॥ सा विद्या या विमुक्तये ॥

INDIAN INSTITUTE OF TECHNOLOGY
DHARWAD

Group:13

CS-612:Lab Report

Assignment 1 - Bayes Classifier

Submitted To:
Prof. Dileep A.D.
Department of Computer Science and Engineering

Submitted By :
Surya Narayan Upadhyay - EE22BT052
Adarsh Gupta - MC22BT001
Krunal Patel - IS22BM022

1 Introduction

In this assignment, we used Bayes Classifier using different types of covariance matrix for each dataset provided. We divided the dataset into train and test used the covariance matrix obtained in training data to predict the class for test data using the Decision Boundary equation.

2 Bayes Classifier Types

Bayes classifiers are a family of probabilistic classifiers based on Bayes' theorem. The type of covariance matrix used in the classifier significantly affects its decision boundary and performance. Below, we describe four types of Bayes classifiers based on different assumptions about the covariance matrix.

2.1 Covariance Matrix: $\sigma^2 I$ (Same for All Classes)

In this case, the covariance matrix for all the classes is the same and is a scaled identity matrix $\sigma^2 I$. The diagonal elements are equal, and all off-diagonal elements are zero. The same variance σ^2 is obtained by averaging the variances of all classes. This model assumes that the features are uncorrelated and have the same variance across all classes.

2.2 Full Covariance Matrix: Σ (Same for All Classes)

Here, the covariance matrix for all classes is the same, but it is a full matrix Σ, which means it can have different values for both variances and covariances between features. The same covariance matrix Σ is obtained by averaging the covariance matrices of all classes. This model allows for correlations between features and results in linear decision boundaries.

2.3 Diagonal Covariance Matrix: Σ_k (Different for Each Class)

In this type, the covariance matrix is diagonal but different for each class, denoted by Σ_k. The diagonal elements represent the variance of each feature within a class, while all off-diagonal elements are zero, indicating no correlation between features. This model assumes that features are uncorrelated within each class but can have different variances across classes. The decision boundaries can vary significantly depending on the feature variances of each class.

2.4 Full Covariance Matrix: Σ_k (Different for Each Class)

In this case, the covariance matrix is fully different for each class, denoted by Σ_k. This means that each class has its own unique covariance structure, with both variances and covariances varying between features. This model allows for the most flexibility, as it can model complex relationships between features within each class.

Each of these classifiers makes different assumptions about the data, and the choice of which to use depends on the specific characteristics of the dataset being analyzed.

3 Data Set - Linear

3.1 Using Covariance Matrix Type 1 (σ^2I)

Provided below is the plot for the decision boundaries of each class for the linearly separable data using the covariance matrix of the type σ^2I.The accuracy on the linear data for the Bayes Classifier using the Decision Boundary formulated by the covariance of type 1 obtained by the training data is very high this suggests the Bayes classfier is the best for such linearly seperable data.
Since the covariance matrix in this case is same for all classes the discrimnant function in this case gives a linear decision boundary.
Since the covarince matrix used is generalized for all three classes and same is used to plot the decision region there is slight error in one of the classes. This is also reflected in the confusion matrix, precision, recall, accuracy and f-score calculated for the test data.

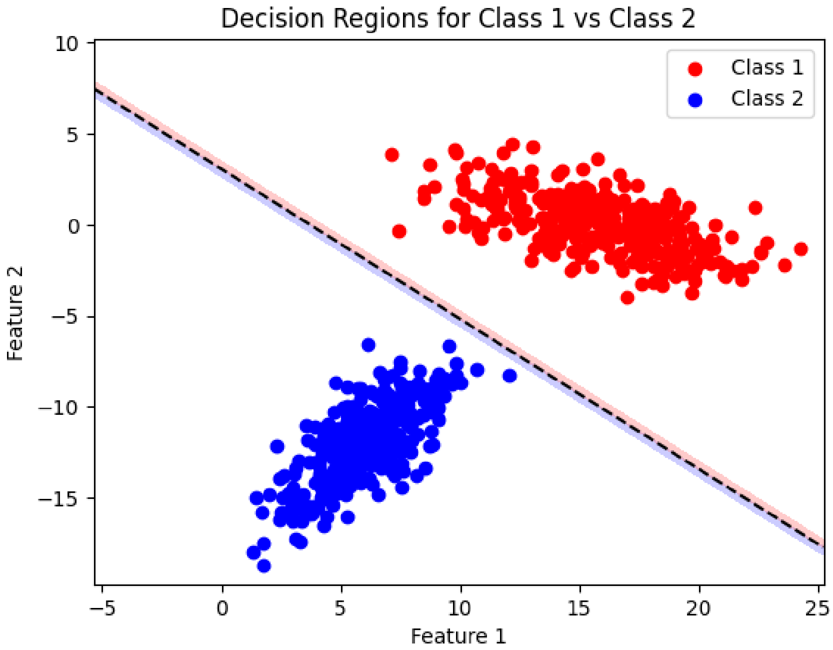


Figure 1: Decision Boundary for Class1 v/s Class2.

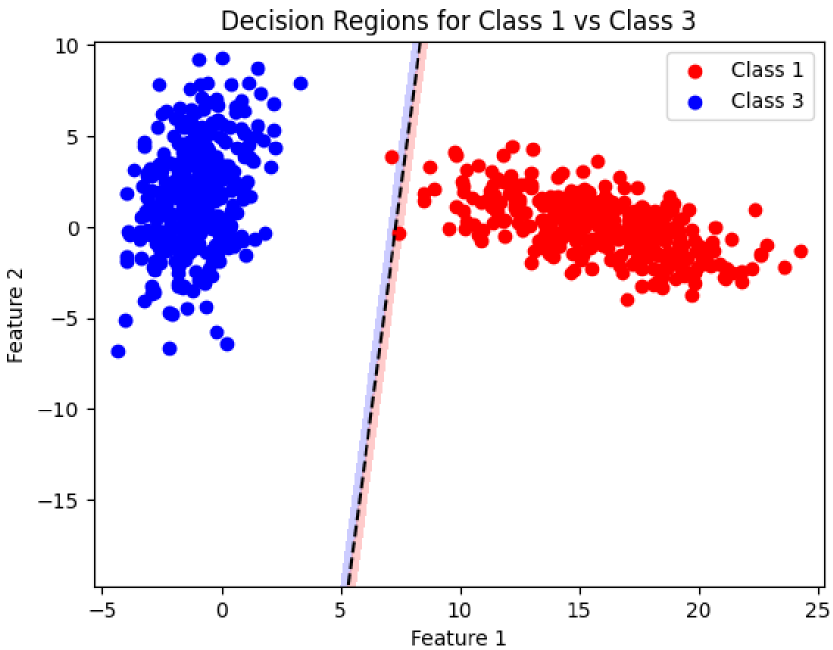


Figure 2: Decision Boundary plot for Class1 v/s Class3.

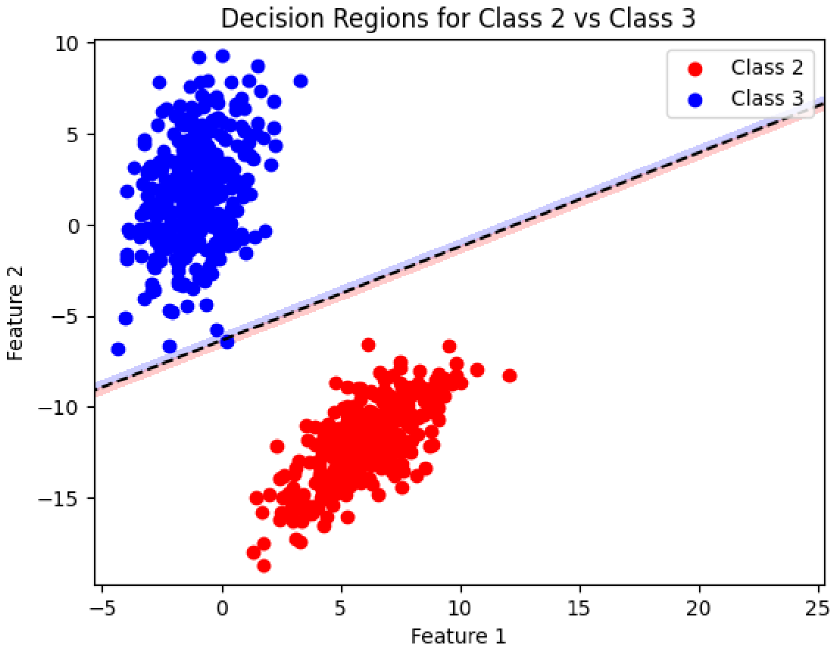


Figure 3: Decision Boundary plot for Class2 v/s Class3.

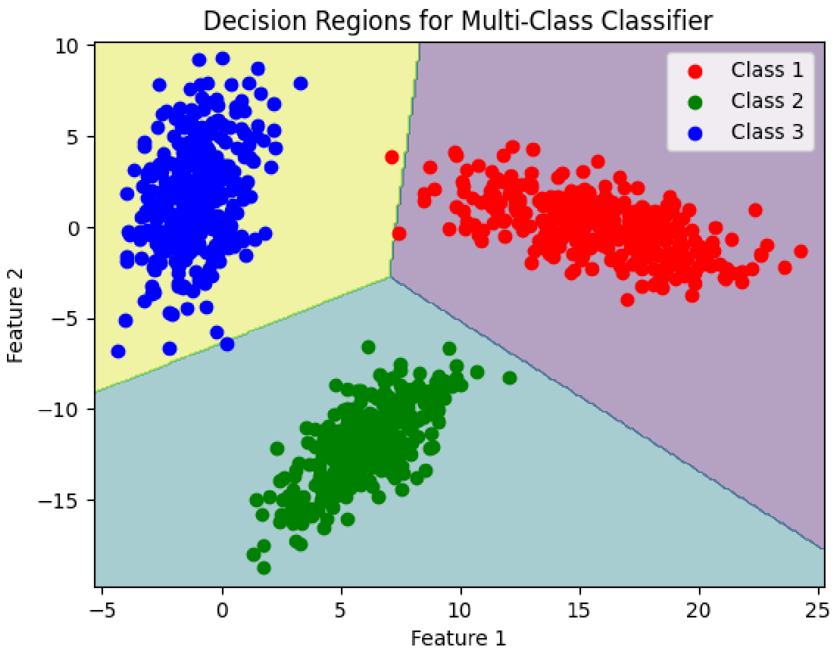


Figure 4: Decision Boundary plot for all three classes.

```
| ClassName        |   Actual_Class1 |   Actual_Class2 |   Actual_Class3 |
|------------------+-----------------+-----------------+-----------------|
| Predicted_class1 |             148 |               0 |               0 |
| Predicted_class2 |               0 |             149 |               0 |
| Predicted_class3 |               1 |               0 |             149 |
Class Accuracies:
 [0.99328859 1.         1.        ]
Class Precisions:
 [1.         1.         0.99333333]
Class Recalls:
 [0.99328859 1.         1.        ]
Class F-Measures:
 [0.996633   1.         0.99665552]
+-----------+----------+
| Metric    |    Value |
+===========+==========+
| Precision | 0.997778 |
+-----------+----------+
| Recall    | 0.997763 |
+-----------+----------+
| F1 Score  | 0.997763 |
+-----------+----------+
```

Figure 5: Confusion Matrix and Accuracy, Precision, Recall, F1-Score for Linear Data.

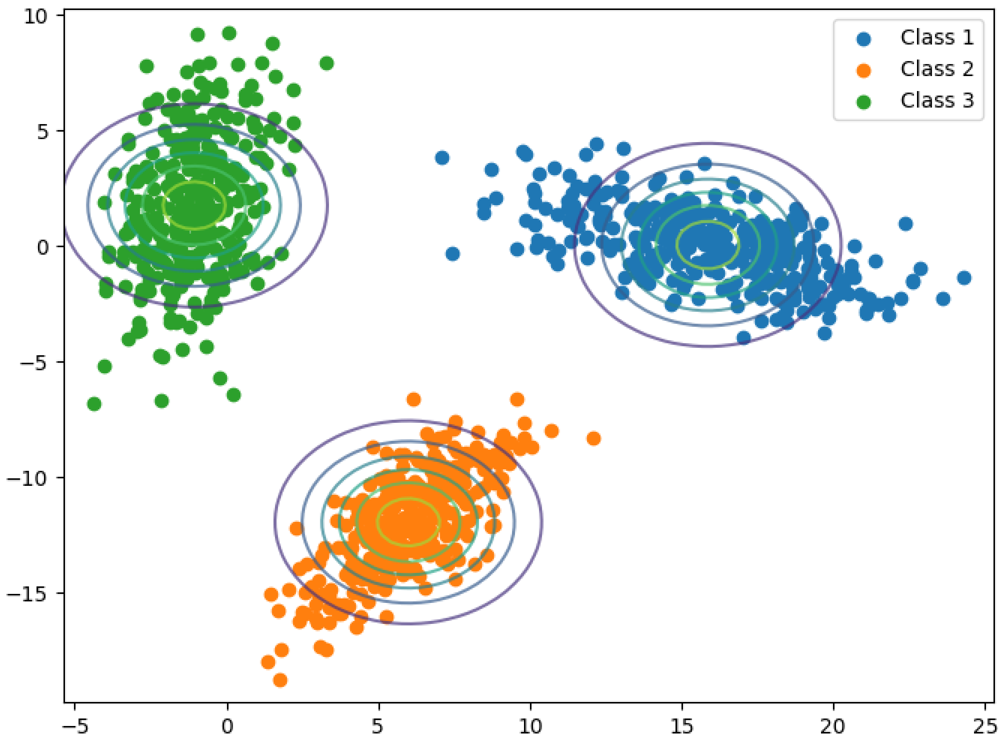


Figure 6: Contour Density Plot for the data.

3.2 Using Covariance Matrix of type 2 (Σ)

Below is provided the similar plots as above with a slight difference of using the covariance matrix as Σ that is obtained by averaging the covariance matrices of all classes.

In theory this does slightly better classification that the type 1, but in this case we don't see any significant improvement because the classification was already quite accurate.

The covariance matrix is same for all in this case too thus we obtain a linear decision boundary here as well.

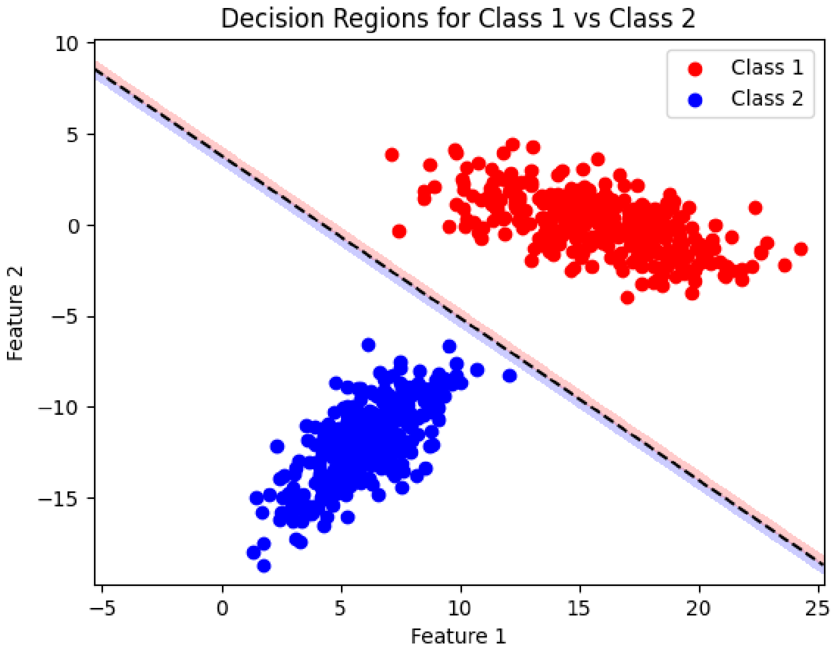


Figure 7: Decision Boundary for Class1 v/s Class2.

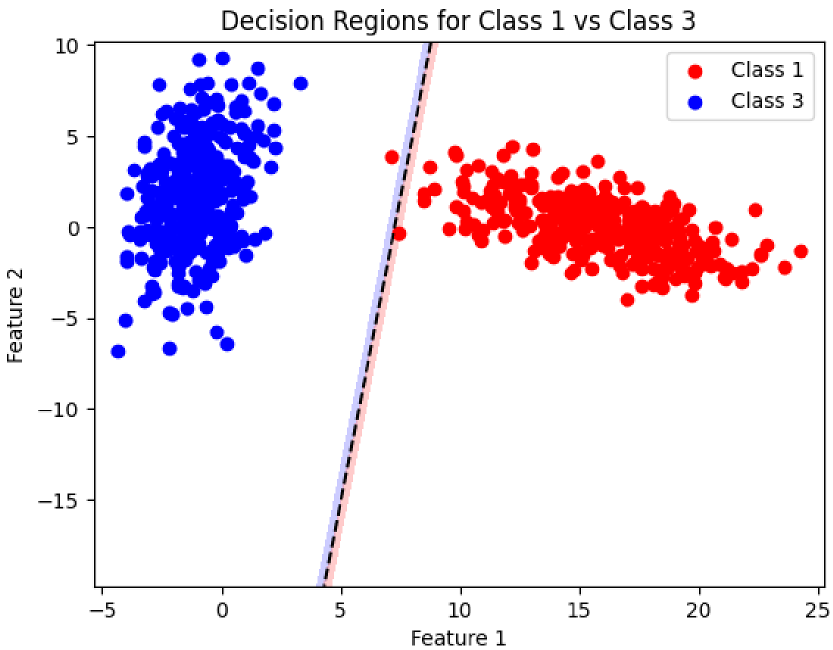


Figure 8: Decision Boundary plot for Class1 v/s Class3.

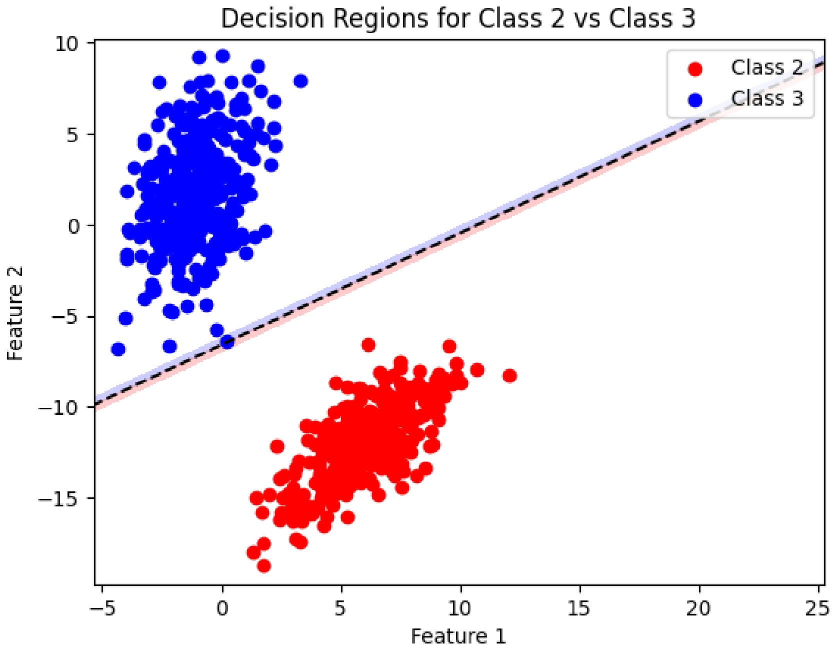


Figure 9: Decision Boundary plot for Class2 v/s Class3.

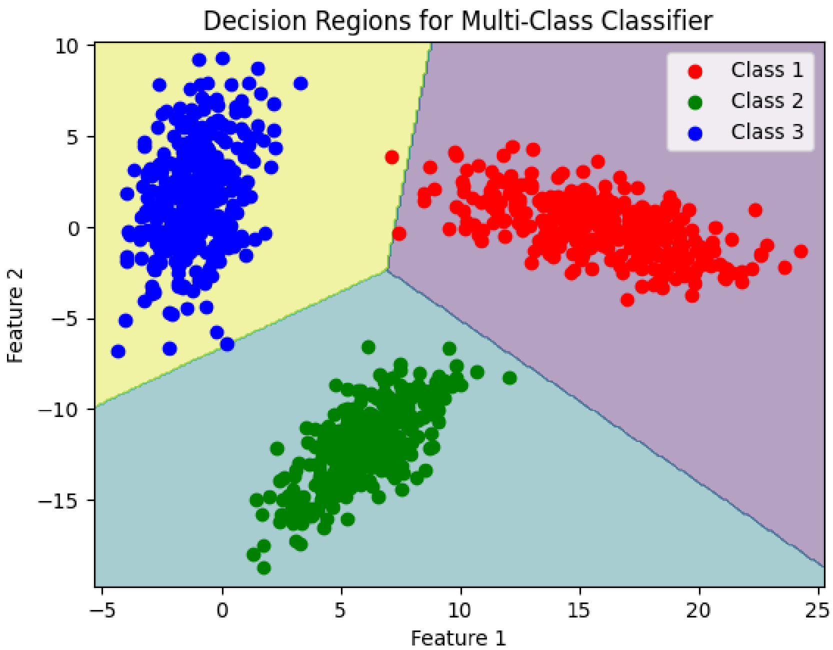


Figure 10: Decision Boundary plot for all three classes.

```
| ClassName        |   Actual_Class1 |   Actual_Class2 |   Actual_Class3 |
|------------------+-----------------+-----------------+-----------------|
| Predicted_class1 |             148 |               0 |               0 |
| Predicted_class2 |               0 |             149 |               0 |
| Predicted_class3 |               1 |               0 |             149 |
Class Accuracies:
 [0.99328859 1.         1.        ]
Class Precisions:
 [1.         1.         0.99333333]
Class Recalls:
 [0.99328859 1.         1.        ]
Class F-Measures:
 [0.996633   1.         0.99665552]
+-----------+----------+
| Metric    |    Value |
+===========+==========+
| Precision | 0.997778 |
+-----------+----------+
| Recall    | 0.997763 |
+-----------+----------+
| F1 Score  | 0.997763 |
+-----------+----------+
```

Figure 11: Confusion Matrix and Accuracy, Precision, Recall, F1-Score for Linear Data.

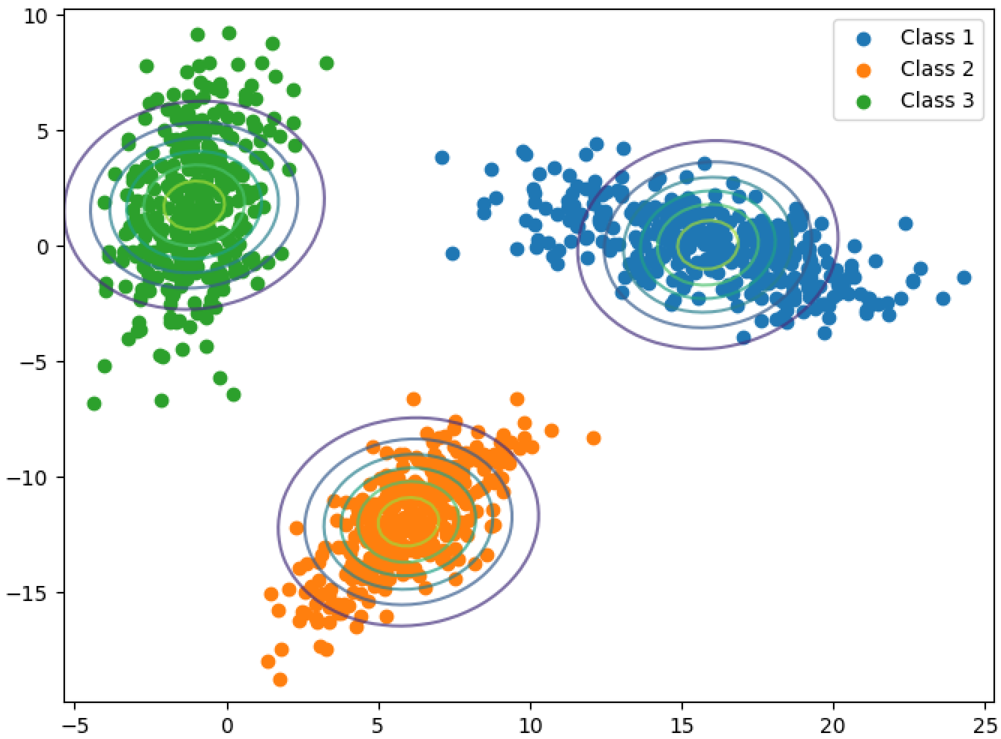


Figure 12: Contour Density Plot for the data.

3.3 Using Covariance Matrix of type 3

Below is provided the plots as above with the difference of using the Diagonal covariance matrix, this covariance matrix is diagonal but different for each class.
We can observe that the decision region plot for the train data is more accurate than the type 2, implying that this covariance matrix actually does a better job than the average covariance matrix.
Since the covariance matrix is different for each class thus, the decision boundary in this case formulated by the discriminant function is non linear.
This also does better classification for test data than the type 2, we can in the confusion matrix that each and every data sample is correctly classified.

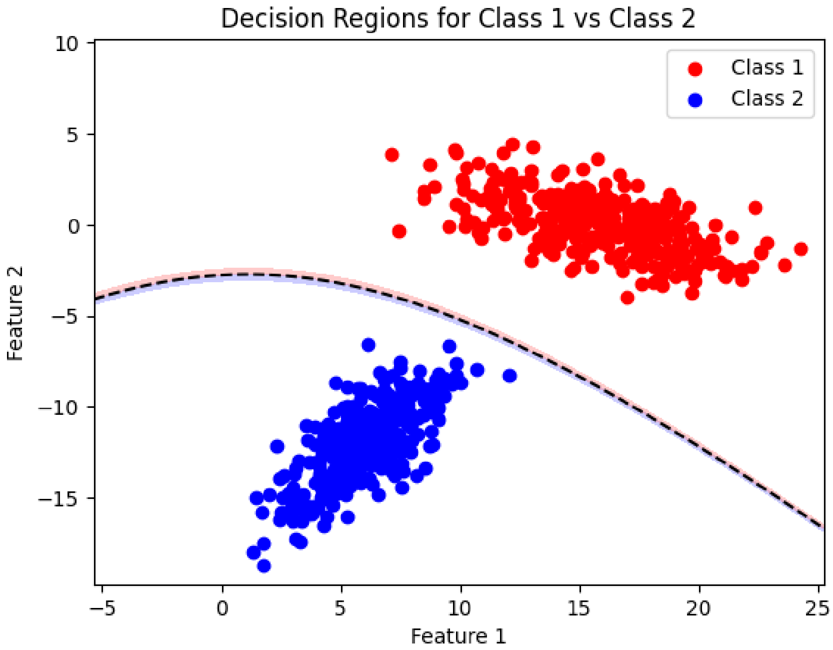


Figure 13: Decision Boundary for Class1 v/s Class2.

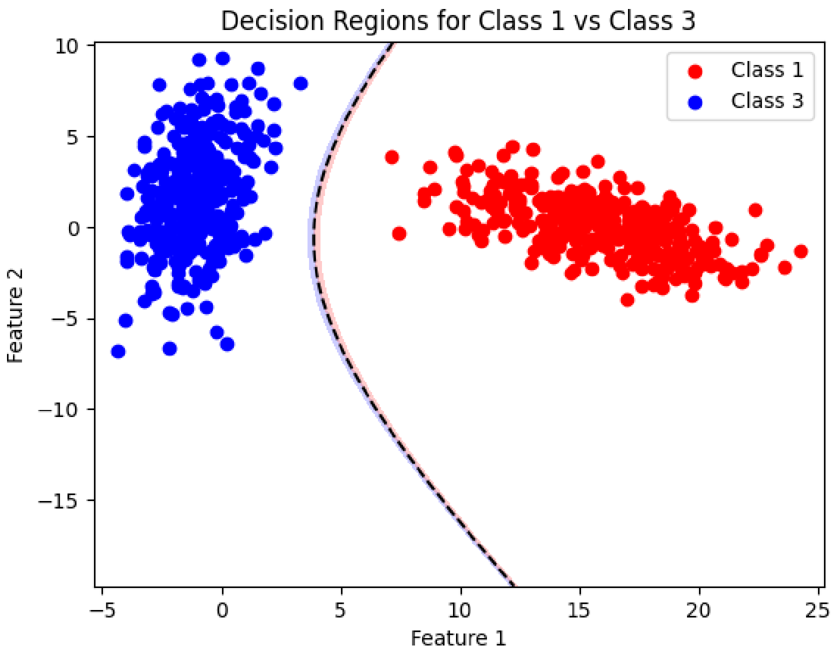


Figure 14: Decision Boundary plot for Class1 v/s Class3.

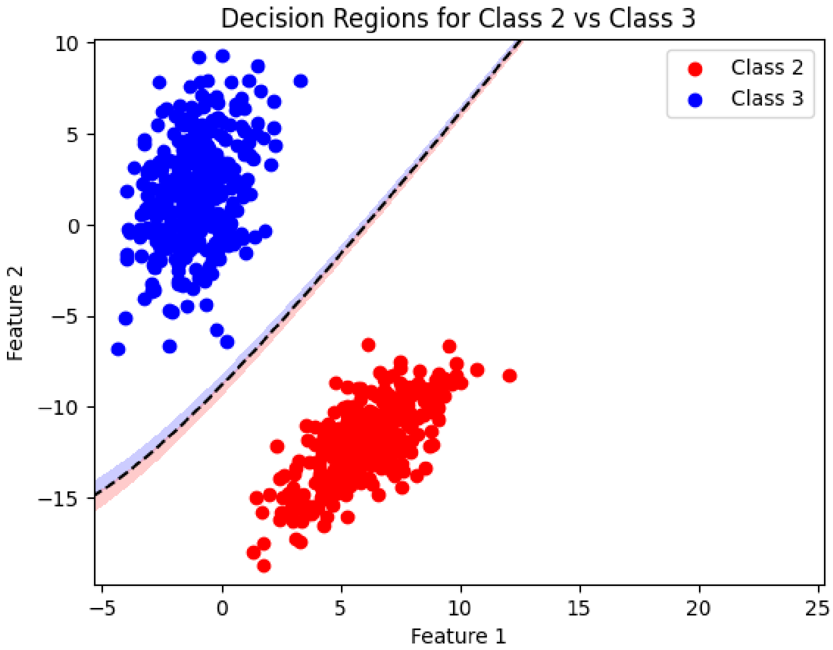


Figure 15: Decision Boundary plot for Class2 v/s Class3.

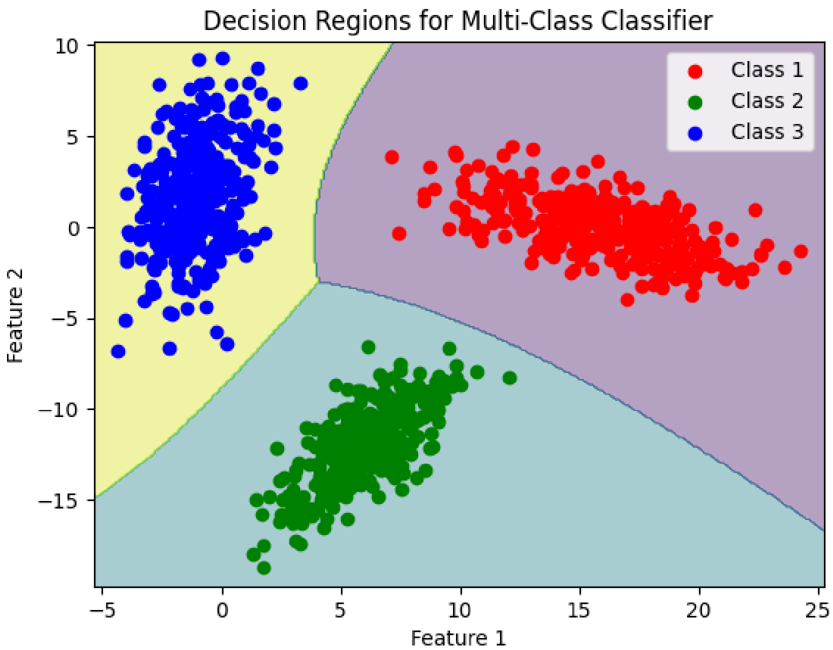


Figure 16: Decision Boundary plot for all three classes.

```
| ClassName        |   Actual_Class1 |   Actual_Class2 |   Actual_Class3 |
|------------------+-----------------+-----------------+-----------------|
| Predicted_class1 |             149 |               0 |               0 |
| Predicted_class2 |               0 |             149 |               0 |
| Predicted_class3 |               0 |               0 |             149 |
Class Accuracies:
 [1. 1. 1.]
Class Precisions:
 [1. 1. 1.]
Class Recalls:
 [1. 1. 1.]
Class F-Measures:
 [1. 1. 1.]
+-----------+---------+
| Metric    |   Value |
+===========+=========+
| Precision |       1 |
+-----------+---------+
| Recall    |       1 |
+-----------+---------+
| F1 Score  |       1 |
+-----------+---------+
```

Figure 17: Confusion Matrix and Accuracy, Precision, Recall, F1-Score for Linear Data.

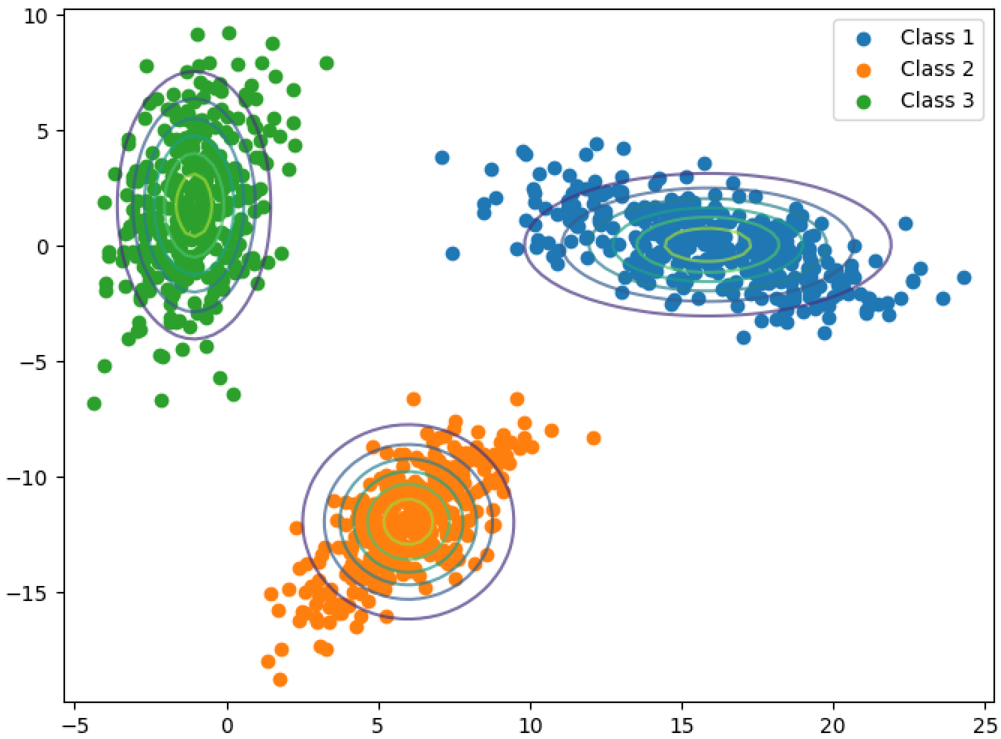


Figure 18: Contour Density Plot for the data.

3.4 Using Covariance Matrix of type 4

Below is provided the plots as above with the difference of using the Complete covariance matrices that are different for each class.
We can observe that the decision region plot for the train data is quite different than type 3, as it tries to be equidistant from all the classes.
In this case we use complete covariance matrices we see the decision boundary is non linear here too. Also, we can see in the confusion matrix that each and every data sample is correctly classified in this type as well.

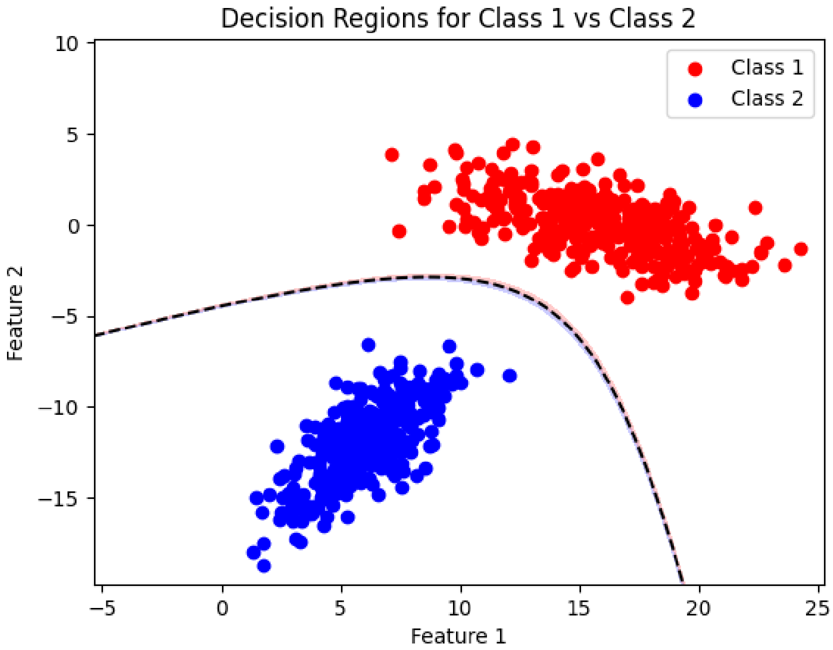


Figure 19: Decision Boundary for Class1 v/s Class2.

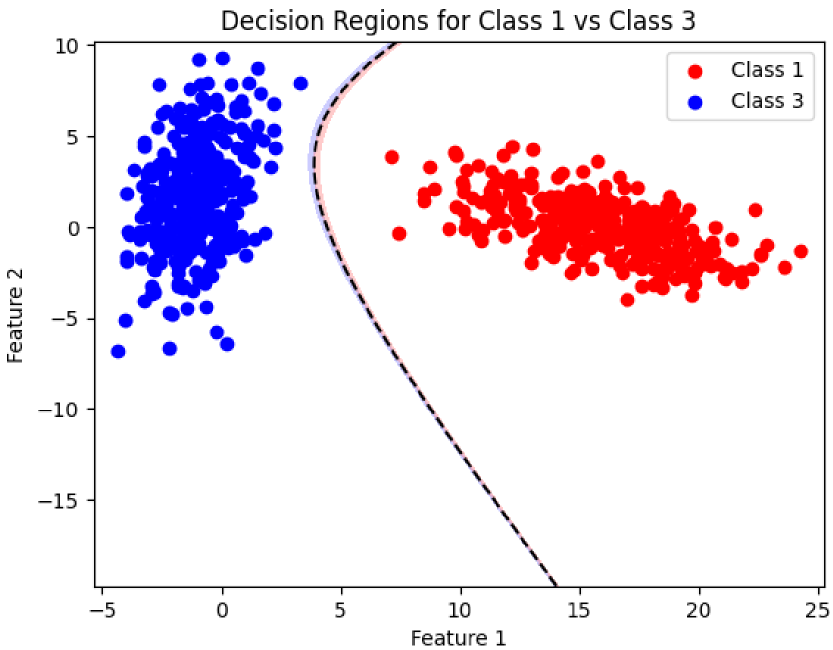


Figure 20: Decision Boundary plot for Class1 v/s Class3.

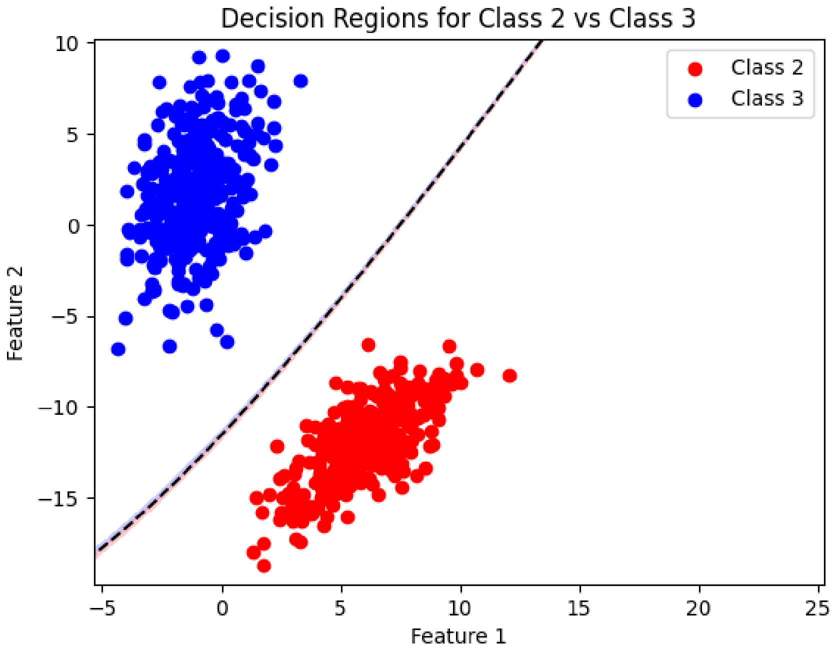


Figure 21: Decision Boundary plot for Class2 v/s Class3.

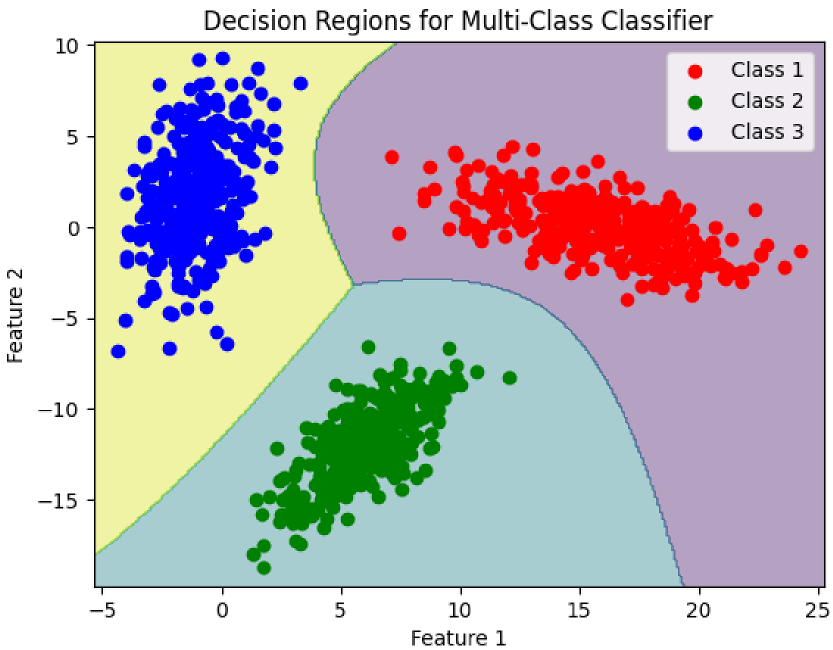


Figure 22: Decision Boundary plot for all three classes.

```
| ClassName        |   Actual_Class1 |   Actual_Class2 |   Actual_Class3 |
|------------------+-----------------+-----------------+-----------------|
| Predicted_class1 |             149 |               0 |               0 |
| Predicted_class2 |               0 |             149 |               0 |
| Predicted_class3 |               0 |               0 |             149 |
Class Accuracies:
 [1. 1. 1.]
Class Precisions:
 [1. 1. 1.]
Class Recalls:
 [1. 1. 1.]
Class F-Measures:
 [1. 1. 1.]
+-----------+---------+
| Metric    |   Value |
+===========+=========+
| Precision |       1 |
+-----------+---------+
| Recall    |       1 |
+-----------+---------+
| F1 Score  |       1 |
+-----------+---------+
```

Figure 23: Confusion Matrix and Accuracy, Precision, Recall, F1-Score for Linear Data.

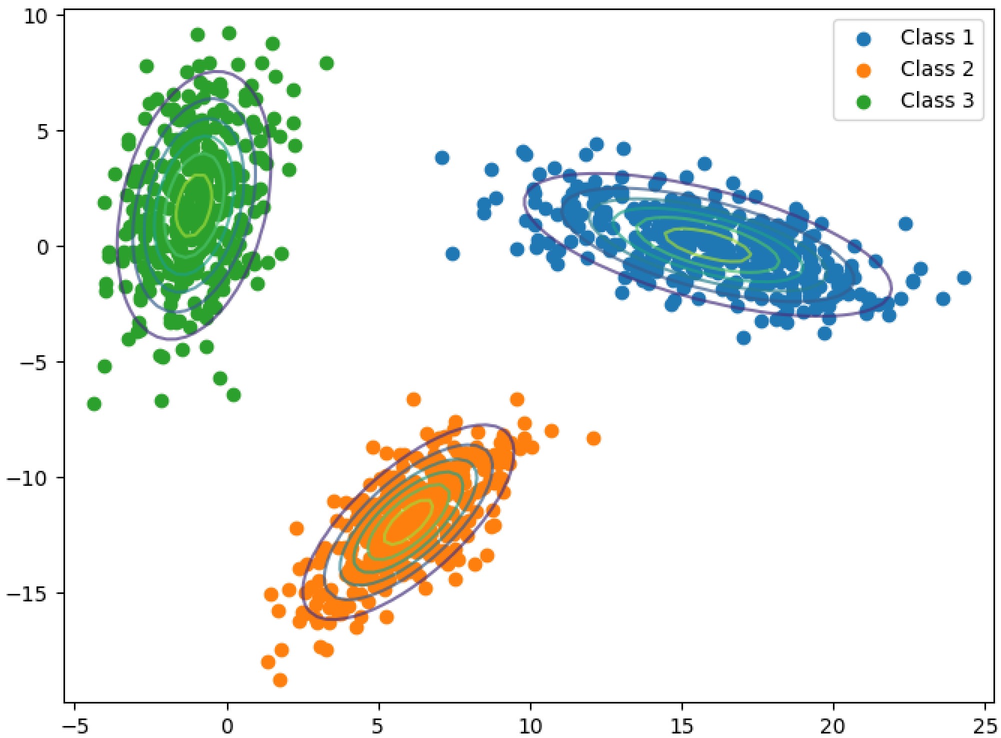


Figure 24: Contour Density Plot for the data.

4 Data Set - Non Linear

4.1 Using Covariance Matrix Type 1 (σ^2I)

Provided below is the plot for the decision regions of both classes of the non-linearly separable data using the covariance matrix of the type σ^2I.The accuracy using the Decision Boundary formulated by the covariance matrix of type 1 is very low this suggests the Bayes classfier formed by this covariance matrix is not good for such non-linearly seperable data.
Since the covariance matrix in this case is same for all classes the discrimnant function in this case gives a linear decision boundary.
This is also reflected in the confusion matrix, precision, recall, accuracy and f-score calculated for the test data.

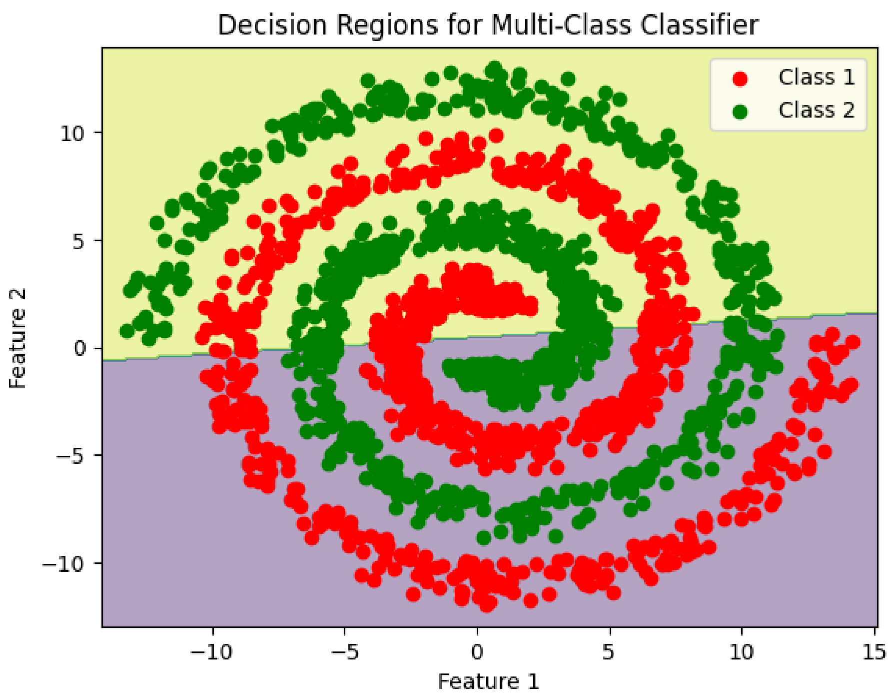


Figure 25: Decision Boundary plot classes.

-
```
| ClassName         |   Actual_Class1 |   Actual_Class2 |
|-------------------+-----------------+-----------------|
| Predicted_class1 |             200 |             190 |
| Predicted_class2 |             190 |             200 |
Class Accuracies:
 [0.51282051 0.51282051]
Class Precisions:
 [0.51282051 0.51282051]
Class Recalls:
 [0.51282051 0.51282051]
Class F-Measures:
 [0.51282051 0.51282051]
+-----------+----------+
| Metric    |    Value |
+===========+==========+
| Precision | 0.512821 |
+-----------+----------+
| Recall    | 0.512821 |
+-----------+----------+
| F1 Score  | 0.512821 |
+-----------+----------+
```

Figure 26: Confusion Matrix and Accuracy,Precision,Recall,F1-score for Non Linear Data.

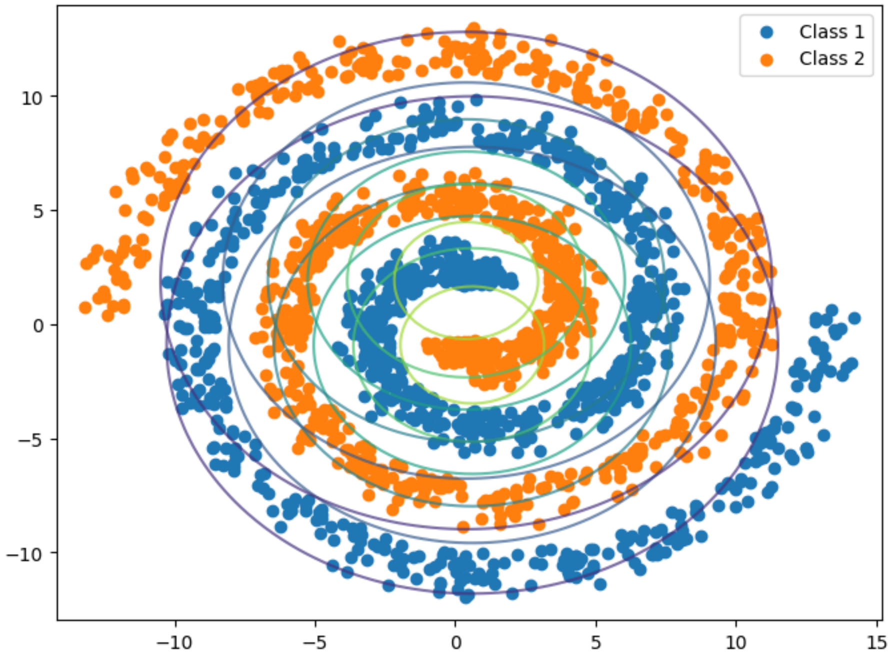


Figure 27: Contour Density Plot for Non Linear Data Set

4.2 Using Covariance Matrix Type 2 (Σ)

Provided below is the plot for the decision regions of both classes of the non-linearly separable data using the covariance matrix of the type 2, that is obtained by averaging the covariance matrices of all classes. The accuracy using the Decision Boundary formulated by the covariance matrix of type 2 is also very low this suggests the Bayes classfier formed by this covariance matrix is not good for such non-linearly seperable data.
Since the covariance matrix in this case is same for all classes the discrimnant function in this case gives a linear decision boundary.
This is also reflected in the confusion matrix, precision, recall, accuracy and f-score calculated for the test data.

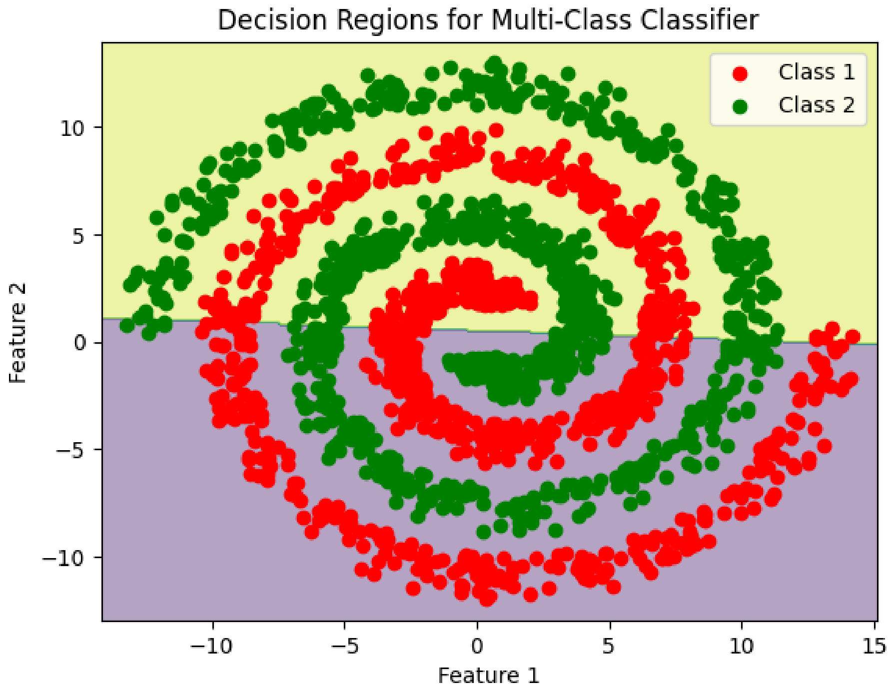
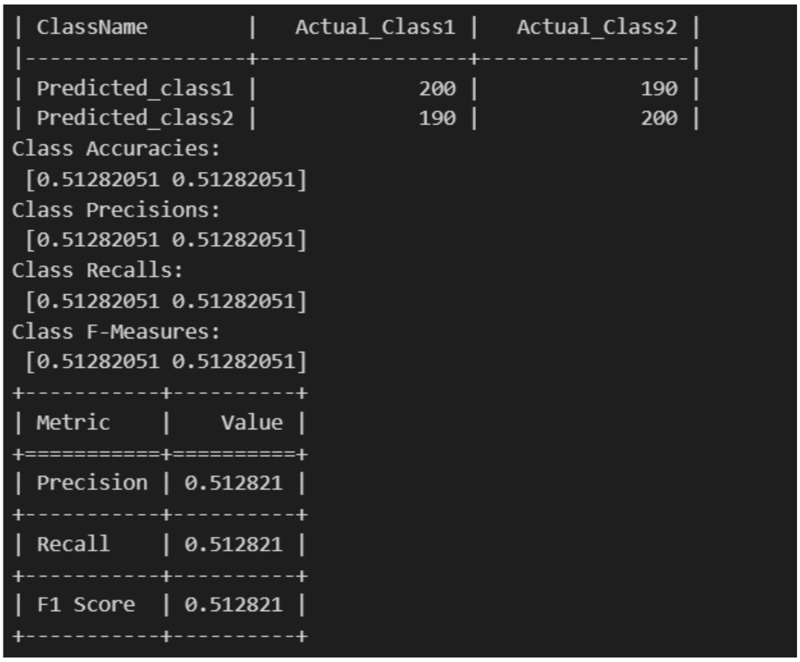


Figure 28: Decision Boundary plot classes.



```
| ClassName        |   Actual_Class1 |   Actual_Class2 |
|------------------+-----------------+-----------------|
| Predicted_class1 |             200 |             190 |
| Predicted_class2 |             190 |             200 |
Class Accuracies:
 [0.51282051 0.51282051]
Class Precisions:
 [0.51282051 0.51282051]
Class Recalls:
 [0.51282051 0.51282051]
Class F-Measures:
 [0.51282051 0.51282051]
+-----------+----------+
| Metric    |    Value |
+===========+==========+
| Precision | 0.512821 |
+-----------+----------+
| Recall    | 0.512821 |
+-----------+----------+
| F1 Score  | 0.512821 |
+-----------+----------+
```

Figure 29: Confusion Matrix and Accuracy,Precision,Recall,F1-score for Non Linear Data.

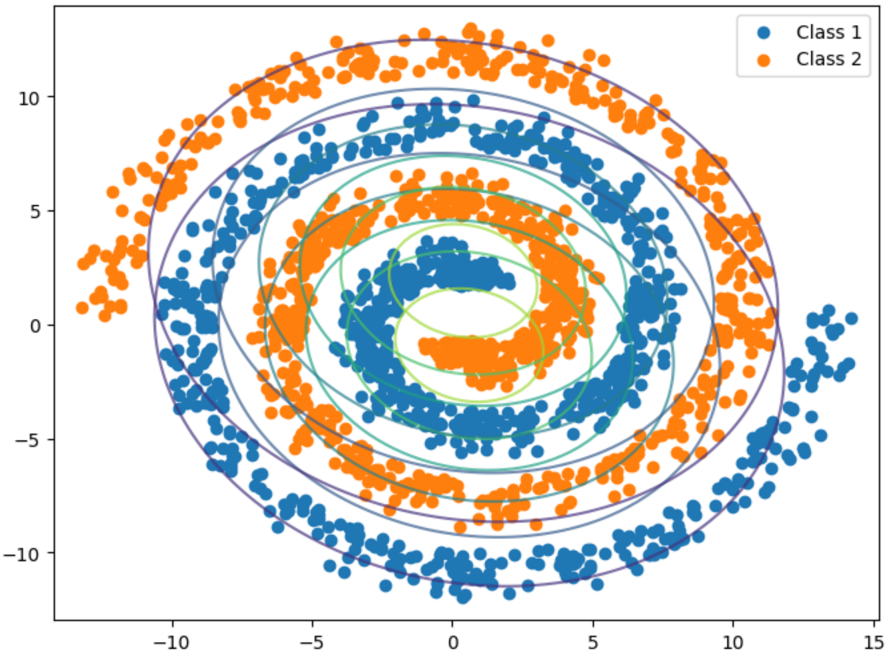


Figure 30: Contour Density Plot for Non Linear Data Set

4.3 Using Covariance Matrix of Type 3

Provided below is the plot for the decision regions of both classes of the non-linearly separable data using the Diagonal co-variance matrix, this covariance matrix is diagonal but different for each class.
In theory, decision boundary formed using these covariance matrices must give non-linear boundary but for the given dataset the distribution of data for both the classes is very similar, which leads to computation of similar covariance matrices. Hence, the decision boundaries formed also appear to be linear.
This is also reflected in the confusion matrix, precision, recall, accuracy and f-score calculated for the test data.

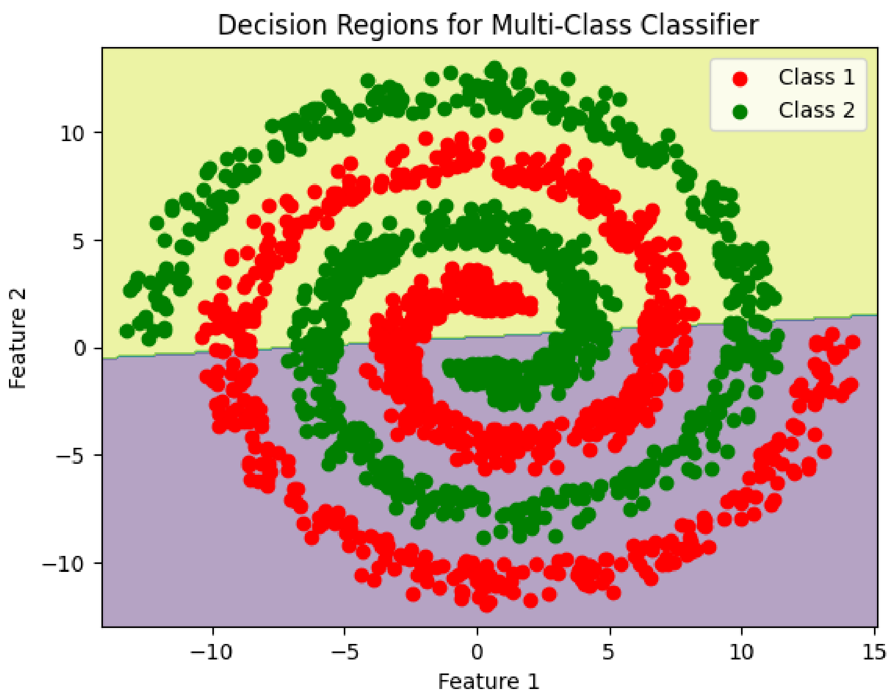


Figure 31: Decision Boundary plot classes.

```
| ClassName        |   Actual_Class1 |   Actual_Class2 |
|------------------+-----------------+-----------------|
| Predicted_class1 |             201 |             189 |
| Predicted_class2 |             189 |             201 |
Class Accuracies:
 [0.51538462 0.51538462]
Class Precisions:
 [0.51538462 0.51538462]
Class Recalls:
 [0.51538462 0.51538462]
Class F-Measures:
 [0.51538462 0.51538462]
+-----------+----------+
| Metric    |    Value |
+===========+==========+
| Precision | 0.515385 |
+-----------+----------+
| Recall    | 0.515385 |
+-----------+----------+
| F1 Score  | 0.515385 |
+-----------+----------+
```

Figure 32: Confusion Matrix and Accuracy,Precision,Recall,F1-score for Non Linear Data.

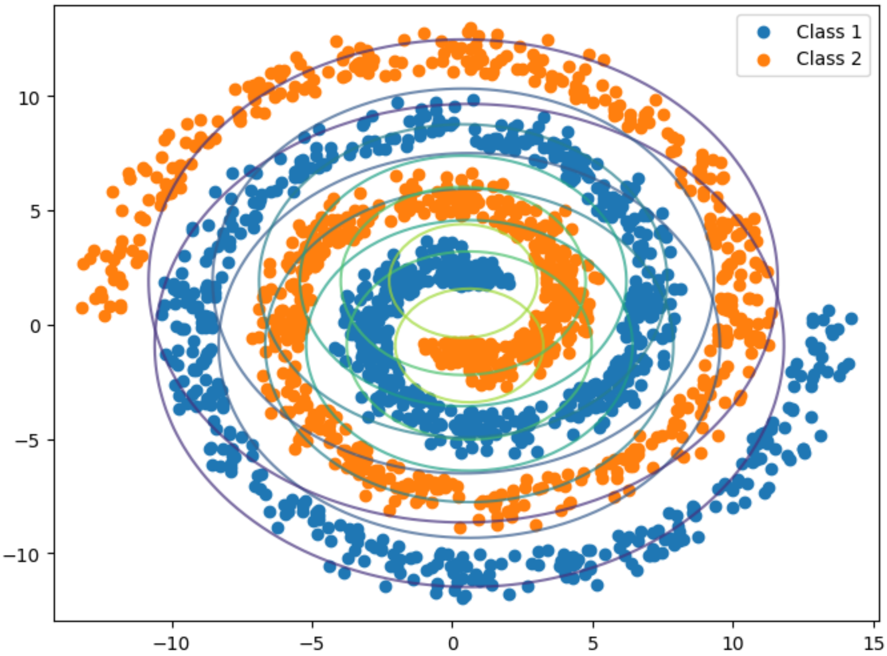


Figure 33: Contour Density Plot for Non Linear Data Set

4.4 Using Covariance Matrix of Type 4

Provided below is the plot for the decision regions of both classes of the non-linearly separable data using the complete co-variance matrix, this covariance matrix is different for each class.
In theory, decision boundary formed using these covariance matrices must give non-linear boundary but for the given dataset the distribution of data for both the classes is very similar, which leads to computation of similar covariance matrices. Hence, the decision boundaries formed also appear to be linear.
This is also reflected in the confusion matrix, precision, recall, accuracy and f-score calculated for the test data.

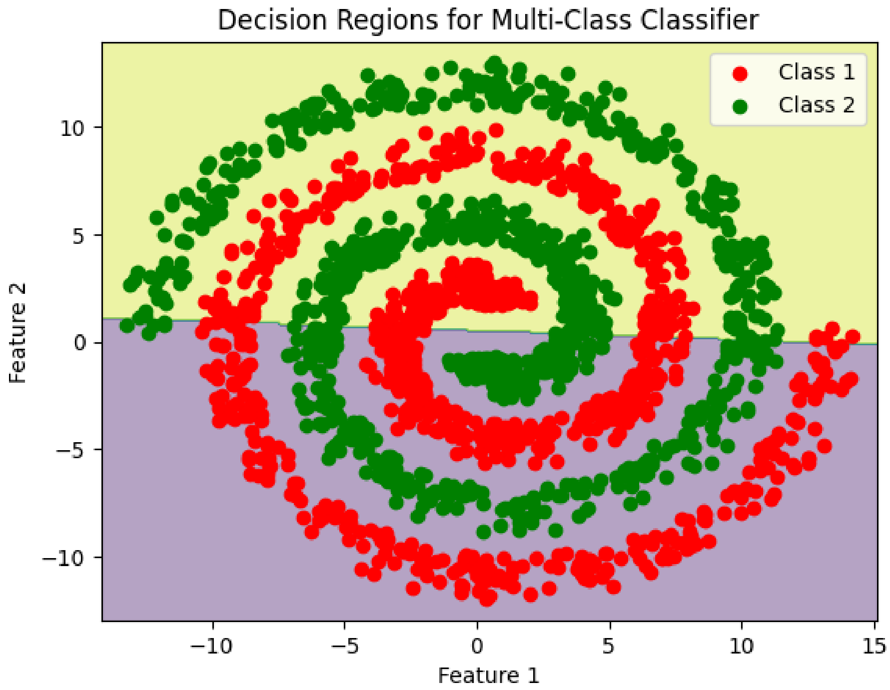


Figure 34: Decision Boundary plot classes.

```
| ClassName       |   Actual_Class1 |   Actual_Class2 |
|-----------------+-----------------+-----------------|
| Predicted_class1 |             200 |             190 |
| Predicted_class2 |             190 |             200 |
Class Accuracies:
 [0.51282051 0.51282051]
Class Precisions:
 [0.51282051 0.51282051]
Class Recalls:
 [0.51282051 0.51282051]
Class F-Measures:
 [0.51282051 0.51282051]
+-----------+----------+
| Metric    |    Value |
+===========+==========+
| Precision | 0.512821 |
+-----------+----------+
| Recall    | 0.512821 |
+-----------+----------+
| F1 Score  | 0.512821 |
+-----------+----------+
```

Figure 35: Confusion Matrix and Accuracy,Precision,Recall,F1-score for Non Linear Data.

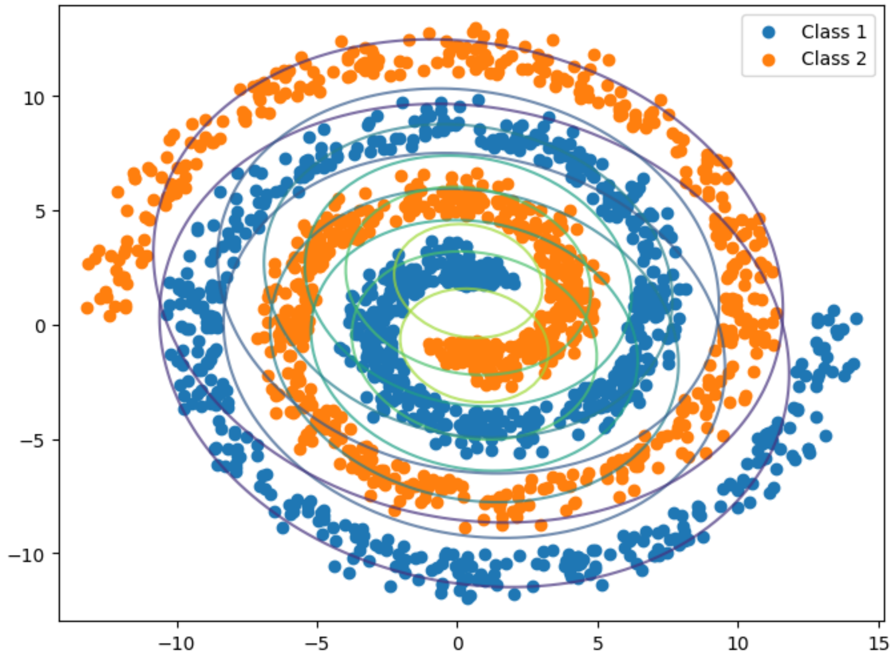


Figure 36: Contour Density Plot for Non Linear Data Set

5 Data Set - Real

recall, accuracy and f-score calculated for the test data.

5.1 Using Covariance Matrix Type 1 (σ^2I)

Provided below are the plot for the decision boundaries of each class for the real data using the covariance matrix of the type 2I.The accuracy on this data for the Bayes Classifier using the Decision Boundary formulated by the covariance of type 1 obtained by the training data is quite good.
Since the covariance matrix in this case is same for all classes the discrimnant function in this case gives a linear decision boundary. As the covarince matrix used is generalized for all three classes and same is used to plot the decision region there are minor errors in the classification, but it can be overcome using better covariance matrix.
This is also reflected in the confusion matrix, precision, recall, accuracy and f-score calculated for the test data.

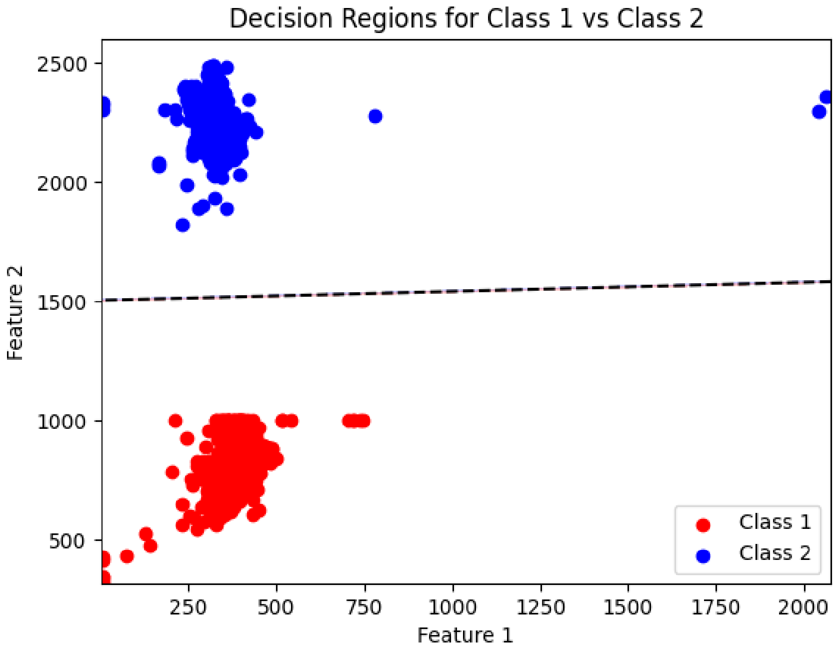


Figure 37: Decision Boundary for Class1 v/s Class2.

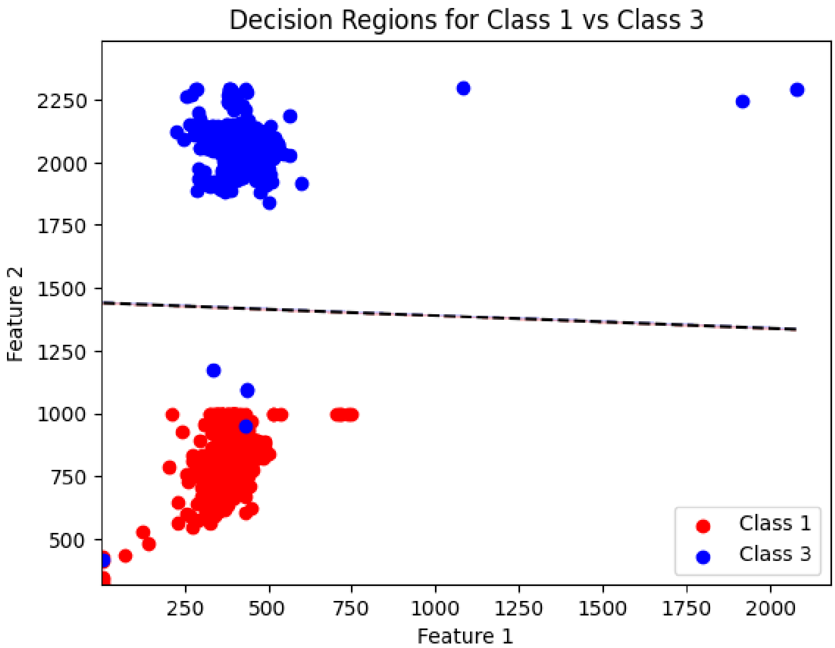


Figure 38: Decision Boundary plot for Class1 v/s Class3.

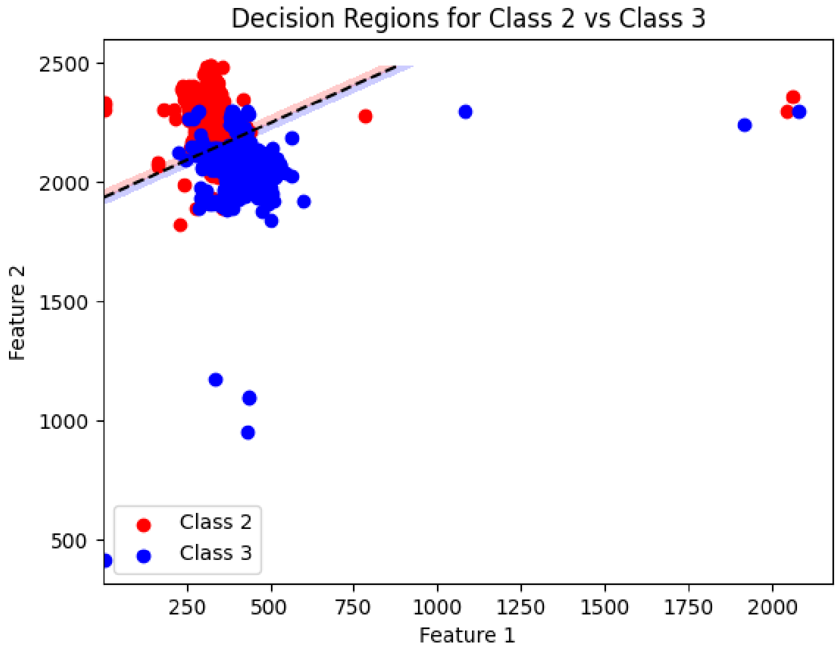


Figure 39: Decision Boundary plot for Class2 v/s Class3.

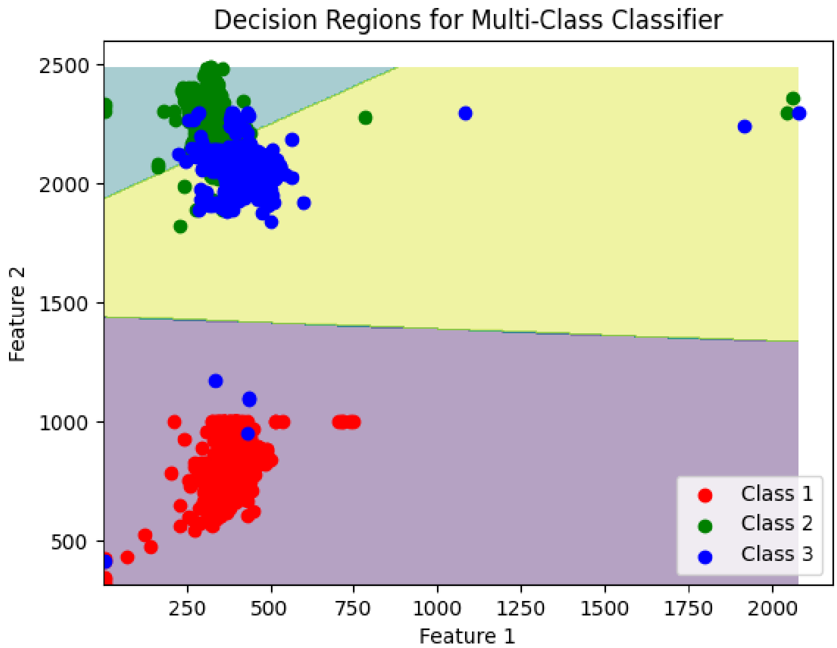


Figure 40: Decision Boundary plot for all three classes.

•

```
| ClassName       |   Actual_Class1 |   Actual_Class2 |   Actual_Class3 |
|-----------------+-----------------+-----------------+-----------------|
| Predicted_class1 |            746 |               3 |              18 |
| Predicted_class2 |              0 |             689 |             349 |
| Predicted_class3 |              0 |              24 |             319 |
Class Accuracies:
 [1.         0.9622905  0.46501458]
Class Precisions:
 [0.9726206  0.66377649 0.93002915]
Class Recalls:
 [1.         0.9622905  0.46501458]
Class F-Measures:
 [0.98612029 0.78563284 0.62001944]
+-----------+----------+
| Metric    |    Value |
+===========+==========+
| Precision | 0.85607  |
+-----------+----------+
| Recall    | 0.816574 |
+-----------+----------+
| F1 Score  | 0.802371 |
+-----------+----------+
```

Figure 41: Confusion Matrix and Accuracy, Precision, Recall, F1-Score for Linear Data.

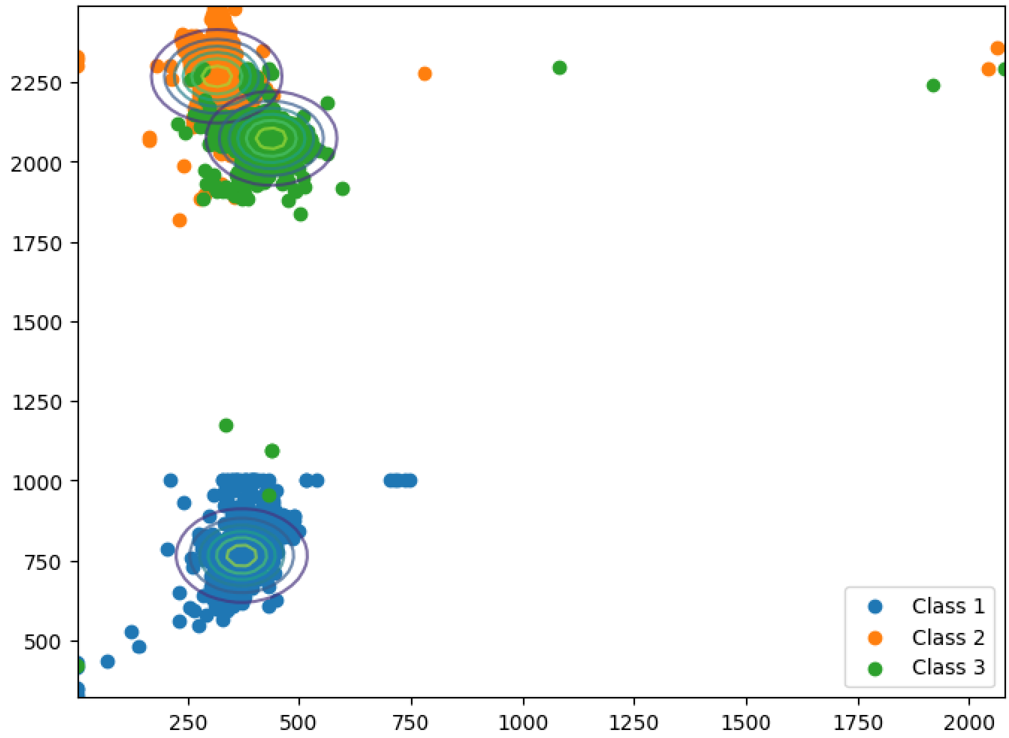


Figure 42: Contour Density Plot for the data.

5.2 Using Covariance Matrix of type 2 (Σ)

Provided below are the plot for the decision boundaries of each class for the real data using using the covariance matrix of the type 2, that is obtained by averaging the covariance matrices of all classes.The accuracy on this data for the Bayes Classifier using the Decision Boundary formulated by the covariance of type 2 obtained by the training data is again quite good.
Since the covariance matrix in this case is same for all classes the discrimnant function in this case gives a linear decision boundary. As the covarince matrix used is generalized for all three classes and same is used to plot the decision region there are minor errors in the classification, but it can be overcome using better covariance matrix.
This is also reflected in the confusion matrix, precision, recall, accuracy and f-score calculated for the test data.
In the confusion matrix and the metrics, we can see that there is slight improvement from that of type 1, implying that this classifier works better than that of type 1.

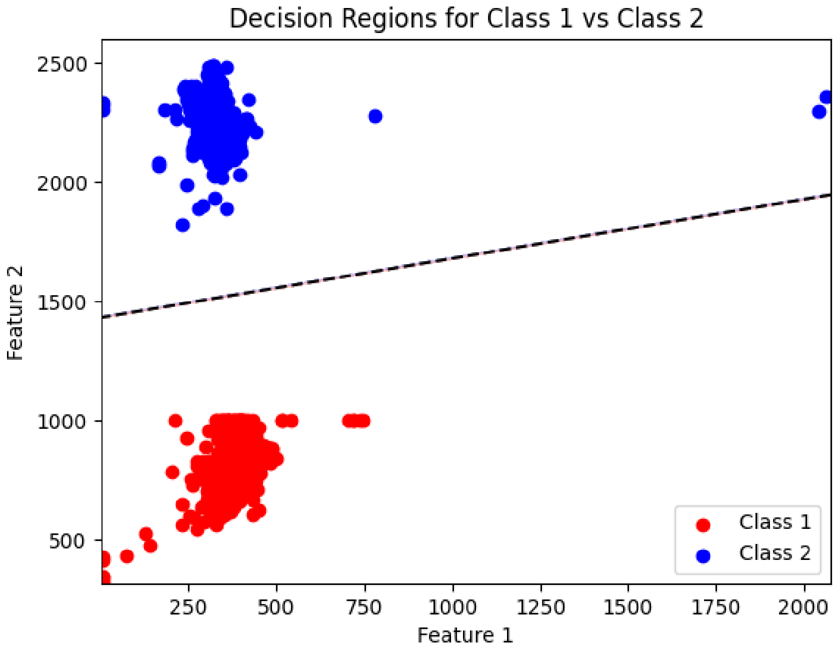


Figure 43: Decision Boundary for Class1 v/s Class2.

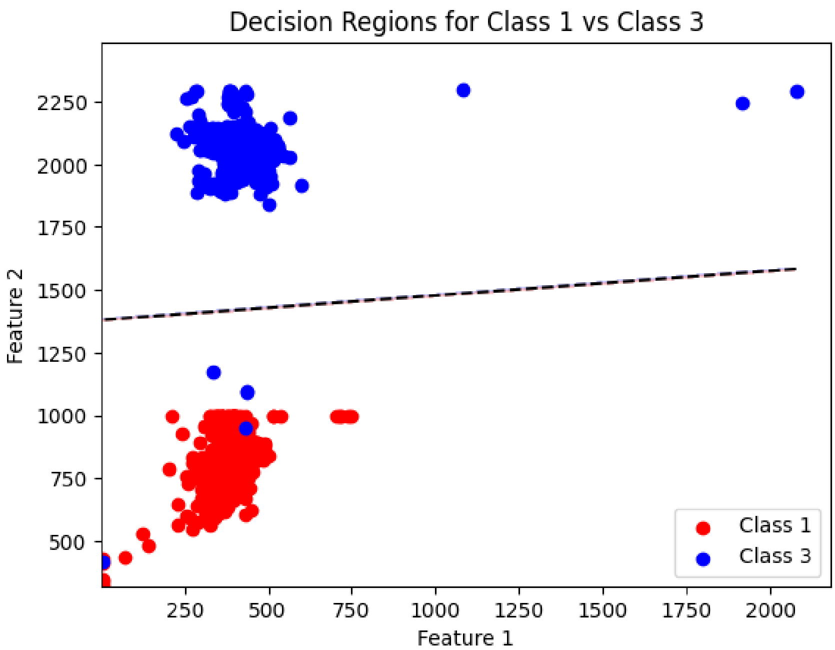


Figure 44: Decision Boundary plot for Class1 v/s Class3.

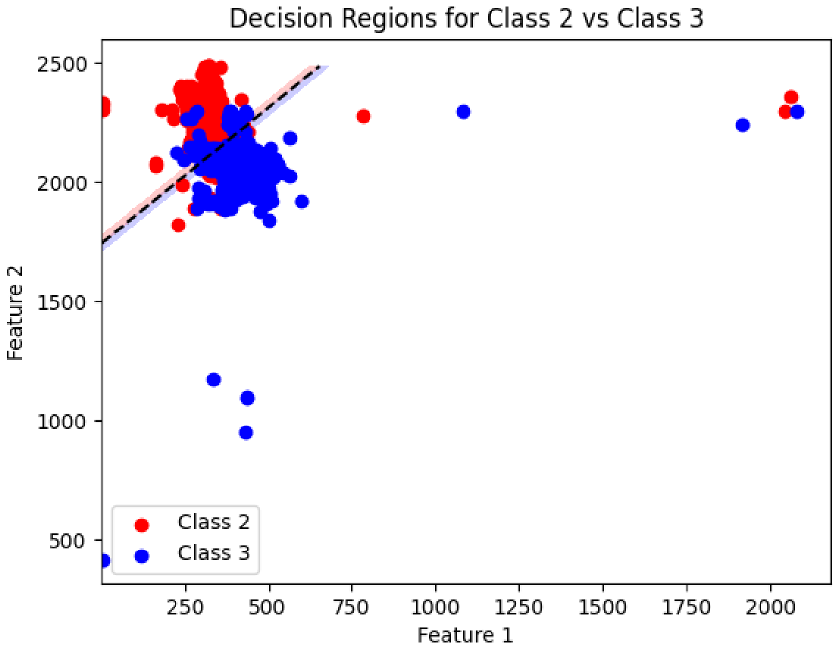


Figure 45: Decision Boundary plot for Class2 v/s Class3.

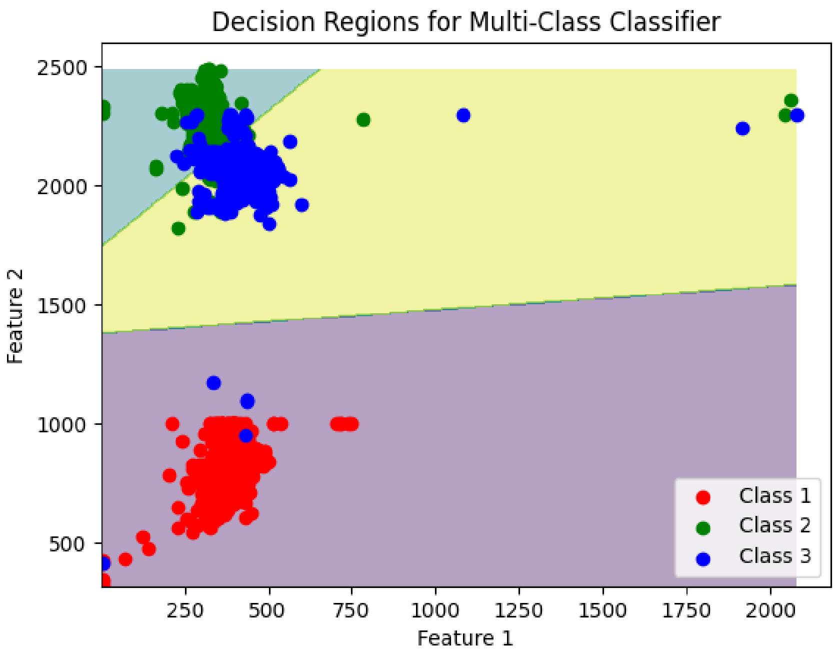


Figure 46: Decision Boundary plot for all three classes.

```
| ClassName        |   Actual_Class1 |   Actual_Class2 |   Actual_Class3 |
|------------------+-----------------+-----------------+-----------------|
| Predicted_class1 |             746 |               3 |              18 |
| Predicted_class2 |               0 |             690 |             346 |
| Predicted_class3 |               0 |              23 |             322 |
Class Accuracies:
 [1.         0.96368715 0.46938776]
Class Precisions:
 [0.9726206  0.66602317 0.93333333]
Class Recalls:
 [1.         0.96368715 0.46938776]
Class F-Measures:
 [0.98612029 0.78767123 0.62463628]
+-----------+----------+
| Metric    |    Value |
+===========+==========+
| Precision | 0.857874 |
+-----------+----------+
| Recall    | 0.818436 |
+-----------+----------+
| F1 Score  | 0.804525 |
+-----------+----------+
```

Figure 47: Confusion Matrix and Accuracy, Precision, Recall, F1-Score for Linear Data.

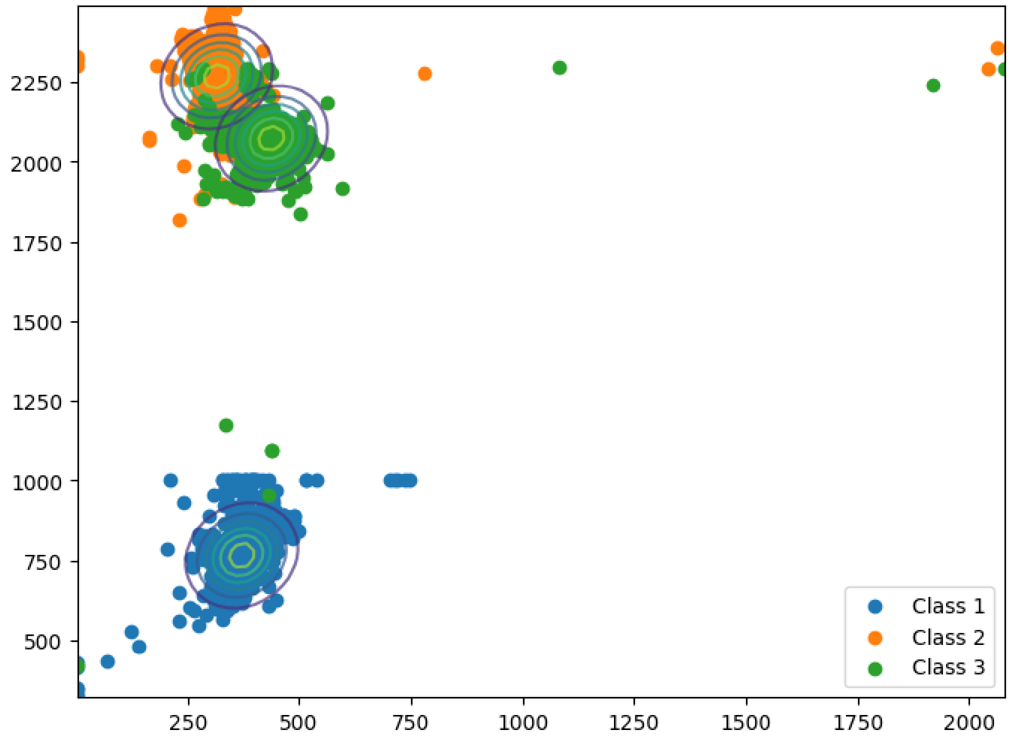


Figure 48: Contour Density Plot for the data.

5.3 Using Covariance Matrix of type 3

Provided below are the plot for the decision boundaries of each class for the real data using the Diagonal co-variance matrix, this covariance matrix is diagonal but different for each class.
The accuracy on this data for the Bayes Classifier using the Decision Boundary formulated by the covariance of type 3 obtained by the training data is again quite good.
Since the covariance matrix in this case is different for all classes the discrimnant function in this case gives a non-linear decision boundary. In the confusion matrix and the metrics, we can see that there is improvement from that of type 2, implying that this classifier works better than that of type 2.

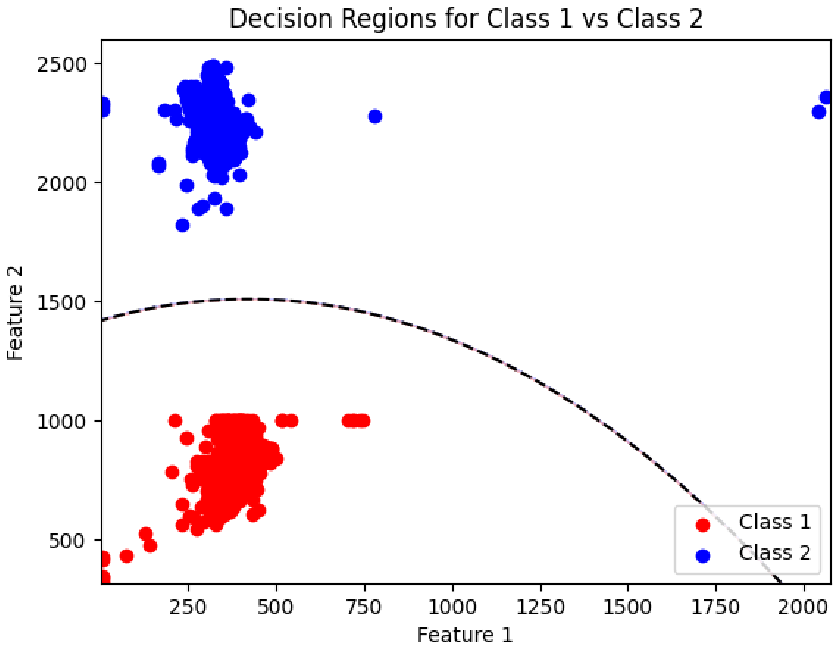


Figure 49: Decision Boundary for Class1 v/s Class2.

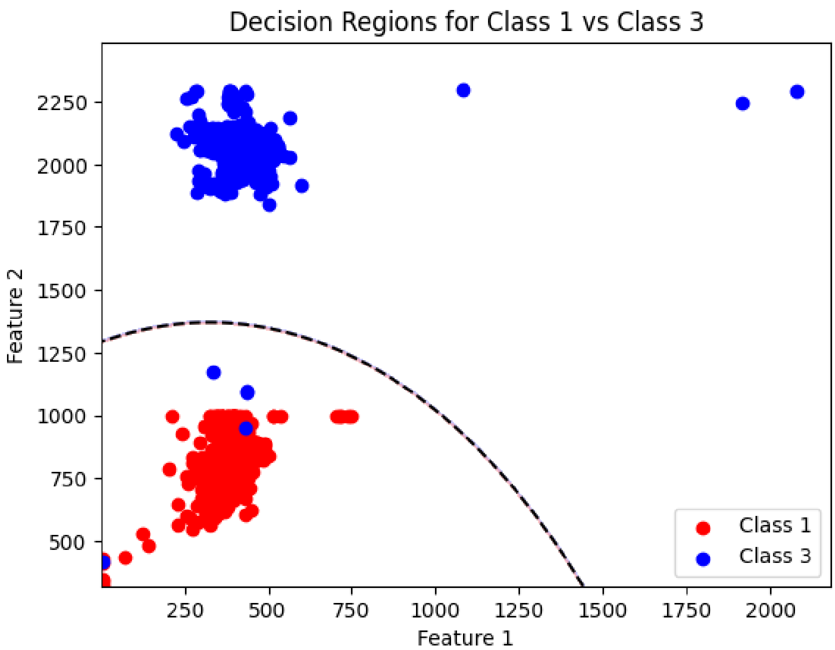


Figure 50: Decision Boundary plot for Class1 v/s Class3.

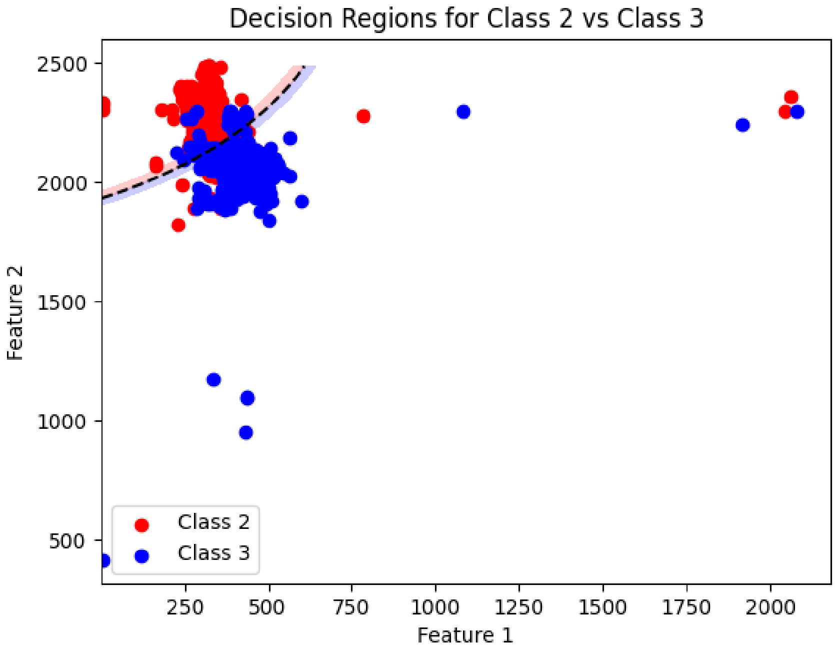


Figure 51: Decision Boundary plot for Class2 v/s Class3.

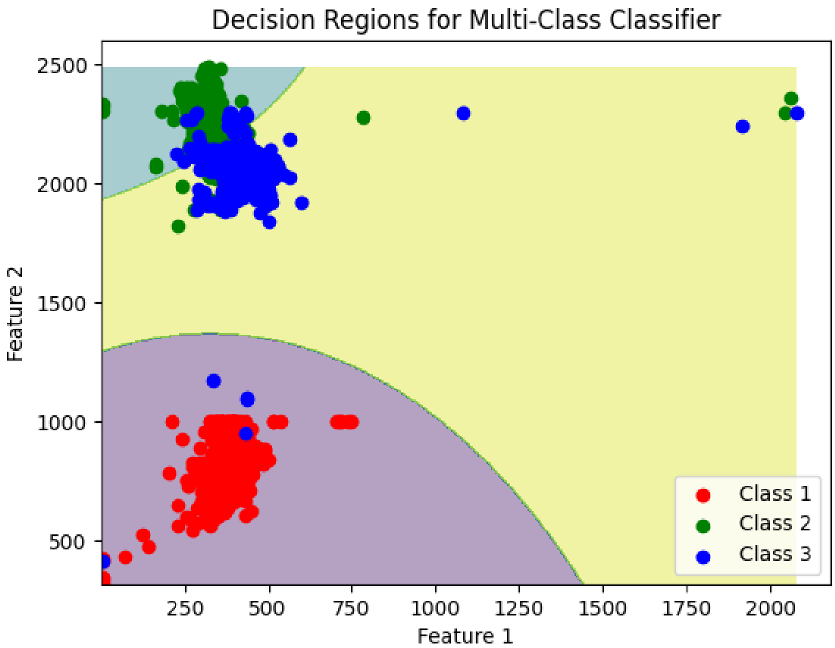


Figure 52: Decision Boundary plot for all three classes.

-

```
| ClassName        |   Actual_Class1 |   Actual_Class2 |   Actual_Class3 |
|------------------+-----------------+-----------------+-----------------|
| Predicted_class1 |             746 |               3 |              16 |
| Predicted_class2 |               0 |             689 |             345 |
| Predicted_class3 |               0 |              24 |             325 |
Class Accuracies:
 [1.         0.9622905  0.47376093]
Class Precisions:
 [0.9751634  0.66634429 0.93123209]
Class Recalls:
 [1.         0.9622905  0.47376093]
Class F-Measures:
 [0.98742555 0.78742857 0.62801932]
+-----------+----------+
| Metric    |    Value |
+===========+==========+
| Precision | 0.858193 |
+-----------+----------+
| Recall    | 0.819367 |
+-----------+----------+
| F1 Score  | 0.805977 |
+-----------+----------+
```

Figure 53: Confusion Matrix and Accuracy, Precision, Recall, F1-Score for Linear Data.

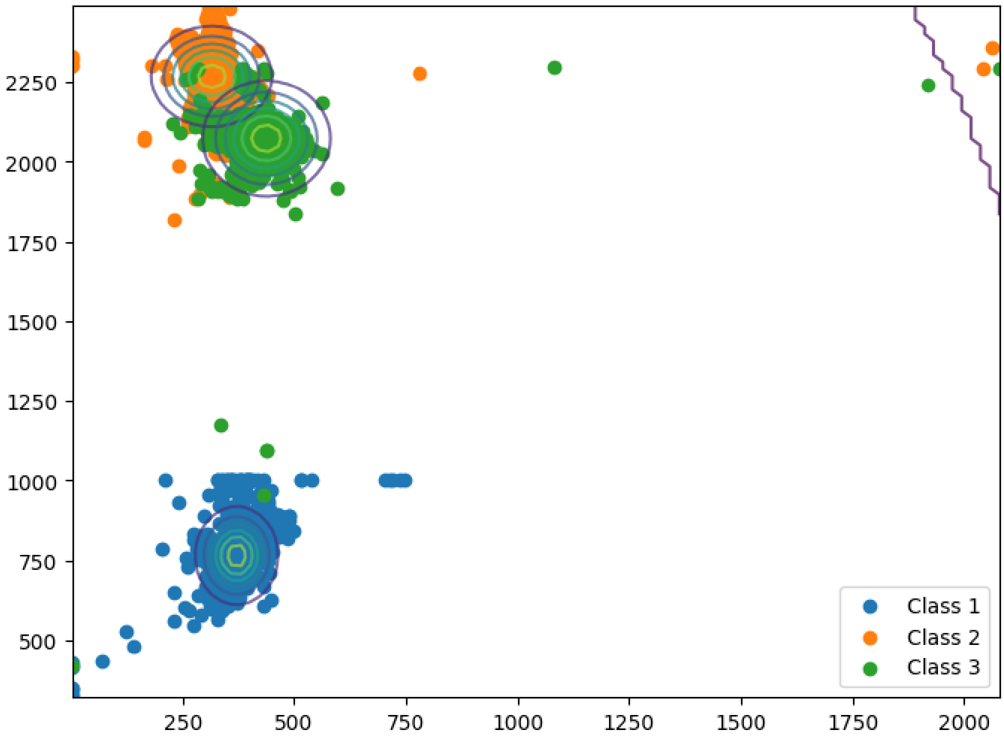


Figure 54: Contour Density Plot for the data.

5.4 Using Covariance Matrix of type 4

Provided below are the plot for the decision boundaries of each class for the real data using the complete co-variance matrix, this covariance matrix is different for each class.
The accuracy on this data for the Bayes Classifier using the Decision Boundary formulated by the covariance of type 4 obtained by the training data is again quite good.
Since the covariance matrix in this case is different for all classes the discrimnant function in this case gives a non-linear decision boundary. In the confusion matrix and the metrics, we can see that there is improvement from that of type 3, implying that this classifier works better than that of type 3.

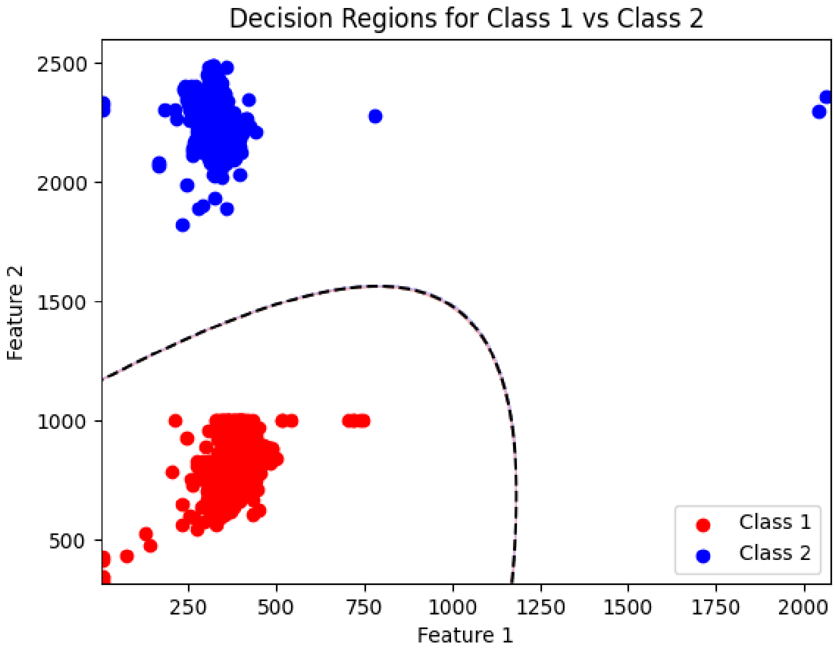


Figure 55: Decision Boundary for Class1 v/s Class2.

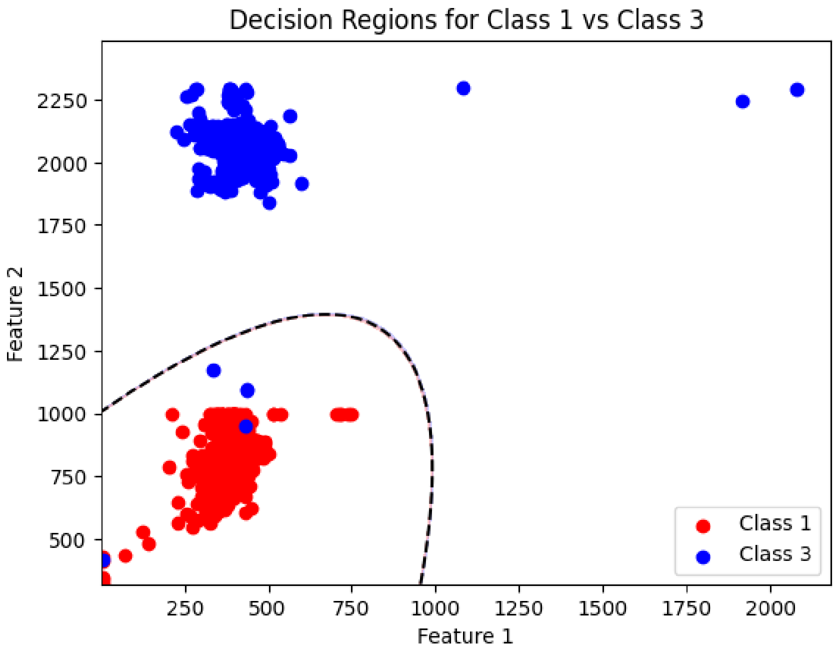


Figure 56: Decision Boundary plot for Class1 v/s Class3.

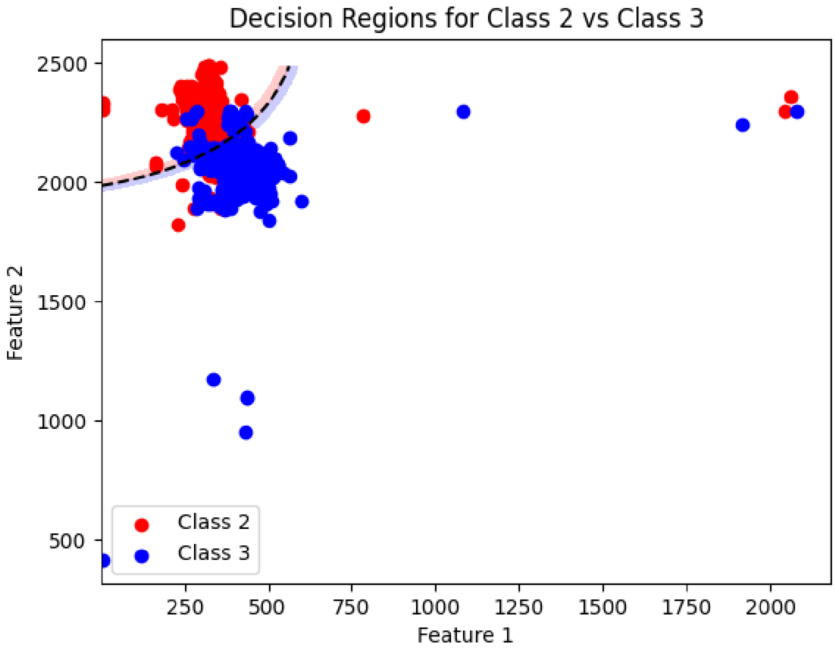


Figure 57: Decision Boundary plot for Class2 v/s Class3.

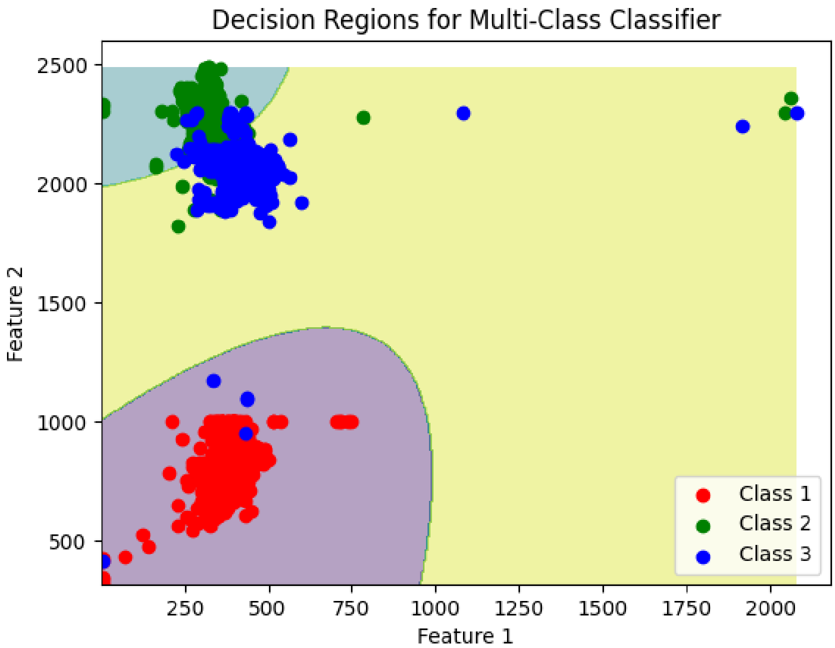


Figure 58: Decision Boundary plot for all three classes.

- ```
  | ClassName        |   Actual_Class1 |   Actual_Class2 |   Actual_Class3 |
  |------------------+-----------------+-----------------+-----------------|
  | Predicted_class1 |             745 |               3 |              15 |
  | Predicted_class2 |               0 |             689 |             346 |
  | Predicted_class3 |               1 |              24 |             325 |
  Class Accuracies:
   [0.99865952 0.9622905  0.47376093]
  Class Precisions:
   [0.97640891 0.66570048 0.92857143]
  Class Recalls:
   [0.99865952 0.9622905  0.47376093]
  Class F-Measures:
   [0.98740888 0.78697887 0.62741313]
  +-----------+----------+
  | Metric    |    Value |
  +===========+==========+
  | Precision | 0.857562 |
  +-----------+----------+
  | Recall    | 0.818901 |
  +-----------+----------+
  | F1 Score  | 0.805628 |
  +-----------+----------+
  ```

Figure 59: Confusion Matrix and Accuracy, Precision, Recall, F1-Score for Linear Data.
  ```

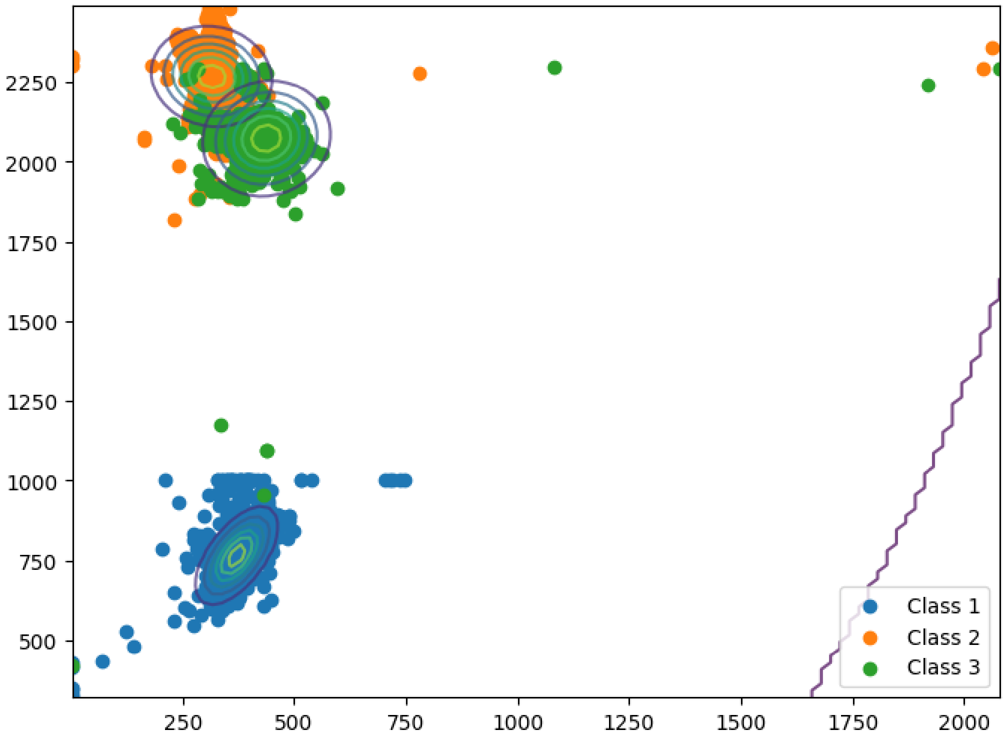


Figure 60: Contour Density Plot for the data.