ASSIGNMENT 2

CS5691 Pattern Recognition and Machine Learning

CS5691 Assignment 2

Team No. 26

Team Members:

CS24M020	Grishma Karekar
CS24M046	Shreyas Rathod

Indian institute of Technology, Madras

1. DATASET 1

1.1. K-nearest Neighbors classifier:

1.1.1. Mathematical Formulation

The K Nearest Neighbor is a statistically non-parametric model that can be used for regression as well as for classification. It assumes that similar things exist in close proximity. Crucial steps in a K-Nearest Neighbor classifier are:

- A distance metric is first specified, the most commonly used metric is the Euclidean distance:

$$d = ||\overrightarrow{x_1} - \overrightarrow{x_2}||$$

where ||.|| denotes the norm function. Other commonly used distance metrics are the Manhattan distance and Cosine similarity. For our application, Euclidean distance is used.

1.1.2. Model performance across different k

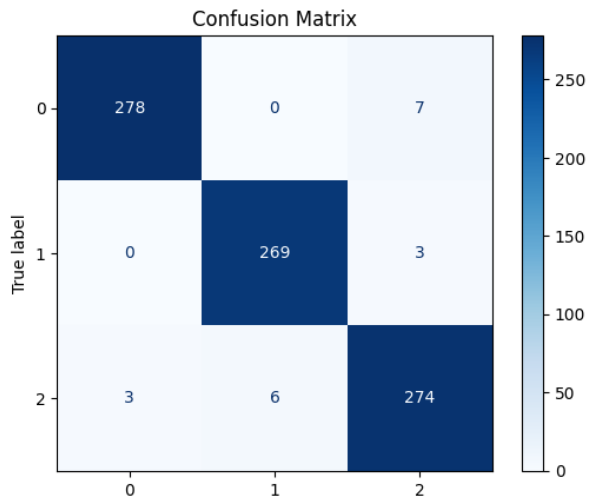
The Model performance is best for value k = 5

The accuracy table and confusion matrix are:

k-value	Train Accuracy	Validation Accuracy	Test Accuracy
1	100	97.5000	96.6667
5	97.7381	99.5833	96.6667
9	97.8571	99.5833	98.3333

***Table 1**: Accuracy table for Dataset 1 – KNN classifier*

Training Confusion Matrix



Testing Confusion Matrix

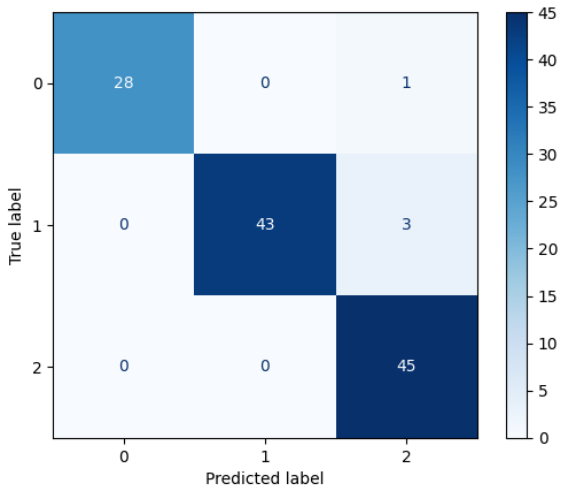


***Figure 1**: Confusion matrix for k = 5, train and test dataset on left and right respectively*

1.1.3. Decision region plots

For Training data ***For Testing data***

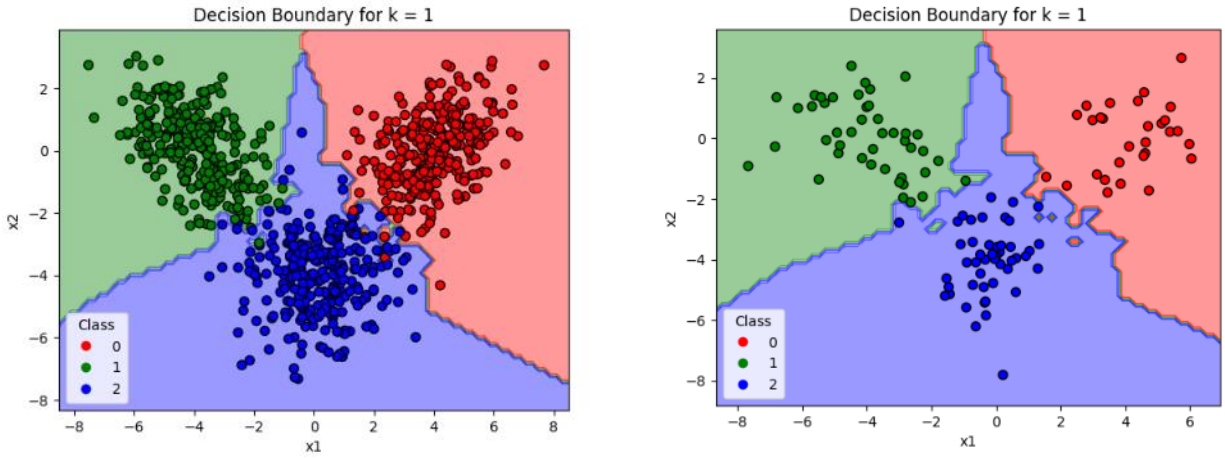


***Figure 2**: Decision region plot for k = 1, superimposed with train and test dataset on left and right respectively*

For Training data ***For Testing data***

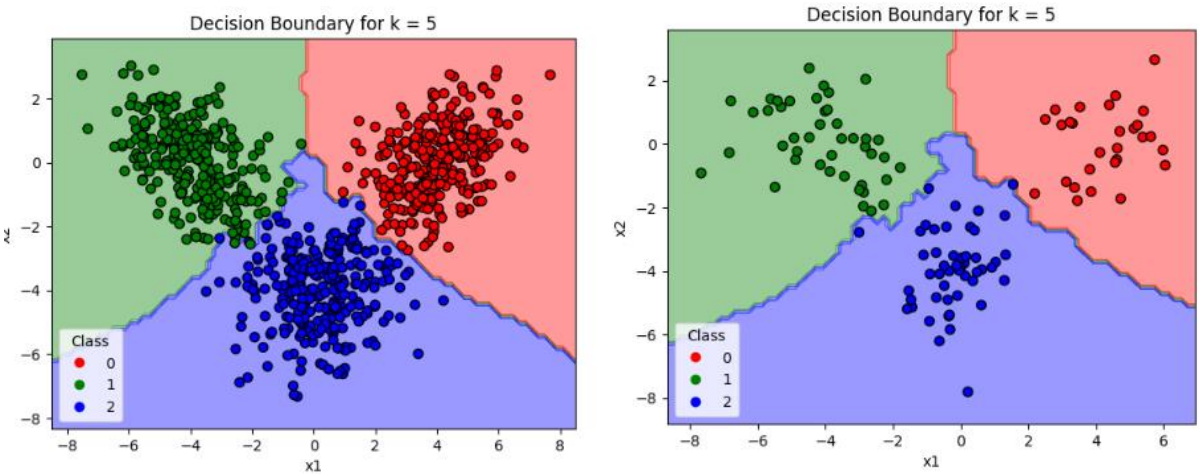


***Figure 3**: Decision region plot for k = 5, superimposed with train and test dataset on left and right respectively*

For Training data *For Testing data*

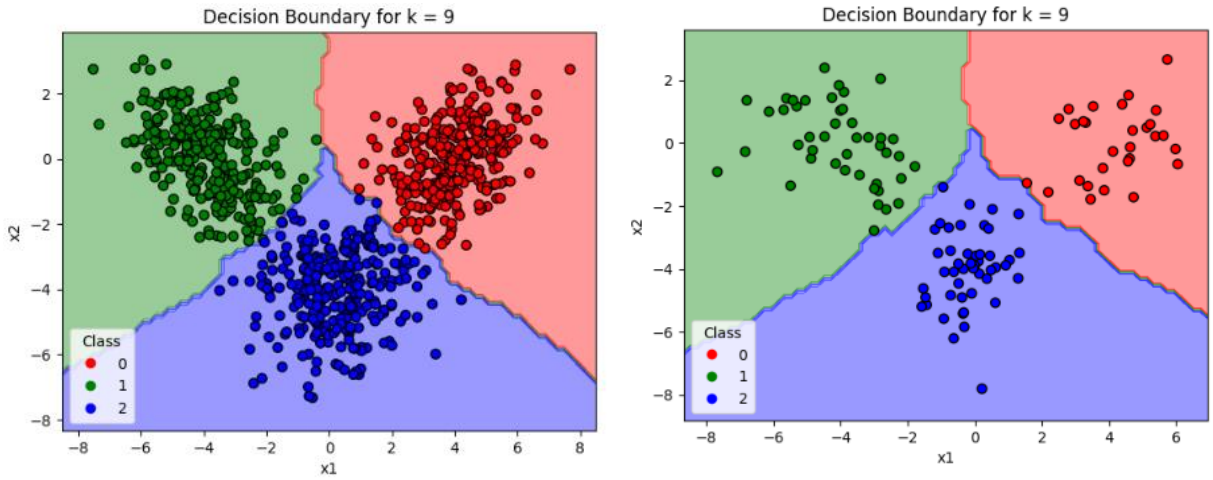


***Figure 4**: Decision region plot for k = 9, superimposed with train and test dataset on left and right respectively*

1.2. Bayes Classifier with a Gaussian distribution for every class:

1.2.1. Accuracy table and Confusion Matrix

The accuracy table and confusion matrix are:

Condition	Train Accuracy	Validation Accuracy	Test Accuracy
$C_i = C_j$	97.8571	98.3333	99.1667
$Ci \neq C_j$	98.0952	98.3333	100

***Table 2**: Accuracy table for Dataset 1: Bayes Classifier*

Training Confusion Matrix *Testing Confusion Matrix*

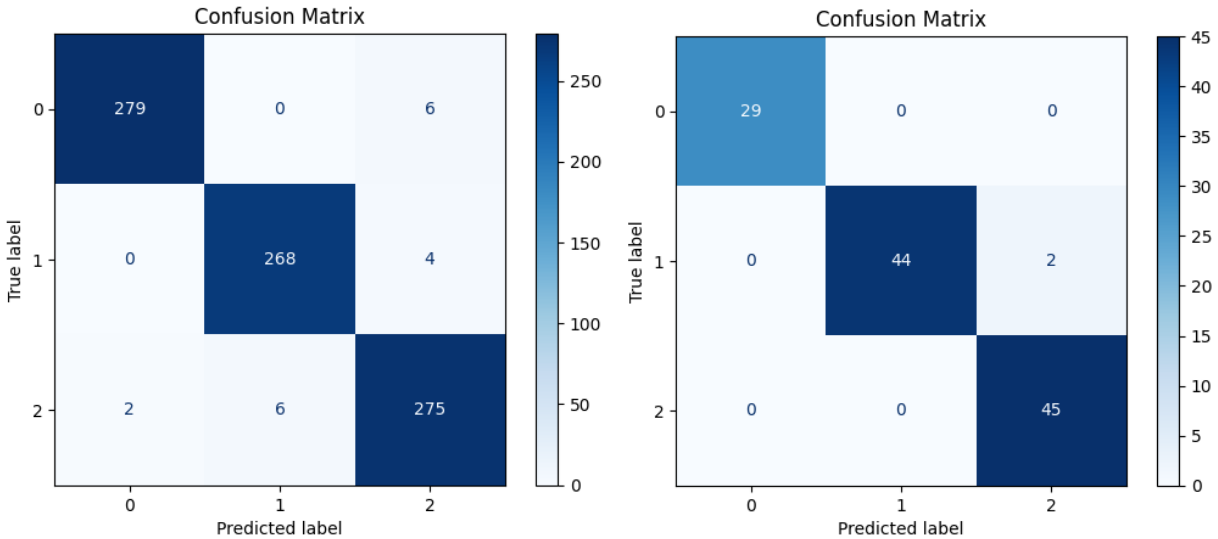


***Figure 5**: Confusion matrix for Gaussian Bayes classifier with same covariance matrix, train and test dataset on left and right respectively*

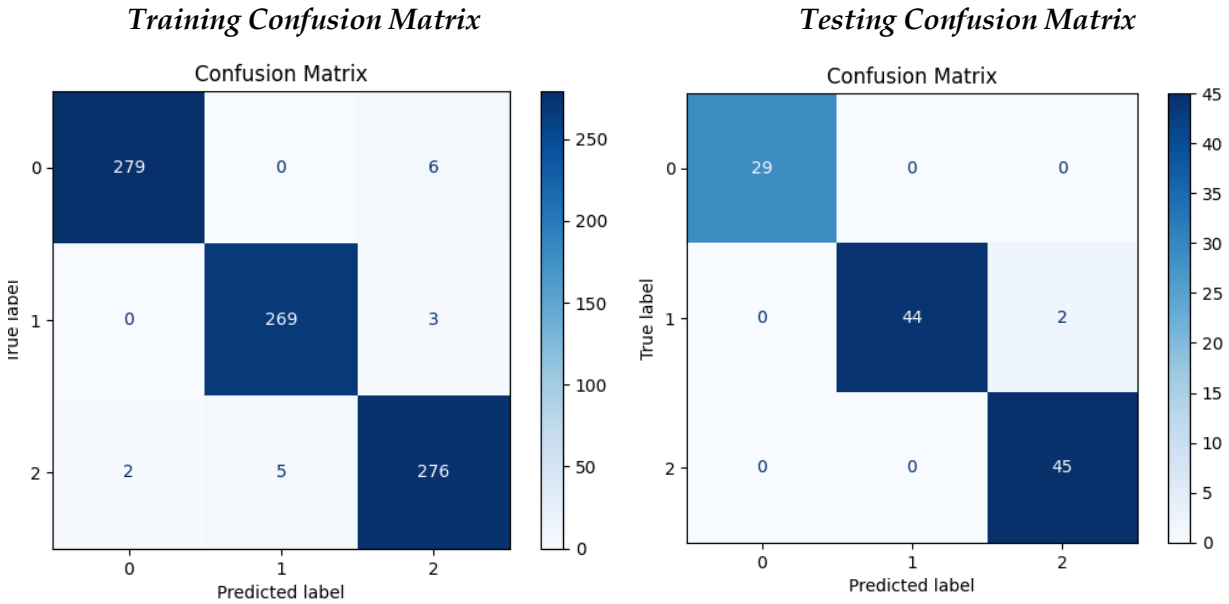


***Figure 6**: Confusion matrix for Gaussian Bayes classifier with different covariance matrix, train and test dataset on left and right respectively*

1.2.2. Decision Boundary Plot for Gaussian Bayes Classifier

- **Decision Boundaries + Same Covariance**:

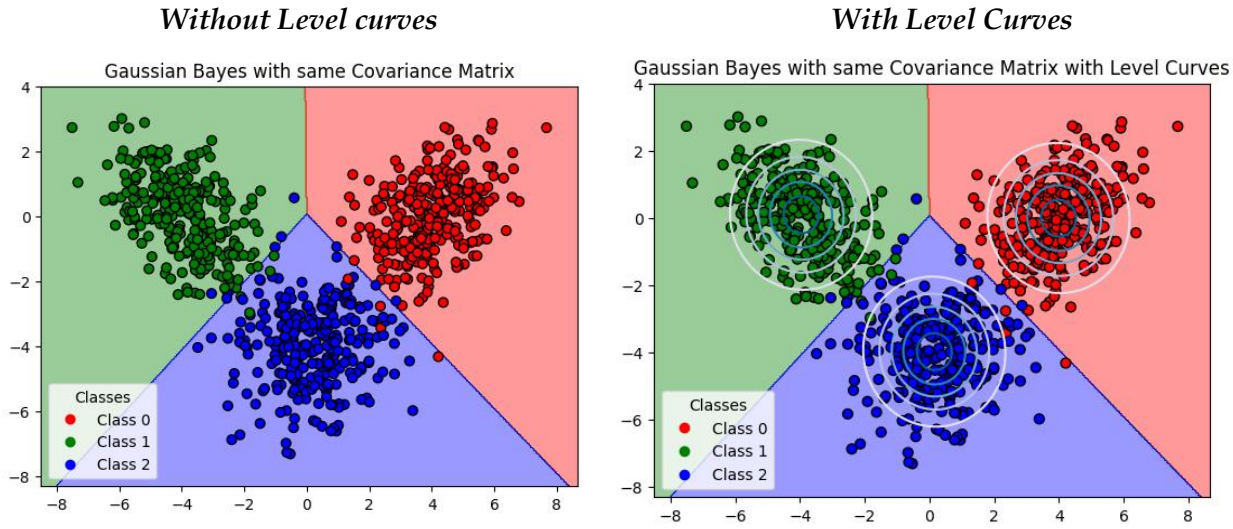


***Figure 7**: Decision Boundary Plot for Gaussian Bayes classifier with same covariance matrix, without and with level curves on left and right respectively*

- **Decision Boundaries + Different Covariance**:

Without Level curves *With Level Curves*

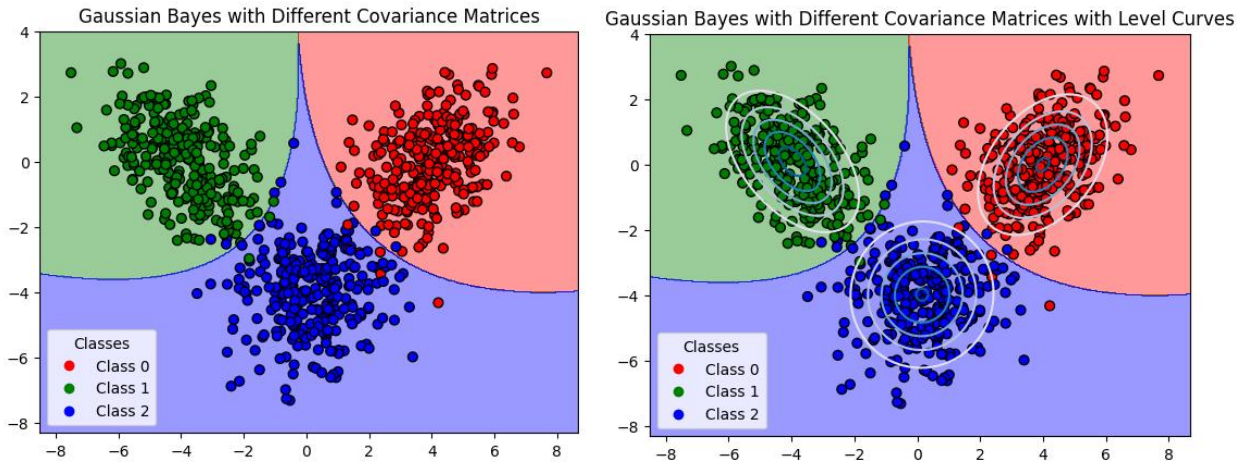


***Figure 8**: Decision Boundary Plot for Gaussian Bayes classifier with different covariance matrix, without and with level curves on left and right respectively*

2. DATASET 2

2.1. K-nearest Neighbors classifier:

2.1.1. Mathematical Formulation

Similar to subsubsection 1.1.1., K nearest neighbor classifier is used to predict class labels for dataset 2.

2.1.2. Model performance across k

The accuracy table is as follows:

k-value	Train Accuracy	Validation Accuracy	Test Accuracy
1	100	100	100
5	100	100	100
9	100	100	100

***Table 3**: Accuracy table for dataset 2 – KNN classifier*

Since model performance is best irrespective of k, the accuracy table, confusion matrix and decision boundary plot are all evaluated using k = 1 as to minimize the run time.

Training Confusion Matrix ***Testing Confusion Matrix***

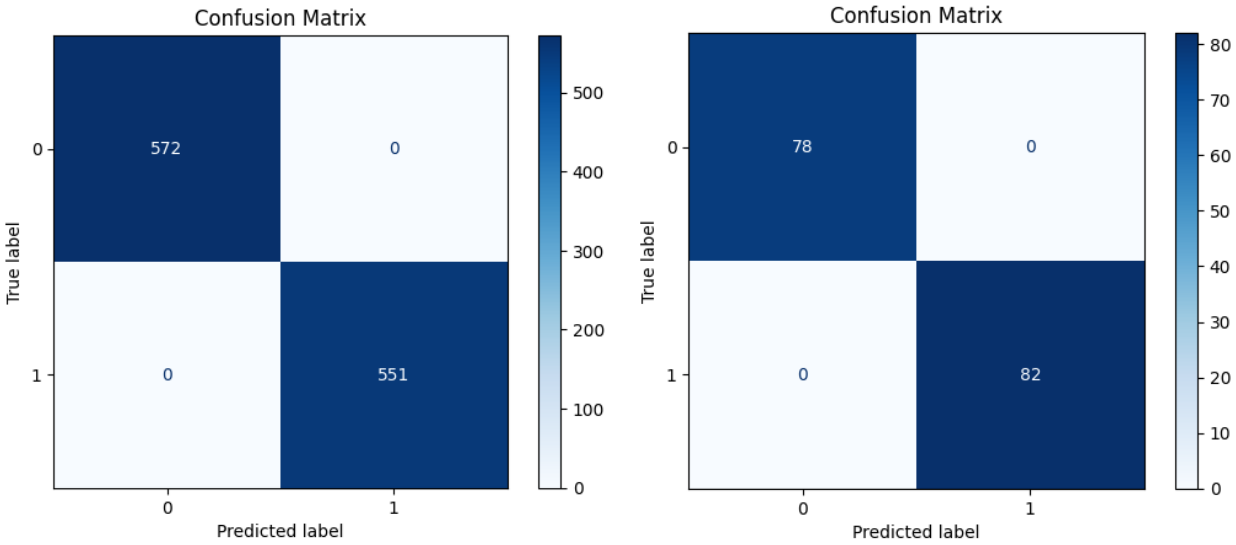


***Figure 9**: Confusion matrix for k = 1, train and test dataset on left and right respectively*

2.1.3. Decision Region Plot

For Training data ***For Testing data***

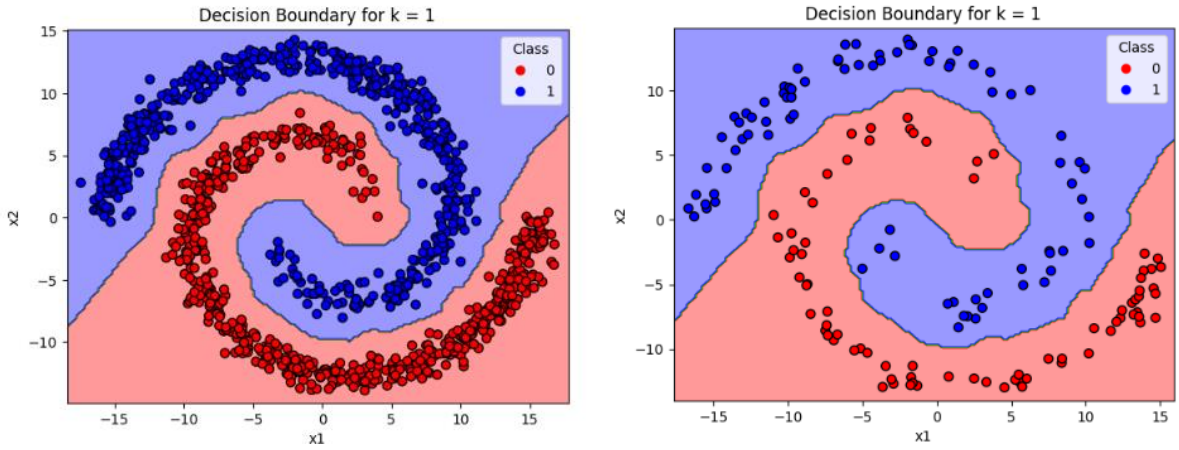


***Figure 10**: Decision region plot for k = 1, superimposed with train and test dataset on left and right respectively*

For Training data ***For Testing data***

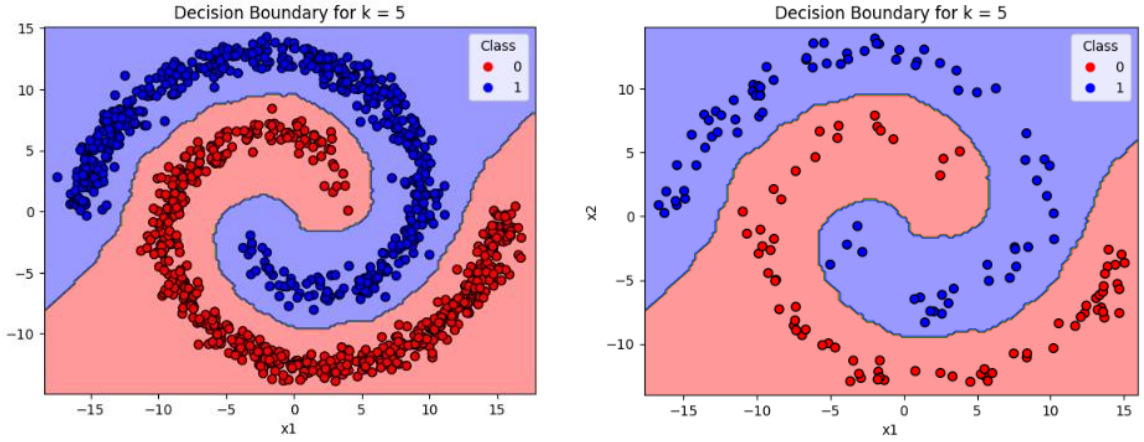


***Figure 11**: Decision region plot for k = 5, superimposed with train and test dataset on left and right respectively*

For Training data ***For Testing data***

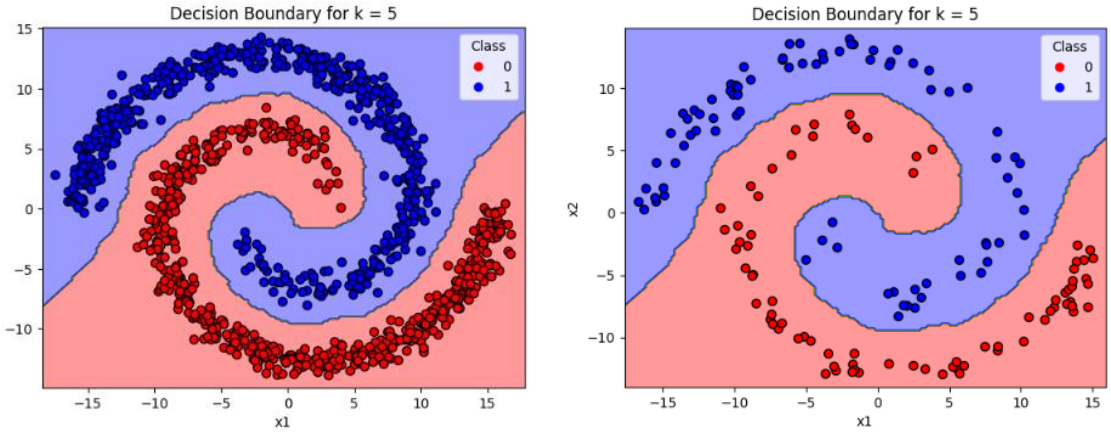


***Figure 12**: Decision region plot for k = 9, superimposed with train and test dataset on left and right respectively*

2.2. K-nearest representatives' classifier: (10 representative per class)

2.2.1. Model performance across k

The accuracy table is as follows:

k-value	Train Accuracy	Validation Accuracy	Test Accuracy
1	99.9110	99.3691	99.3750
3	95.7257	94.0063	94.3750
5	84.4167	84.5426	82.5000

***Table 4**: Accuracy table for dataset 2 – KNR classifier*

Best Model is when k = 1, Confusion matrix and decision boundary plot are all evaluated using k = 1 as to minimize the run time.

Training Confusion Matrix ***Testing Confusion Matrix***

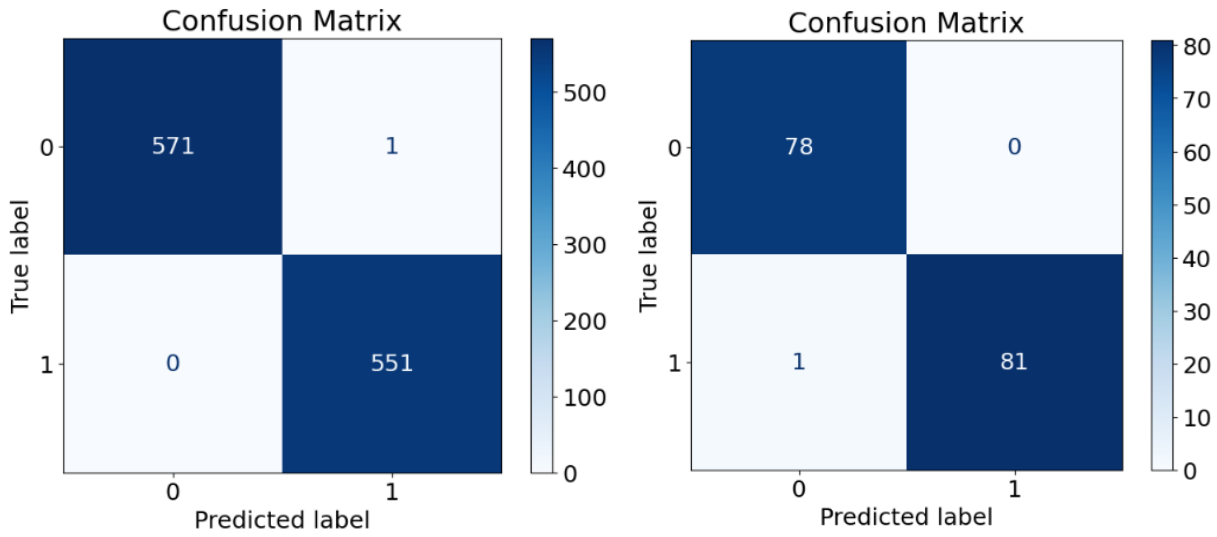


***Figure 13**: Confusion matrix for k = 1, train and test dataset on left and right respectively*

2.2.2. Decision Boundary Plot:

For Training data ***For Testing data***

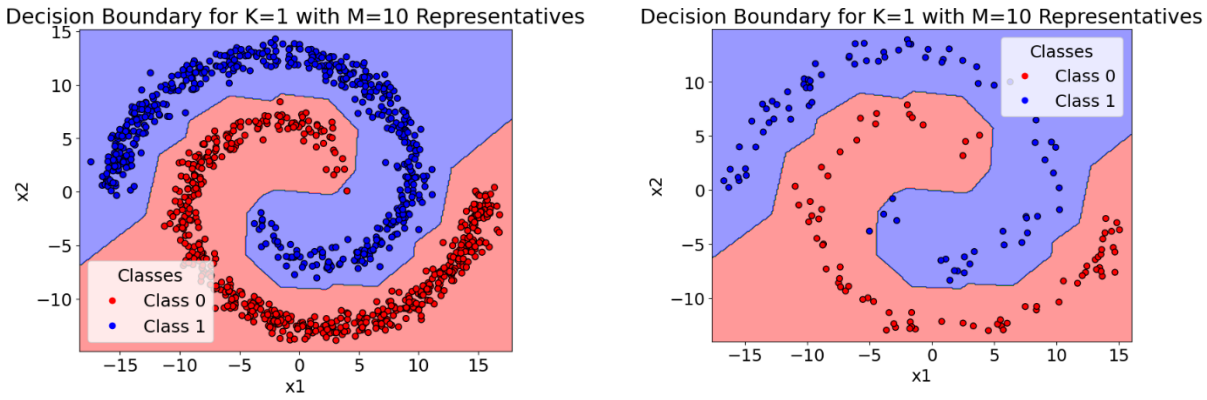


***Figure 14**: Decision region plot for k = 1, superimposed with train and test dataset on left and right respectively*

For Training data ***For Testing data***

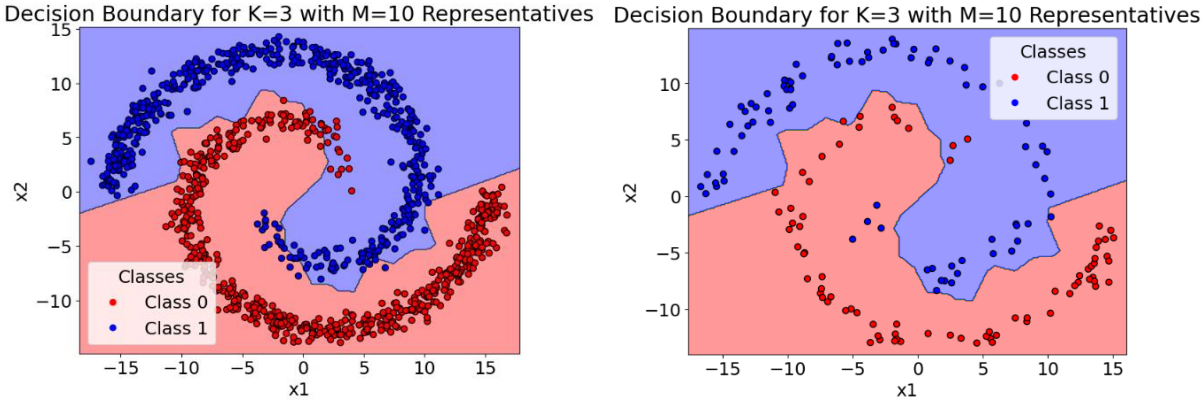


***Figure 15**: Decision region plot for k = 3, superimposed with train and test dataset on left and right respectively*

For Training data ***For Testing data***

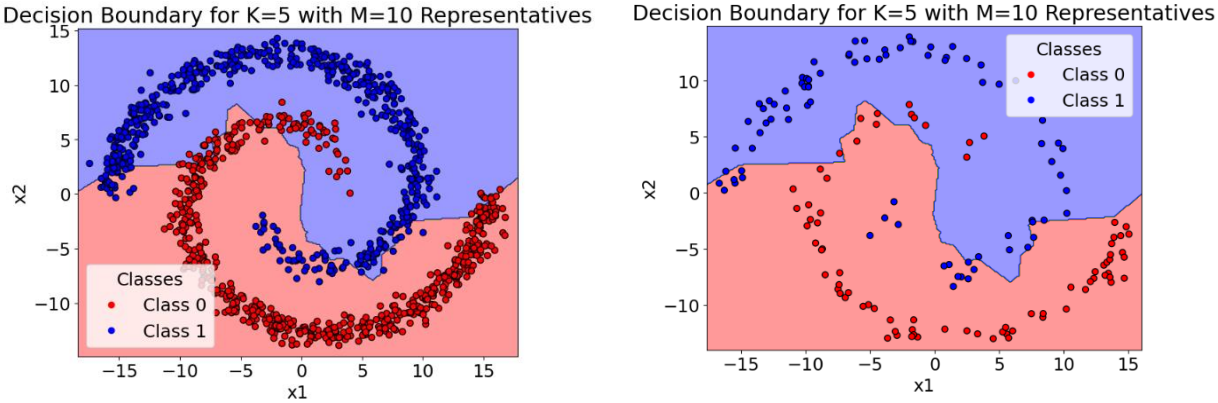


***Figure 16**: Decision region plot for k = 5, superimposed with train and test dataset on left and right respectively*

2.3. Bayes Classifier with a Gaussian Distribution for every Class

2.3.1. Accuracy table and Confusion Matrix

The accuracy table and confusion matrix are:

Condition	Train Accuracy	Validation Accuracy	Test Accuracy
$C_i = C_j$	76.5806	76.3407	76.8750
$Ci \neq C_j$	76.4915	76.3407	76.8750

***Table 5**: Accuracy table for Dataset 2: Bayes Classifier*

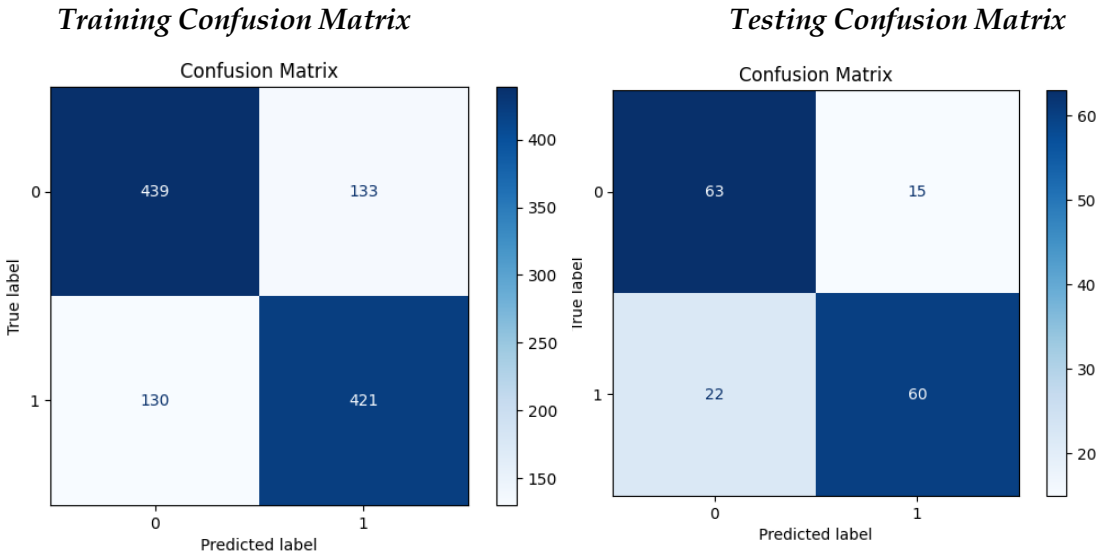


***Figure 17**: Confusion matrix for Gaussian Bayes classifier with same covariance matrix, train and test dataset on left and right respectively*

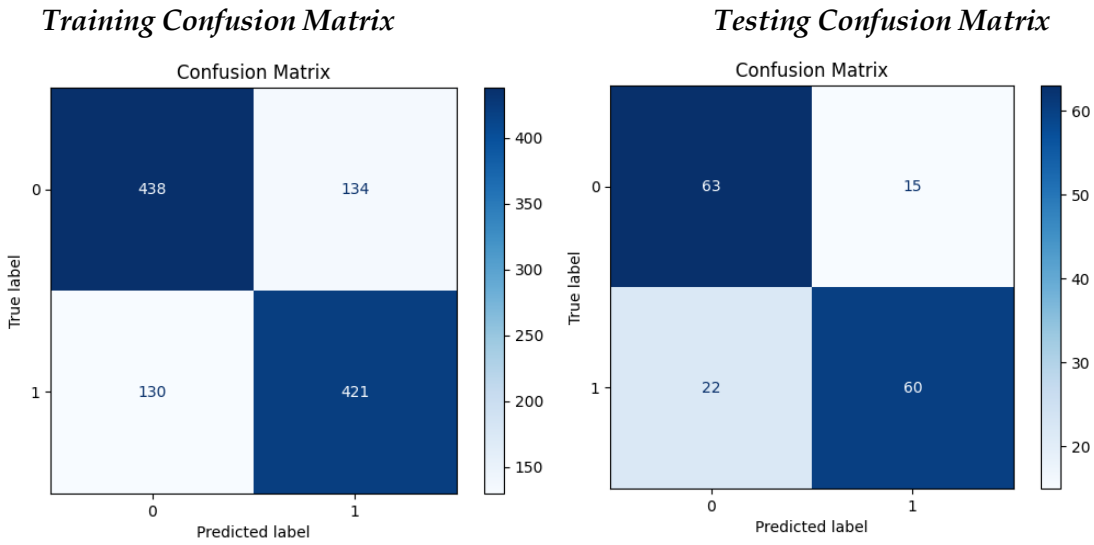


***Figure 18**: Confusion matrix for Gaussian Bayes classifier with different covariance matrix, train and test dataset on left and right respectively*

2.3.2. Decision Boundary Plot for Gaussian Bayes Classifier

- **Decision Boundaries + Same Covariance**:

Without Level curves *With Level Curves*

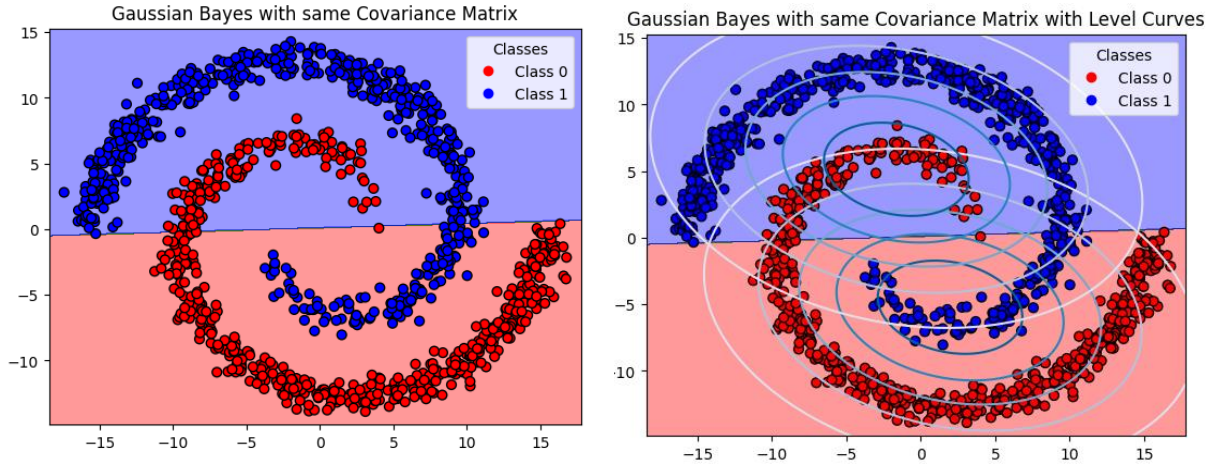


***Figure 19**: Decision Boundary Plot for Gaussian Bayes classifier with same covariance matrix, without and with level curves on left and right respectively*

- **Decision Boundaries + Different Covariance**:

Without Level curves *With Level Curves*

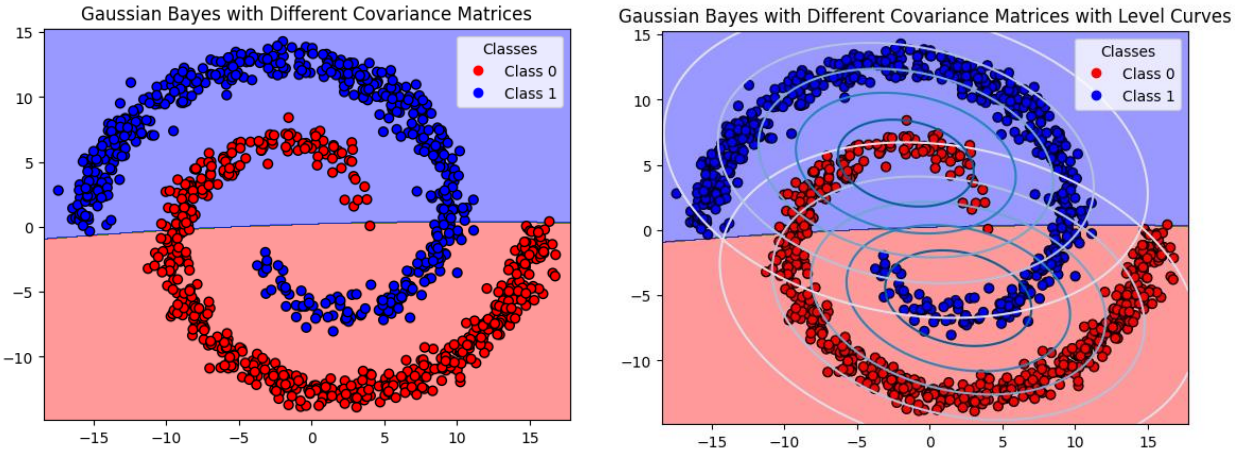


***Figure 20**: Decision Boundary Plot for Gaussian Bayes classifier with different covariance matrix, without and with level curves on left and right respectively*

2.4. Naïve-Bayes Classifier with a Gaussian Distribution for every class

2.4.1. Accuracy table and Confusion Matrix

The accuracy table and confusion matrix are:

Condition	Train Accuracy	Validation Accuracy	Test Accuracy
$C_i = C_j$	74.3544	73.1861	75
$Ci \neq C_j$	74.3544	76.5016	75

***Table 6**: Accuracy table for Dataset 2: Bayes Classifier*

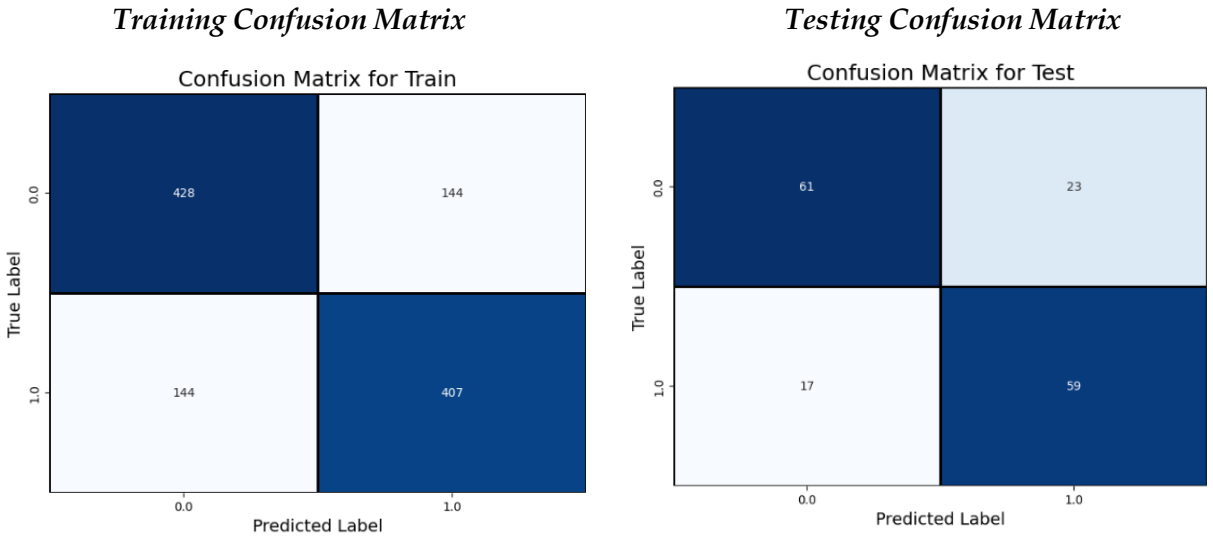


***Figure 21**: Confusion matrix for Naive Bayes classifier with same covariance matrix, train and test dataset on left and right respectively*

***Figure 22**: Confusion matrix for Naive Bayes classifier with different covariance matrix, train and test dataset on left and right respectively*

2.4.2. Decision Boundary Plot for Naive Bayes Classifier

- **Decision Boundaries + Same Covariance**:

Without Level curves ***With Level Curves***

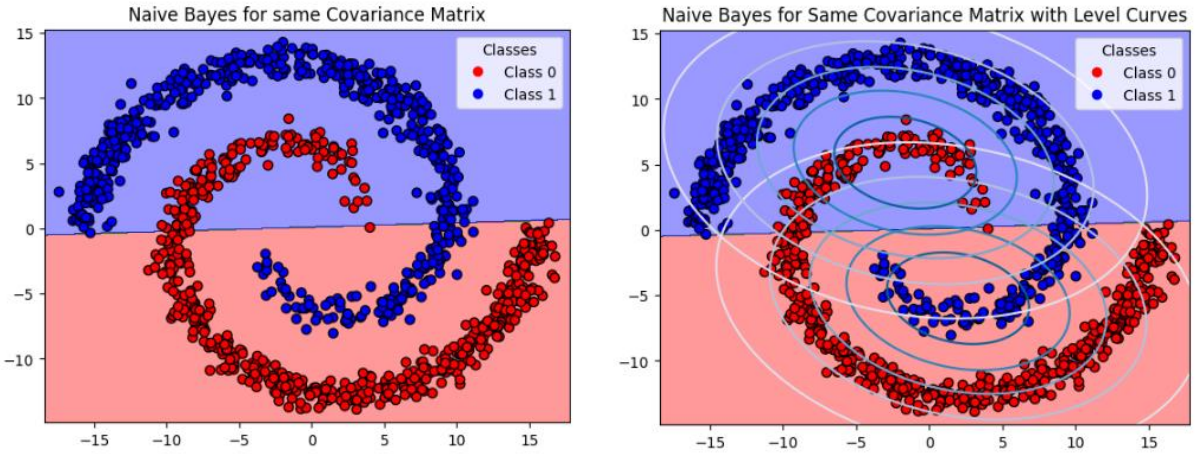


***Figure 23**: Decision Boundary Plot for Naive Bayes classifier with same covariance matrix, without and with level curves on left and right respectively*

- **Decision Boundaries + Different Covariance**:

Without Level curves ***With Level Curves***

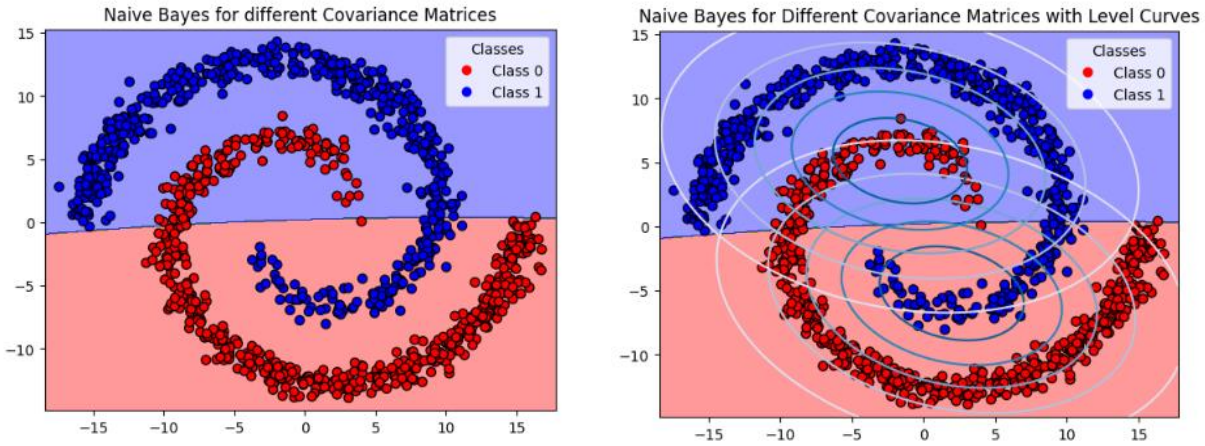


***Figure 24**: Decision Boundary Plot for Naive Bayes classifier with different covariance matrix, without and with level curves on left and right respectively*

3. DATASET 3

3.1. K-nearest Neighbors classifier:

3.1.1. Mathematical Formulation

Similar to subsubsection 1.1.1., K nearest neighbor classifier is used to predict class labels for dataset 3.

3.1.2. Accuracy table and Confusion Matrix

The Model performance is best for value k = 15

The accuracy table and confusion matrix are:

k-value	Train Accuracy	Validation Accuracy	Test Accuracy
1	100	52	50.5000
9	71.9048	55.6666	54.3336
15	66.6190	56	56.6666

***Table 7**: Accuracy table for Dataset 3 – KNN classifier*

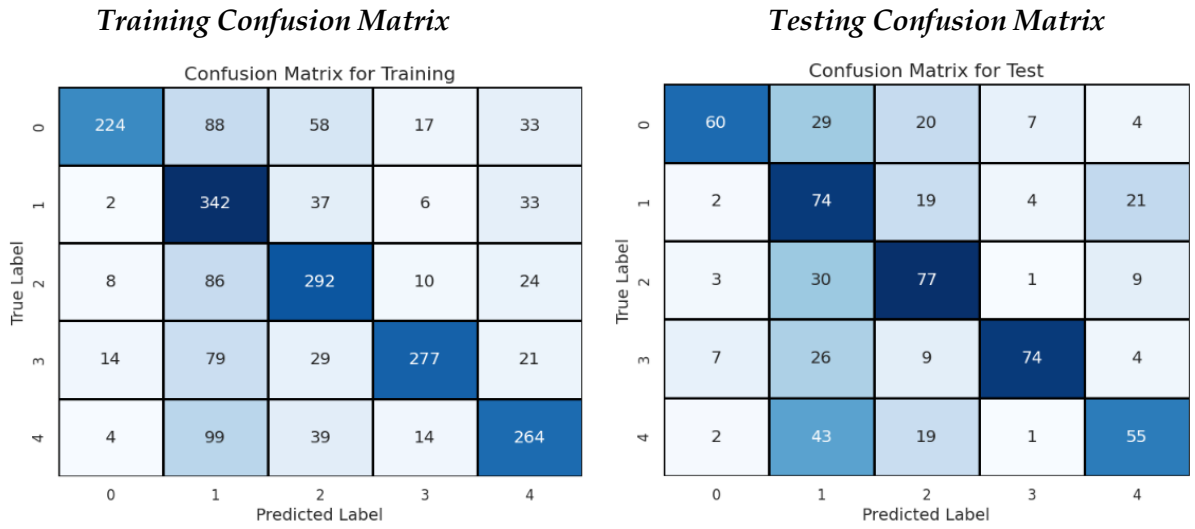


***Figure 25**: Confusion matrix for k = 15, train and test dataset on left and right respectively*

3.2. K-nearest representatives' classifier: (10 representative per class)

3.2.1. Accuracy table and Confusion Matrix

The Model performance is best for value k = 5, M = 10

The accuracy table and confusion matrix are:

k-value	Train Accuracy	Validation Accuracy	Test Accuracy
1	67.9371	54.1806	57.5960
5	59.7427	59.5318	59.9332
9	57.1701	54.1806	55.9265

***Table 8**: Accuracy table for Dataset 3 – KNR classifier*

Training Confusion Matrix ***Testing Confusion Matrix***

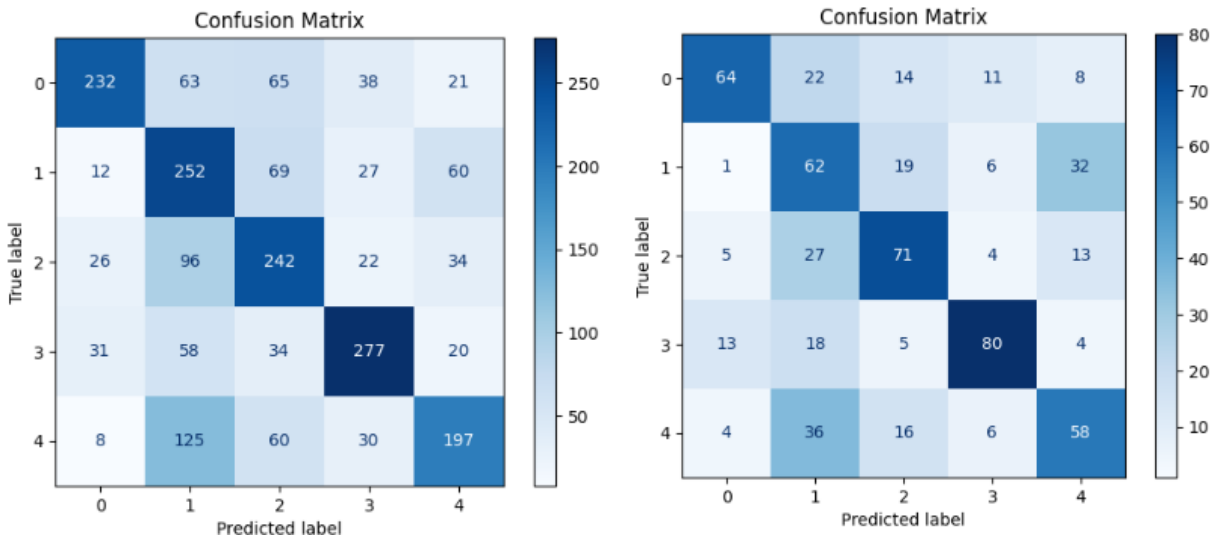


***Figure 26**: Confusion matrix for k = 15, train and test dataset on left and right respectively*

3.3. Bayes Classifier with a Gaussian Distribution for every Class

3.3.1. Accuracy Table and Confusion Matrix

Sr. No.	Train Accuracy	Validation Accuracy	Test Accuracy
1	64.7451	50.5016	49.5826

***Table 9**: Accuracy table for Dataset 3 – Bayes classifier*

Training Confusion Matrix ***Testing Confusion Matrix***

***Figure 27**: Confusion matrix of train and test dataset on left and right respectively*

3.4. Naïve-Bayes Classifier with a Gaussian Distribution for every class

3.4.1. Accuracy table and Confusion Matrix

The accuracy table and confusion matrix are:

Condition	Train Accuracy	Validation Accuracy	Test Accuracy
$C_i = C_j$	56.3130	58.1940	56.0930
$Ci \neq C_j$	59.1230	57.1910	57.4290

***Table 10**: Accuracy table for Dataset 3: Naïve Bayes Classifier*

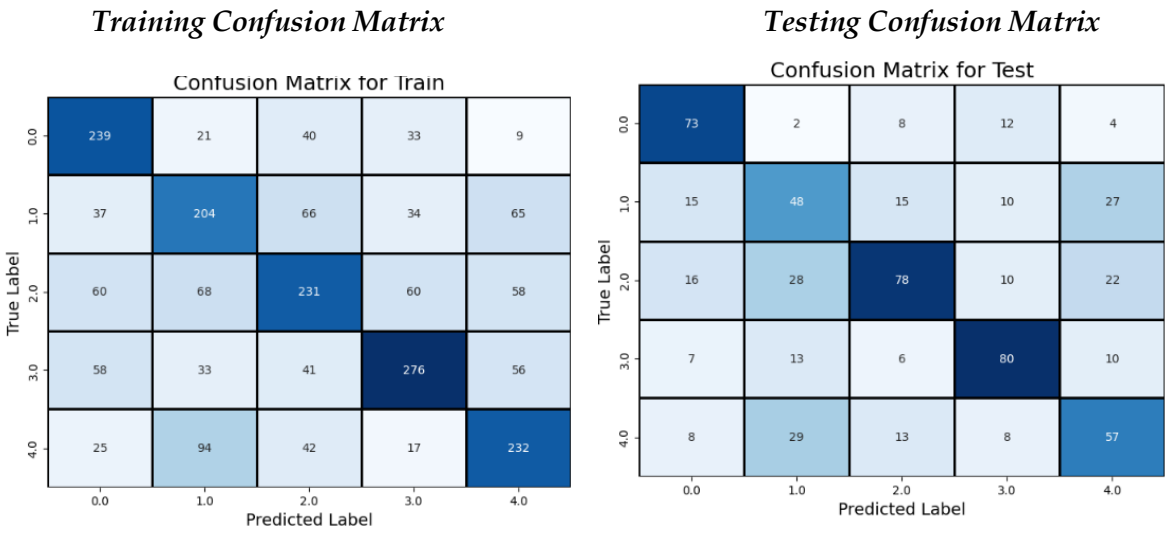


***Figure 21**: Confusion matrix for Naive Bayes classifier <u>with same covariance matrix</u>, train and test dataset on left and right respectively*

Training Confusion Matrix ***Testing Confusion Matrix***

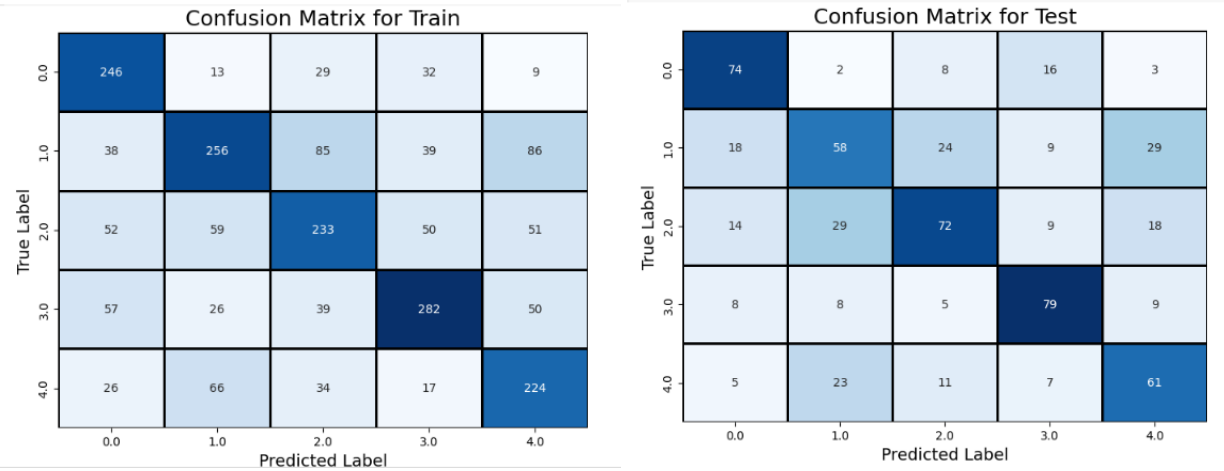


Figure 22*: Confusion matrix for Naive Bayes classifier with different covariance matrix, train and test dataset on left and right respectively*