Assignment1 Report

**Subject- Pattern Recognition Course code- CS669 Submitted to- Dr. A.D Dileep Department- CSE (SCEE) Institute- IIT MANDI**

**Group Number-18**

**Group Members Details- Rohit Roy: S22048**

**Abhishek Singh Rawat: T23191 Arpit Dua: T23192**

# Introduction:

To understand the task in hand some context needs to be laid which is as follows:

# Problem statement:

There are 3 datasets given to us in text format. Each dataset has further been divided into 3 classes.

These 3 classes may or may not be of the same size.

# Context of the problem:

Each data point which can also be referred to as sample is a 2-dimensional entity (which essentially means 2 features are there for each sample).

The primary task is to plot and observe the nature of decision boundaries and regions when we use the Bayes Decision classifier to classify an arbitrary data point and see that for all the data points in the graph which eventually gives a graph having the training data of the classes and colored regions equal to the number of the classes indicating that whichever point falls in a specific region is being predicted into the that class by the model which comes under this region.

Indirectly speaking any point in a specific decision region is an indication that most likely that point is being predicted by the model to come from that particular class (consequently the distribution associated with that class) which comes under that particular decision region.

What we want to do is see that for a given test data which class has the highest-class posterior probability for that data point.

We decide a specific color for any point that is being predicted into a certain class by our model and color that point.

When we do the same process for infinite samples, we will obtain the so-called decision region plot in which any point falling into a specific color region has been predicted by our model to belong to that particular class whose training data when plotted comes mostly under this region only.

# The following are some assumptions/salient points:

* The data of each class is coming from a distribution, what it essentially means is that every d dimensional data point of a certain class is nothing but equivalent to that of picking a random point from a d dimensional distribution (assumed to be gaussian in our case) which in turn is characterized by its parameters.
* We have assumed the distribution to be gaussian as:
  1. Normal distribution imitates most of the real-world situations (as a consequence of the central limit theorem).
  2. It is one of the most well-known distributions and has stood the test of the time.
  3. There is plenty of literature and results available for gaussian distribution to work with so it is easy and reliable to work with.
  4. The parameters and their expression for the gaussian distribution are simple to work with and visualize.
* We assume that the mathematical process (like maximum likelihood estimation etc.) to estimate the optimal mathematical formula for parameters of a specific distribution has been done and given to us.
* We also assume that the training/modelling part has been done and we have a discriminative model which can now be used for testing.
* The task in hand is essentially a classification task (supervised learning).
* We are using Bayes Decision Classifier for that purpose.

# Brief about Discriminative Models:

* In discriminative model predictions are made on unseen data based on conditional probabilities.
* The discriminative model are models used in statistical classification primarily for supervised machine learning.
* Another name of discriminative models is conditional models since they learn the boundaries between the classes or labels in the dataset.
* Discriminative models try to model the decision boundary between classes in a classification problem. The aim is to learn a function that maps inputs to binary outputs, indicating the class label of the input. Maximum likelihood estimation is often used to estimate the parameters of the discriminative model, such as the coefficients of a logistic regression model or the weights of a neural network.
* The discriminative model refers to a class of models used in **Statistical Classification**, mainly used for supervised machine learning. These types of models are also known

as **conditional models** since they learn the boundaries between classes or labels in a dataset.

* Discriminant function is learnt from the data only.
* Statistical ML algorithms use statistics of the data to classify and in this process of trying to classify test data samples correctly these ML algorithms indirectly learn the decision boundary.
* Data of each class is represented by a statistic.

.



A distribution characterised by its parameters. We assume that the training/modelling part has been done

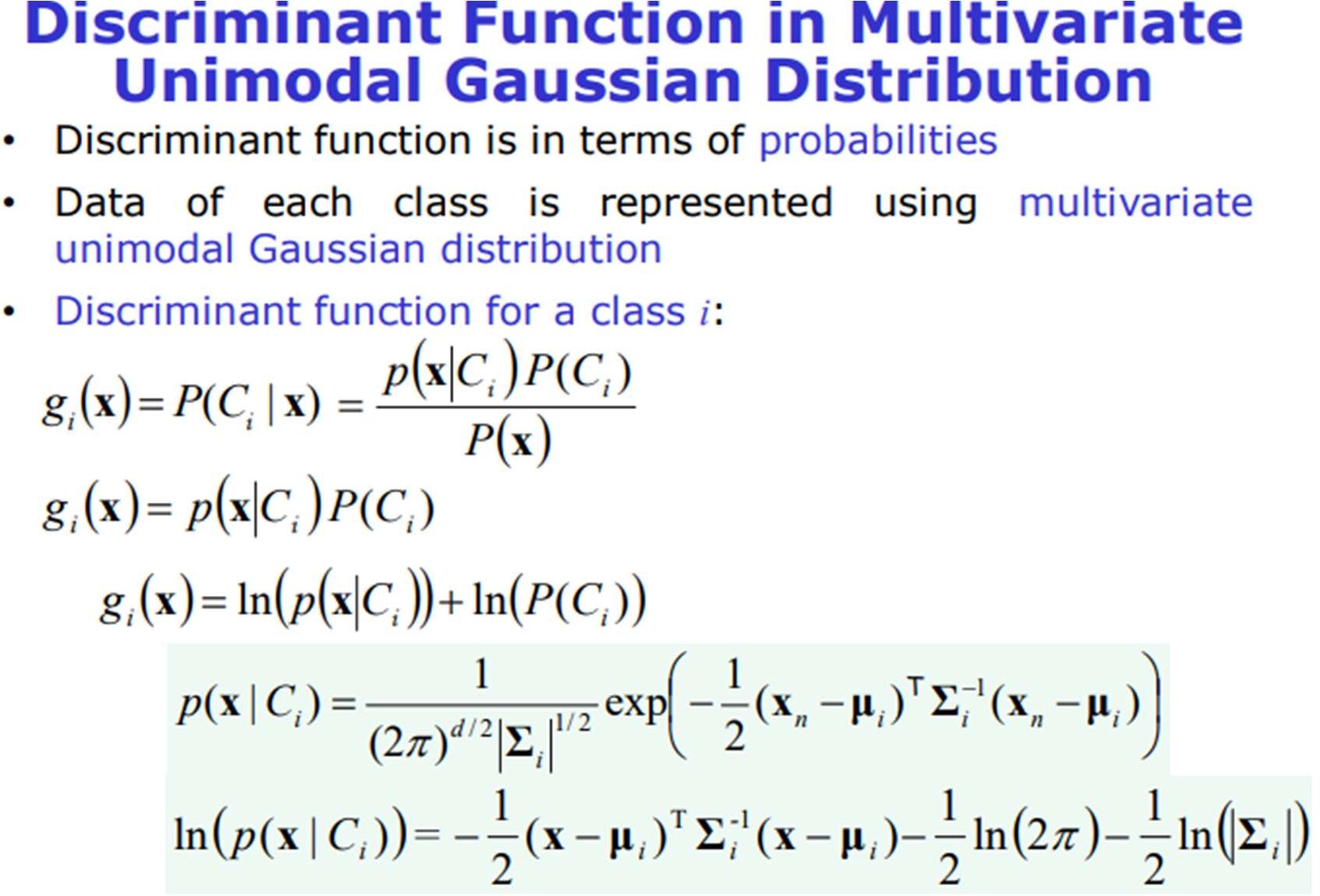
as well.

Samples/data points of a class are nothing but d dimensional data points picked from a

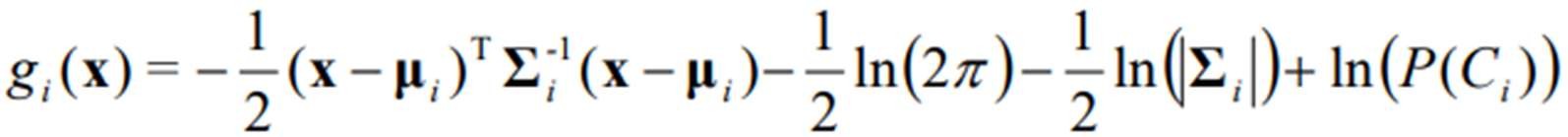
specific distribution.

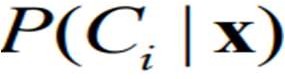
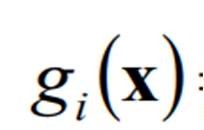
This phase is assumed to be done and the result of it have directly been used.

1. Calculate the parameter values for each of the class for a certain dataset. As discussed earlier that the exact mathematical formula to calculate these parameters is done via methods like maximum likelihood estimation and it is assumed to be given.
2. We use Bayes decision classifier. We need to predict that given an arbitrary sample data point/test data point which may have come from one of the training classes (indirectly speaking from a distribution of the class) or maybe from some other unknown distribution what is the probability that it belongs to a certain class.
3. Since we have the evidence here so bayes theorem is the one which comes to the mind naturally.
4. When the distribution with which we are representing a class is indeed the true distribution of the class the bayes classifier is the one with minimum error.
5. Here we have removed the denominator term as it is the probability of the evidence which will essentially be the same term in the formula of individual discriminant functions and hence will not help in discriminating between a set of classes. Also, it is the only term independent of class i, hence not considering in all the cases is a sensible thing to do.
6. We have taken the log both sides because:
   * We don’t want values to become small so that calculation remains easy and problem of significance doesn’t come.
   * We don’t want to deal with the exponential term in the formula of the likelihood function and taking log of it solves it.
   * We take the natural log by default.
   * A general choice of a monotonically increasing function is good for our usecase and log base e satisfies that.



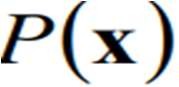
The following is the final result for the discriminant function for a certain class ci

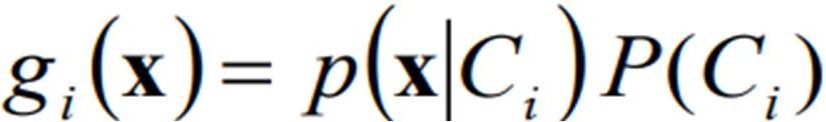


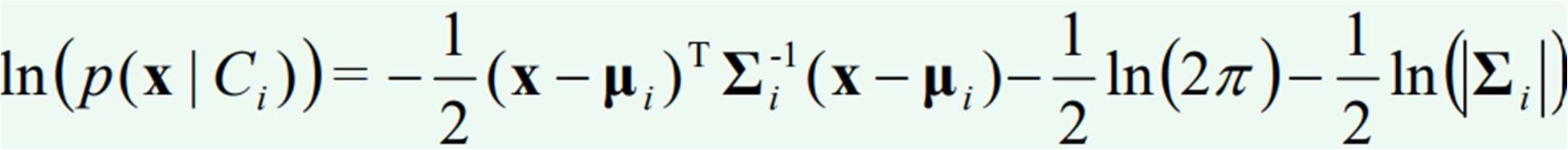
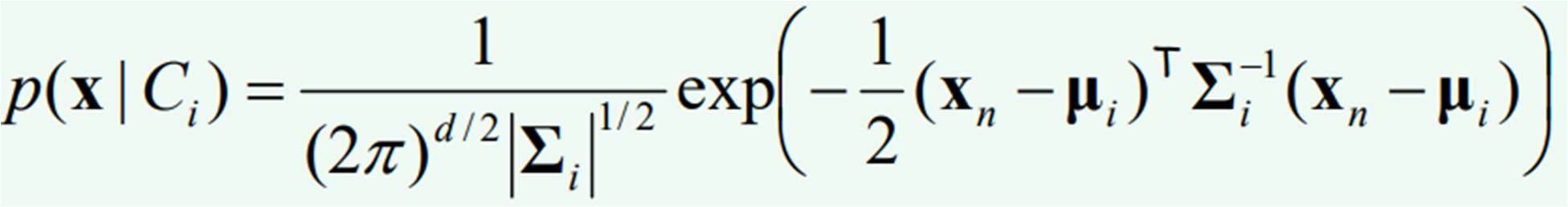
 =Class Posterior Probability (The Probability that given a random sample x the probability of it coming from class C).

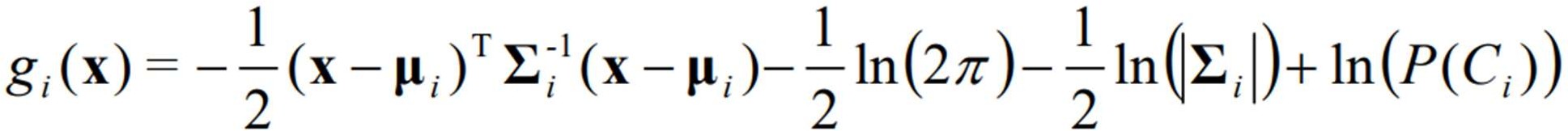
= The equation of the discriminant function for a certain class i.

 =likelihood of a sample coming from a class ci  =Prior of a certain class.

= Evidence or total probability



It is necessary to take the log likelihood so that we don’t have



The above one is the equation of the discriminant function of a class i.

# Results

1. **Results for Dataset1 (LSD)**

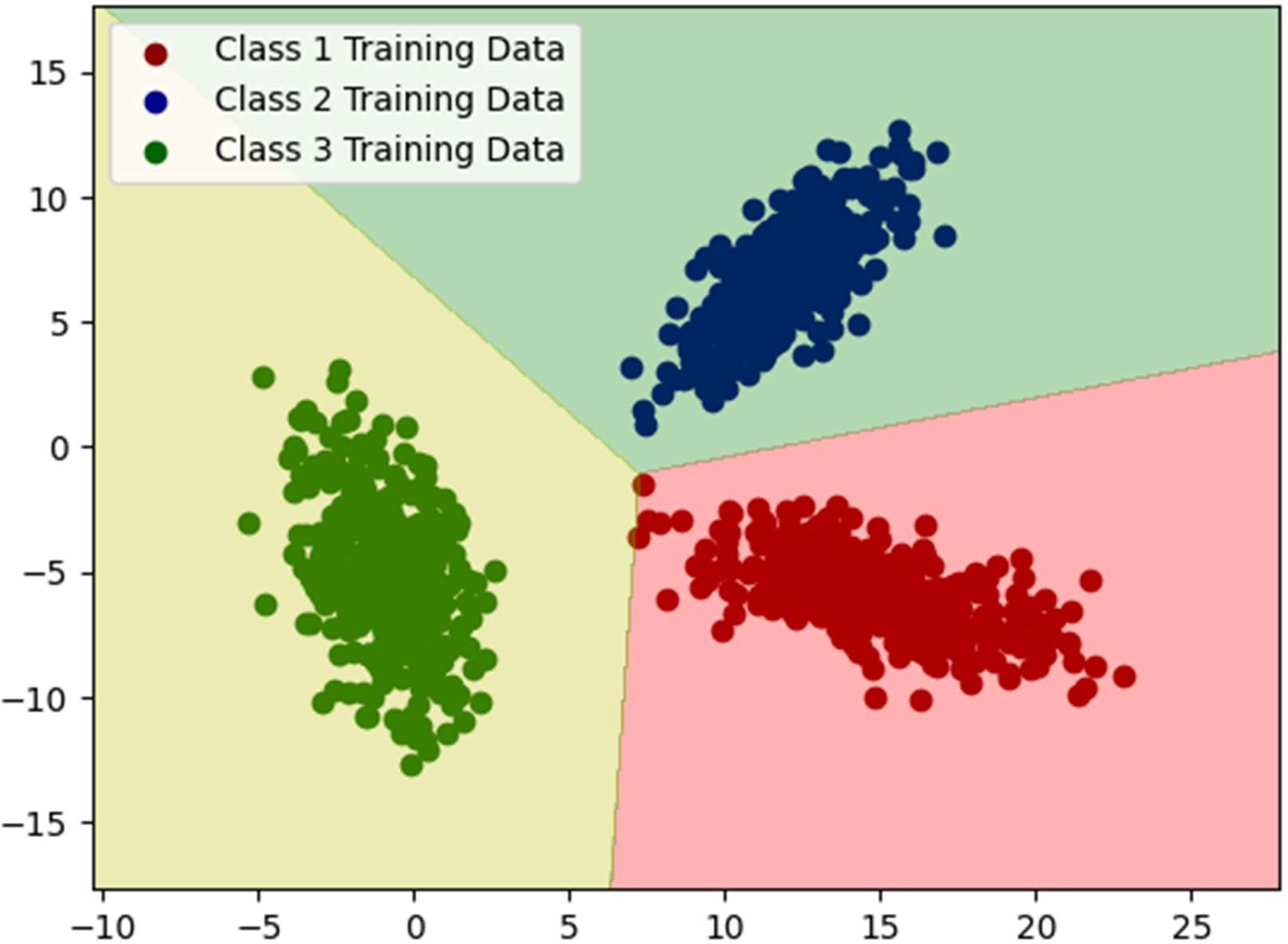
# General plot of train and test data for Dataset1 without considering any case as of now

* The 1st graph for the Dataset1 comprises of the training data and testing data (70% and 30% respectively) superimposed on each other. Label has been attached for the correspondence.
* The respective means of all the three classes have also been plotted.



# Dataset1 (LSD) Case1 Cumulative Decision Region

* The graph consists of cumulative decision region for case1 for Dataset1, the final decision region where training data of all the three classes has been plotted and the region has come after considering each minute pixel as a data point. The posterior probabilities have been calculated for all three classes for that data point and whichever one is the highest the data point will go into that class or colored with a specific color with training data of the class being superimposed in the decision region.

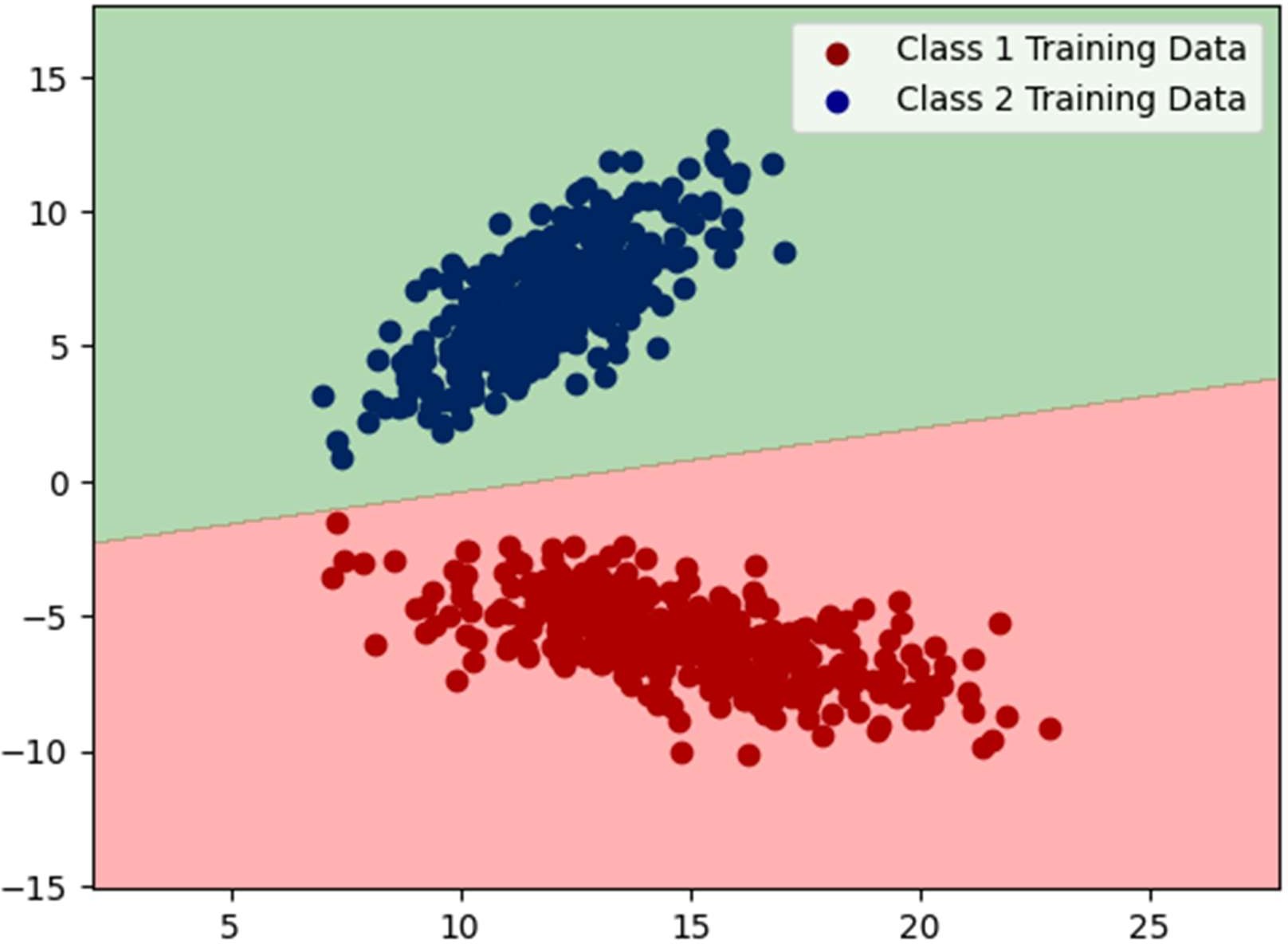


# Decision Region plot for combination of classes

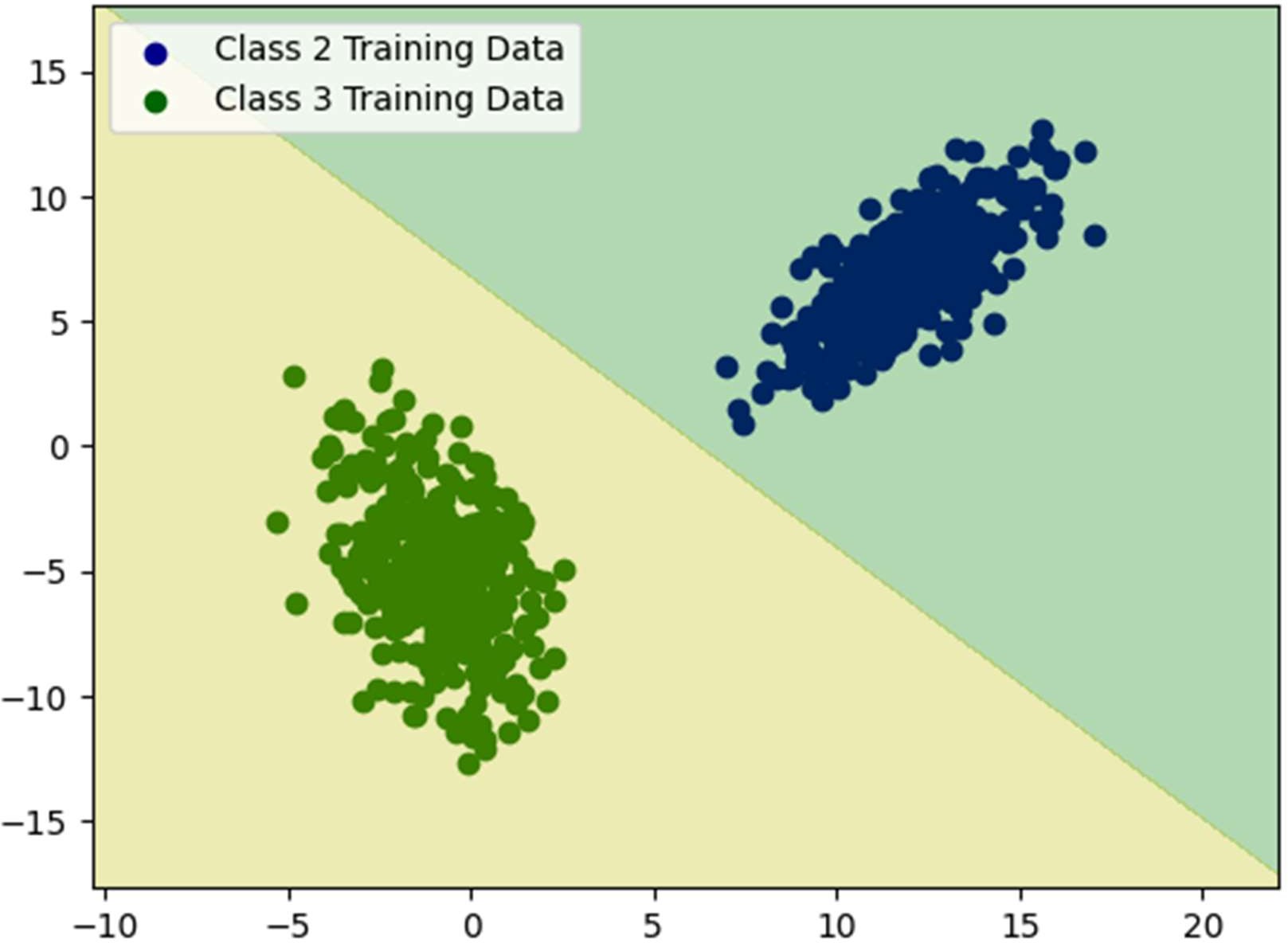
What we do in this section is simply plotting decision region just like above except that we consider 2 classes only instead of three so there will be 3 graphs having decision regions.

# Dataset1 (LSD) Case1 Decision Region for class1 and class2

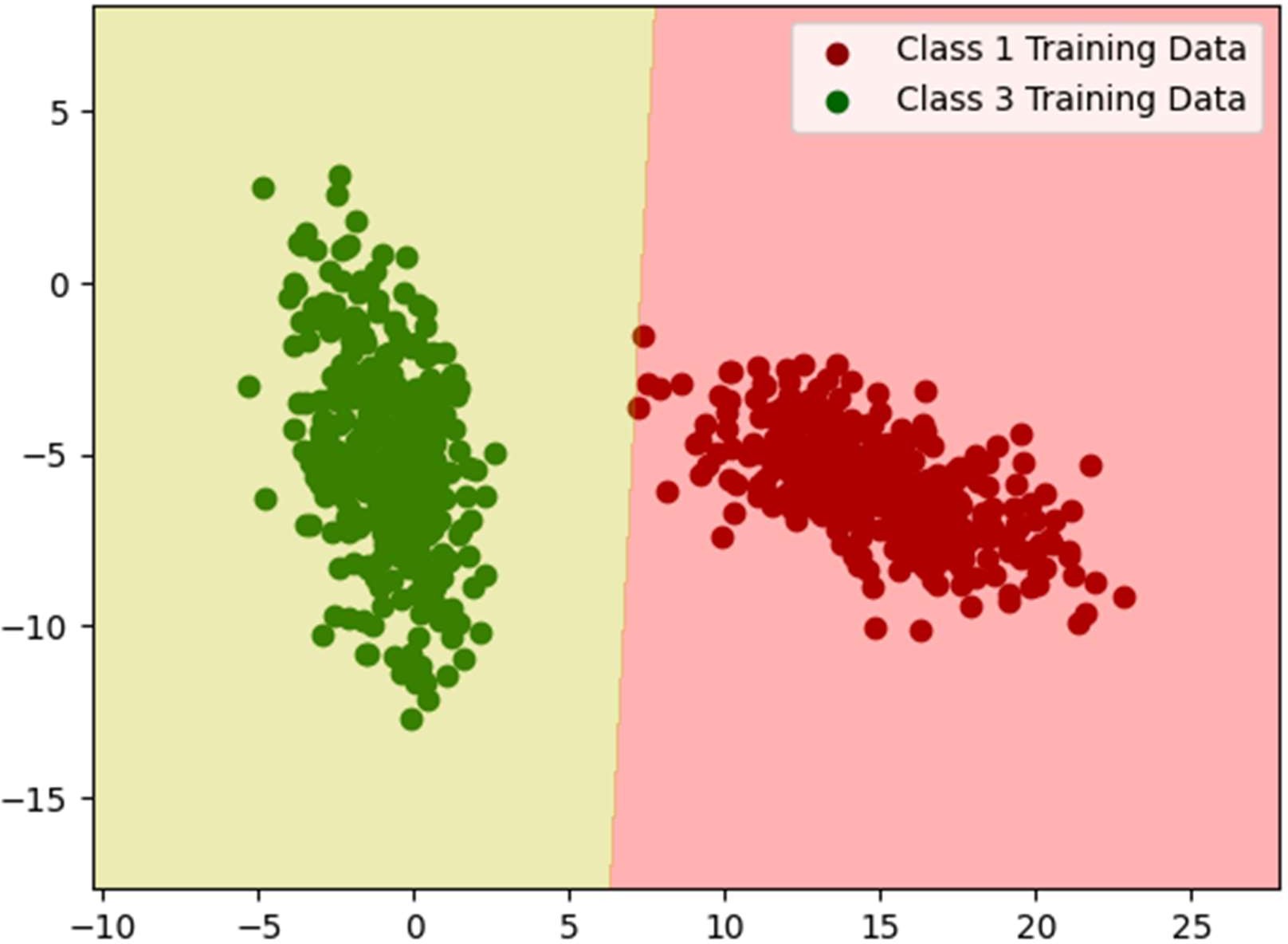
The following is the graph for Dataset1 (LSD) Case1 Decision Region for class1 and class2



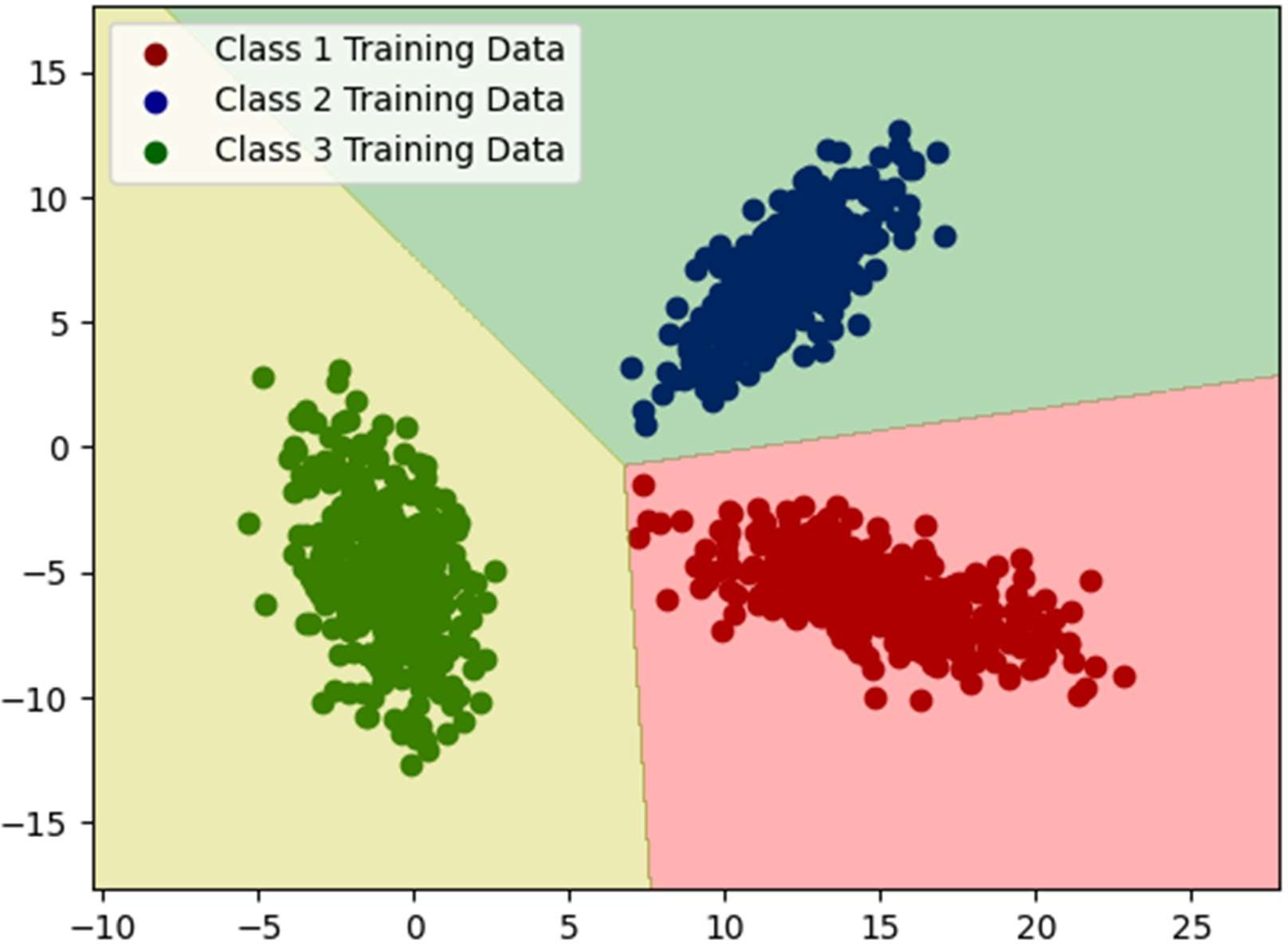
The following is the graph for Dataset1 (LSD) Case1 Decision Region for class2 and class3



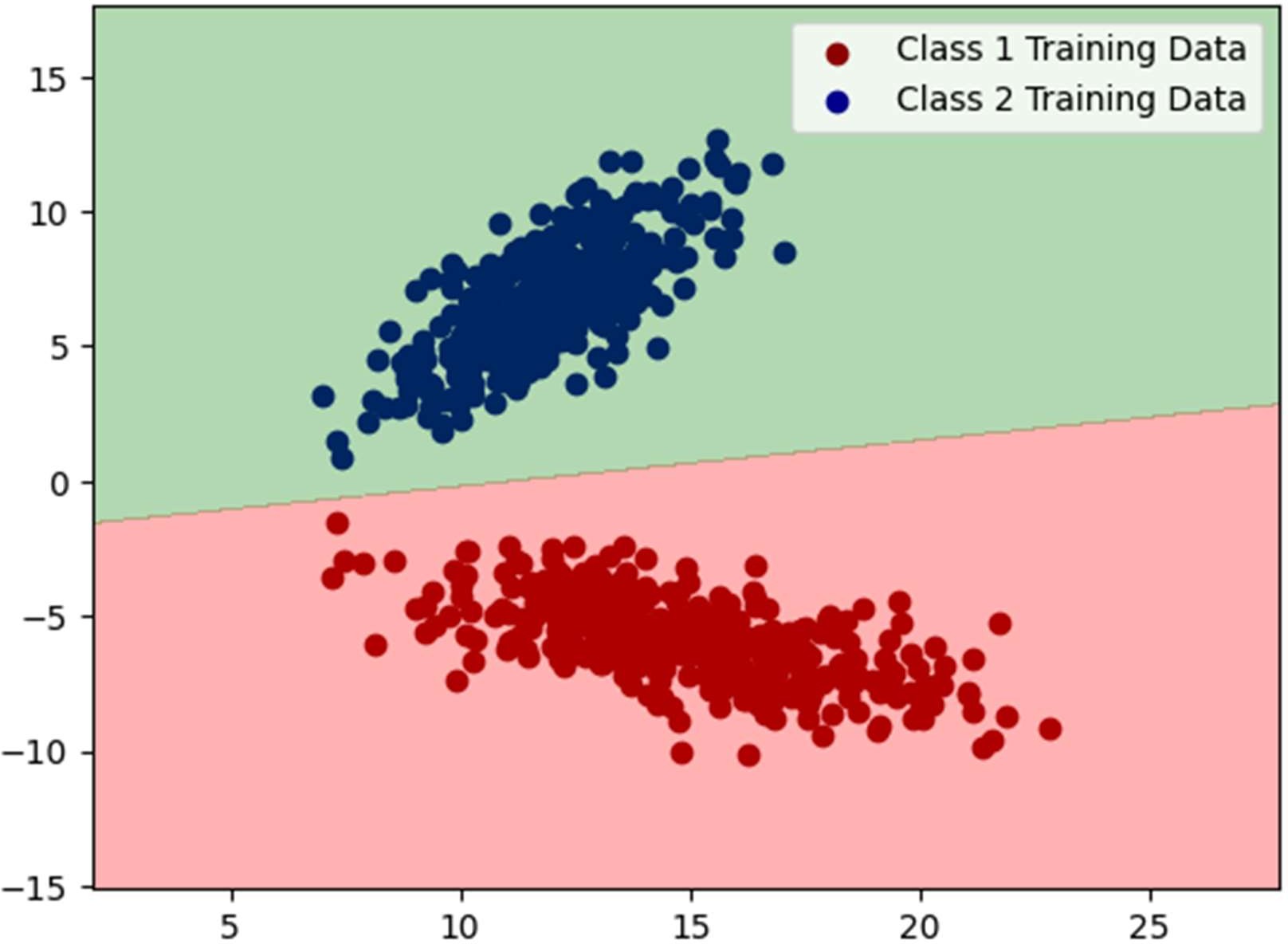
The following is the graph for Dataset1 (LSD) Case1 Decision Region for class3 and class1.



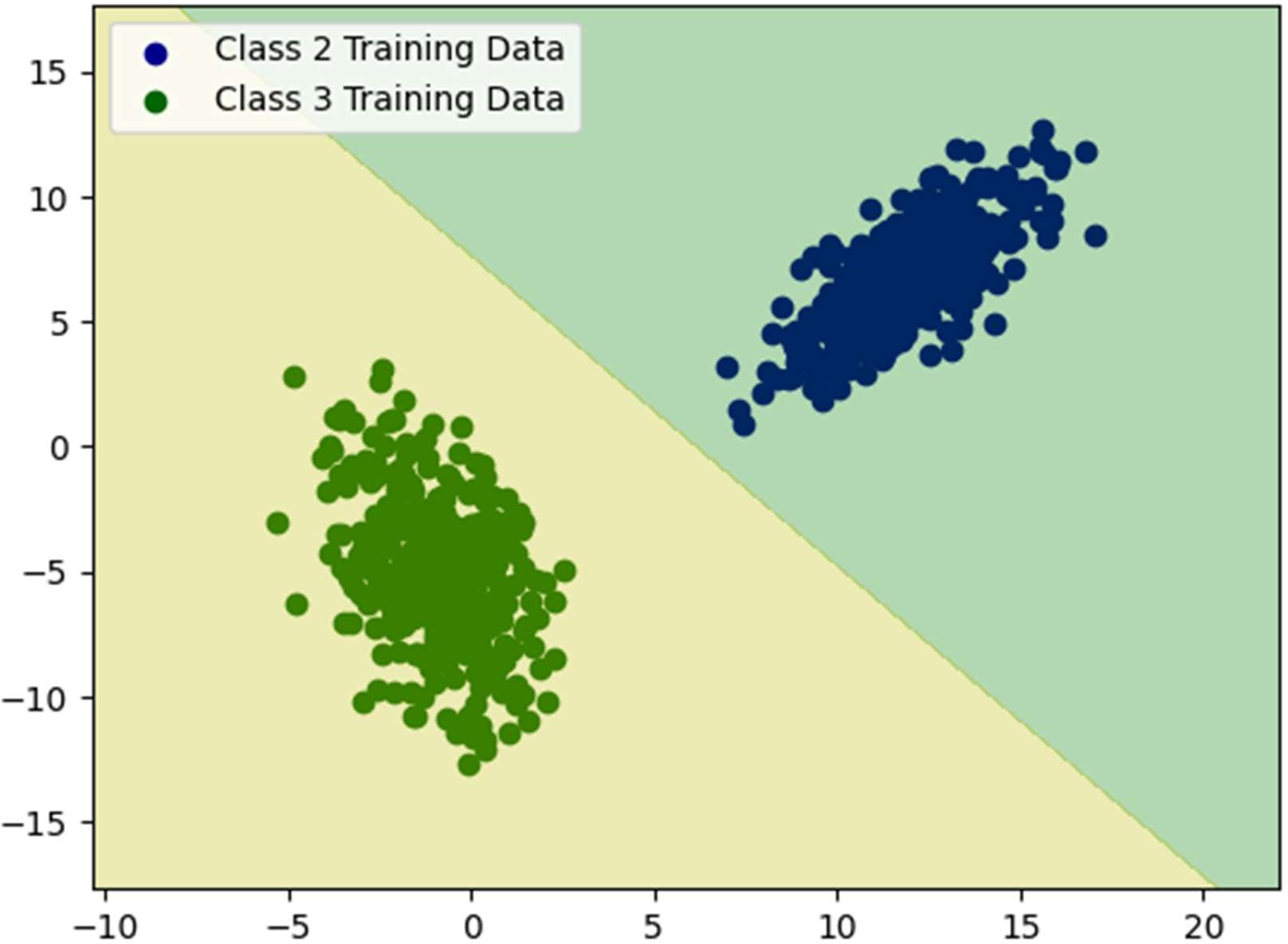
* The graph consists of cumulative decision region for case2 for Dataset1, the final decision region where training data of all the three classes has been plotted and the region has come after considering each minute pixel as a data point. The posterior probabilities have been calculated for all three classes for that data point and whichever one is the highest the data point will go into that class or colored with a specific color with training data of the class being superimposed in the decision region.



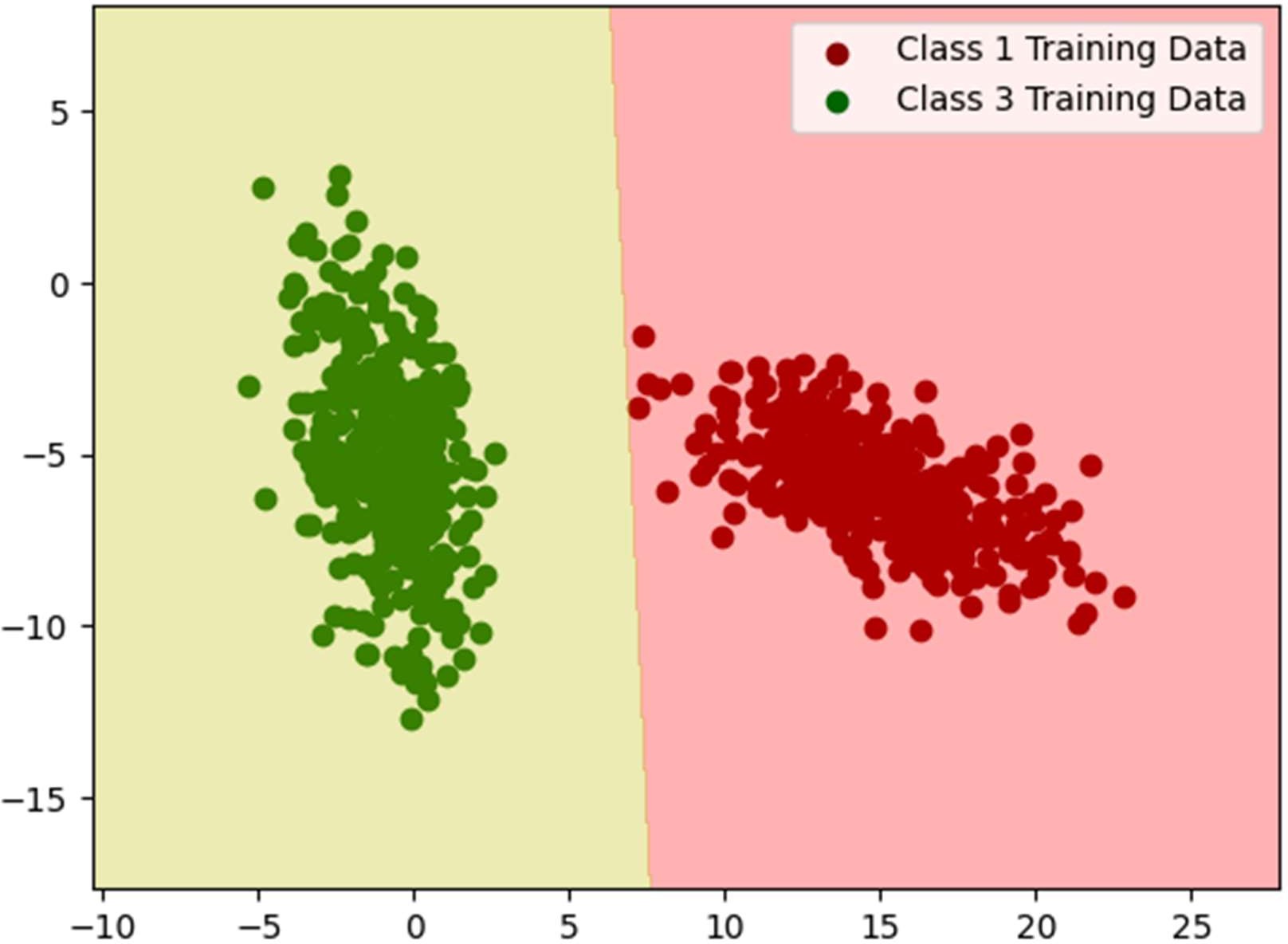
The following is the graph for Dataset1 (LSD) Case2 Decision Region for class1 and class2.



The following is the graph for Dataset1 (LSD) Case2 Decision Region for class3 and class2.

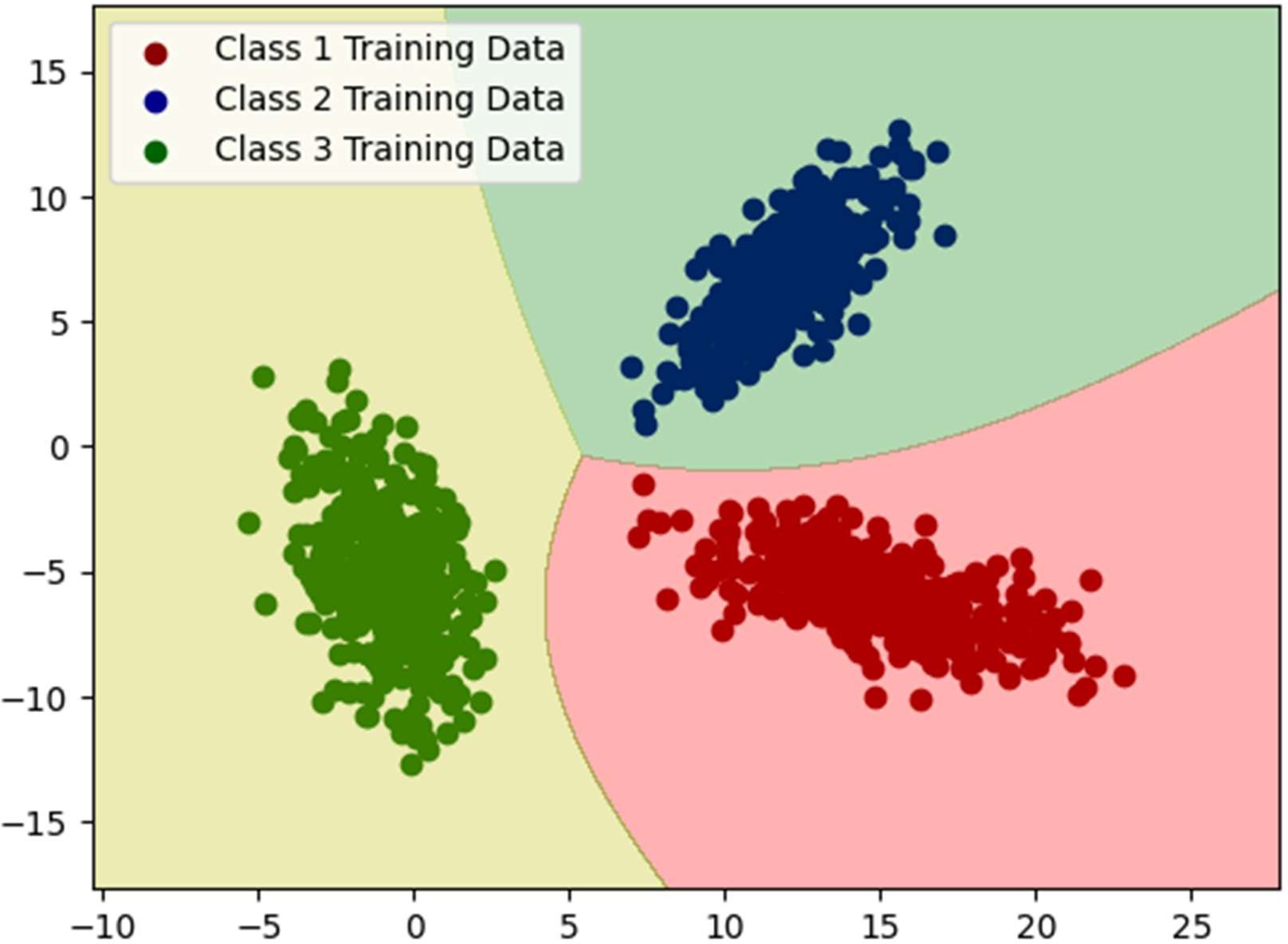


The following is the graph for Dataset1 (LSD) Case2 Decision Region for class3 and class1.

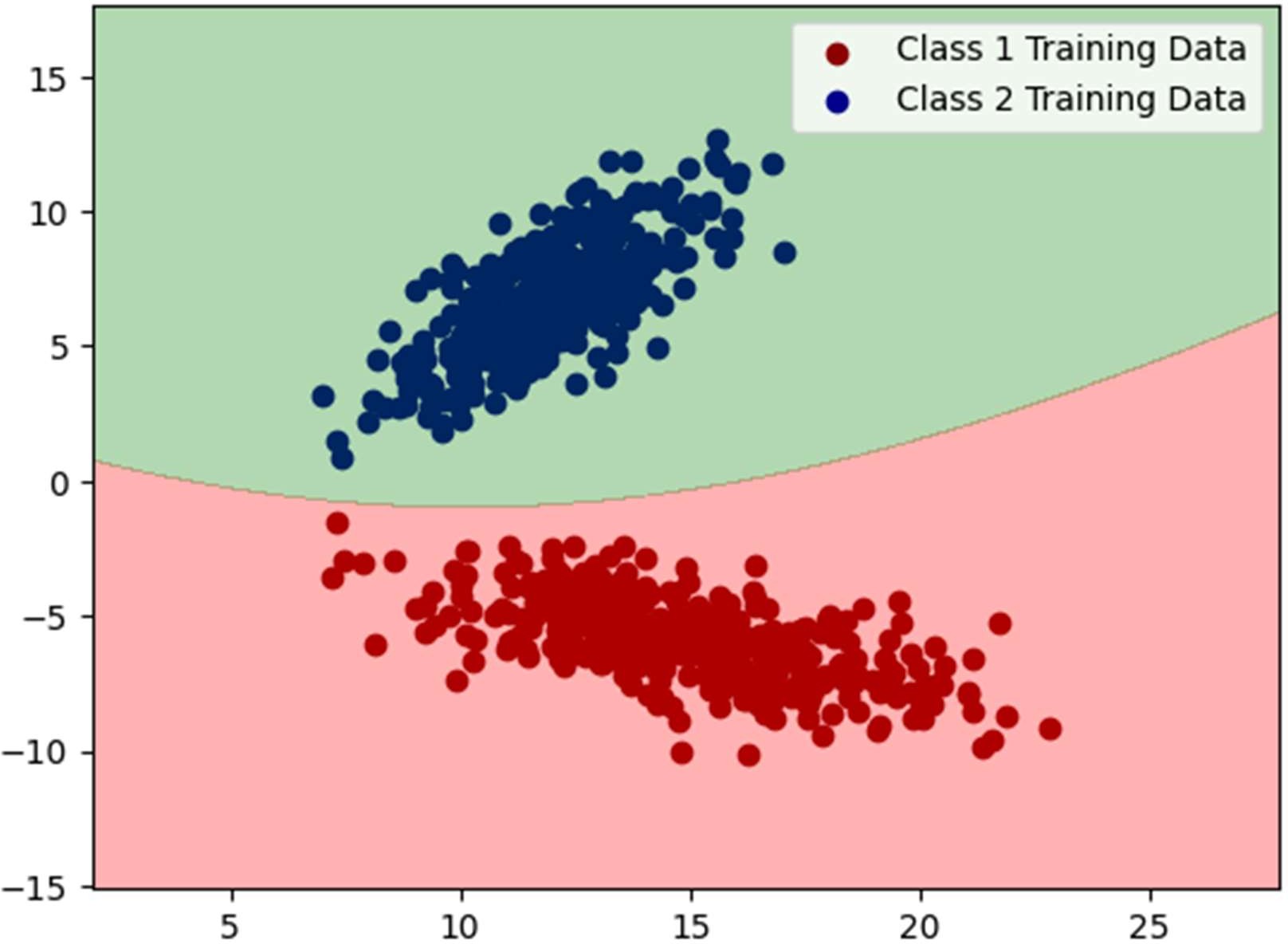


# Dataset1 (LSD) Case3 Cumulative Decision Region

* The graph consists of cumulative decision region for case3 for Dataset1, the final decision region where training data of all the three classes has been plotted and the region has come after considering each minute pixel as a data point. The posterior probabilities have been calculated for all three classes for that data point and whichever one is the highest the data point will go into that class or colored with a specific color with training data of the class being superimposed in the decision region.



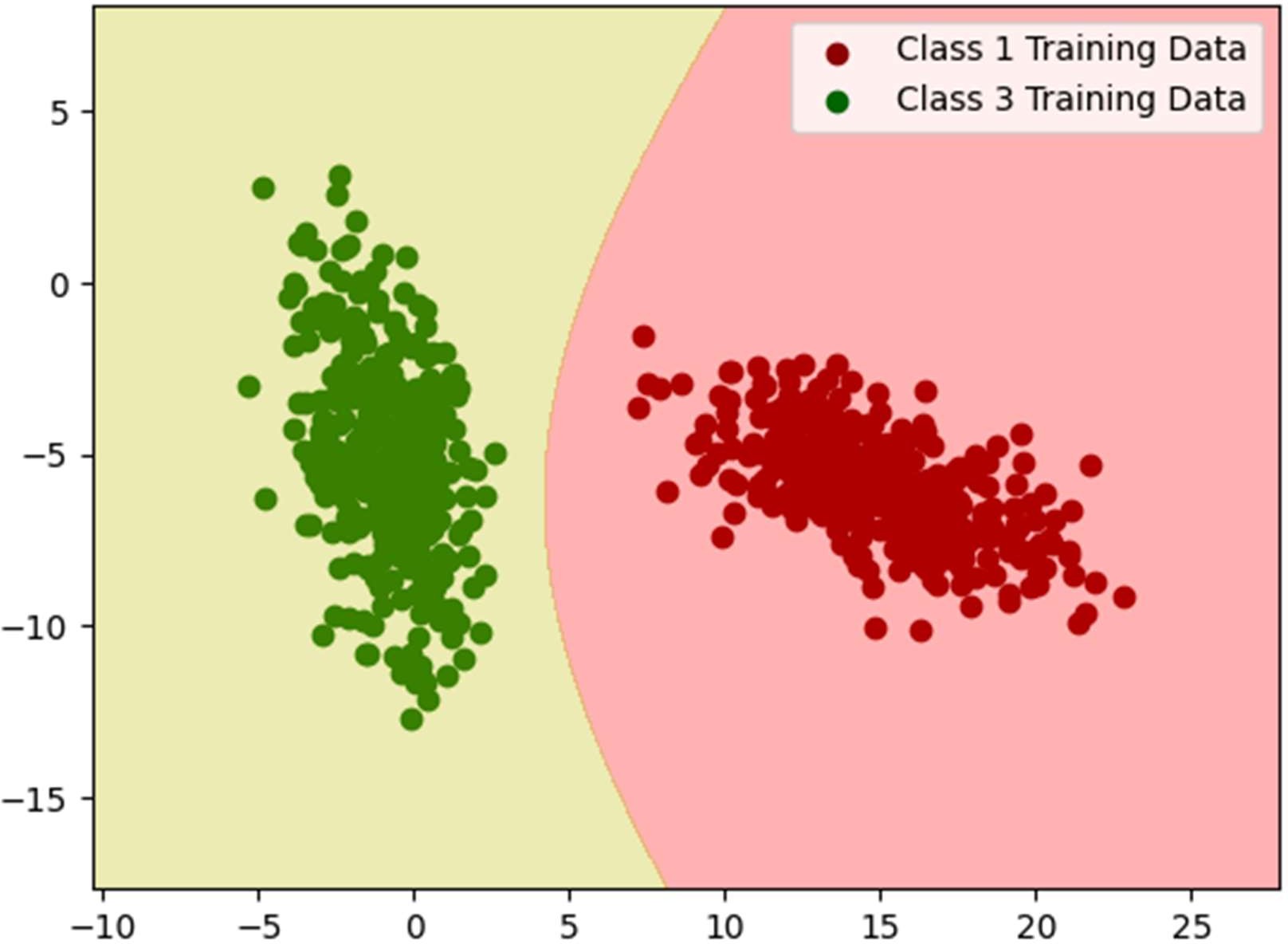
The following is the graph for Dataset1 (LSD) Case2 Decision Region for class1 and class2.



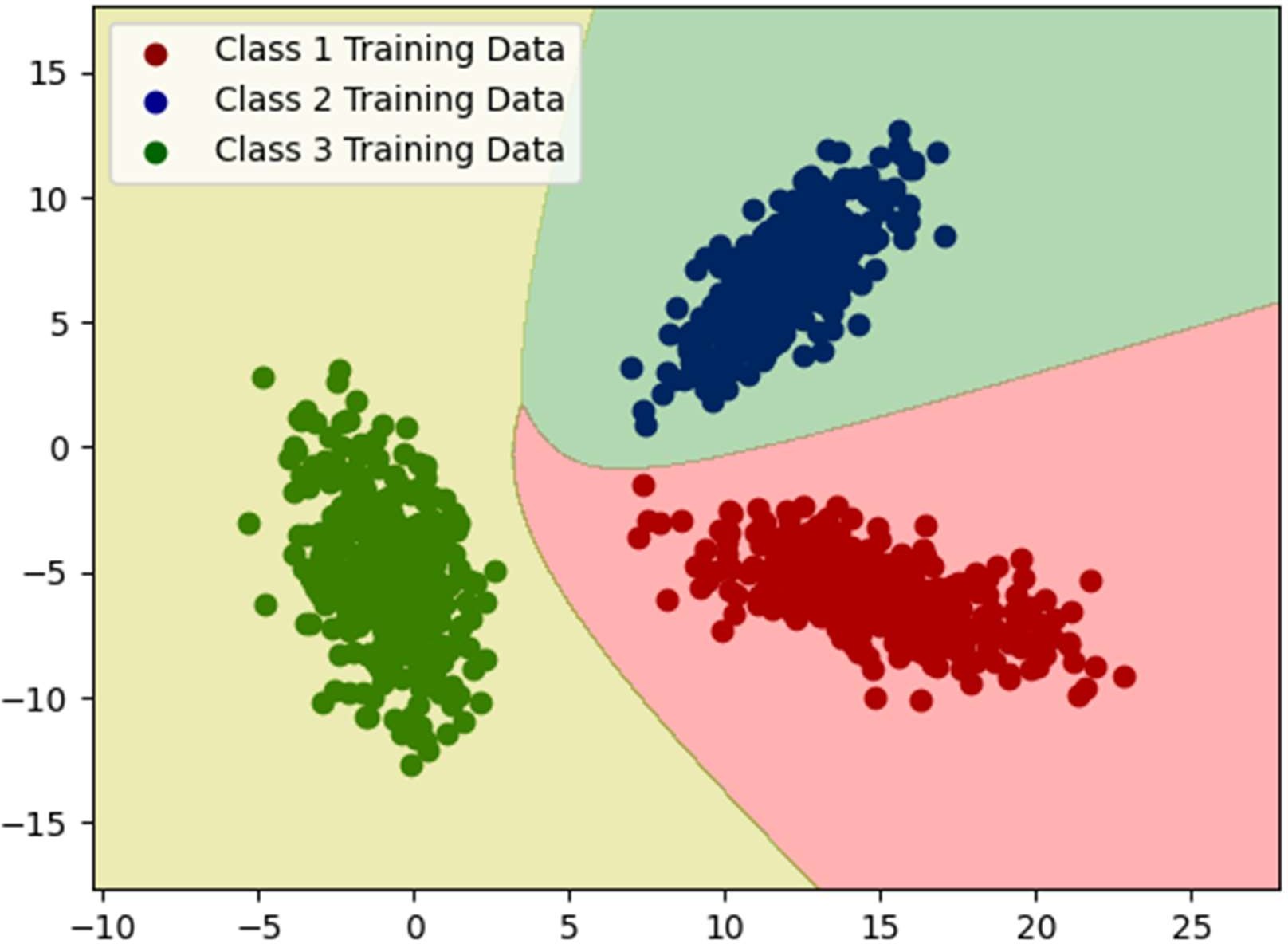
The following is the graph for Dataset1 (LSD) Case2 Decision Region for class3 and class2.



The following is the graph for Dataset1 (LSD) Case2 Decision Region for class3 and class1.

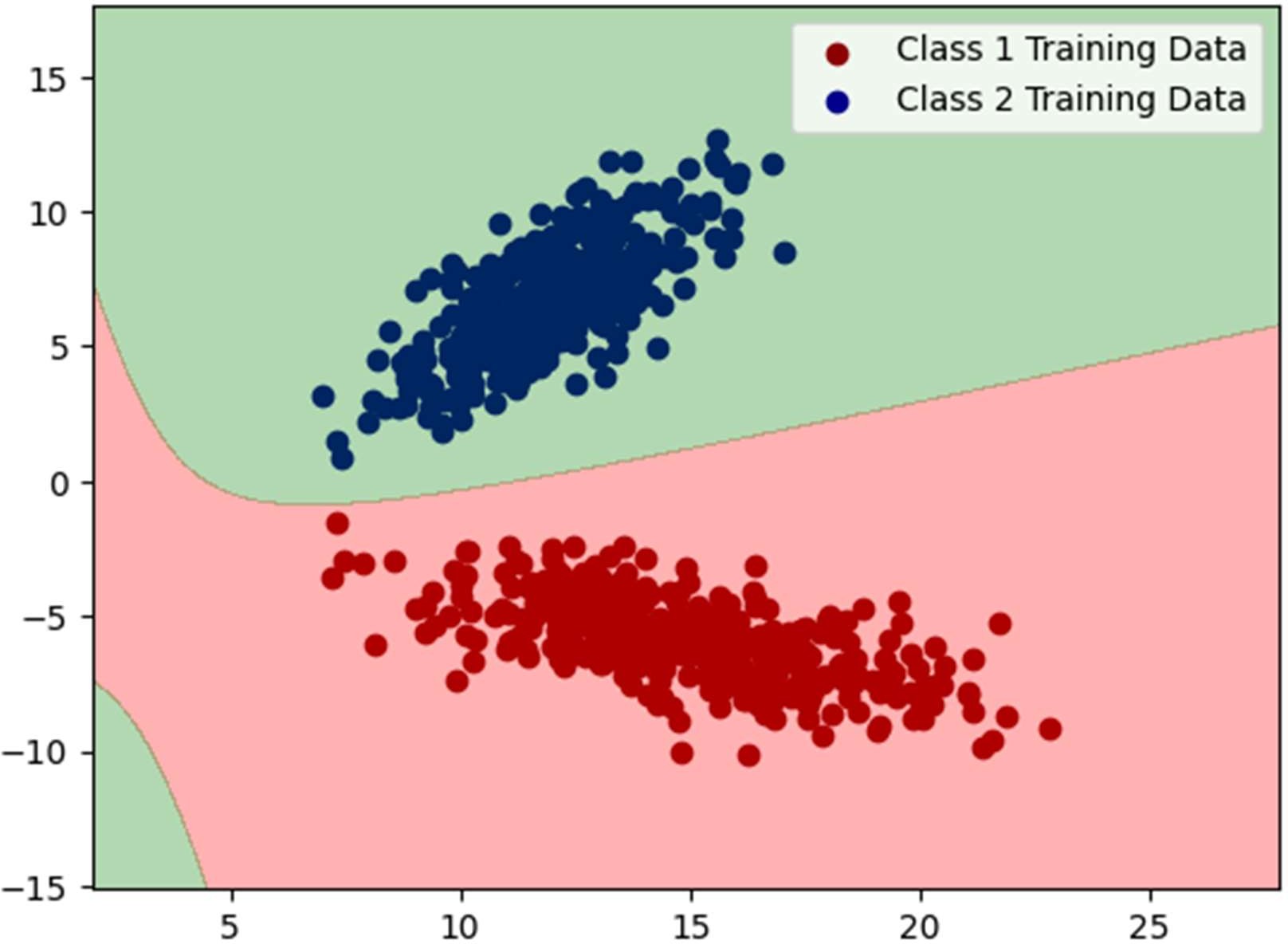


* The graph consists of cumulative decision region for case4 for Dataset1, the final decision region where training data of all the three classes has been plotted and the region has come after considering each minute pixel as a data point. The posterior probabilities have been calculated for all three classes for that data point and whichever one is the highest the data point will go into that class or colored with a specific color with training data of the class being superimposed in the decision region.

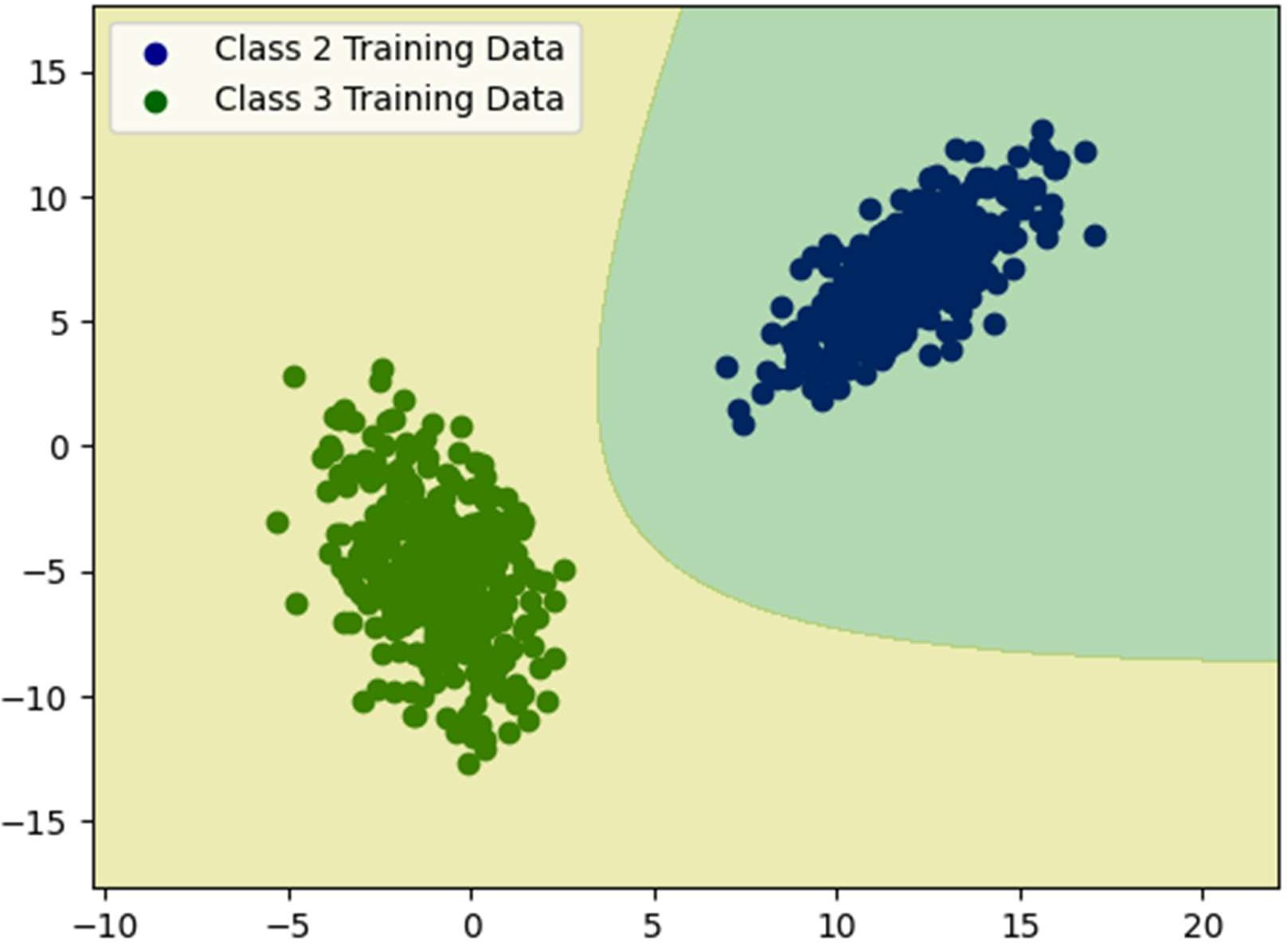


# Dataset1 (LSD) Case4 Decision Region for class1 and class2

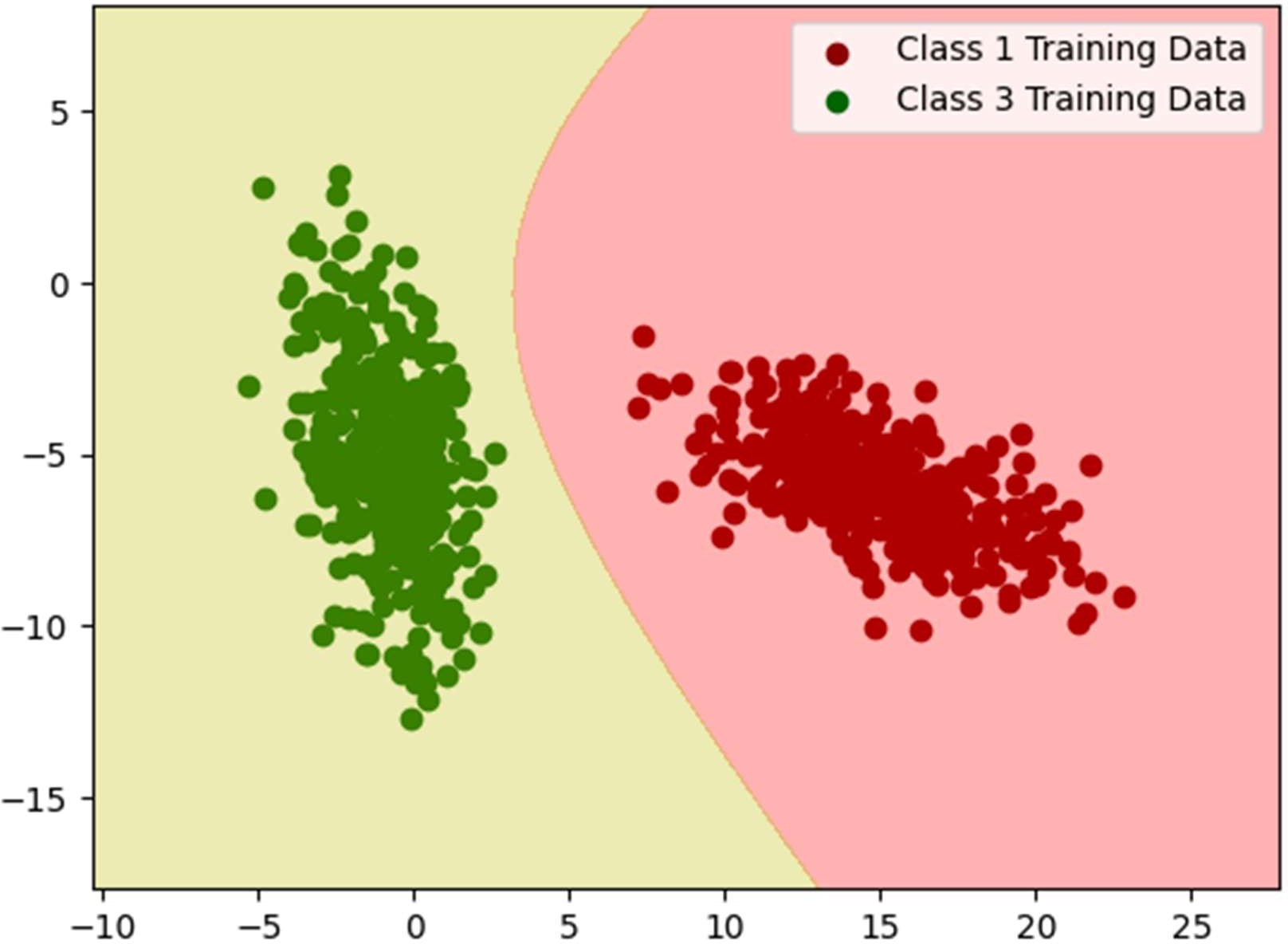
The following is the graph for Dataset1 (LSD) Case2 Decision Region for class1 and class2.



The following is the graph for Dataset1 (LSD) Case2 Decision Region for class3 and class2.

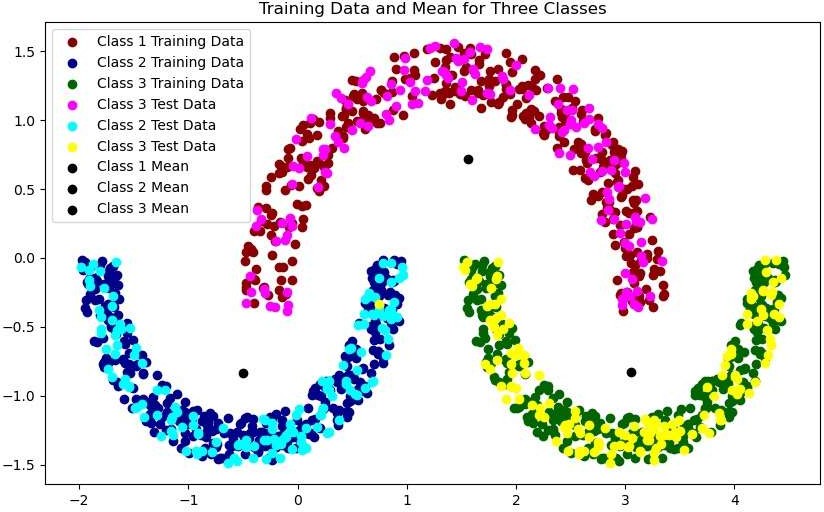


The following is the graph for Dataset1 (LSD) Case2 Decision Region for class3 and class1.



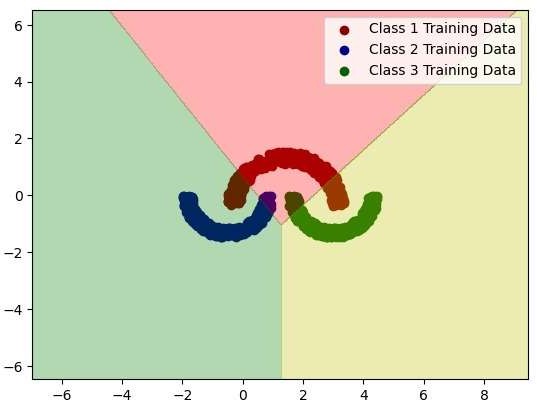
# General plot of train and test data for Dataset2 without considering any case as of now

* The 1st graph for the Dataset2 comprises of the training data and testing data (70% and 30% respectively) superimposed on each other. Label has been attached for the correspondence.
* The respective means of all the three classes have also been plotted.



# Dataset2 (NLS) Case1 Cumulative Decision Region

* The graph consists of cumulative decision region for case1 for Dataset2, the final decision region where training data of all the three classes has been plotted and the region has come after considering each minute pixel as a data point. The posterior probabilities have been calculated for all three classes for that data point and whichever one is the highest the data point will go into that class or colored with a specific color with training data of the class being superimposed in the decision region.

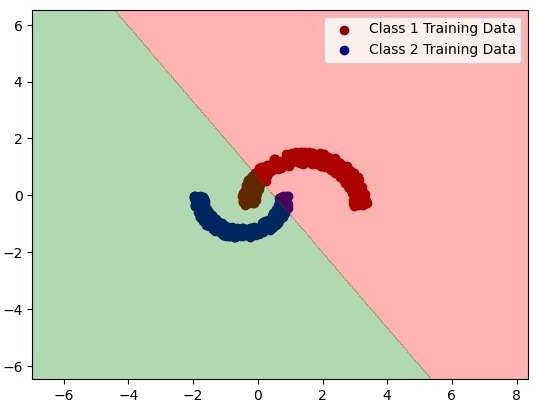


# Decision Region plot for combination of classes

What we do in this section is simply plotting decision region just like above except that we consider 2 classes only instead of three so there will be 3 graphs having decision regions.

# Dataset2 (NLS) Case1 Decision Region for class1 and class2

The following is the graph for Dataset2 (NLS) Case1 Decision Region for class1 and class2



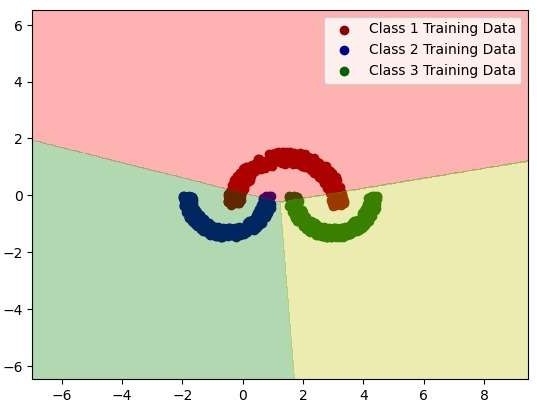
The following is the graph for Dataset2 (NLS) Case1 Decision Region for class2 and class3



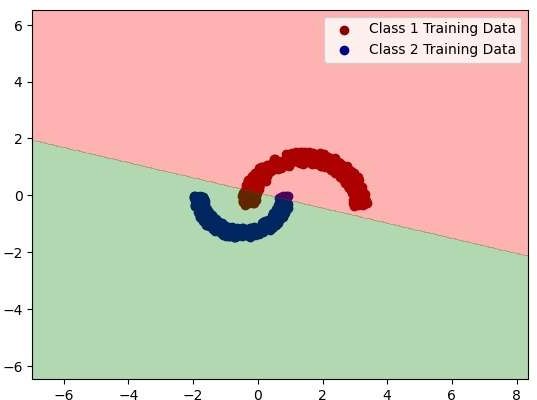
The following is the graph for Dataset2 (NLS) Case1 Decision Region for class3 and class1.



* The graph consists of cumulative decision region for case2 for Dataset2, the final decision region where training data of all the three classes has been plotted and the region has come after considering each minute pixel as a data point. The posterior probabilities have been calculated for all three classes for that data point and whichever one is the highest the data point will go into that class or colored with a specific color with training data of the class being superimposed in the decision region.



The following is the graph for Dataset2 (NLS) Case2 Decision Region for class1 and class2.



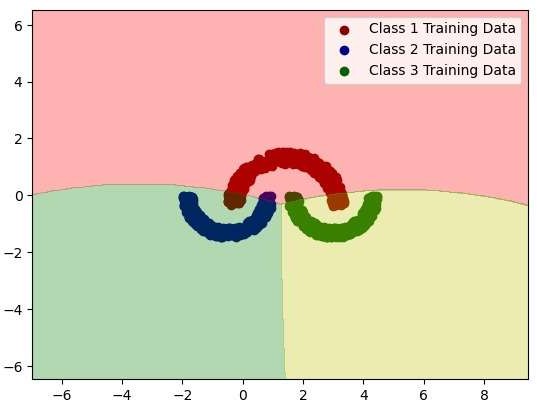
The following is the graph for Dataset2 (NLS) Case2 Decision Region for class3 and class2.



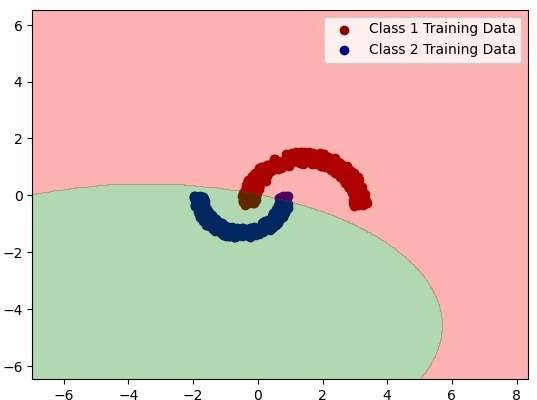
The following is the graph for Dataset2 (NLS) Case2 Decision Region for class3 and class1.



* The graph consists of cumulative decision region for case3 for Dataset2, the final decision region where training data of all the three classes has been plotted and the region has come after considering each minute pixel as a data point. The posterior probabilities have been calculated for all three classes for that data point and whichever one is the highest the data point will go into that class or colored with a specific color with training data of the class being superimposed in the decision region.



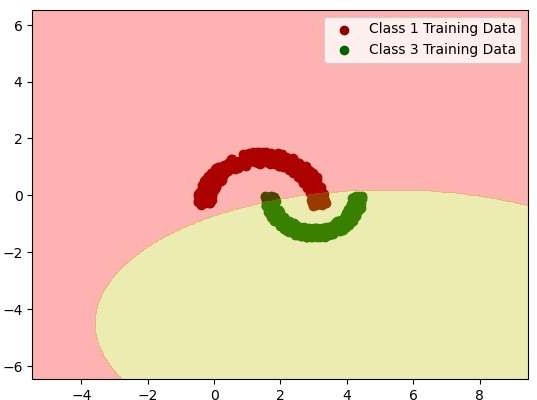
The following is the graph for Dataset2 (NLS) Case2 Decision Region for class1 and class2.



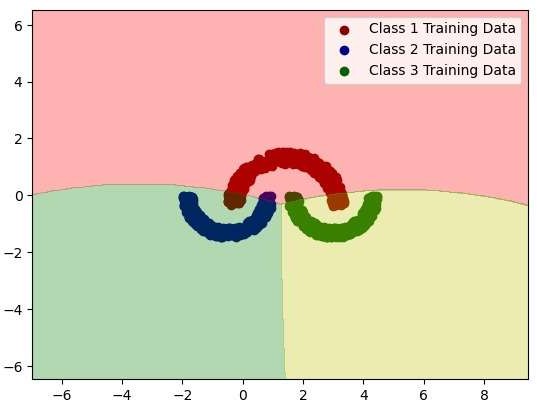
The following is the graph for Dataset2 (NLS) Case2 Decision Region for class3 and class2.



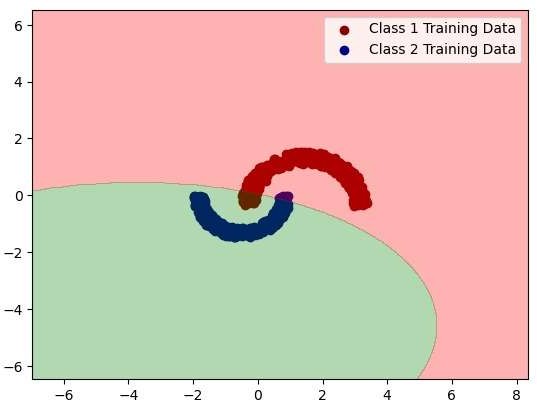
The following is the graph for Dataset2 (NLS) Case2 Decision Region for class3 and class1.



* The graph consists of cumulative decision region for case4 for Dataset2, the final decision region where training data of all the three classes has been plotted and the region has come after considering each minute pixel as a data point. The posterior probabilities have been calculated for all three classes for that data point and whichever one is the highest the data point will go into that class or colored with a specific color with training data of the class being superimposed in the decision region.



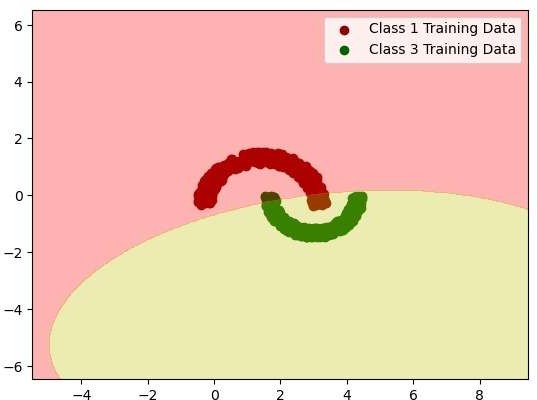
The following is the graph for Dataset2 (NLS) Case2 Decision Region for class1 and class2.



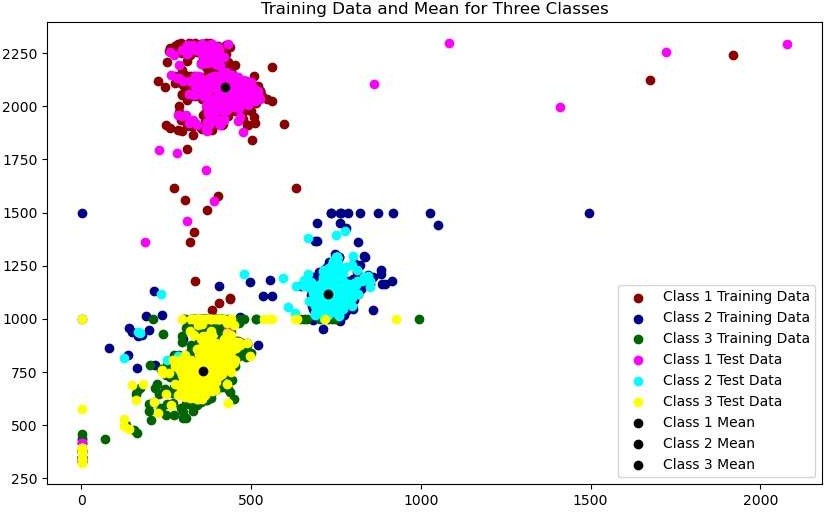
The following is the graph for Dataset2 (NLS) Case2 Decision Region for class3 and class2.

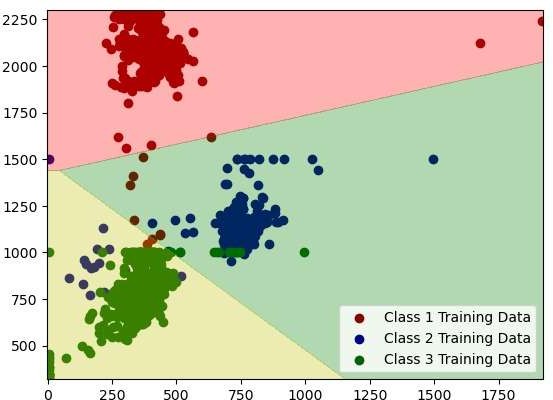


The following is the graph for Dataset2 (NLS) Case2 Decision Region for class3 and class1.



# General plot of train and test data for Dataset3 without considering any case as of now

* The 1st graph for the Dataset3 comprises of the training data and testing data (70% and 30% respectively) superimposed on each other. Label has been attached for the correspondence.
* The respective means of all the three classes have also been plotted.
* The graph consists of cumulative decision region for case1 for Dataset3, the final decision region where training data of all the three classes has been plotted and the region has come after considering each minute pixel as a data point. The posterior probabilities have been calculated for all three classes for that data point and whichever one is the highest the data point will go into that class or colored with a specific color with training data of the class being superimposed in the decision region.

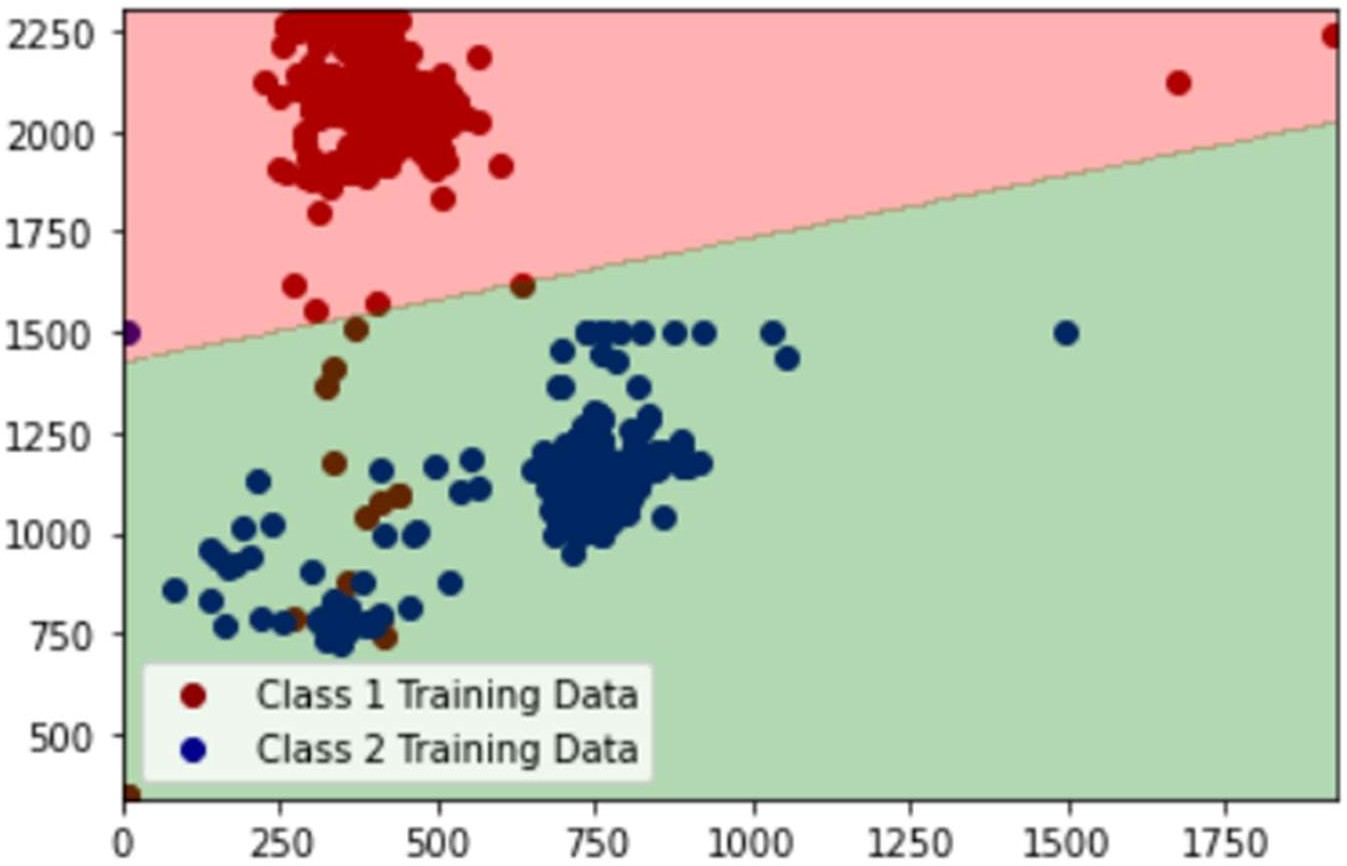


# Decision Region plot for combination of classes

What we do in this section is simply plotting decision region just like above except that we consider 2 classes only instead of three so there will be 3 graphs having decision regions.

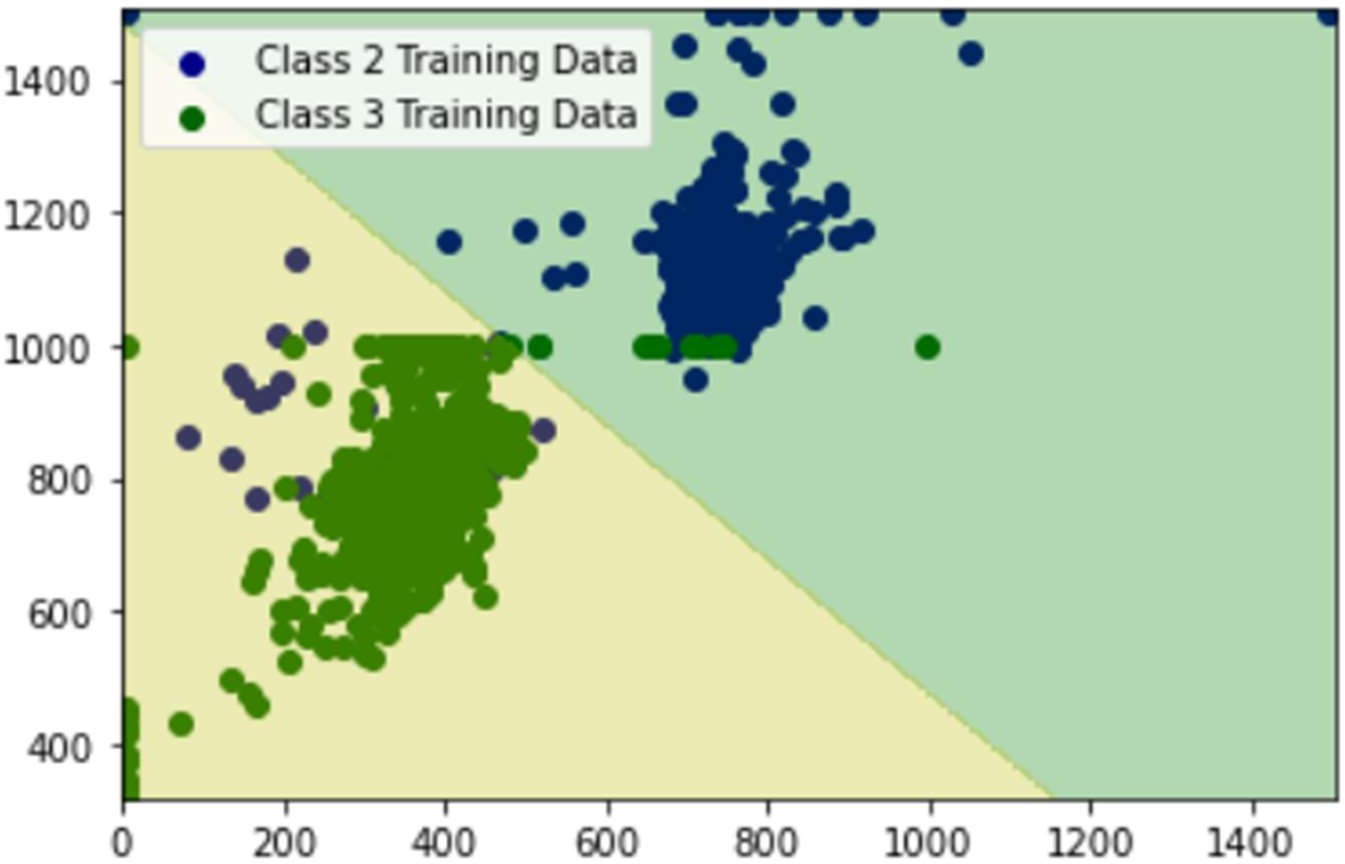
# Dataset3 (rwd) Case1 Decision Region for class1 and class2

The following is the graph for Dataset3 (rwd) Case1 Decision Region for class1 and class2



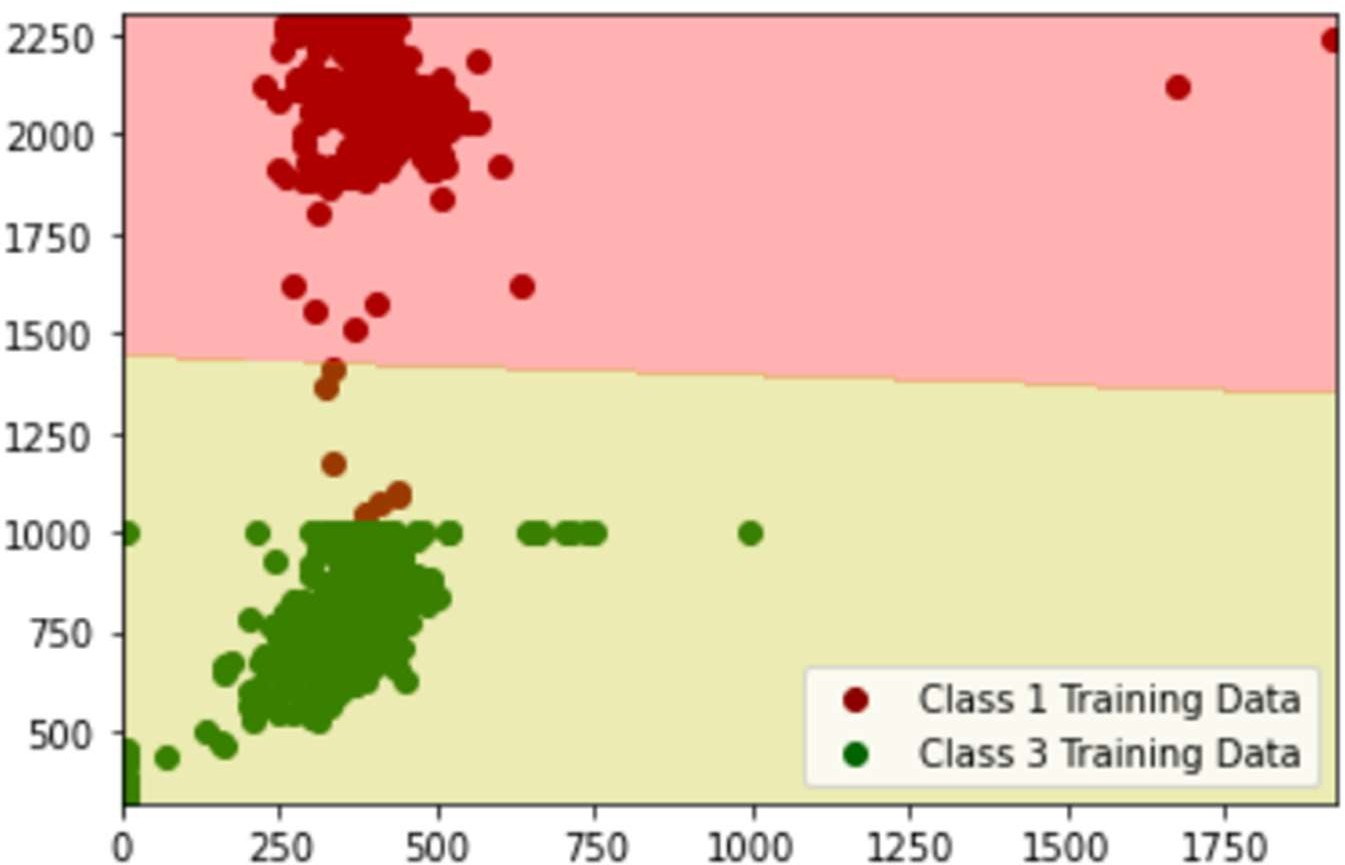
# Dataset3 (RWD) Case1 Decision Region for class2 and class3

The following is the graph for Dataset3 (RWD) Case1 Decision Region for class2 and class3



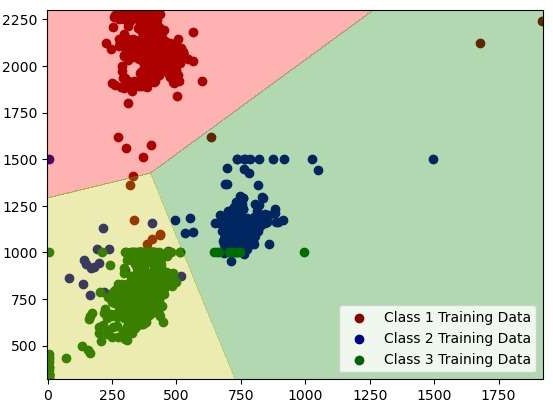
# Dataset3 (RWD) Case1 Decision Region for class3 and class1

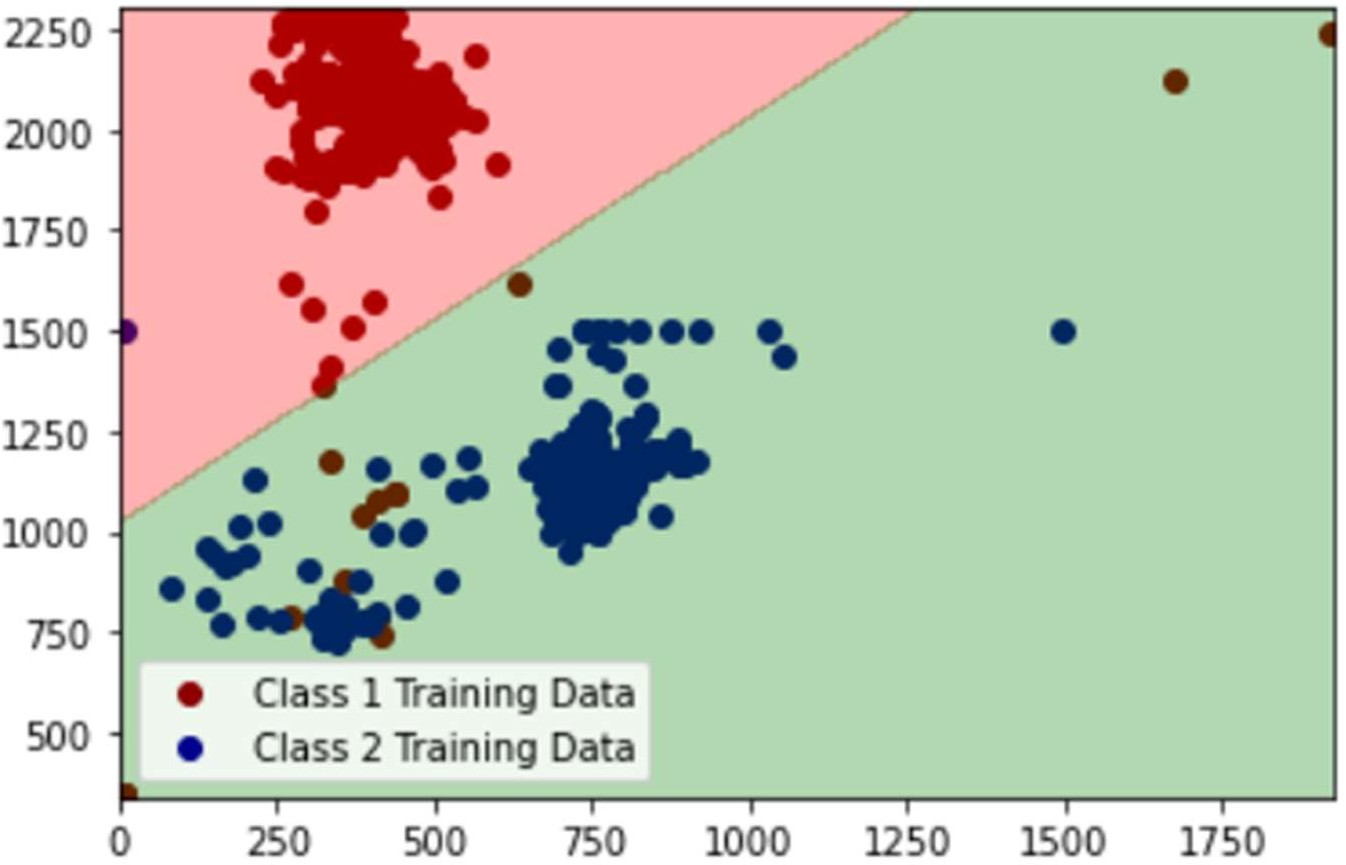
The following is the graph for Dataset3 (RWD) Case1 Decision Region for class3 and class1.



# Dataset3 (RWD) Case2 Cumulative Decision Region

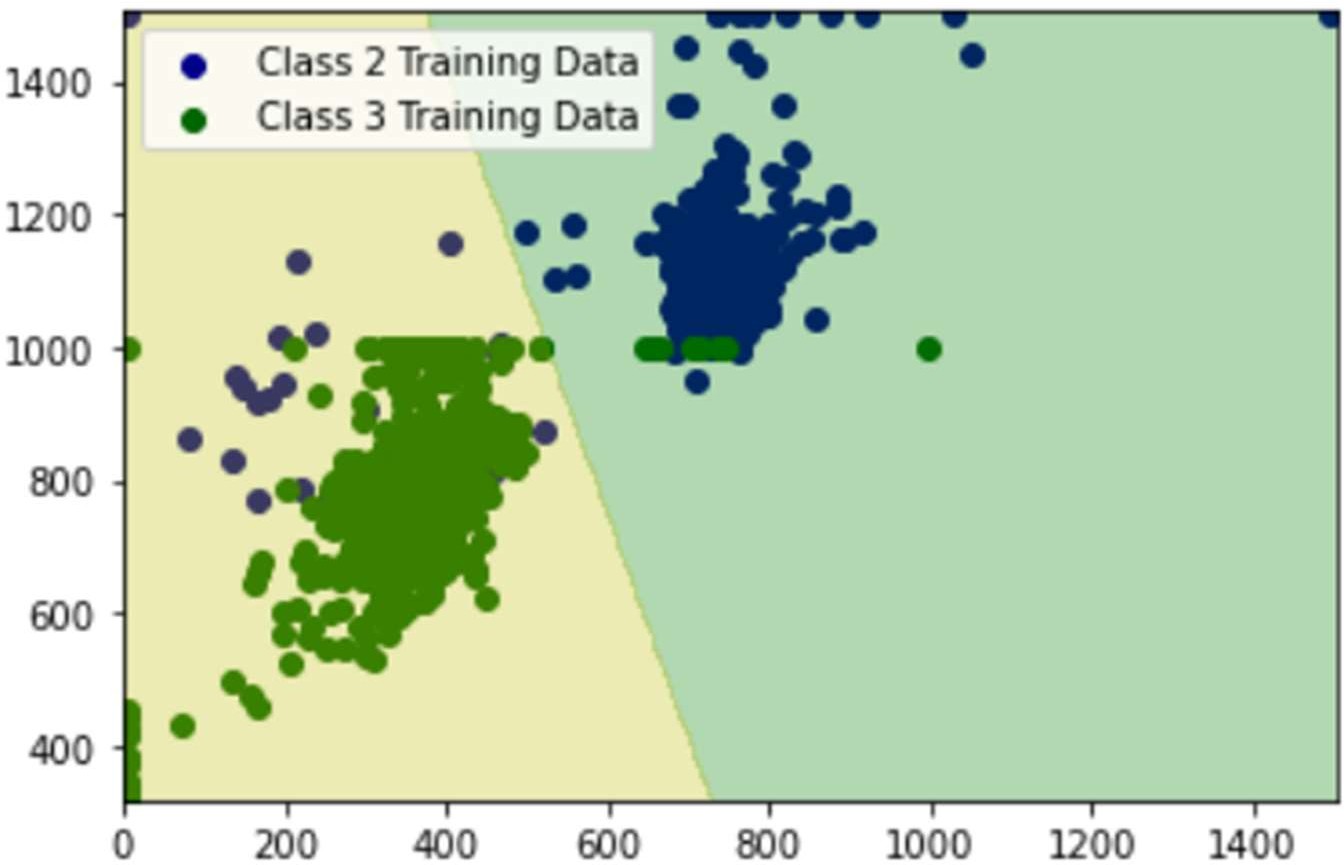
* The graph consists of cumulative decision region for case2 for Dataset3, the final decision region where training data of all the three classes has been plotted and the region has come after considering each minute pixel as a data point. The posterior probabilities have been calculated for all three classes for that data point and whichever one is the highest the data point will go into that class or colored with a specific color with training data of the class being superimposed in the decision region.

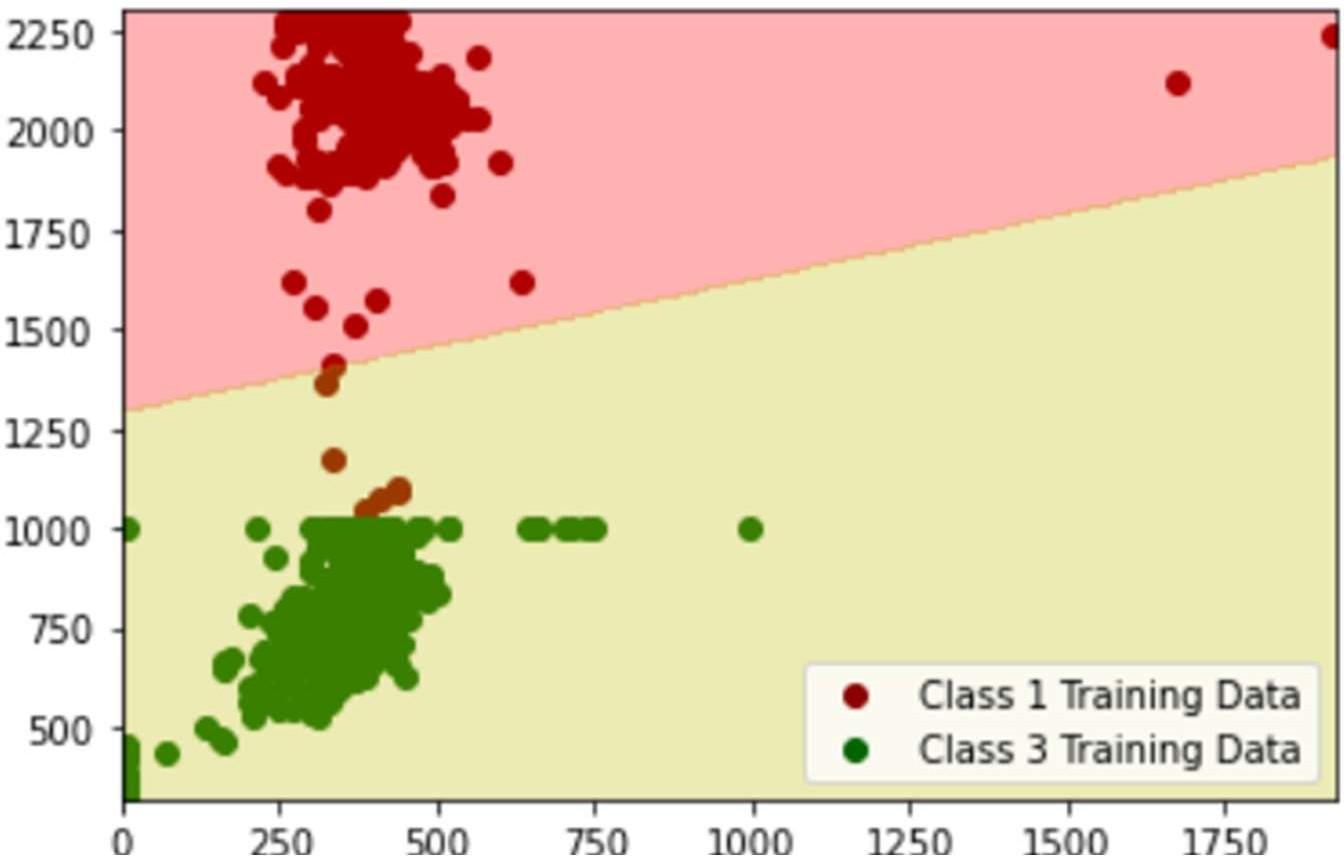




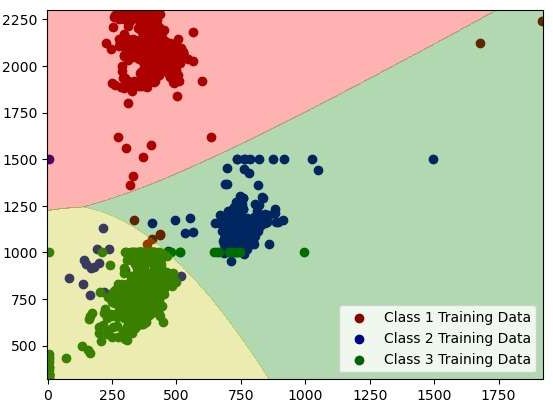
# Dataset3 (RWD) Case2 Decision Region for class3 and class2

The following is the graph for Dataset3 (RWD) Case2 Decision Region for class3 and class2.

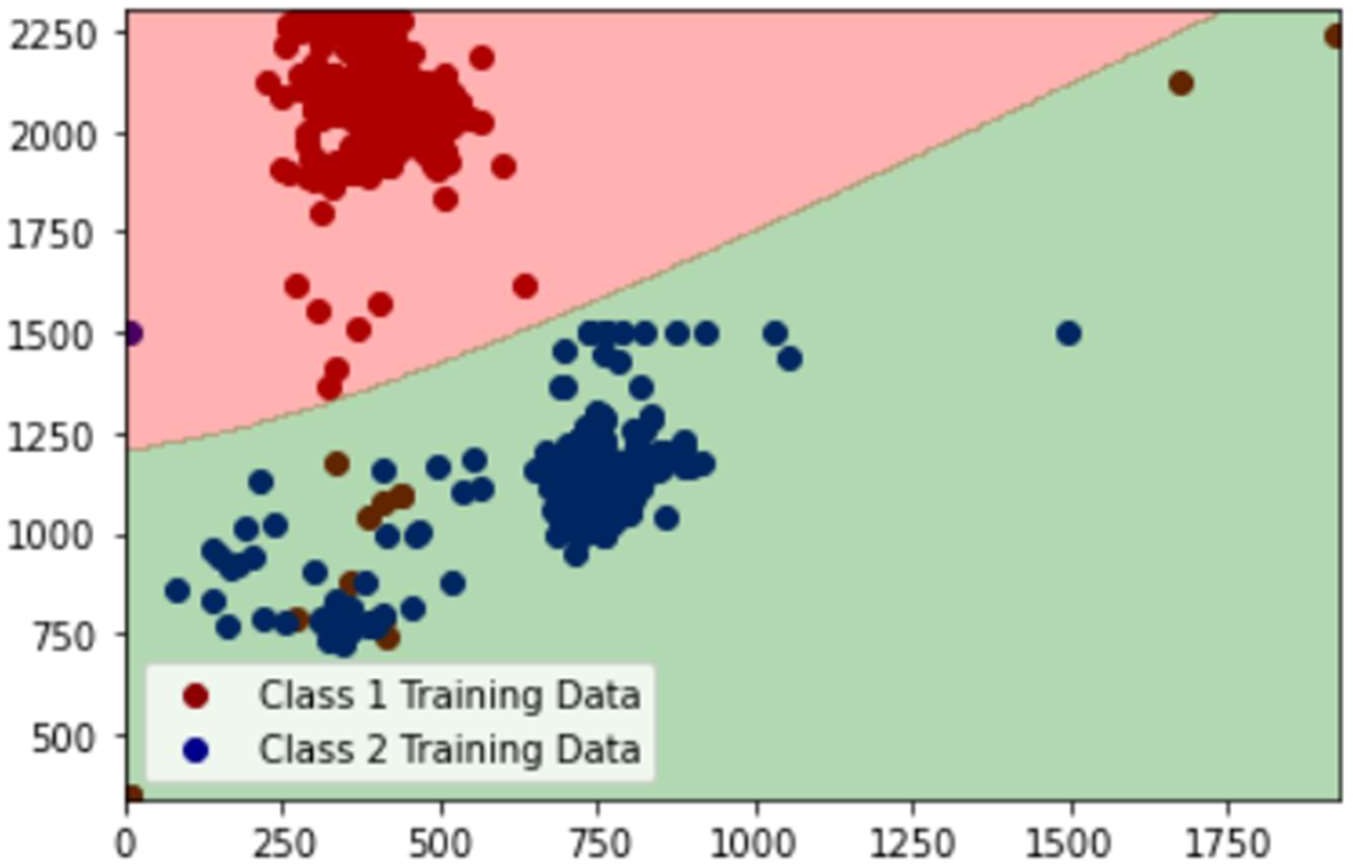




* The graph consists of cumulative decision region for case3 for Dataset3, the final decision region where training data of all the three classes has been plotted and the region has come after considering each minute pixel as a data point. The posterior probabilities have been calculated for all three classes for that data point and whichever one is the highest the data point will go into that class or colored with a specific color with training data of the class being superimposed in the decision region.

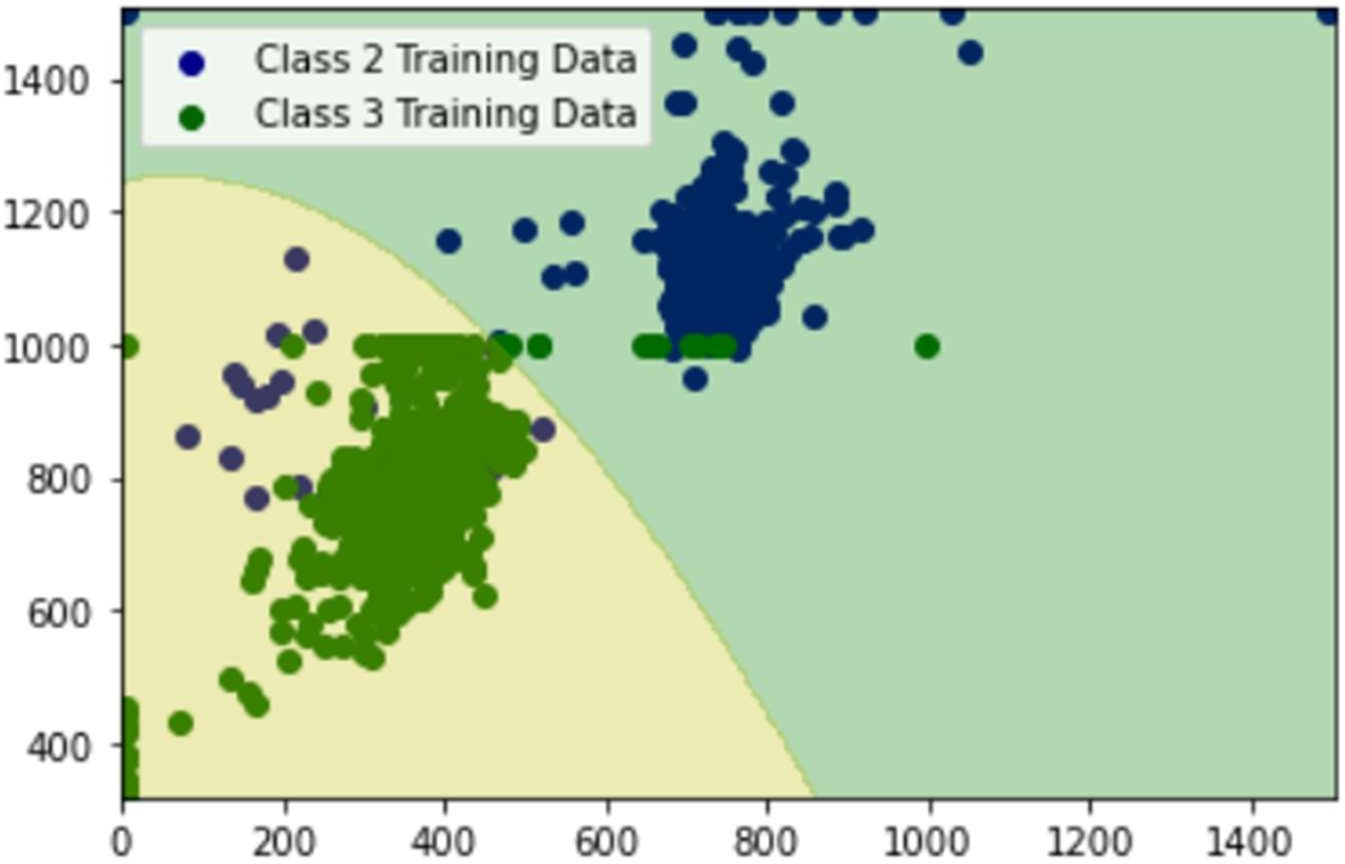


The following is the graph for Dataset3 (RWD) Case2 Decision Region for class1 and class2.



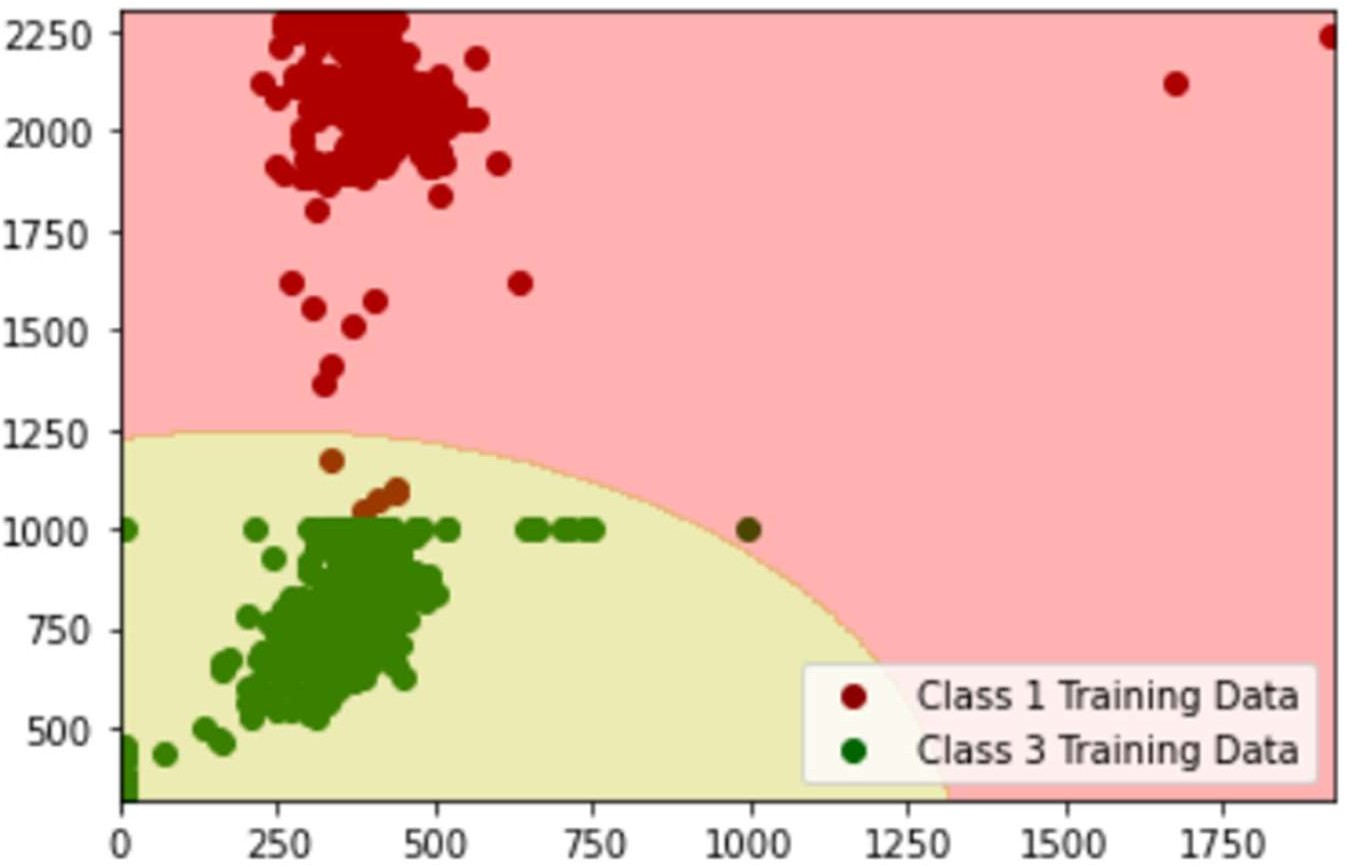
# Dataset3 (RWD) Case3 Decision Region for class3 and class2

The following is the graph for Dataset3 (RWD) Case2 Decision Region for class3 and class2.



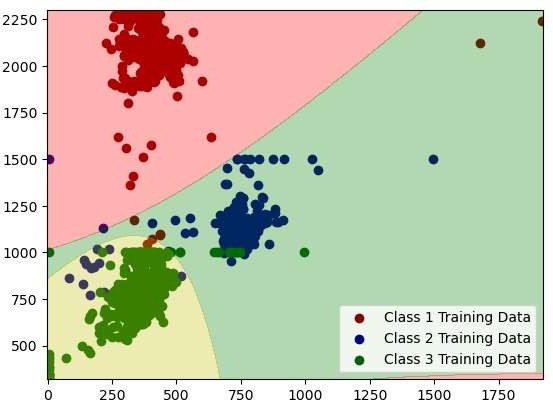
# Dataset3 (RWD) Case3 Decision Region for class3 and class1

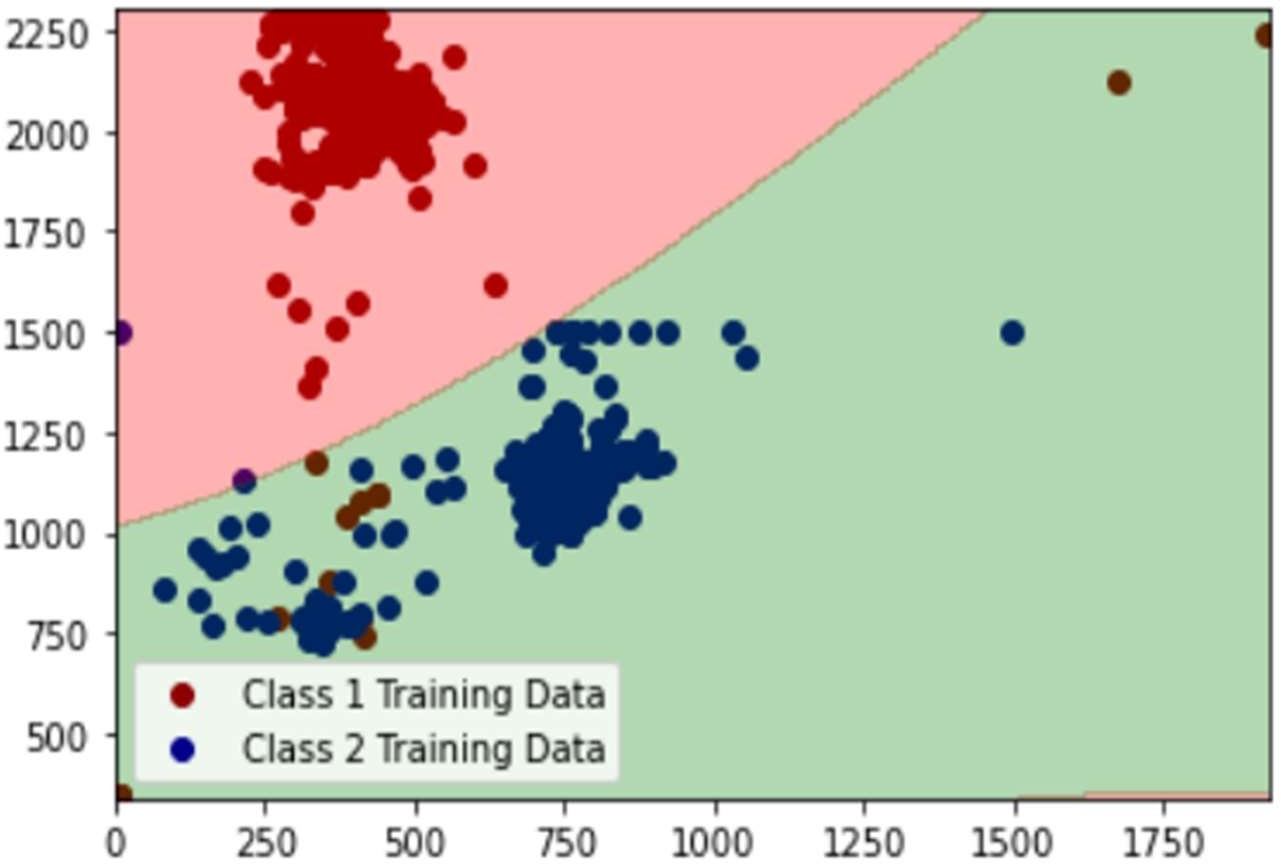
The following is the graph for Dataset3 (RWD) Case2 Decision Region for class3 and class1.



# Dataset3 (RWD) Case4 Cumulative Decision Region

* The graph consists of cumulative decision region for case4 for Dataset3, the final decision region where training data of all the three classes has been plotted and the region has come after considering each minute pixel as a data point. The posterior probabilities have been calculated for all three classes for that data point and whichever one is the highest the data point will go into that class or colored with a specific color with training data of the class being superimposed in the decision region.

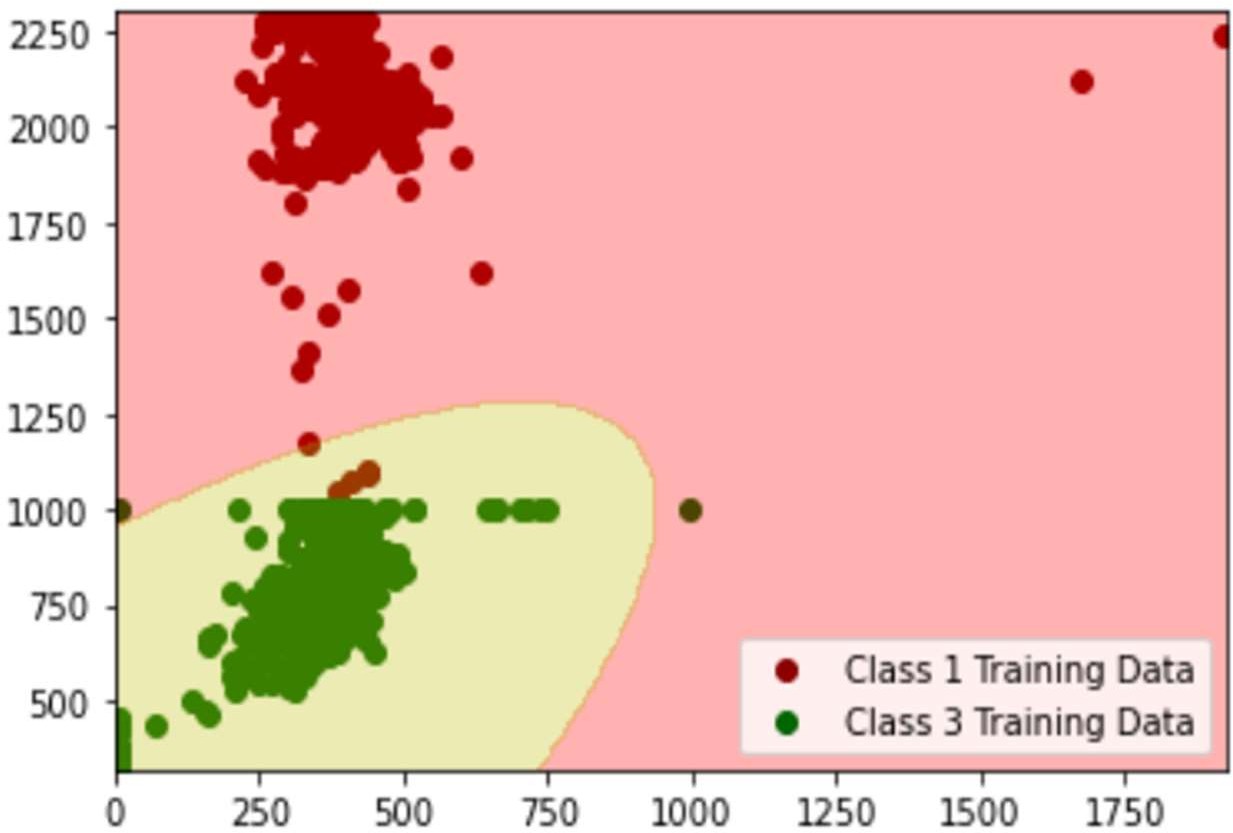




# Dataset3 (RWD) Case4 Decision Region for class3 and class2

The following is the graph for Dataset3 (RWD) Case2 Decision Region for class3 and class2.





# Results for the performance measures

**Dataset1 (linearly separable)**

* 1. Case1:

Confusion matrix:

|  |  |  |  |
| --- | --- | --- | --- |
|  | Actually C1 | Actually C2 | Actually C3 |
| Predicted C1 | 150 | 0 | 0 |
| Predicted C2 | 0 | 150 | 0 |
| Predicted C3 | 0 | 0 | 1 |

Below is the accuracy, precision, recall and f1 score for the three classes of case1 dataset1:

Class 1:

1.0

1.0

1.0

1.0

Class 2:

1.0

1.0

1.0

1.0

Class 3:

1.0

1.0

1.0

1.0

Avg accuracy: 1.0

Avg precision: 1.0

Avg recall: 1.0

Avg f1: 1.0

* 1. Case2:

Confusion matrix:

|  |  |  |  |
| --- | --- | --- | --- |
|  | Actually C1 | Actually C2 | Actually C3 |
| Predicted C1 | 150 | 0 | 0 |
| Predicted C2 | 0 | 150 | 0 |
| Predicted C3 | 0 | 0 | 1 |

Below is the accuracy, precision, recall and f1 score for the three classes of case2 dataset1:

Class 1:

1.0

1.0

1.0

1.0

Class 2:

1.0

1.0

1.0

1.0

Class 3:

1.0

1.0

1.0

1.0

Avg accuracy: 1.0

Avg precision: 1.0

Avg recall: 1.0

Avg f1: 1.0

* 1. Case3

Confusion matrix:

|  |  |  |  |
| --- | --- | --- | --- |
|  | Actually C1 | Actually C2 | Actually C3 |
| Predicted C1 | 150 | 0 | 0 |
| Predicted C2 | 0 | 150 | 0 |
| Predicted C3 | 0 | 0 | 1 |

Below is the accuracy, precision, recall and f1 score for the three classes of case3 dataset1:

Class 1:

1.0

1.0

1.0

1.0

Class 2:

1.0

1.0

1.0

1.0

Class 3:

1.0

1.0

1.0

1.0

Avg accuracy: 1.0

Avg precision: 1.0

Avg recall: 1.0

Avg f1: 1.0

* 1. Case4:

Below is the accuracy, precision, recall and f1 score for the three classes of case4 dataset1:

Class 1:

1.0

1.0

1.0

1.0

Class 2:

1.0

1.0

1.0

1.0

Class 3:

1.0

1.0

1.0

1.0

Avg accuracy: 1.0

Avg precision: 1.0

Avg recall: 1.0

Avg f1: 1.0

# Dataset2 (Non linearly separable)

* + 1. Case1 confusion matrix:

|  |  |  |  |
| --- | --- | --- | --- |
|  | Actually C1 | Actually C2 | Actually C3 |
| Predicted C1 | 99 | 15 | 22 |
| Predicted C2 | 21 | 135 | 0 |
| Predicted C3 | 30 | 0 | 129 |

Below is the accuracy, precision, recall and f1 score for the three classes of case1 dataset2:

Class 1:

0.8048780487804879

0.7279411764705882

0.66

0.6923076923076923

Class 2:

0.9201773835920177

0.8653846153846154

0.9

0.8823529411764707

Class 3:

0.88470066518847

0.8113207547169812

0.8543046357615894

0.832258064516129

Avg accuracy: 0.8699186991869919

Avg precision: 0.801548848857395

Avg recall: 0.8047682119205298

Avg f1: 0.8023062326667642

* + 1. Case2 confusion matrix:

|  |  |  |  |
| --- | --- | --- | --- |
|  | Actually C1 | Actually C2 | Actually C3 |
| Predicted C1 | 125 | 5 | 7 |
| Predicted C2 | 11 | 145 | 1 |
| Predicted C3 | 14 | 0 | 143 |

Below is the accuracy, precision, recall and f1 score for the three classes of case2 dataset2:

Class 1:

0.917960088691796

0.9124087591240876

0.8333333333333334

0.8710801393728222

Class 2:

0.9623059866962306

0.9235668789808917

0.9666666666666667

0.9446254071661238

Class 3:

0.9512195121951219

0.910828025477707

0.9470198675496688

0.9285714285714285

Avg accuracy: 0.9438285291943829

Avg precision: 0.9156012211942288

Avg recall: 0.915673289183223

Avg f1: 0.9147589917034581

* + 1. Case3 confusion matrix:

|  |  |  |  |
| --- | --- | --- | --- |
|  | Actually C1 | Actually C2 | Actually C3 |
| Predicted C1 | 126 | 7 | 10 |
| Predicted C2 | 10 | 143 | 1 |
| Predicted C3 | 14 | 0 | 140 |

Below is the accuracy, precision, recall and f1 score for the three classes of case3 dataset2:

Class 1:

0.9090909090909091

0.8811188811188811

0.84

0.8600682593856654

Class 2:

0.9600886917960089

0.9285714285714286

0.9533333333333334

0.9407894736842105

Class 3:

0.9445676274944568

0.9090909090909091

0.9271523178807947

0.9180327868852459

Avg accuracy: 0.9379157427937915

Avg precision: 0.9062604062604063

Avg recall: 0.9068285504047093

Avg f1: 0.9062968399850407

* + 1. Case4 Confusion matrix:

|  |  |  |  |
| --- | --- | --- | --- |
|  | Actually C1 | Actually C2 | Actually C3 |
| Predicted C1 | 126 | 7 | 10 |
| Predicted C2 | 10 | 143 | 1 |
| Predicted C3 | 14 | 0 | 140 |

Below is the accuracy, precision, recall and f1 score for the three classes of case3 dataset2:

Class 1:

0.9090909090909091

0.8811188811188811

0.84

0.8600682593856654

Class 2:

0.9600886917960089

0.9285714285714286

0.9407894736842105

Class 3:

0.9445676274944568

0.9090909090909091

0.9271523178807947

0.9180327868852459

Avg accuracy: 0.9379157427937915

Avg precision: 0.9062604062604063

Avg recall: 0.9068285504047093

Avg f1: 0.9062968399850407

# Dataset3 (Real world data)

1. Case1 confusion matrix:

|  |  |  |  |
| --- | --- | --- | --- |
|  | Actually C1 | Actually C2 | Actually C3 |
| Predicted C1 | 679 | 0 | 0 |
| Predicted C2 | 2 | 636 | 7 |
| Predicted C3 | 7 | 14 | 740 |

Below is the accuracy, precision, recall and f1 score for the three classes of case1 dataset3:

Class 1:

0.99568345323741

1.0

0.9869186046511628

0.9934162399414778

Class 2:

0.988968824940048

0.986046511627907

0.9784615384615385

0.9822393822393823

Class 3:

0.9865707434052757

0.9906291834002677

0.9814323607427056

Avg accuracy: 0.9904076738609113

Avg precision: 0.9861504140818385

Avg recall: 0.9853364421709897

Avg f1: 0.9856959943078553

1. Case2:

confusion matrix:

|  |  |  |  |
| --- | --- | --- | --- |
|  | Actually C1 | Actually C2 | Actually C3 |
| Predicted C1 | 678 | 0 | 0 |
| Predicted C2 | 3 | 636 | 7 |
| Predicted C3 | 7 | 14 | 740 |

Below is the accuracy, precision, recall and f1 score for the three classes of case2 dataset3:

Class 1:

0.9952038369304557

1.0

0.9854651162790697

0.9926793557833089

Class 2:

0.9884892086330935

0.9845201238390093

0.9784615384615385

0.9814814814814815

Class 3:

0.9865707434052757

0.9724047306176085

0.9906291834002677

0.9814323607427056

Avg accuracy: 0.9900879296562749

Avg precision: 0.985641618152206

Avg recall: 0.9848519460469586

Avg f1: 0.9851977326691653

|  |  |  |  |
| --- | --- | --- | --- |
|  | Actually C1 | Actually C2 | Actually C3 |
| Predicted C1 | 678 | 0 | 0 |
| Predicted C2 | 3 | 636 | 7 |
| Predicted C3 | 7 | 14 | 740 |

Below is the accuracy, precision, recall and f1 score for the three classes of case3 dataset3:

Class 1:

0.9952038369304557

1.0

0.9854651162790697

0.9926793557833089

Class 2:

0.9884892086330935

0.9845201238390093

0.9784615384615385

0.9814814814814815

Class 3:

0.9865707434052757

0.9724047306176085

0.9906291834002677

0.9814323607427056

Avg accuracy: 0.9900879296562749

Avg precision: 0.985641618152206

Avg recall: 0.9848519460469586

Avg f1: 0.9851977326691653

|  |  |  |  |
| --- | --- | --- | --- |
|  | Actually C1 | Actually C2 | Actually C3 |
| Predicted C1 | 678 | 0 | 0 |
| Predicted C2 | 3 | 637 | 8 |
| Predicted C3 | 7 | 13 | 739 |

Below is the accuracy, precision, recall and f1 score for the three classes of case4 dataset3:

Class 1:

0.9952038369304557

1.0

0.9854651162790697

0.9926793557833089

Class 2:

0.9884892086330935

0.9830246913580247

0.98

0.9815100154083205

Class 3:

0.9865707434052757

0.9736495388669302

0.9892904953145917

0.9814077025232404

Avg accuracy: 0.9900879296562749

Avg precision: 0.9855580767416517

Avg recall: 0.9849185371978871

Avg f1: 0.9851990245716232

* General observation is that whenever we have the same covariance matrices for all the classes we obtain a linear decision boundary.