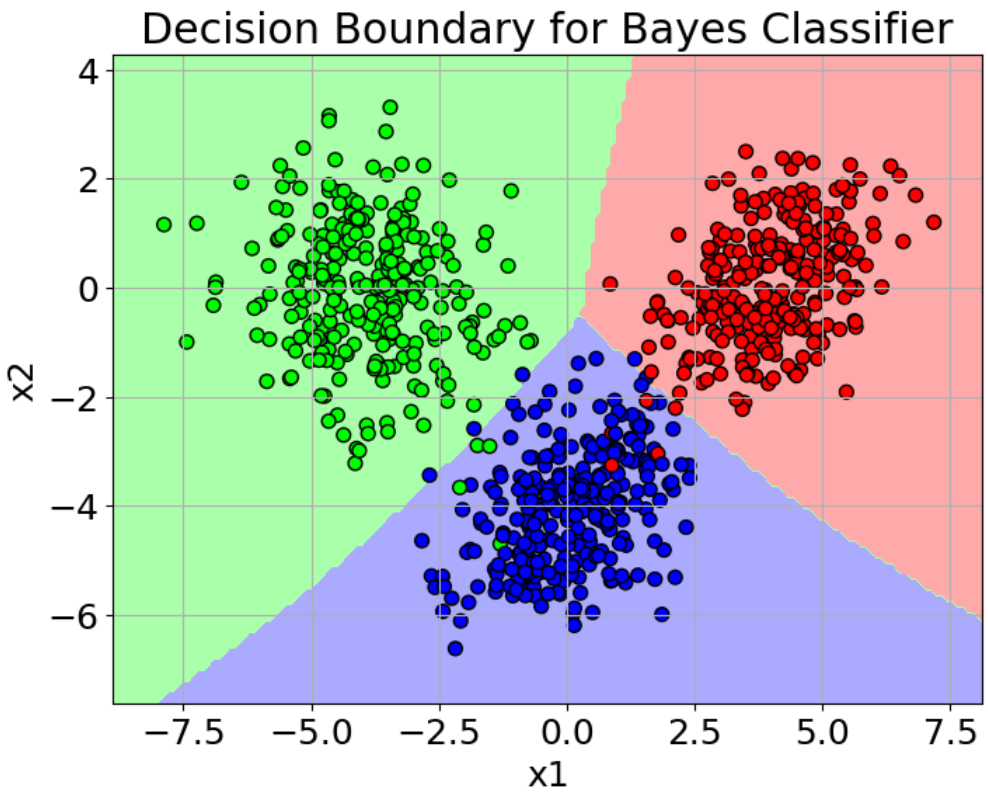
Dataset	Size	Accuracy
Train	840	0.983333
Validation	240	0.991667
Test	120	0.991667

Confusion Matrix for the best configuration of the model, on training data and test data:

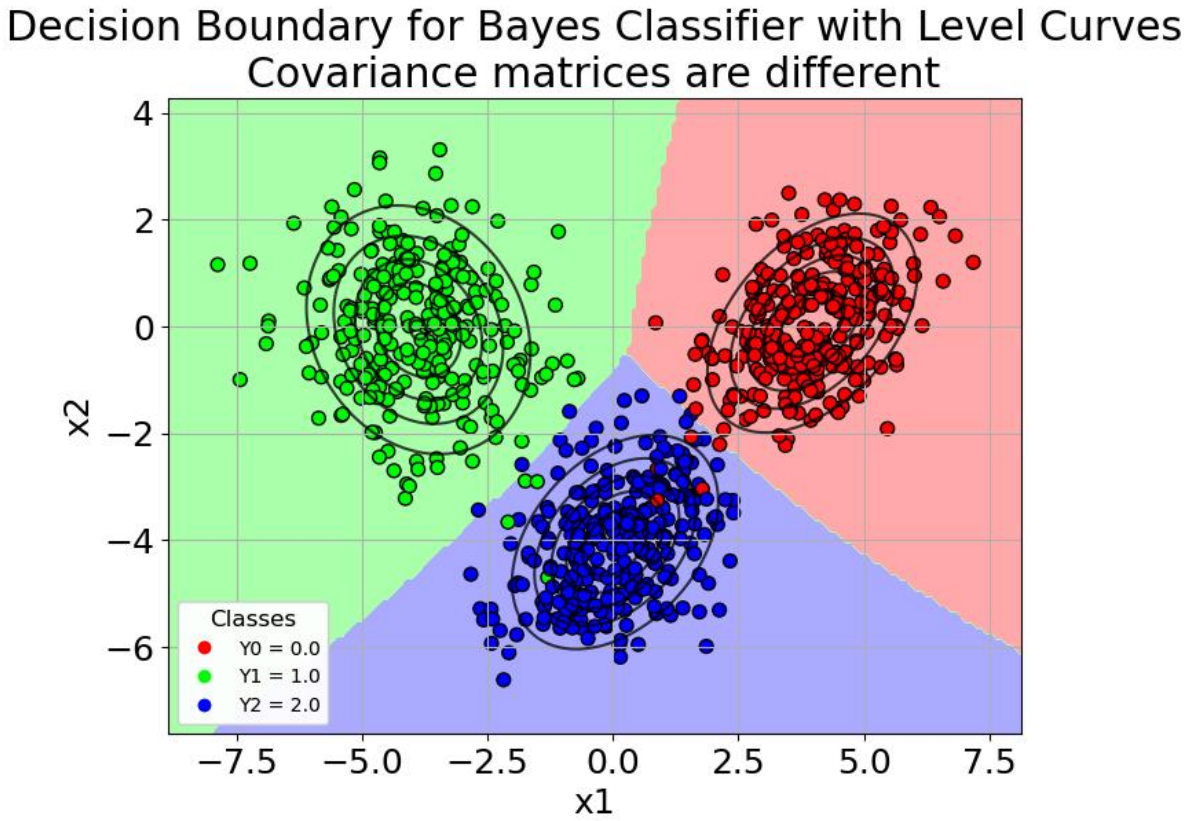




Decision Region Plot:



Decision Region Plot with Level curves on Training Data:



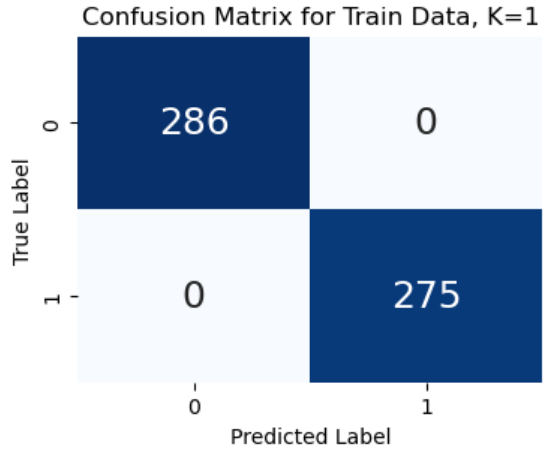
Dataset-2:

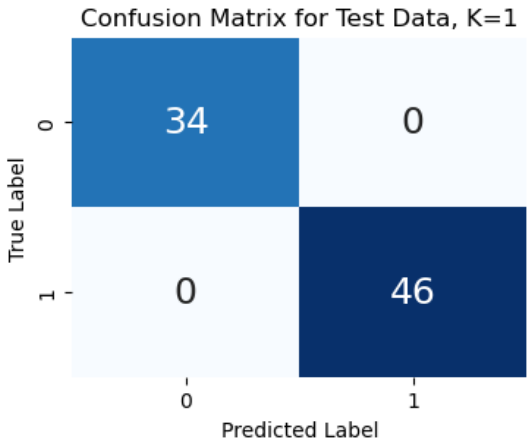
1. K-Nearest Neighbour Classifier for K=1, K=5 and K=9

Table of Classification Accuracies of the Model for Training data, Validation data, and Test data:

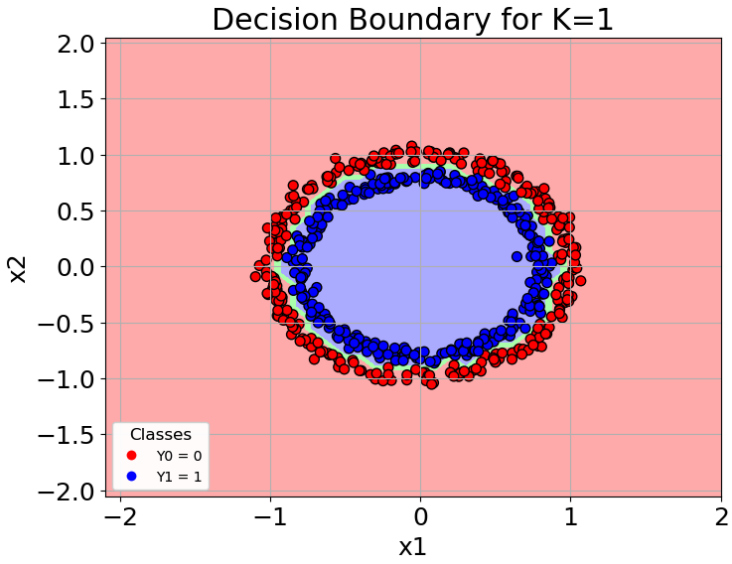
k-value	Accuracy train	Accuracy Val	Accuracy test
1	1.000000	1.000000	1.0
5	1.000000	0.981132	1.0
9	0.994652	0.993711	1.0

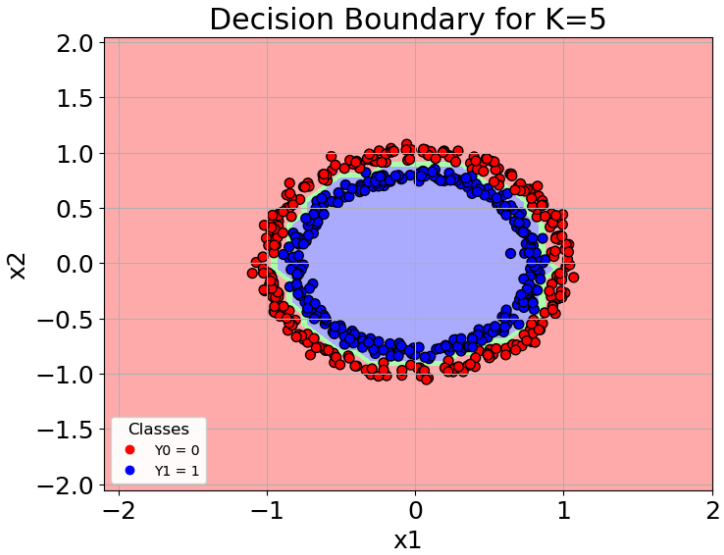
Confusion Matrix for the best configuration of the model, on training data and test data:

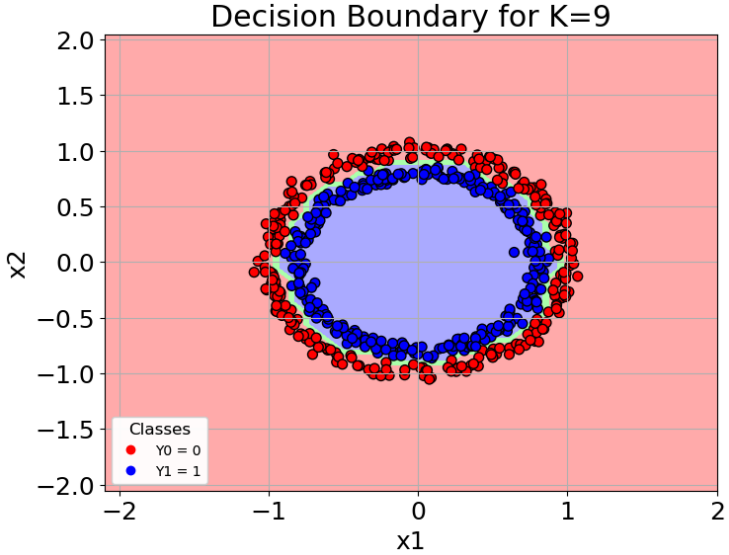




Decision Region Plots:





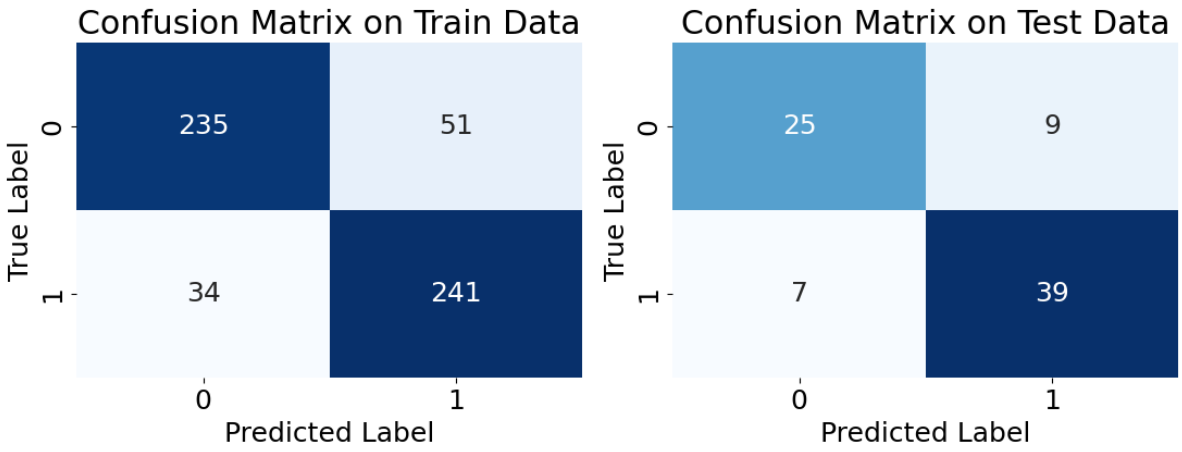


2. K-nearest representatives classifier, for K=1, K=3 and K=5: Use 10 representatives per class

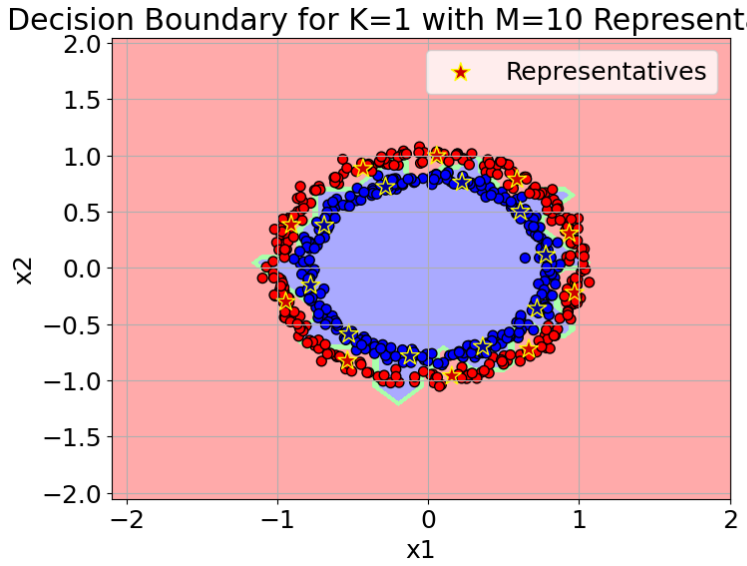
Table of Classification Accuracies of the Model for Training data, Validation data, and Test data:

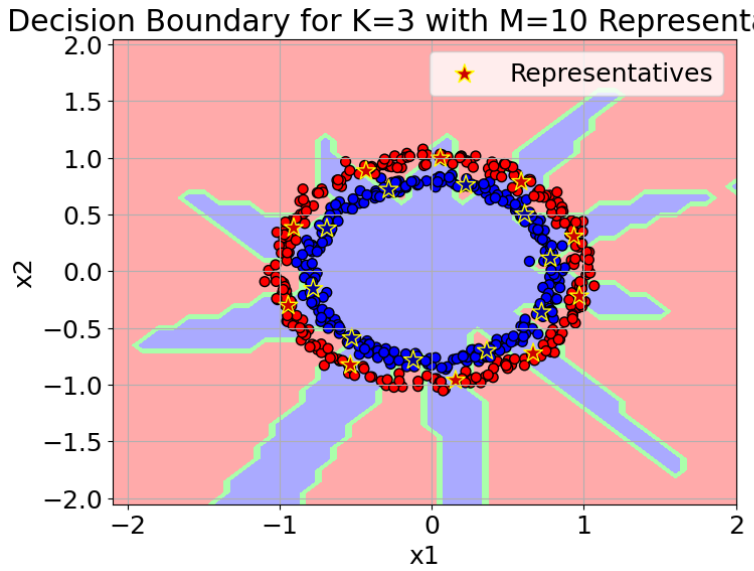
k-value	Accuracy Train	Accuracy Val	Accuracy Test
1	0.848485	0.798742	0.8000
3	0.582888	0.635220	0.6625
5	0.597148	0.572327	0.5875

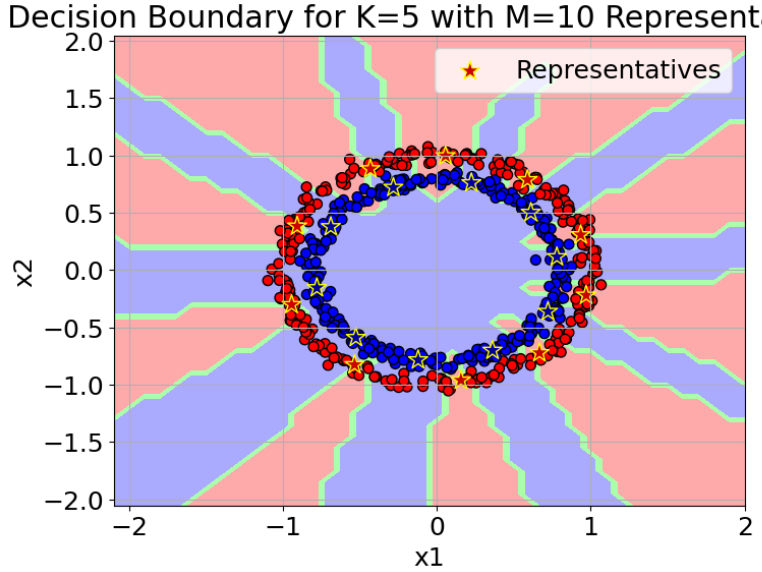
Confusion Matrix for the best configuration of the model, on training data and test data:



Decision Region Plots:







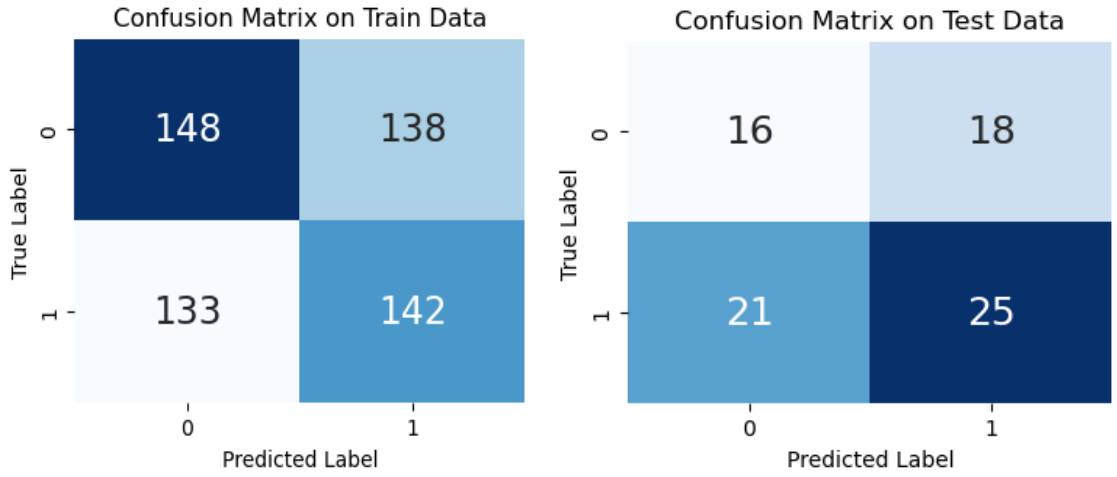
3. Bayes Classifier with a Gaussian Distribution for every class

a. Covariance matrices for all the classes are the same

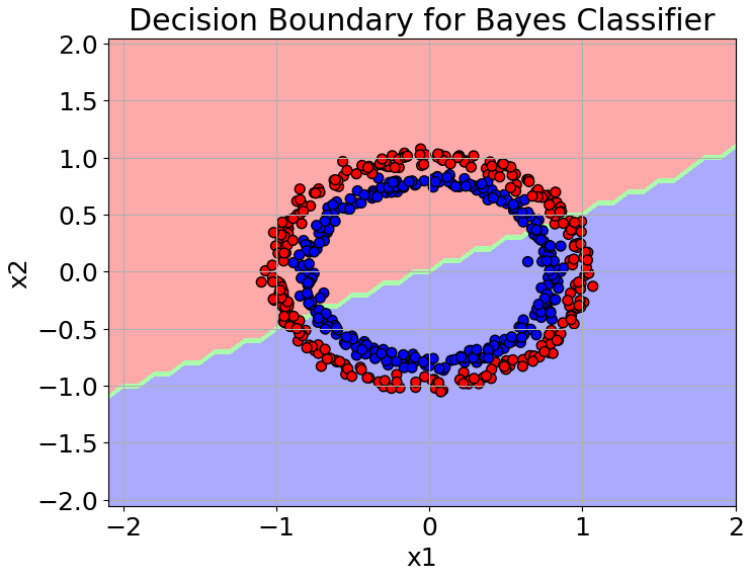
Table of Classification Accuracies of the Model for Training data, Validation data, and Test data:

Dataset	Size	Accuracy
Train	561	0.516934
Validation	159	0.427673
Test	80	0.512500

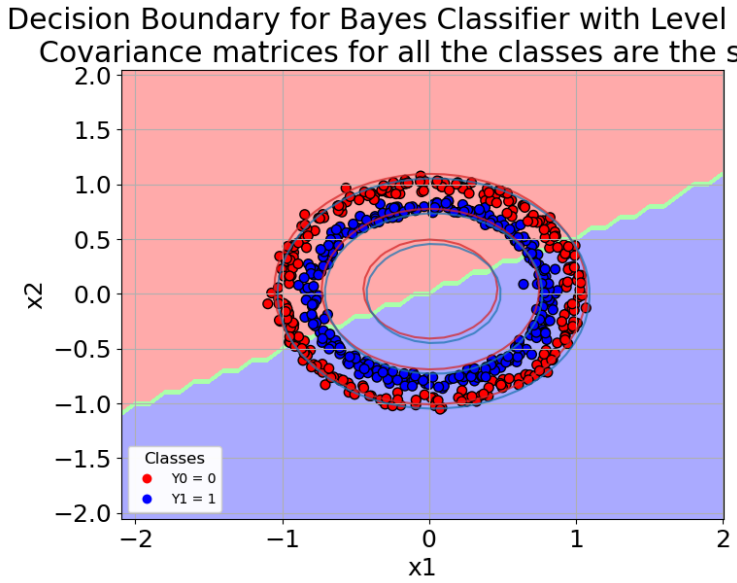
Confusion Matrix for the best configuration of the model, on training data and test data:



Decision Region Plot:



Decision Region Plot with Level curves on Training Data:

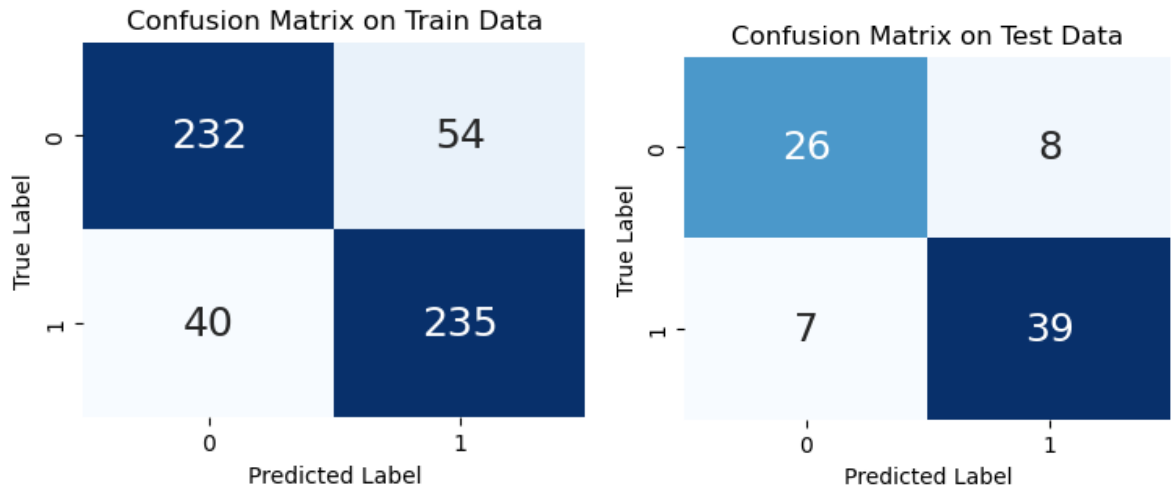


b. Covariance matrices are different

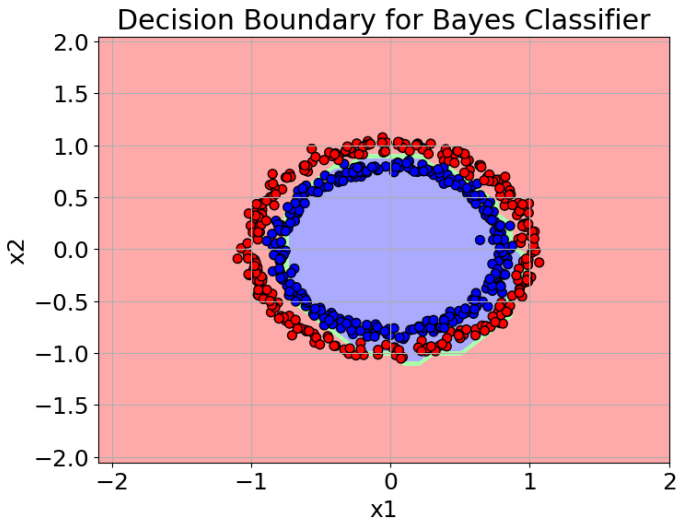
Table of Classification Accuracies of the Model for Training data, Validation data, and Test data:

Dataset	Size	Accuracy
Train	561	0.832442
Validation	159	0.786164
Test	80	0.812500

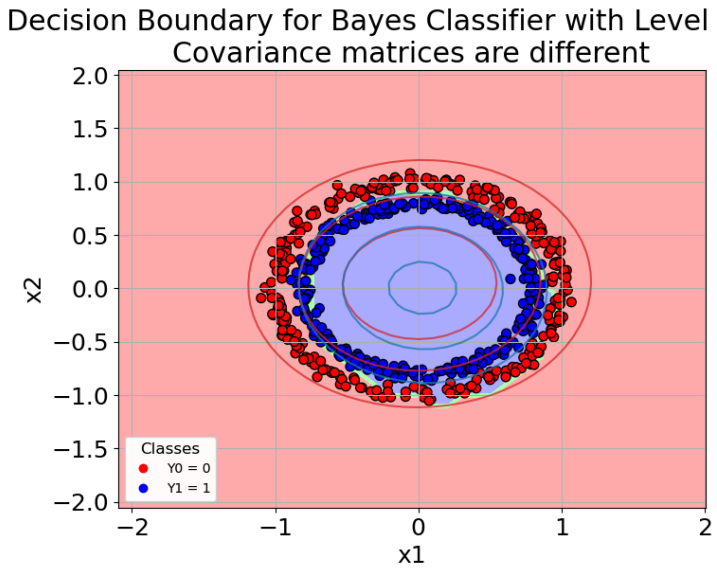
Confusion Matrix for the best configuration of the model, on training data and test data:



Decision Region Plot:



Decision Region Plot with Level curves on Training Data:



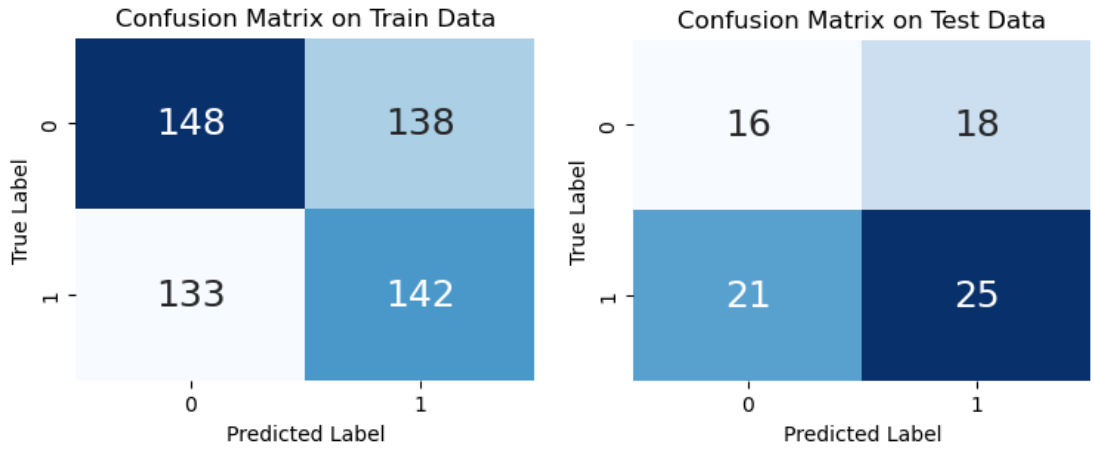
4. Naive-Bayes classifier with a Gaussian distribution for every class:

a. Covariance matrices for all the classes are the same

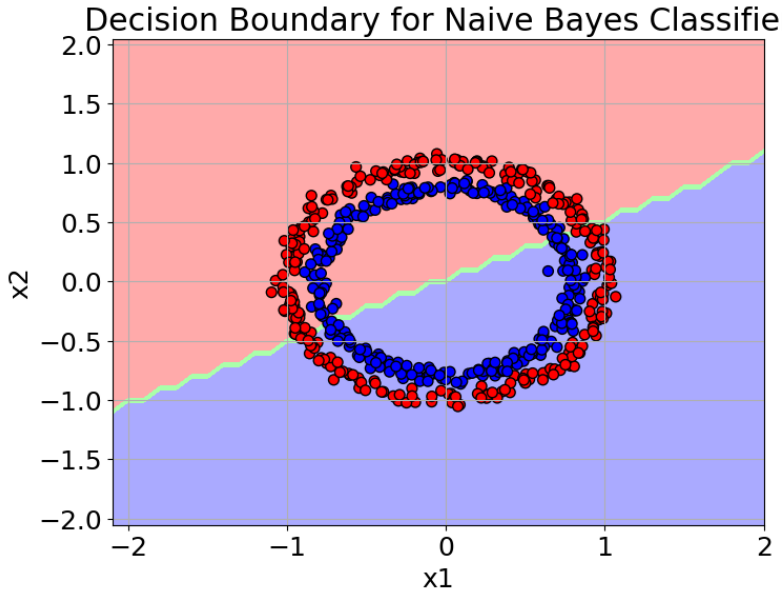
Table of Classification Accuracies of the Model for Training data, Validation data, and Test data:

Dataset	Size	Accuracy
Train	561	0.516934
Validation	159	0.427673
Test	80	0.512500

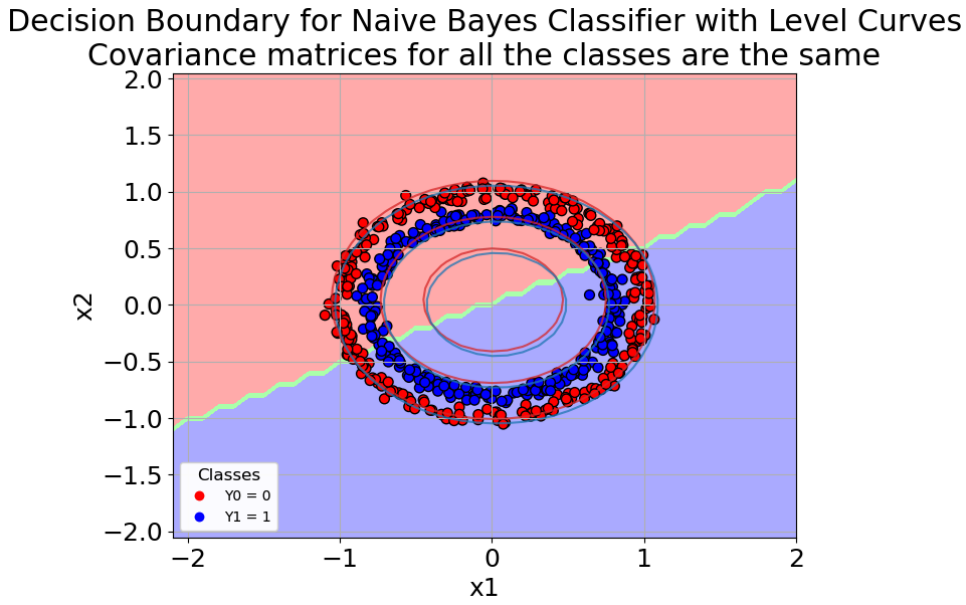
Confusion Matrix for the best configuration of the model, on training data and test data:



Decision Region Plot:



Decision Region Plot with Level curves on Training Data:

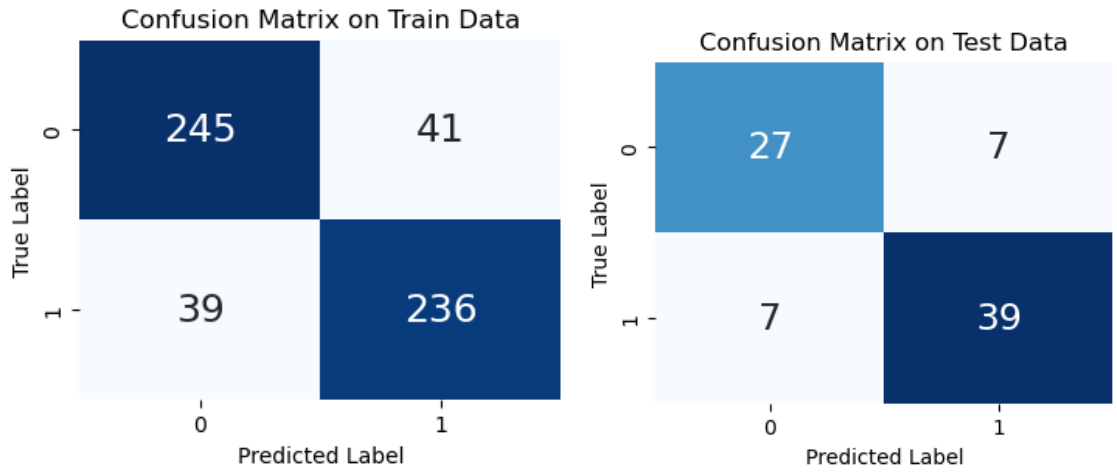


b. Covariance matrices are different

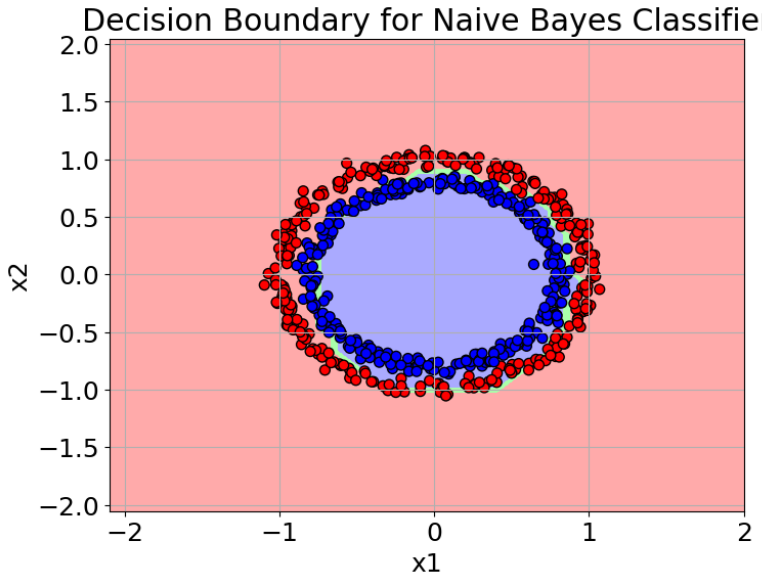
Table of Classification Accuracies of the Model for Training data, Validation data, and Test data:

Dataset	Size	Accuracy
Train	561	0.857398
Validation	159	0.773585
Test	80	0.825000

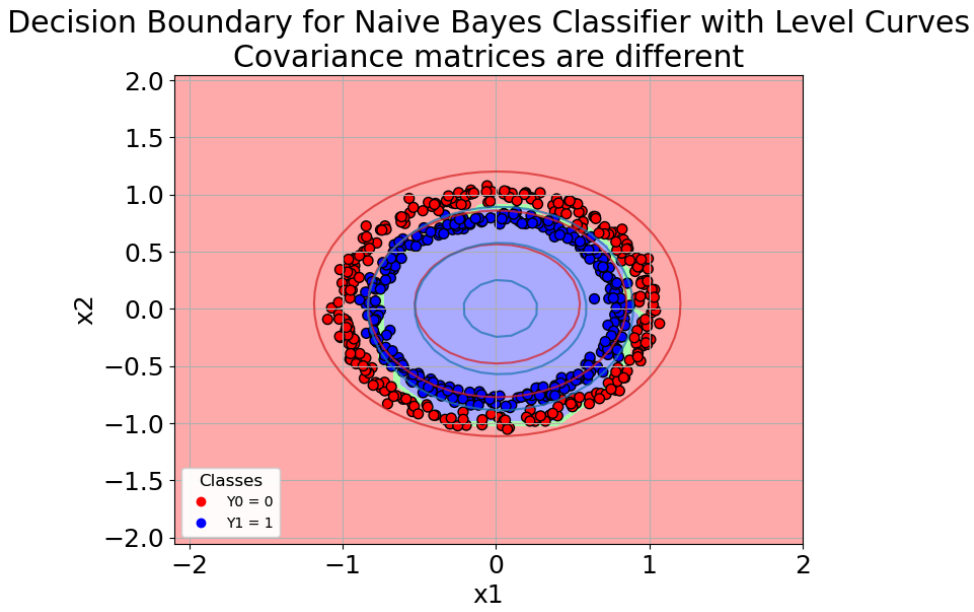
Confusion Matrix for the best configuration of the model, on training data and test data:



Decision Region Plot:



Decision Region Plot with Level curves on Training Data:



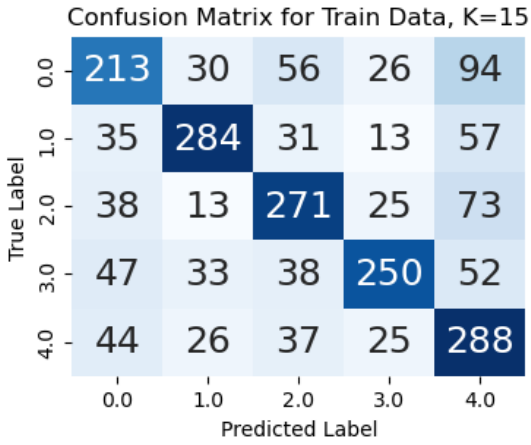
Dataset 3:

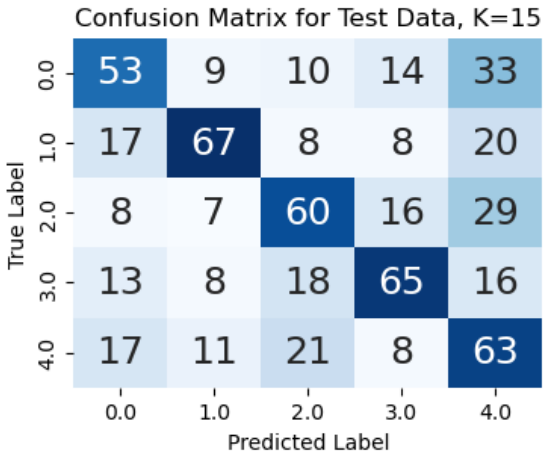
1. K-nearest neighbours classifier, for K=1, K=9 and K=15

Table of Classification Accuracies of the Model for Training data, Validation data, and Test data:

k-value	Accuracy train	Accuracy Val	Accuracy test
1	1.000000	0.387960	0.405676
9	0.669366	0.418060	0.490818
15	0.622201	0.421405	0.514190

Confusion Matrix for the best configuration of the model, on training data and test data:



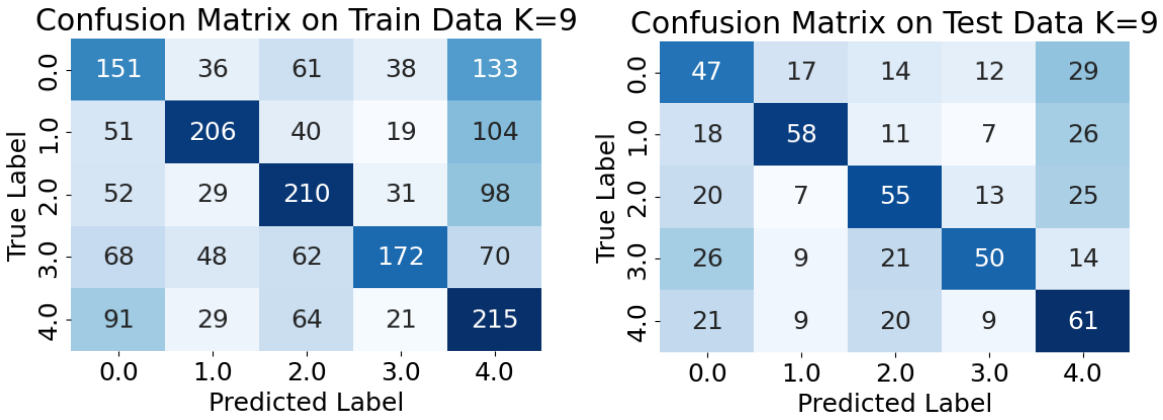


2. K-nearest representatives classifier, for K=1, K=5 and K=9: Use 10 representatives per class

Table of Classification Accuracies of the Model for Training data, Validation data, and Test data:

k-value	Accuracy Train	Accuracy Val	Accuracy Test
1	0.597427	0.384615	0.485810
5	0.501191	0.444816	0.479132
9	0.454502	0.481605	0.452421

Confusion Matrix for the best configuration of the model, on training data and test data:

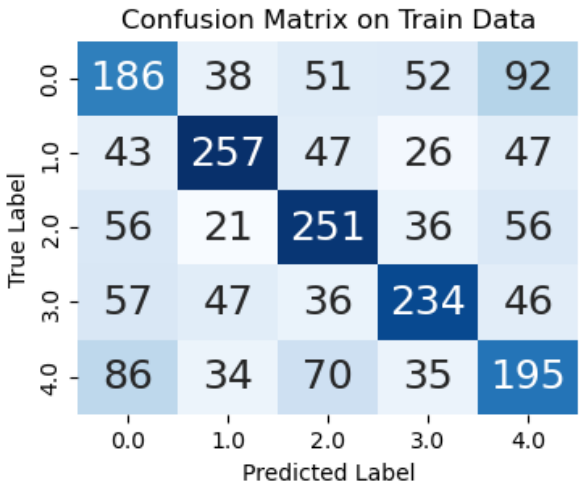


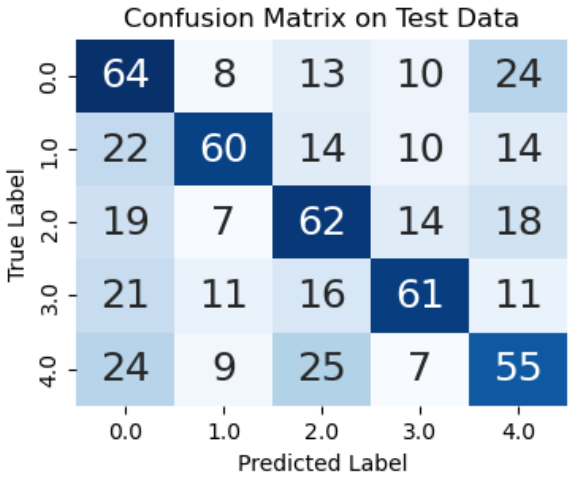
3. Bayes classifier with a Gaussian distribution for every class

Table of Classification Accuracies of the Model for Training data, Validation data, and Test data:

Dataset	Size	Accuracy
Train	2099	0.535017
Validation	299	0.451505
Test	599	0.504174

Confusion Matrix for the best configuration of the model, on training data and test data:





4. Naive-Bayes classifier with a Gaussian distribution for every class

Table of Classification Accuracies of the Model for Training data, Validation data, and Test data:

Index	Type	Count	Score
0	Train	2099	0.476894
1	Validation	299	0.461538
2	Test	599	0.494157

Confusion Matrix for the best configuration of the model, on training data and test data: