Assignment 01

Group 07

Introduction to Machine Learning (CS401)

Submitted by

BEERAVOLU SAI KIRAN 17CSE1007

PRASHANT KARHANA 17CSE1017

Under the Supervision of

Dr. Veena Thenkanidiyoor

Associate Professor, Computer Science and Engineering

Department of Computer Science and Engineering

National Institute of Technology Goa

October-November, 2020

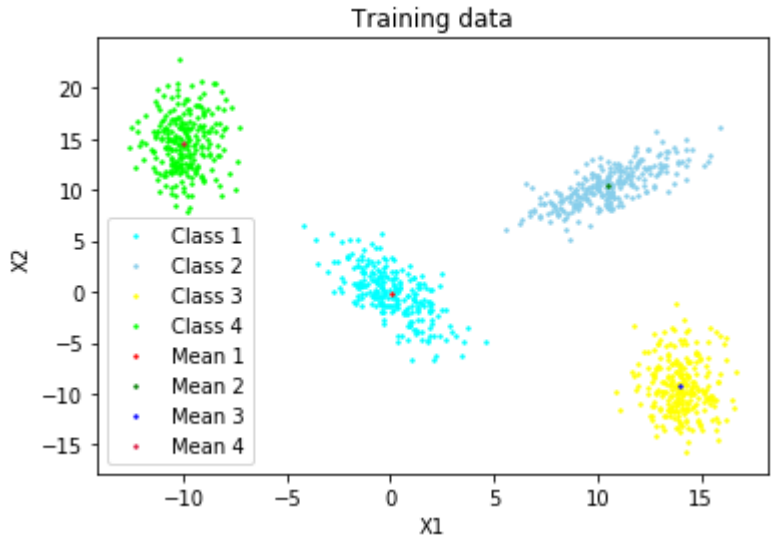
Contents-

Classification

1.1 Linearly Separable Data

1.1.1 Bayes Classifier with Same Diagonal Covariance Matrix

1.1.1.1 Plot of training data with mean displayed in different colour



1.1.1.2 Classification accuracy, precision for every class, mean precision, recall for every class, mean recall, F-measure for every class and mean F-measure on test data

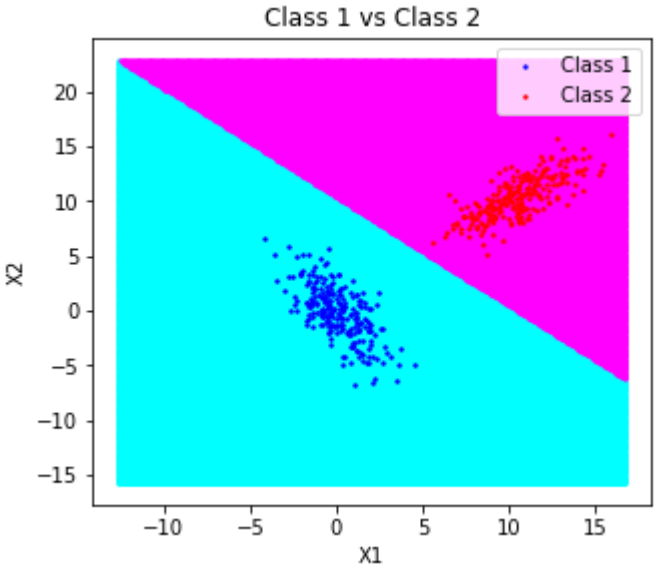
Accuracy:[100.0, 100.0, 100.0, 100.0]
Precision:[1.0, 1.0, 1.0, 1.0]
Recall:[1.0, 1.0, 1.0, 1.0]
F_measure:[1.0, 1.0, 1.0, 1.0]

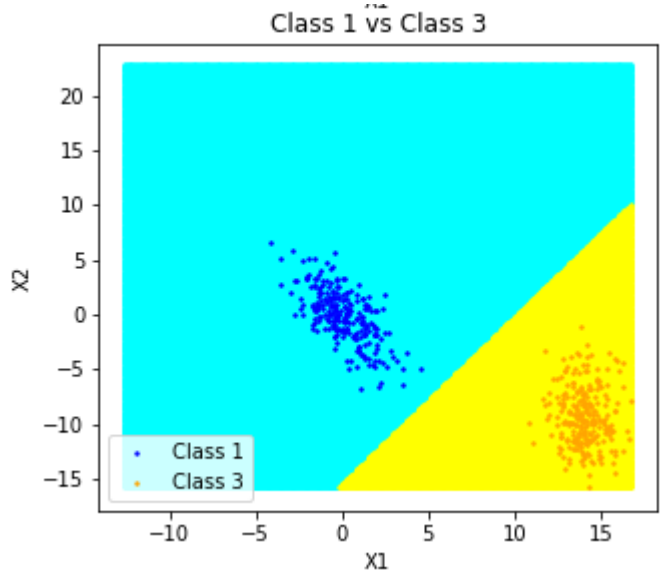
Mean Accuracy: 100.0
Mean Precision: 1.0
Mean Recall: 1.0
Mean F_measure: 1.0

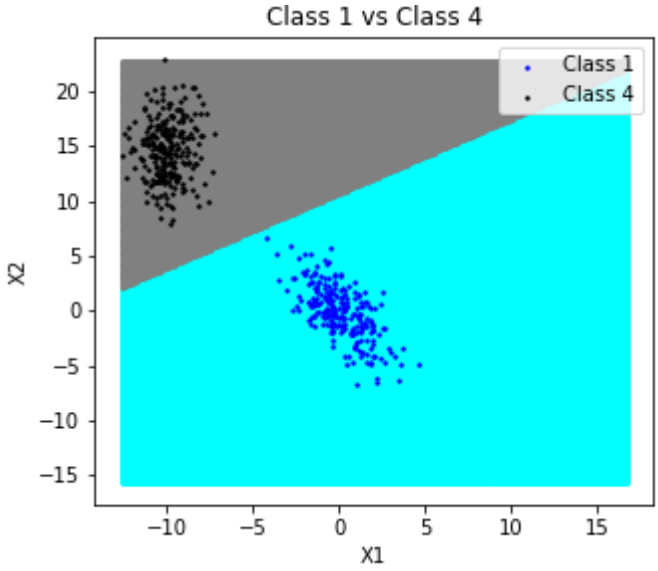
1.1.1.3 Confusion matrix based on the performance for test data. The entries in confusion matrix must be made in percentage.

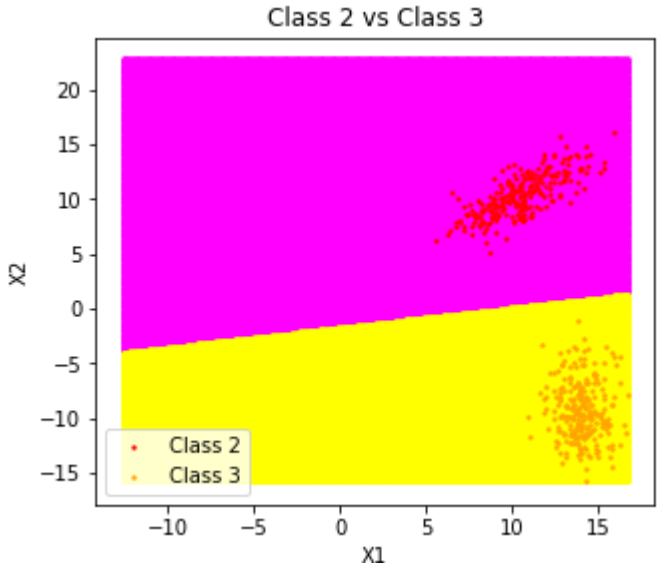
	True 1	True 2	True 3	True 4
Predicted 1	100.0	0.0	0.0	0.0
Predicted 2	0.0	100.0	0.0	0.0
Predicted 3	0.0	0.0	100.0	0.0
Predicted 4	0.0	0.0	0.0	100.0

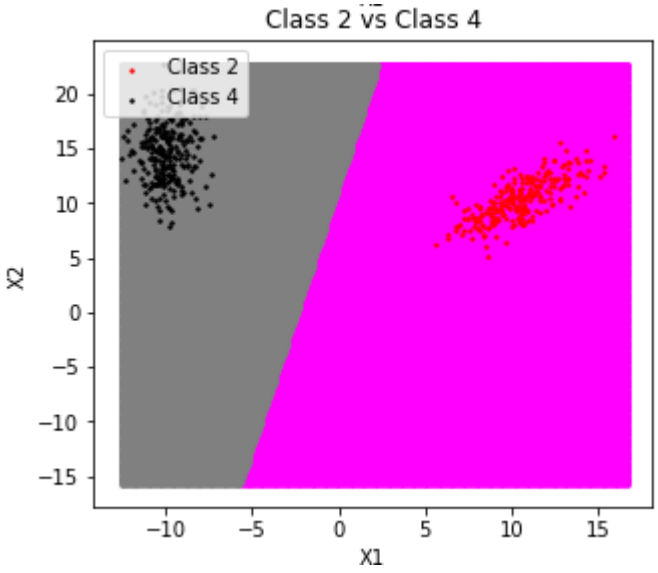
1.1.1.4 Decision region plot for every pair of classes with the respective training data superimposed

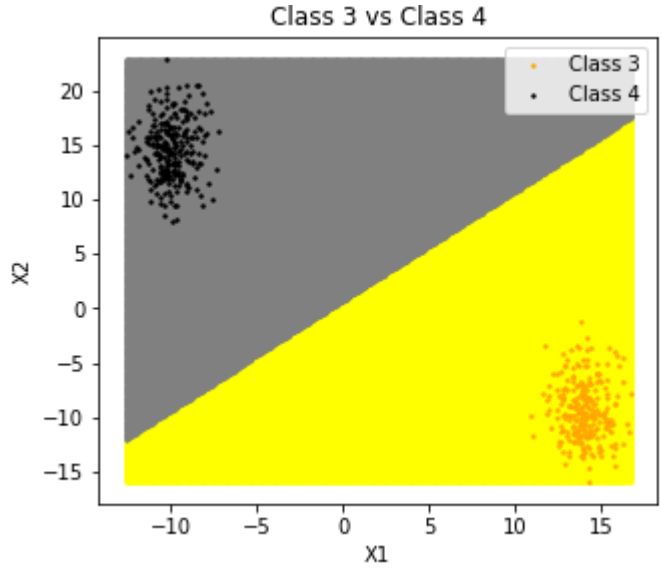




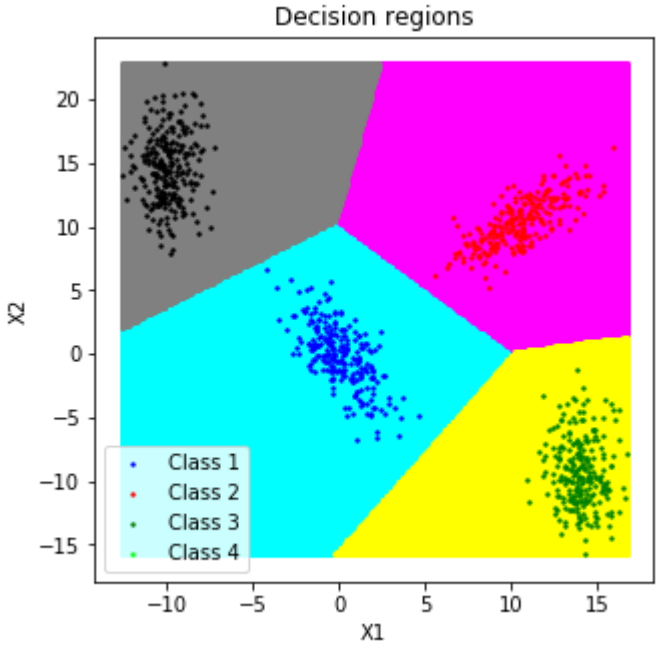



Class 2 vs Class 3
Class 2
Class 3
X1
X2

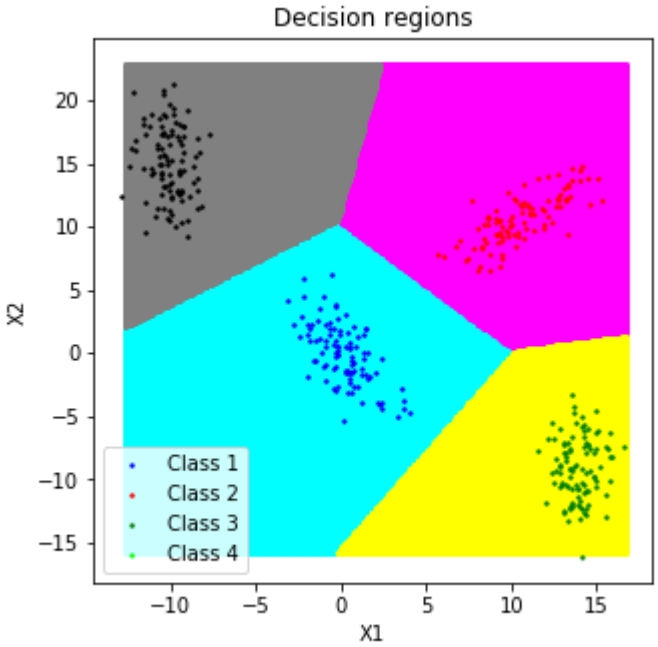

Class 2 vs Class 4
Class 2
Class 4
X1
X2


Class 3 vs Class 4
Class 3
Class 4
X1
X2

1.1.1.5 Decision region plot for all the classes together with the training data superposed

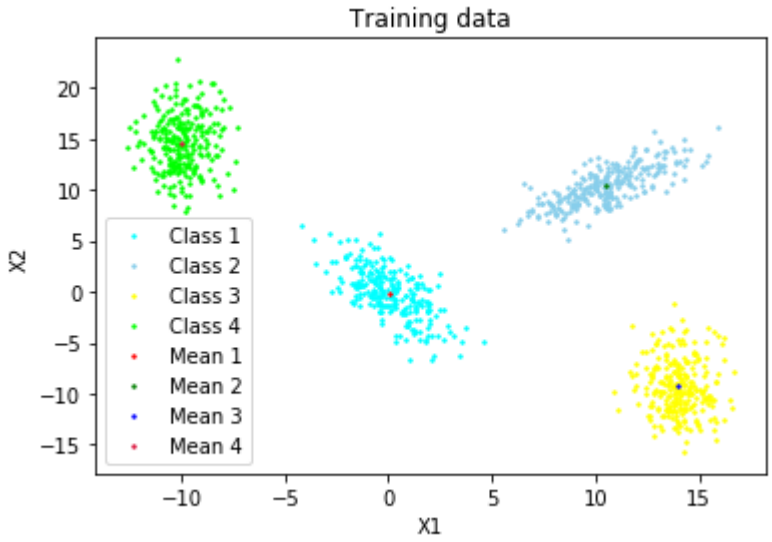


1.1.1.6 Decision region plot for all the classes together with the testing data superimposed



1.1.2 Bayes Classifier with Same Full Covariance Matrix

1.1.2.1 Plot of training data with mean displayed in different colour



1.1.2.2 Classification accuracy, precision for every class, mean precision, recall for every class, mean recall, F-measure for every class and mean F-measure on test data

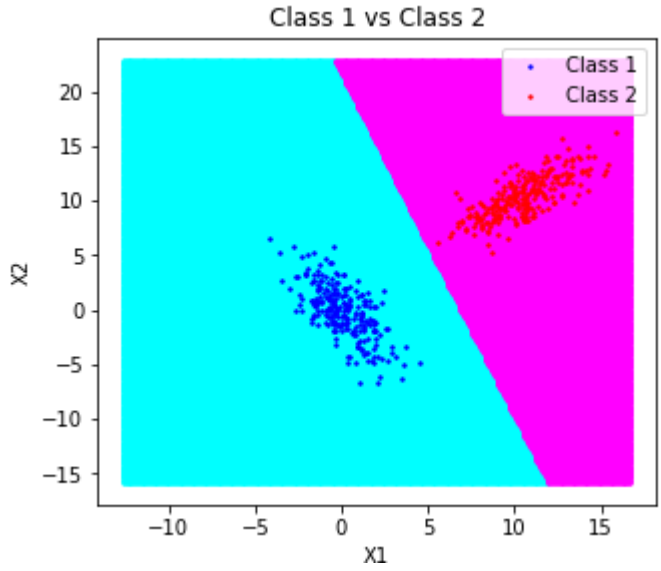
Accuracy:[100.0, 100.0, 100.0, 100.0]
Precision:[1.0, 1.0, 1.0, 1.0]
Recall:[1.0, 1.0, 1.0, 1.0]
F_measure:[1.0, 1.0, 1.0, 1.0]

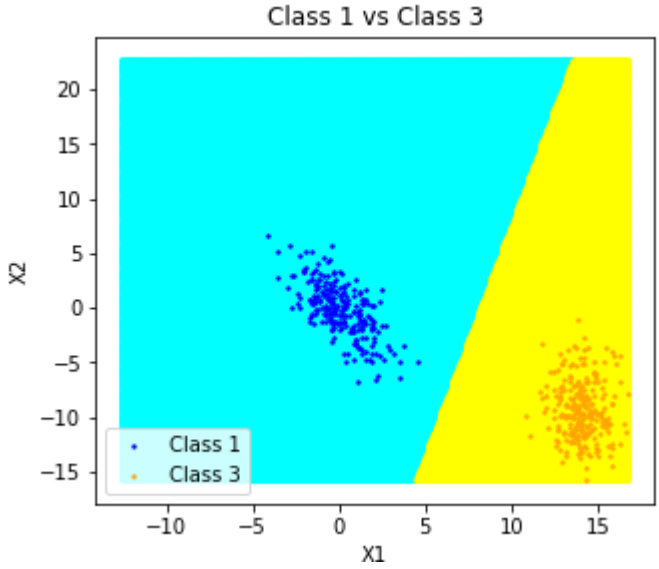
Mean Accuracy: 100.0
Mean Precision: 1.0
Mean Recall: 1.0
Mean F_measure: 1.0

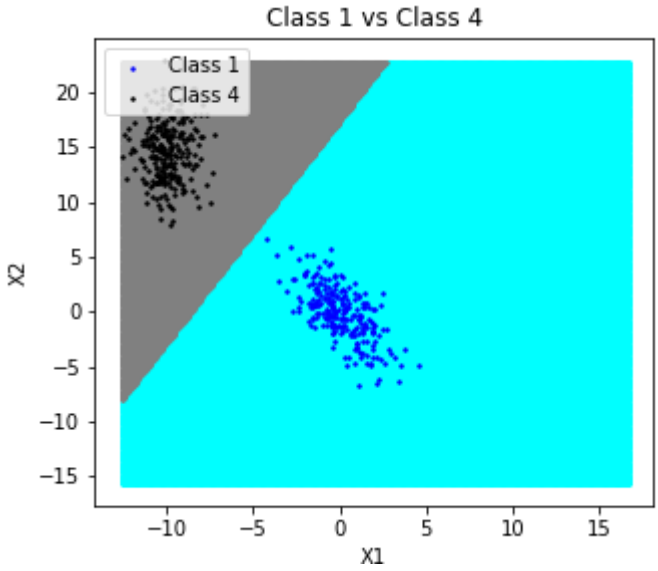
1.1.2.3 Confusion matrix based on the performance for test data. The entries in confusion matrix must be made in percentage.

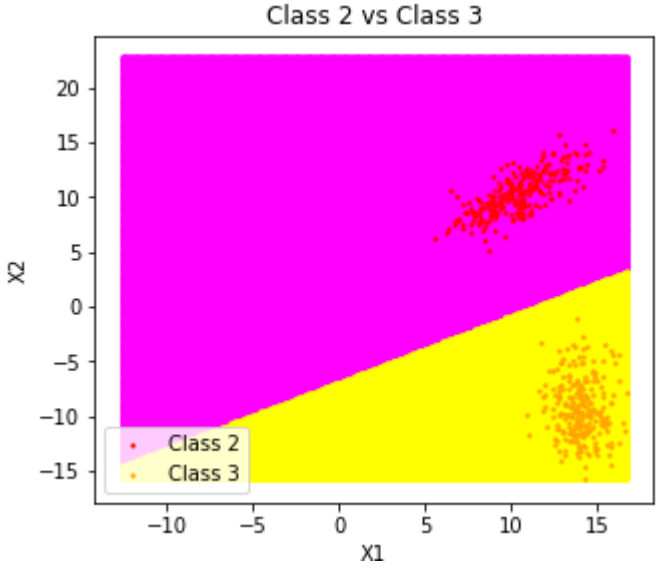
	True 1	True 2	True 3	True 4
Predicted 1	100.0	0.0	0.0	0.0
Predicted 2	0.0	100.0	0.0	0.0
Predicted 3	0.0	0.0	100.0	0.0
Predicted 4	0.0	0.0	0.0	100.0

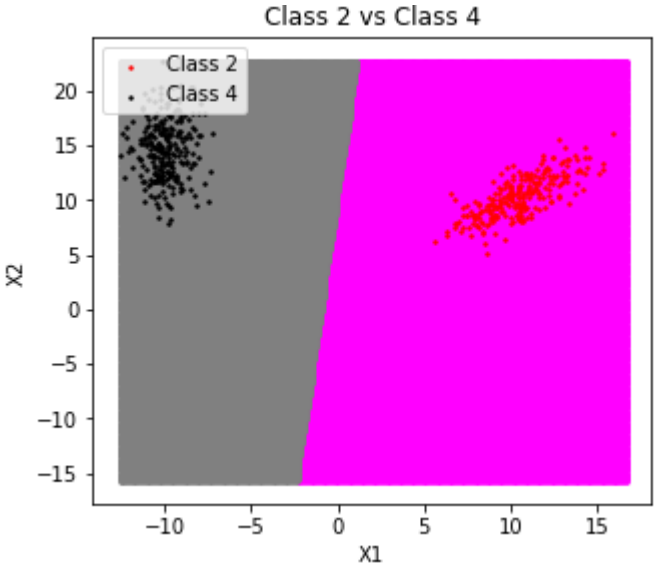
1.1.2.4 Decision region plot for every pair of classes with the respective training data superimposed

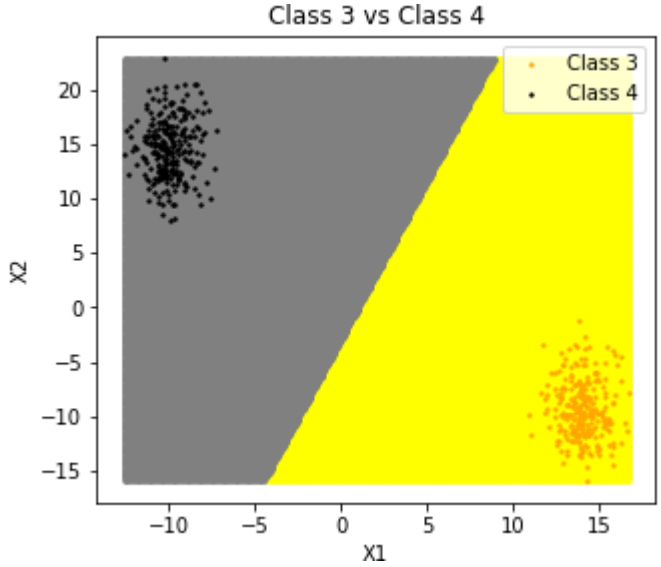




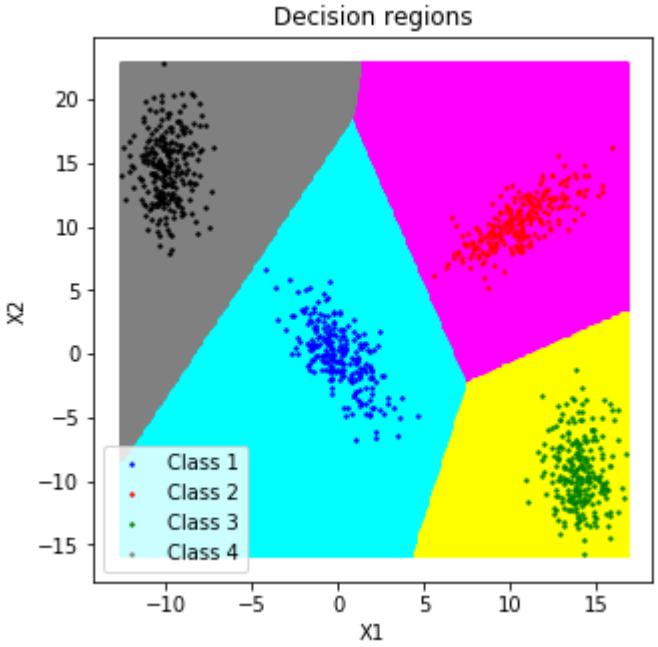



Class 2 vs Class 3
Class 2
Class 3
X1
X2

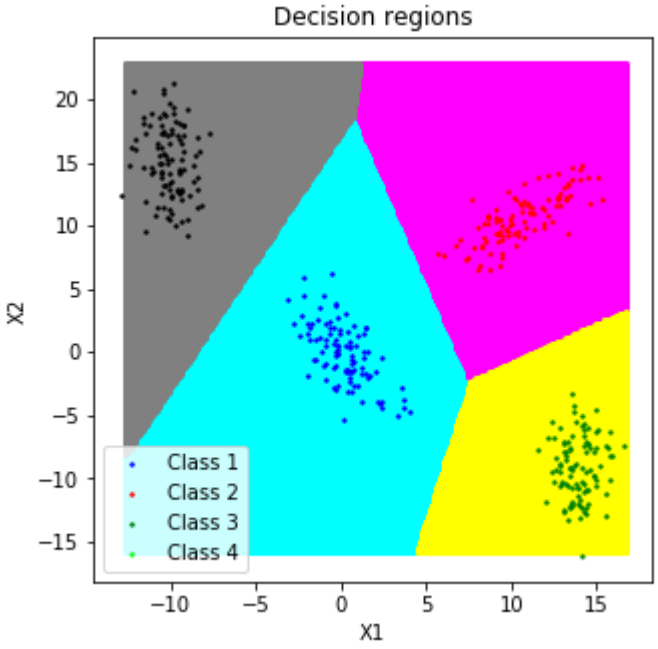

Class 2 vs Class 4
Class 2
Class 4
X1
X2


Class 3 vs Class 4
Class 3
Class 4
X1
X2

1.1.2.5 Decision region plot for all the classes together with the training data superposed

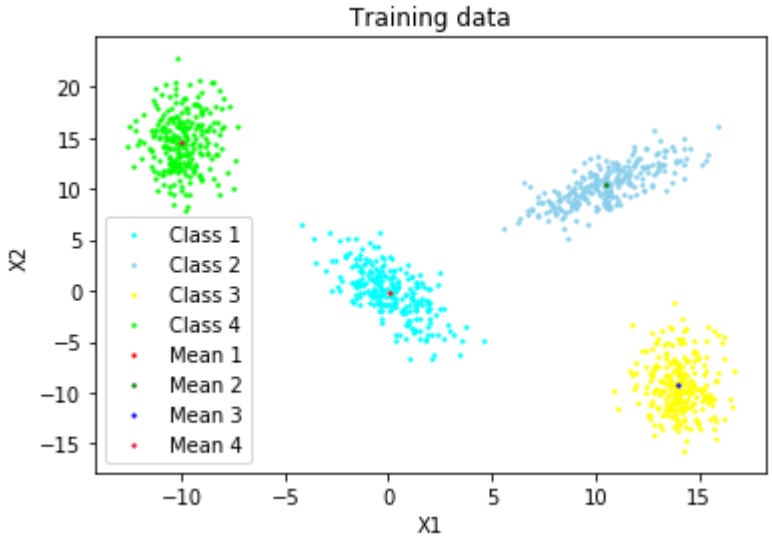


1.1.2.6 Decision region plot for all the classes together with the testing data superimposed



1.1.3 Bayes Classifier with Different Diagonal Covariance Matrix

1.1.3.1 Plot of training data with mean displayed in different colour



1.1.3.2 Classification accuracy, precision for every class, mean precision, recall for every class, mean recall, F-measure for every class and mean F-measure on test data

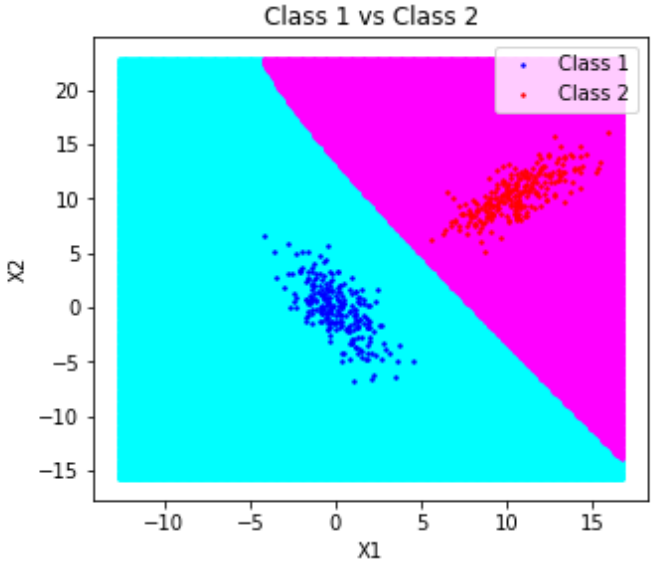
Accuracy:[100.0, 100.0, 100.0, 100.0]
Precision:[1.0, 1.0, 1.0, 1.0]
Recall:[1.0, 1.0, 1.0, 1.0]
F_measure:[1.0, 1.0, 1.0, 1.0]

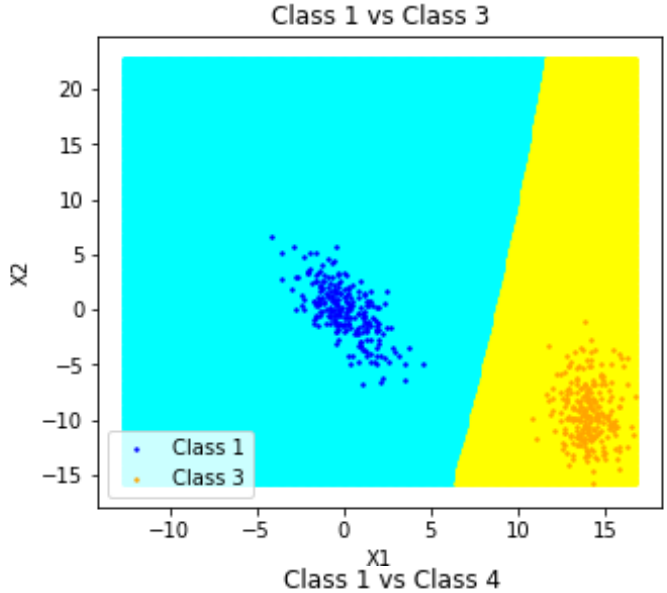
Mean Accuracy: 100.0
Mean Precision: 1.0
Mean Recall: 1.0
Mean F_measure: 1.0

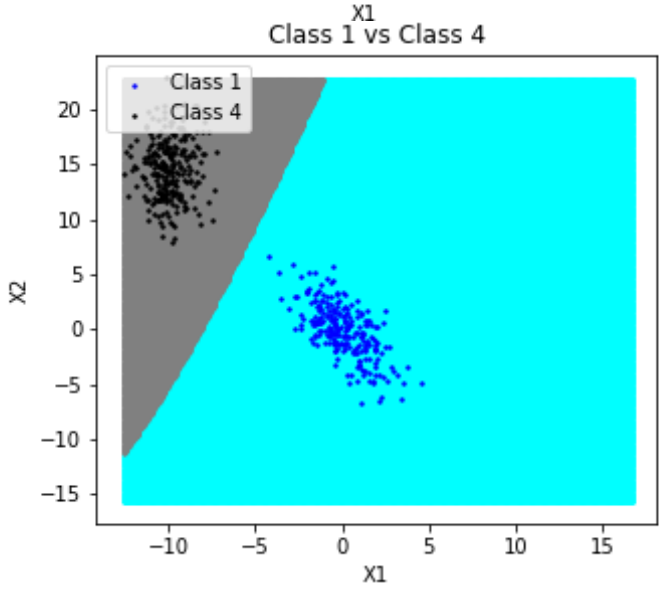
1.1.3.3 Confusion matrix based on the performance for test data. The entries in confusion matrix must be made in percentage.

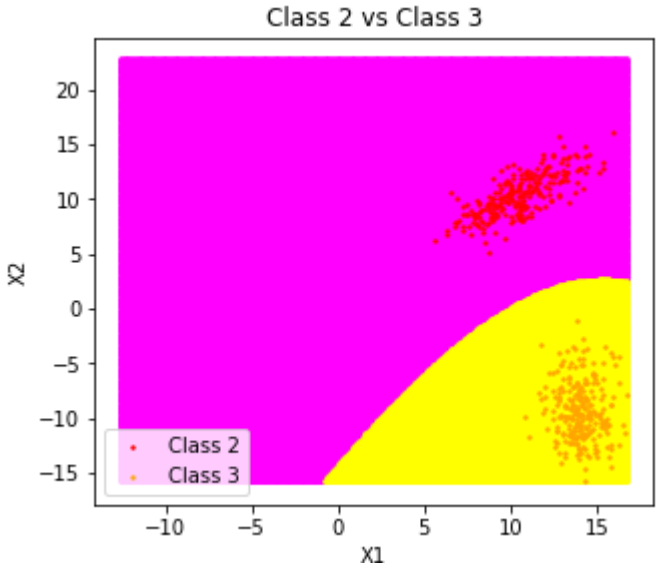
	True 1	True 2	True 3	True 4
Predicted 1	100.0	0.0	0.0	0.0
Predicted 2	0.0	100.0	0.0	0.0
Predicted 3	0.0	0.0	100.0	0.0
Predicted 4	0.0	0.0	0.0	100.0

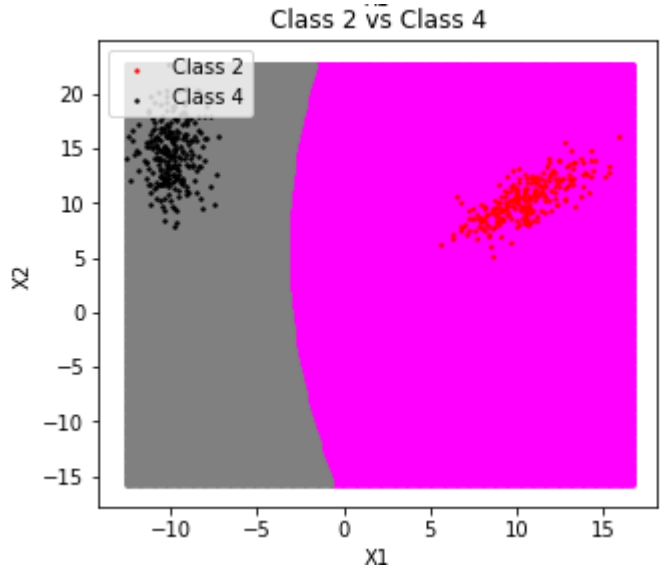
1.1.3.4 Decision region plot for every pair of classes with the respective training data superimposed

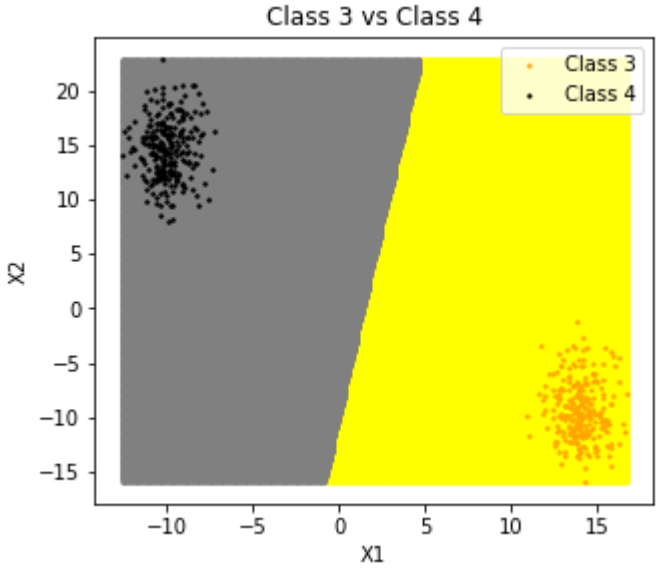




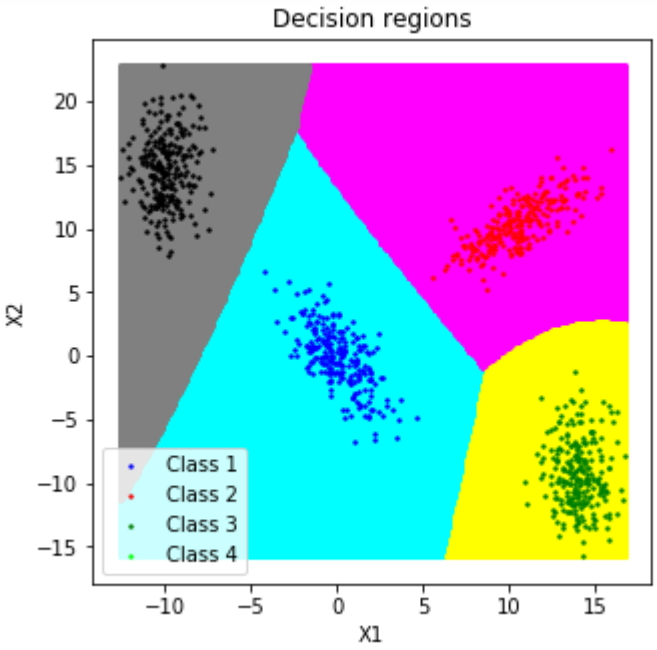



Class 2 vs Class 3
Class 2
Class 3
X1
X2

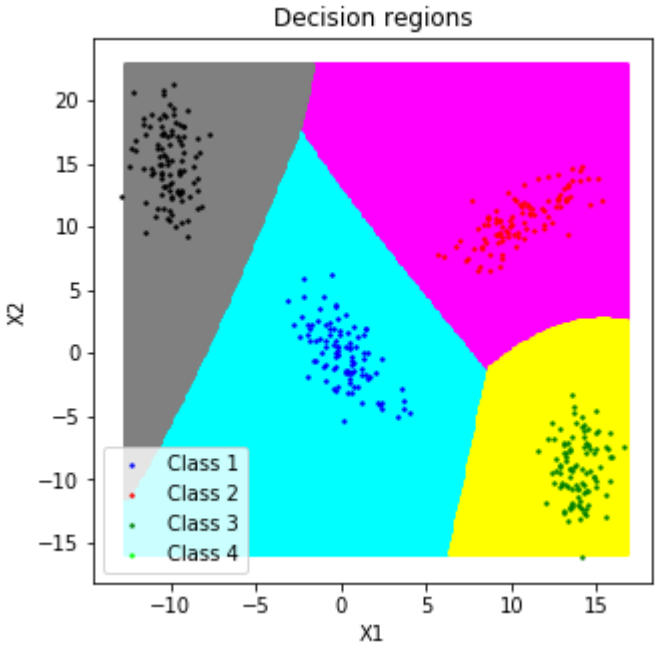

Class 2 vs Class 4
Class 2
Class 4
X1
X2


Class 3 vs Class 4
Class 3
Class 4
X1
X2

1.1.3.5 Decision region plot for all the classes together with the training data superposed

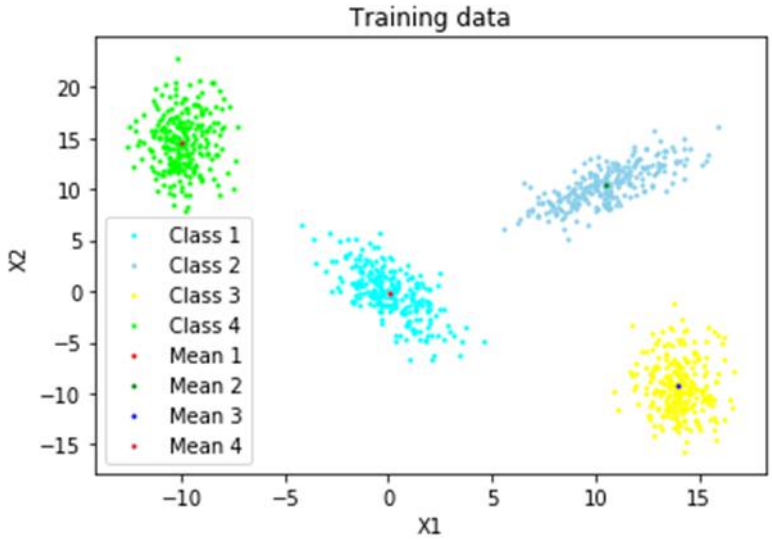


1.1.3.6 Decision region plot for all the classes together with the testing data superimposed



1.1.4 Bayes Classifier with Different Full Covariance Matrix

1.1.4.1 Plot of training data with mean displayed in different colour



1.1.4.2 Classification accuracy, precision for every class, mean precision, recall for every class, mean recall, F-measure for every class and mean F-measure on test data

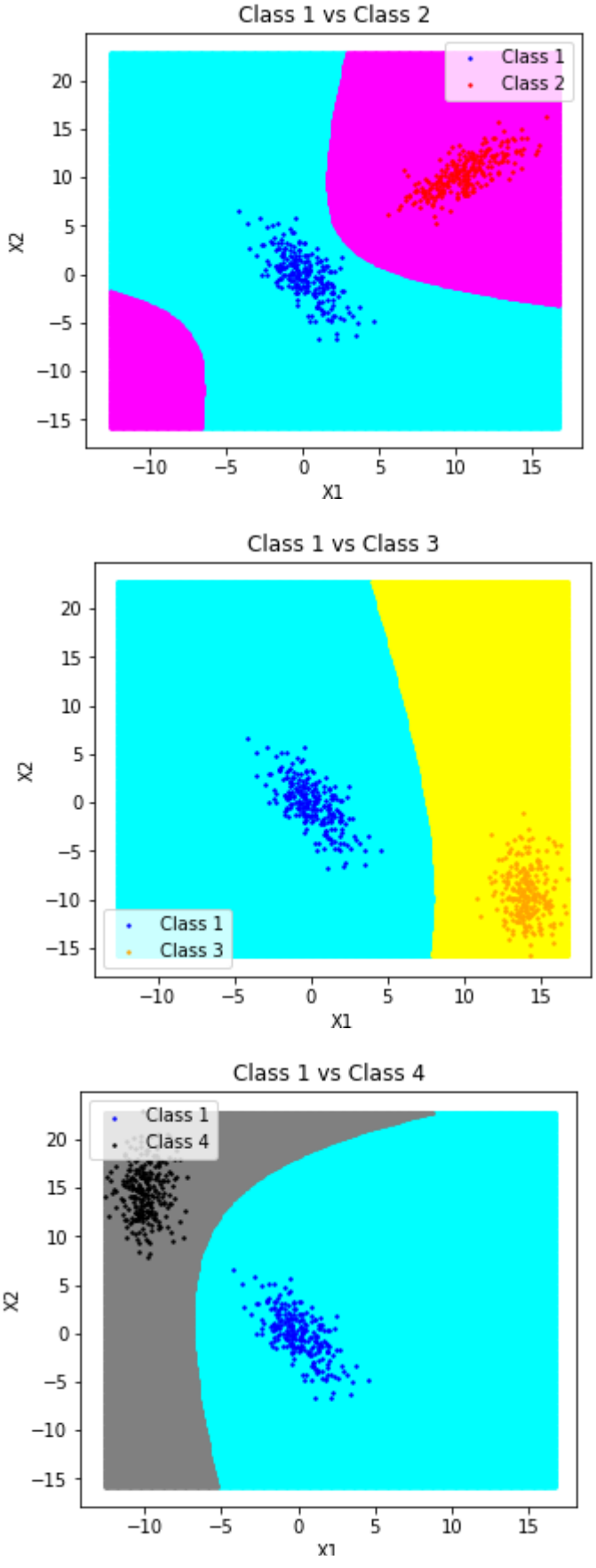
Accuracy:[100.0, 100.0, 100.0, 100.0]
Precision:[1.0, 1.0, 1.0, 1.0]
Recall:[1.0, 1.0, 1.0, 1.0]
F_measure:[1.0, 1.0, 1.0, 1.0]

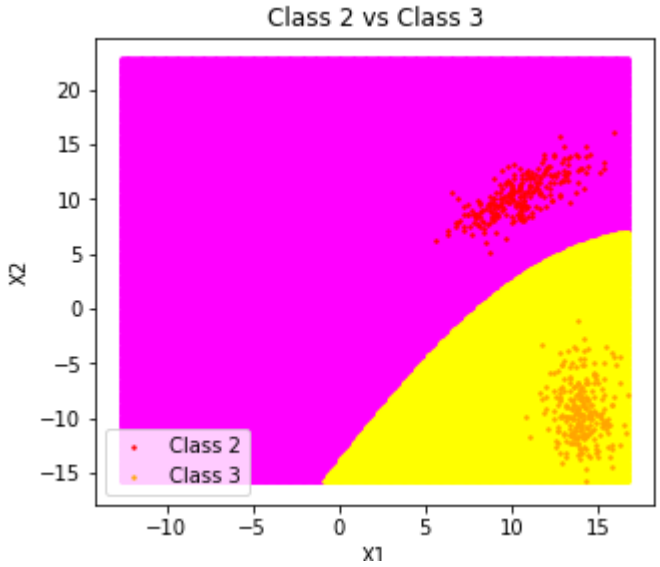
Mean Accuracy: 100.0
Mean Precision: 1.0
Mean Recall: 1.0
Mean F_measure: 1.0

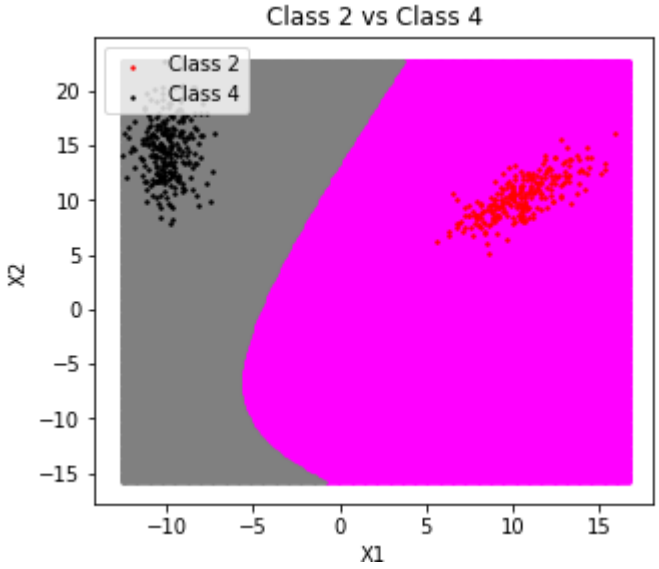
1.1.4.3 Confusion matrix based on the performance for test data. The entries in confusion matrix must be made in percentage.

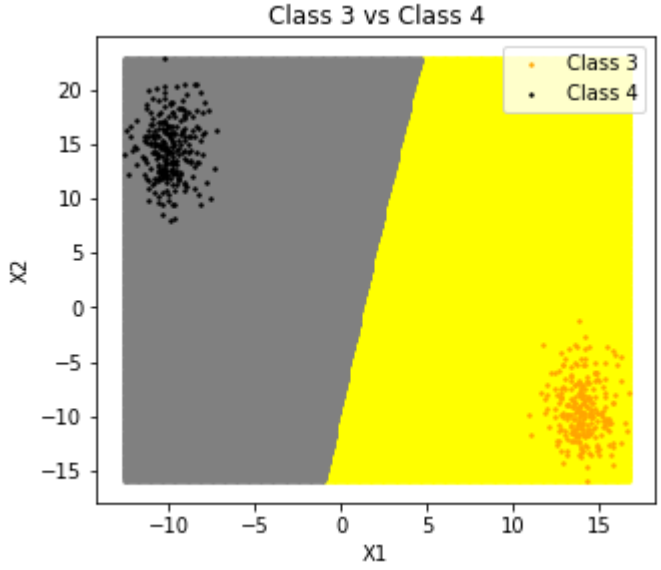
	True 1	True 2	True 3	True 4
Predicted 1	100.0	0.0	0.0	0.0
Predicted 2	0.0	100.0	0.0	0.0
Predicted 3	0.0	0.0	100.0	0.0
Predicted 4	0.0	0.0	0.0	100.0

1.1.4.4 Decision region plot for every pair of classes with the respective training data superimposed

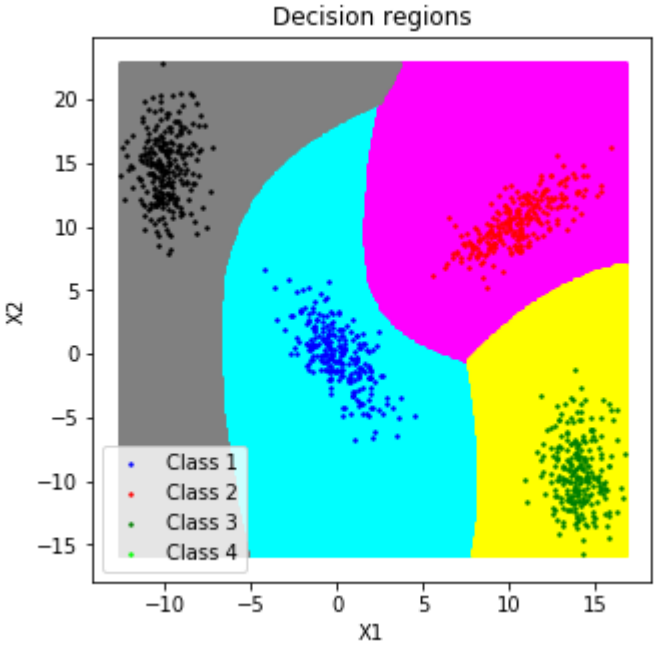



Class 2 vs Class 3
Class 2
Class 3
X1
X2

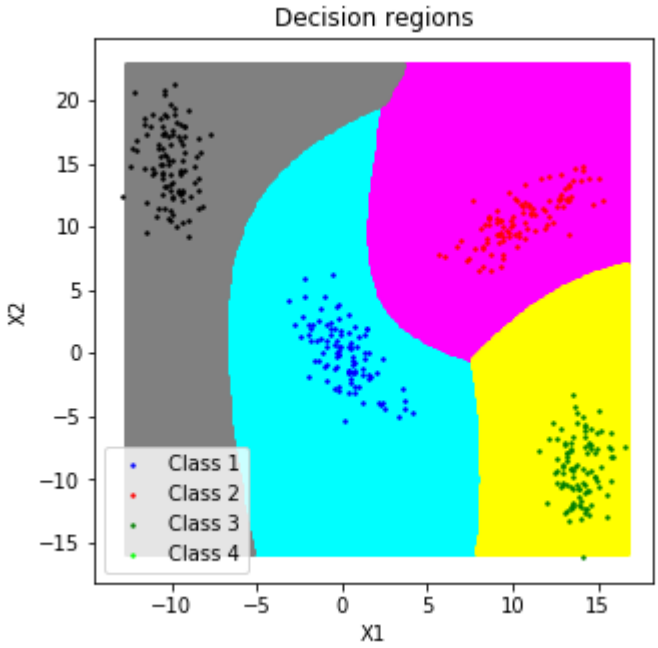

Class 2 vs Class 4
Class 2
Class 4
X1
X2


Class 3 vs Class 4
Class 3
Class 4
X1
X2

1.1.4.5 Decision region plot for all the classes together with the training data superposed



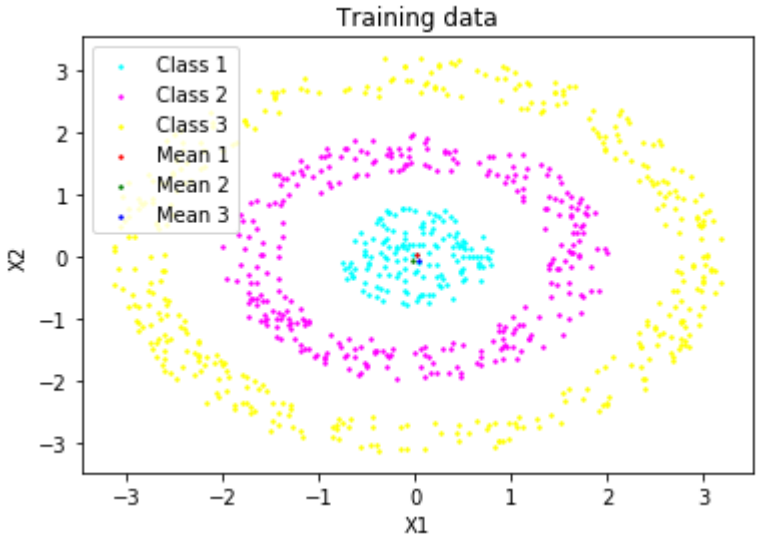
1.1.4.6 Decision region plot for all the classes together with the testing data superimposed



1.2 Non-Linearly Separable Data

1.2.1 Bayes Classifier with Same Diagonal Covariance Matrix

1.2.1.1 Plot of training data with mean displayed in different colour



1.2.1.2 Classification accuracy, precision for every class, mean precision, recall for every class, mean recall, F-measure for every class and mean F-measure on test data

Accuracy: [0.0, 0.0, 100.0]
Precision: [nan, nan, 0.47058823529411764]
Recall: [0.0, 0.0, 1.0]
F_measure: [nan, nan, 0.6399999999999999]

Mean Accuracy: 33.333333333333336
Mean Precision: nan
Mean Recall: 0.3333333333333333
Mean F_measure: nan

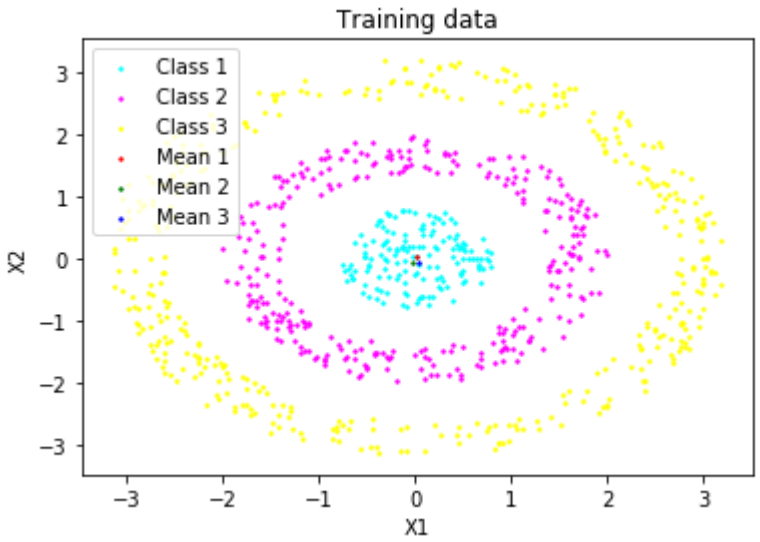
1.2.1.3 Confusion matrix based on the performance for test data. The entries in confusion matrix must be made in percentage.

	True 1	True 2	True 3
Predicted 1	0.0	0.0	0.0
Predicted 2	0.0	0.0	0.0
Predicted 3	60.0	120.0	160.0

Since all the means are nearly at the centre and all the classes are having same covariance matrix(linear boundary) we cannot draw a linear boundary and the classifier predicts as all points belong to class 3 so we can't use this for classifying this type of data.

1.2.2 Bayes Classifier with Same Full Covariance Matrix

1.2.2.1 Plot of training data with mean displayed in different colour



1.2.2.2 Classification accuracy, precision for every class, mean precision, recall for every class, mean recall, F-measure for every class and mean F-measure on test data

Accuracy: [0.0, 0.0, 100.0] Mean Accuracy: 33.333333333333336
Precision: [nan, nan, 0.47058823529411764] Mean Precision: nan
Recall: [0.0, 0.0, 1.0] Mean Recall: 0.3333333333333333
F_measure: [nan, nan, 0.6399999999999999] Mean F_measure: nan

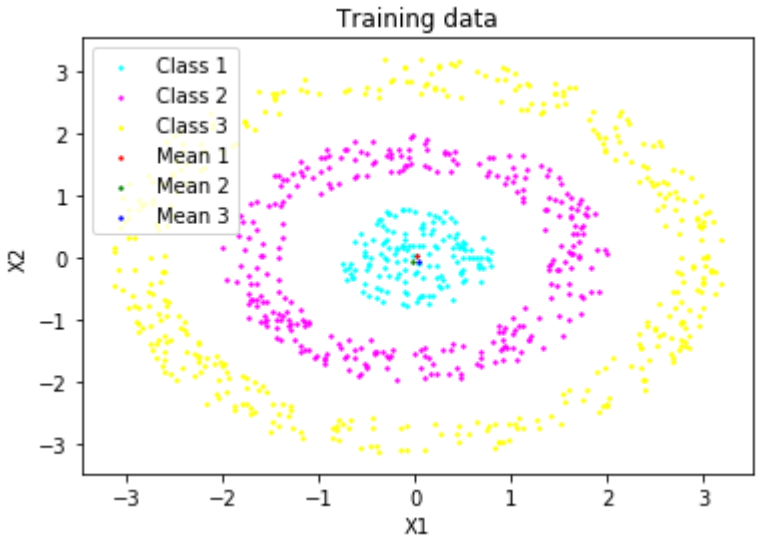
1.2.2.3 Confusion matrix based on the performance for test data. The entries in confusion matrix must be made in percentage.

	True 1	True 2	True 3
Predicted 1	0.0	0.0	0.0
Predicted 2	0.0	0.0	0.0
Predicted 3	60.0	120.0	160.0

Since all the means are nearly at the centre and all the classes are having same covariance matrix(linear boundary) we cannot draw a linear boundary and the classifier predicts as all points belong to class 3 so we can't use this for classifying this type of data.

1.2.3 Bayes Classifier with Different Diagonal Covariance Matrix

1.2.3.1 Plot of training data with mean displayed in different colour



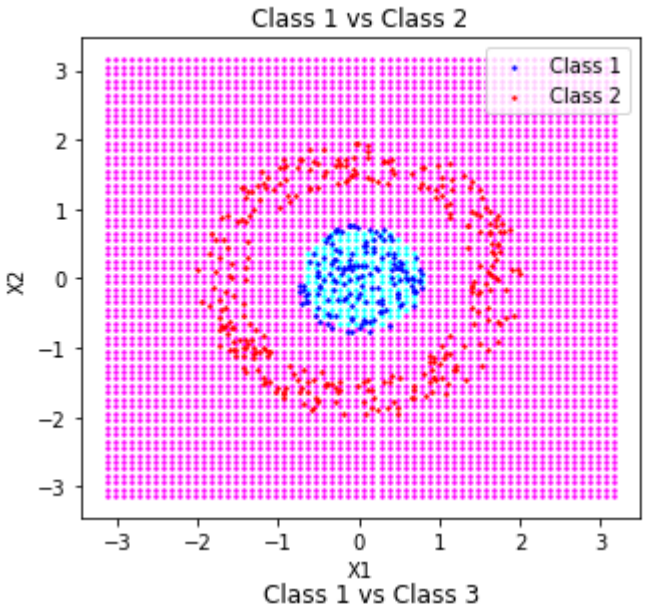
1.2.3.2 Classification accuracy, precision for every class, mean precision, recall for every class, mean recall, F-measure for every class and mean F-measure on test data

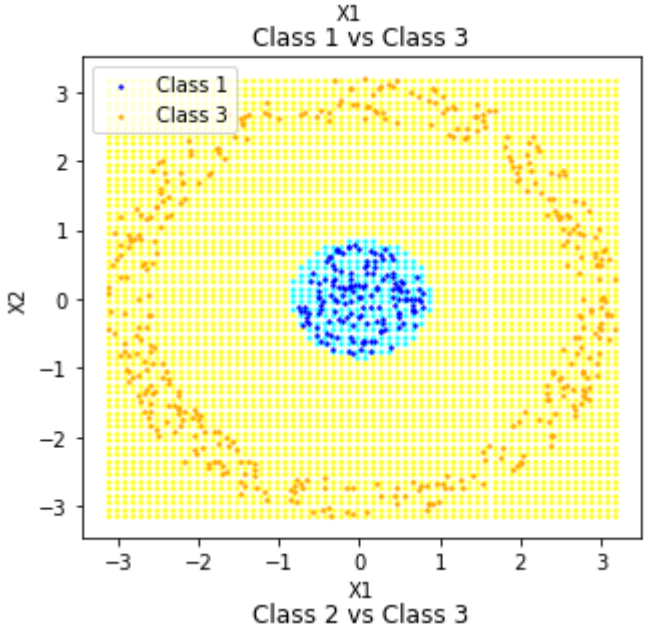
Accuracy: [86.66666666666667, 71.66666666666667, 100.0]
Mean Accuracy: 86.11111111111113
Precision: [1.0, 0.9148936170212766, 0.8247422680412371]
Mean Precision: 0.9132119616875046
Recall: [0.8666666666666667, 0.7166666666666667, 1.0]
Mean Recall: 0.8611111111111112
F_measure: [0.9285714285714286, 0.8037383177570094, 0.903954802259887]
Mean F_measure: 0.8787548495294417

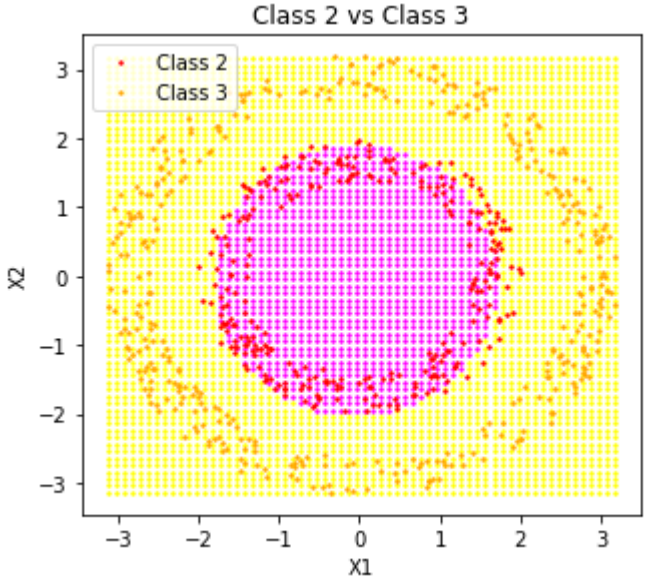
1.2.3.3 Confusion matrix based on the performance for test data. The entries in confusion matrix must be made in percentage.

	True 1	True 2	True 3
Predicted 1	52.0	0.0	0.0
Predicted 2	8.0	86.0	0.0
Predicted 3	0.0	34.0	160.0

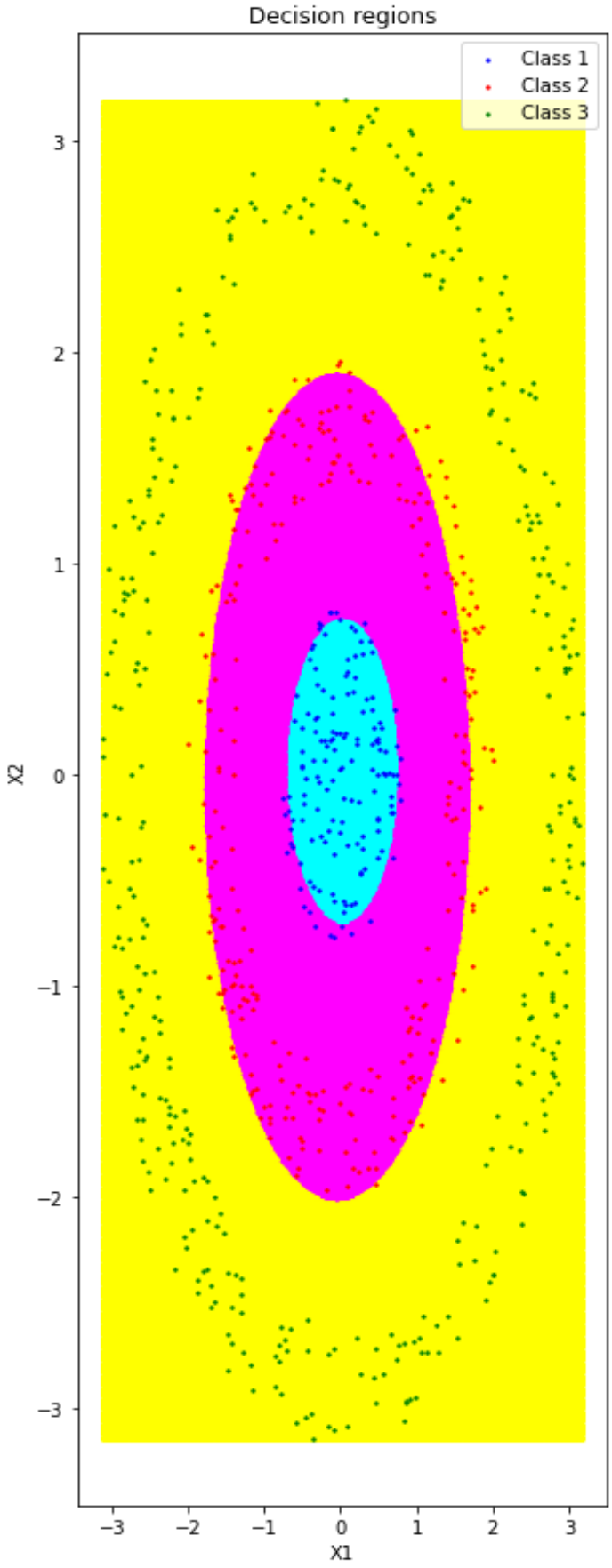
1.2.3.4 Decision region plot for every pair of classes with the respective training data superimposed



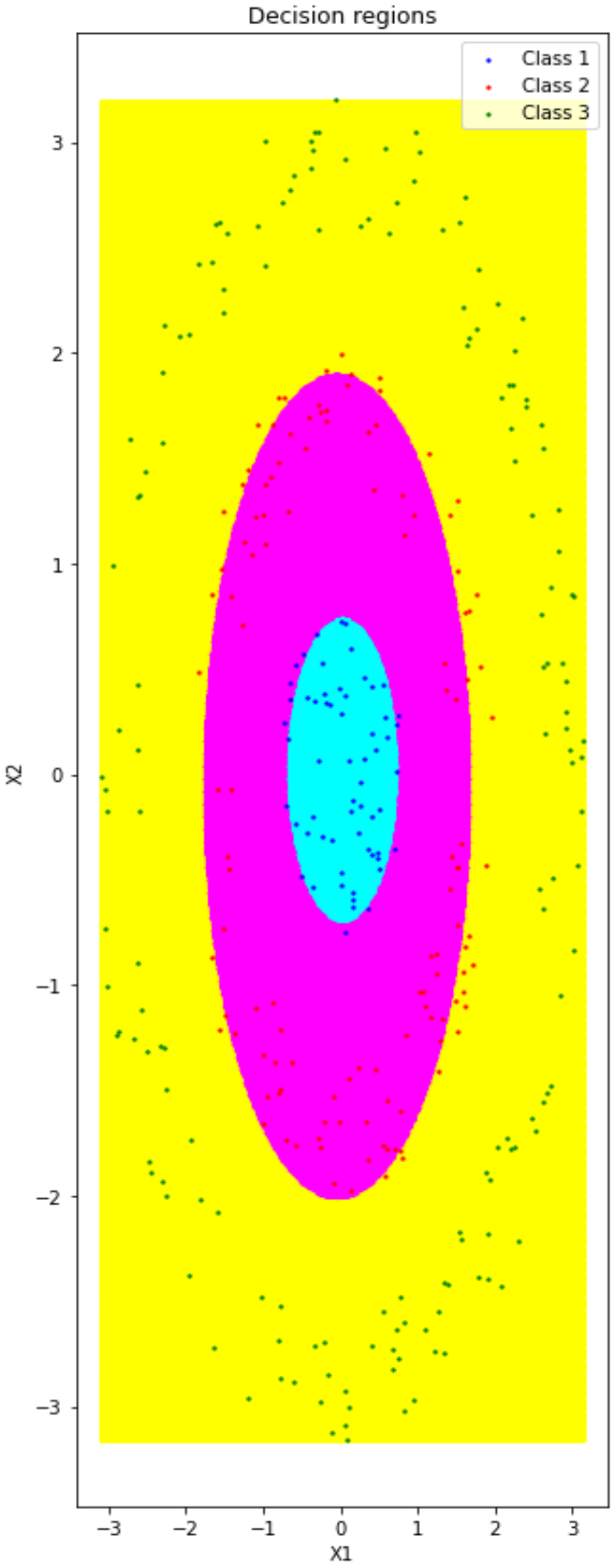




1.2.3.5 Decision region plot for all the classes together with the training data superposed

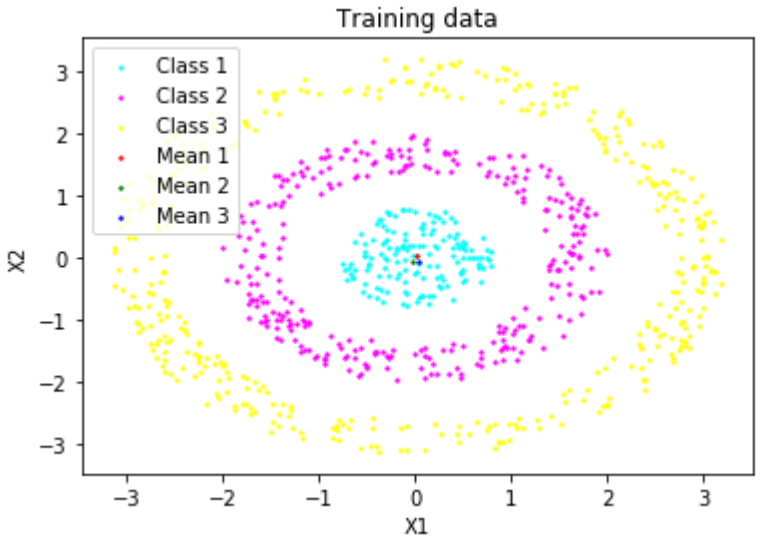


1.2.3.6 Decision region plot for all the classes together with the testing data superimposed



1.2.4 Bayes Classifier with Different Full Covariance Matrix

1.2.4.1 Plot of training data with mean displayed in different colour



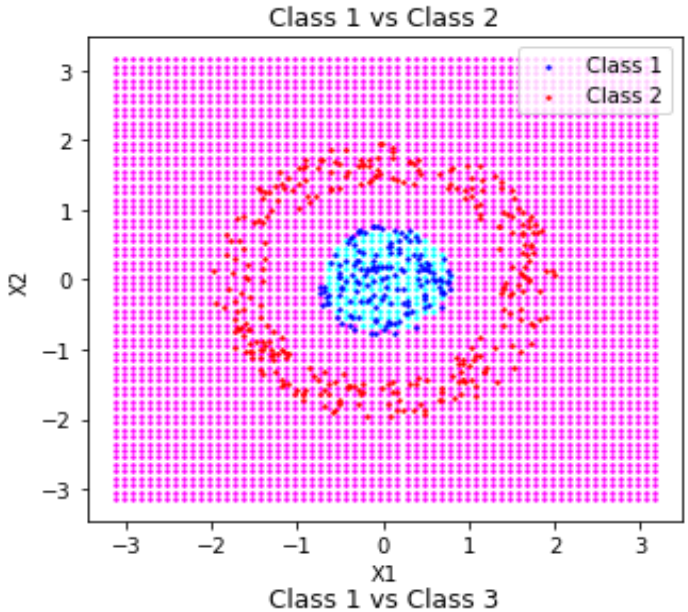
1.2.4.2 Classification accuracy, precision for every class, mean precision, recall for every class, mean recall, F-measure for every class and mean F-measure on test data

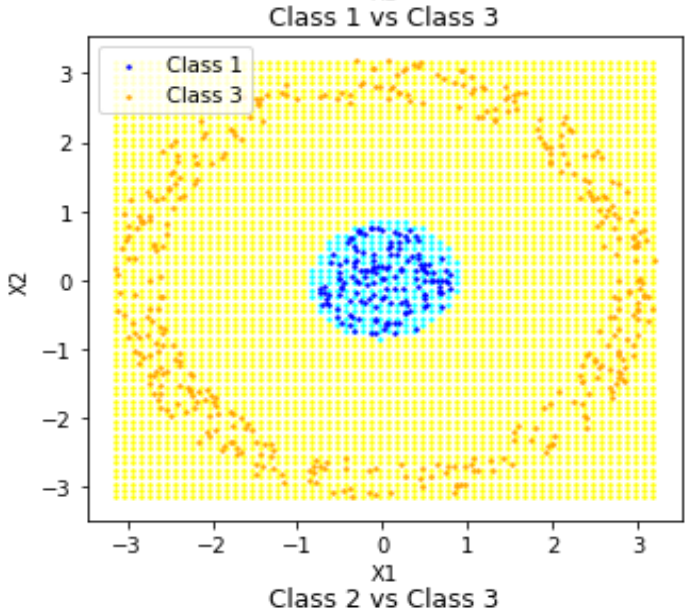
Accuracy: [83.33333333333333, 70.83333333333333, 100.0]
Mean Accuracy: 84.72222222222221}
Precision: [1.0, 0.8947368421052632, 0.8205128205128205]
Mean Precision: 0.9050832208726947
Recall: [0.8333333333333334, 0.7083333333333334, 1.0]
Mean Recall: 0.8472222222222223
F_measure: [0.9090909090909091, 0.7906976744186046, 0.9014084507042254]
Mean F_measure: 0.8670656780712463

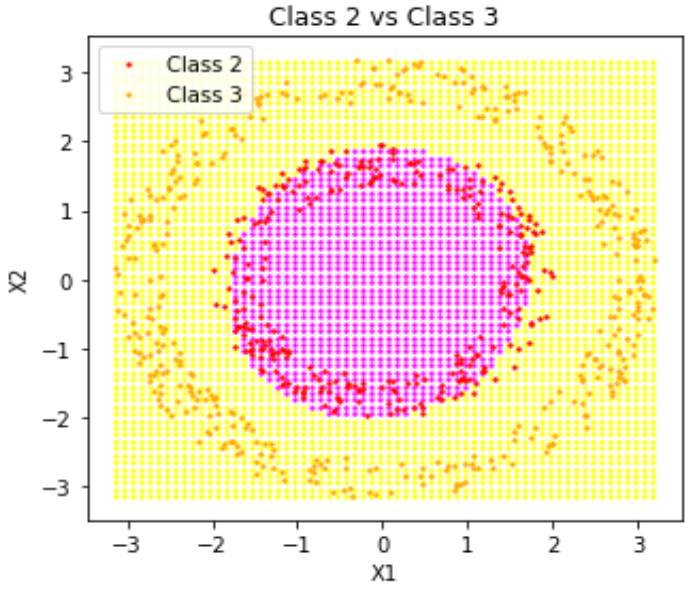
1.2.4.3 Confusion matrix based on the performance for test data. The entries in confusion matrix must be made in percentage.

	True 1	True 2	True 3
Predicted 1	50.0	0.0	0.0
Predicted 2	10.0	85.0	0.0
Predicted 3	0.0	35.0	160.0

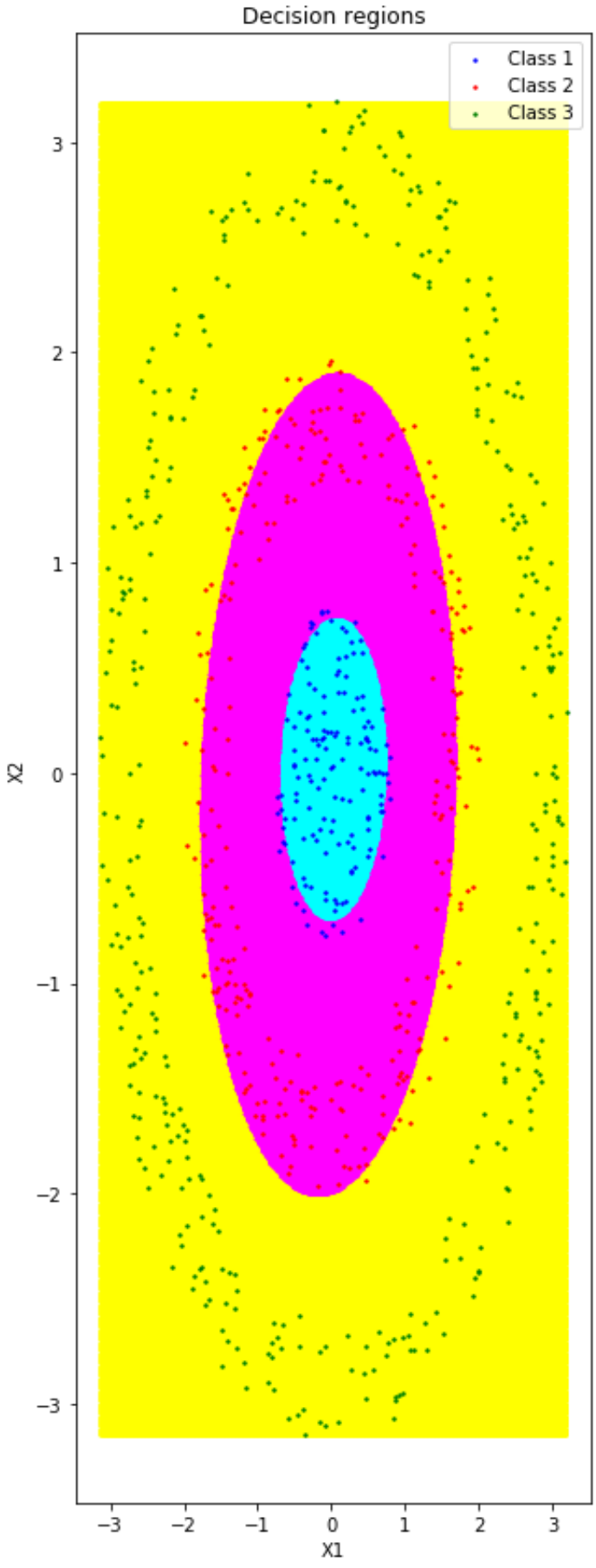
1.2.4.4 Decision region plot for every pair of classes with the respective training data superimposed



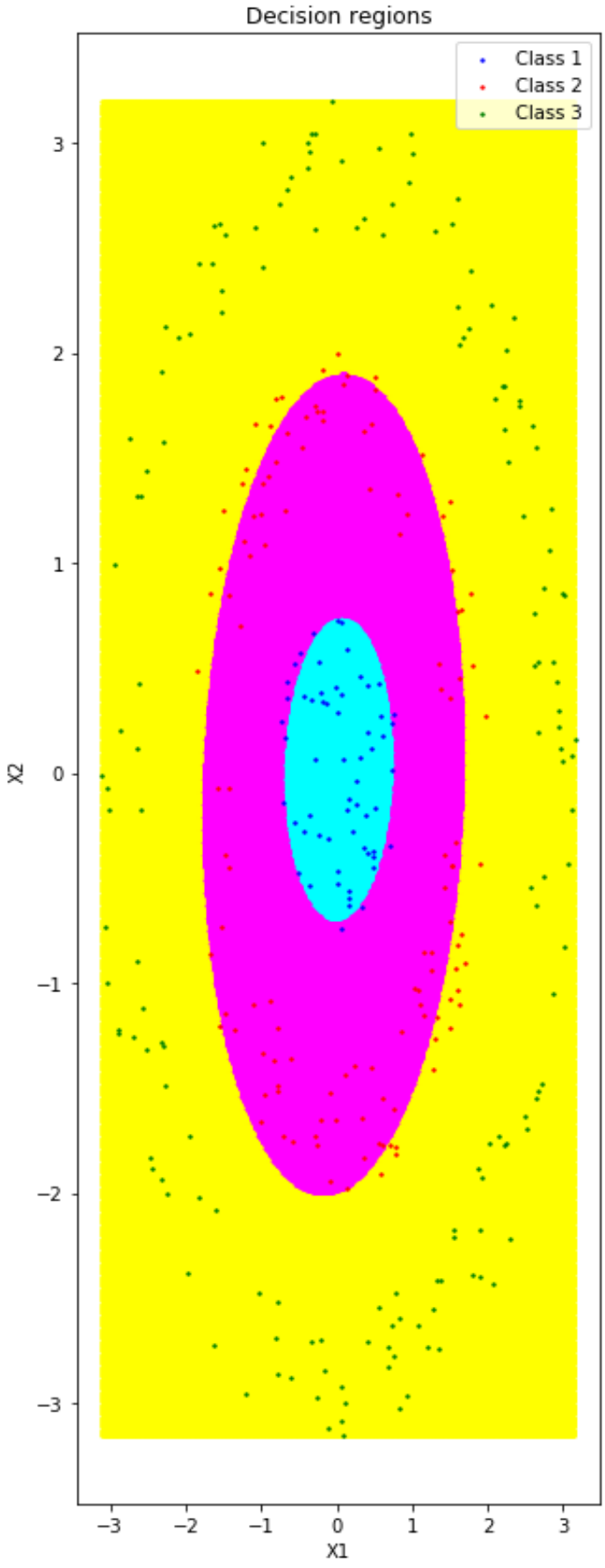




1.2.4.5 Decision region plot for all the classes together with the training data superposed



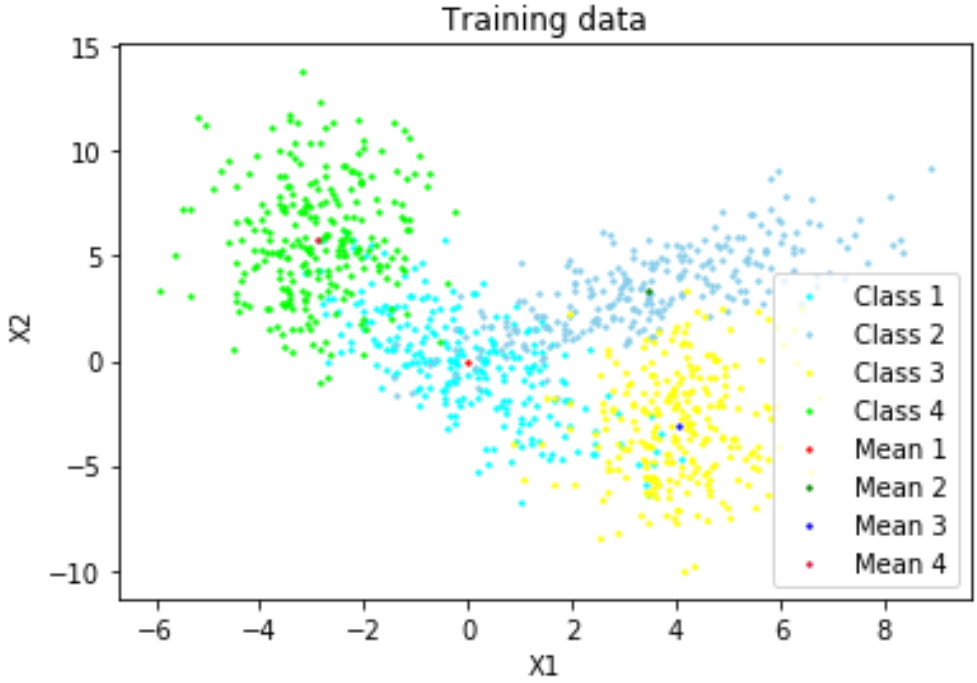
1.2.4.6 Decision region plot for all the classes together with the testing data superimposed



1.3 Overlapping Data Separable Data

1.3.1 Bayes Classifier with Same Diagonal Covariance Matrix

1.3.1.1 Plot of training data with mean displayed in different colour



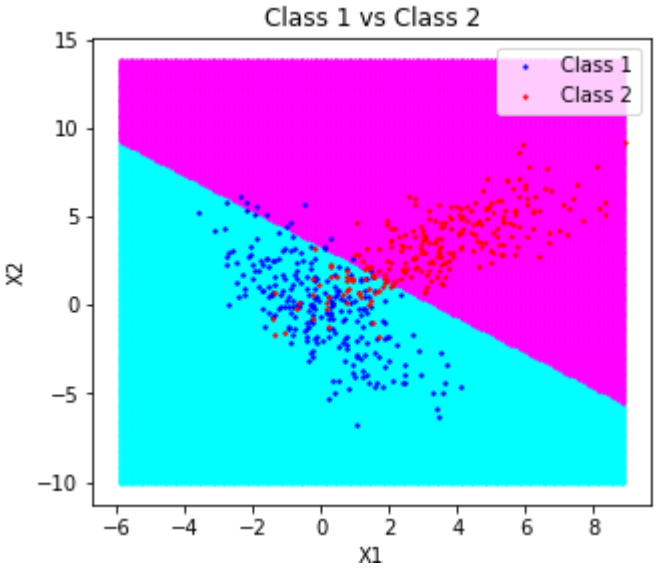
1.3.1.2 Classification accuracy, precision for every class, mean precision, recall for every class, mean recall, F-measure for every class and mean F-measure on test data

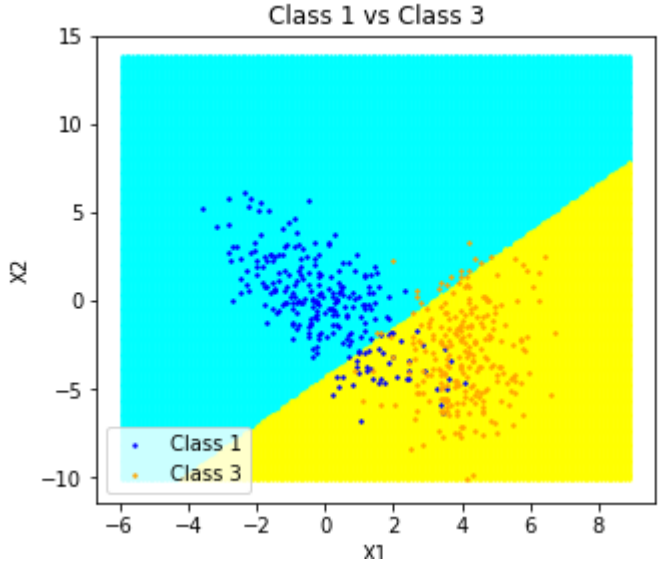
Accuracy: [75.0, 79.0, 88.0, 95.0] Mean Accuracy: 84.25
Precision: [0.7211538461538461, 0.8876404494382022, 0.8888888888888888, 0.8796296296296297]
Mean Precision: 0.8443282035276417
Recall: [0.75, 0.79, 0.88, 0.95] Mean Recall: 0.8425
F_measure: [0.7352941176470588, 0.8359788359788359, 0.8844221105527638, 0.9134615384615385]
Mean F_measure: 0.8422891506600492

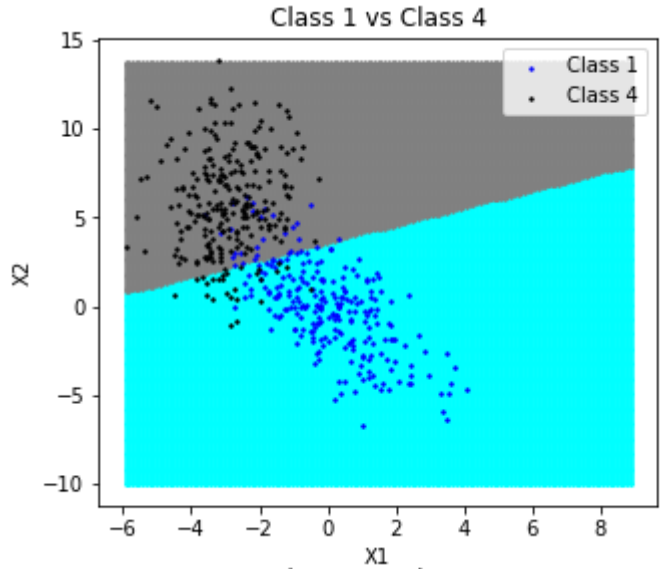
1.3.1.3 Confusion matrix based on the performance for test data. The entries in confusion matrix must be made in percentage.

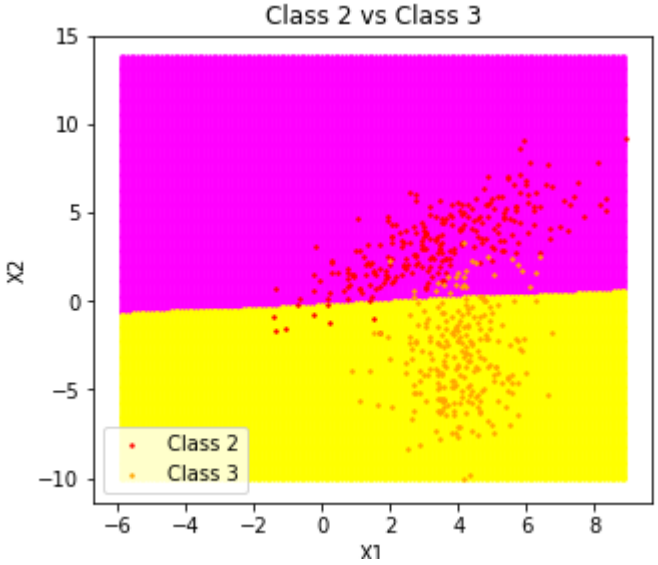
	True 1	True 2	True 3	True 4
Predicted 1	75.0	20.0	4.0	5.0
Predicted 2	2.0	79.0	8.0	0.0
Predicted 3	11.0	0.0	88.0	0.0
Predicted 4	12.0	1.0	0.0	95.0

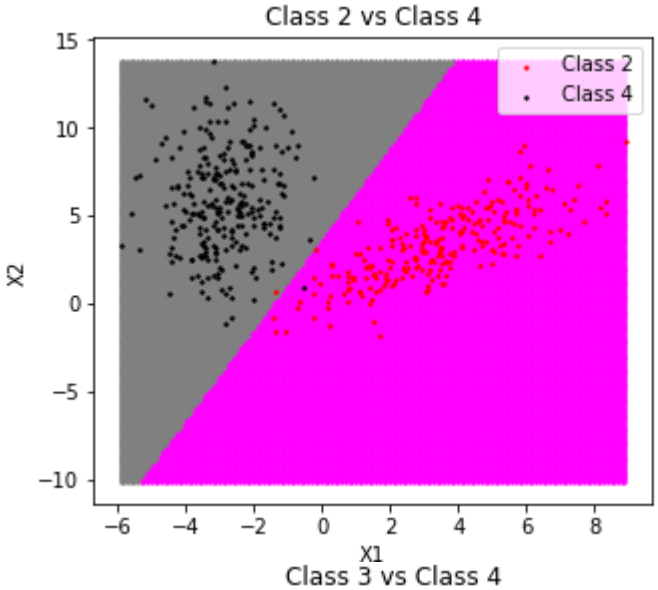
1.3.1.4 Decision region plot for every pair of classes with the respective training data superimposed

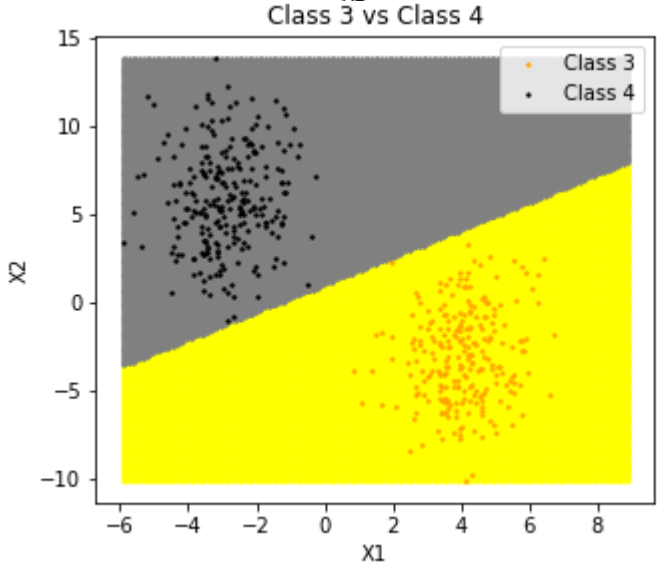




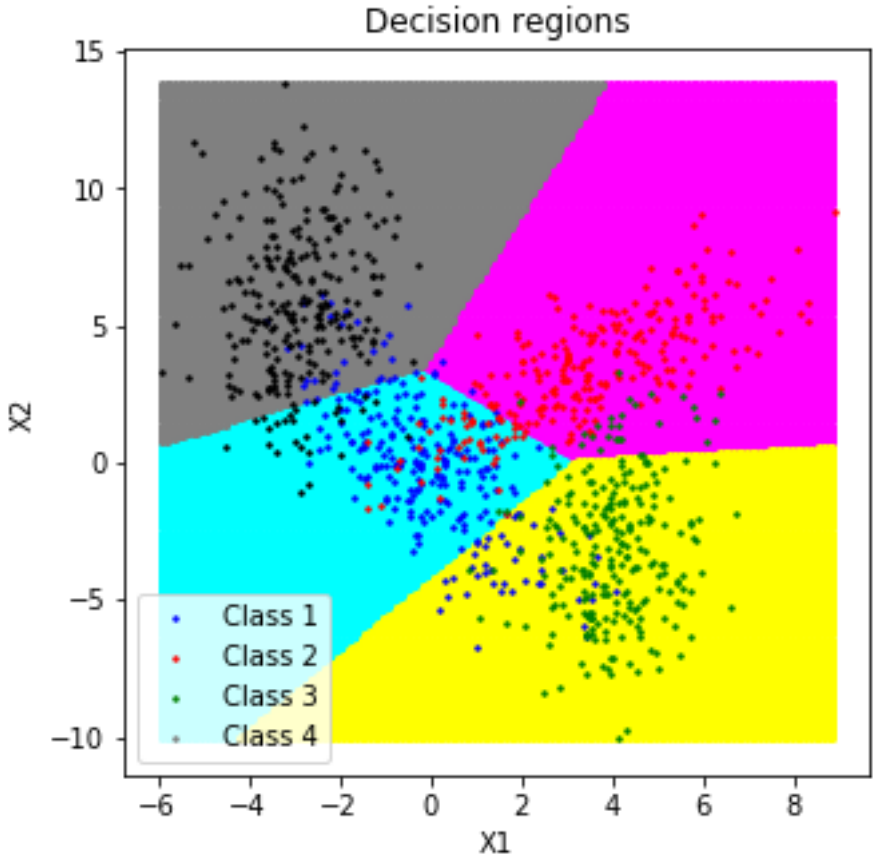



Class 2 vs Class 3
Class 2
Class 3
X1
X2

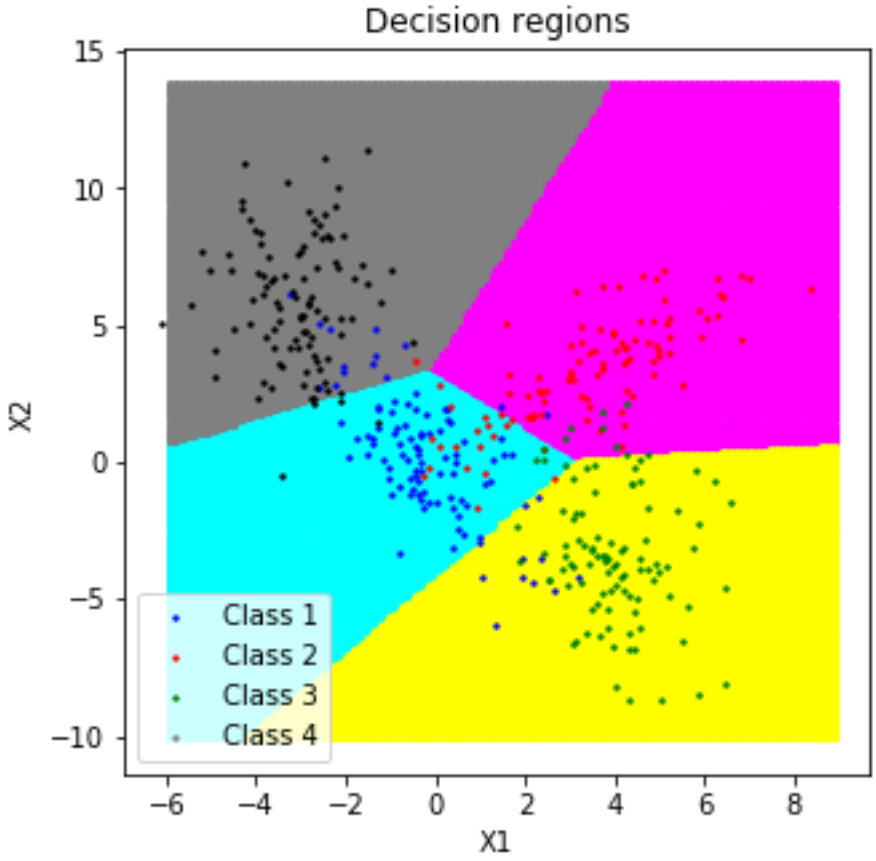

Class 2 vs Class 4
Class 2
Class 4
X1
X2


Class 3 vs Class 4
Class 3
Class 4
X1
X2

1.3.1.5 Decision region plot for all the classes together with the training data superposed

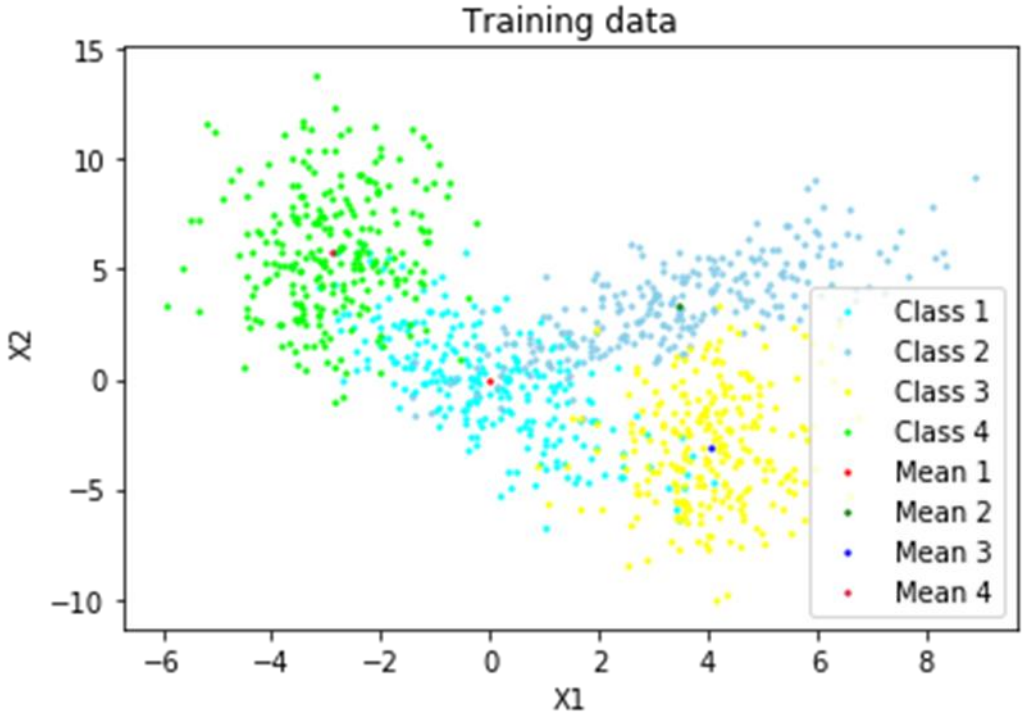


1.3.1.6 Decision region plot for all the classes together with the testing data superimposed



1.3.2 Bayes Classifier with Same Full Covariance Matrix

1.3.2.1 Plot of training data with mean displayed in different colour



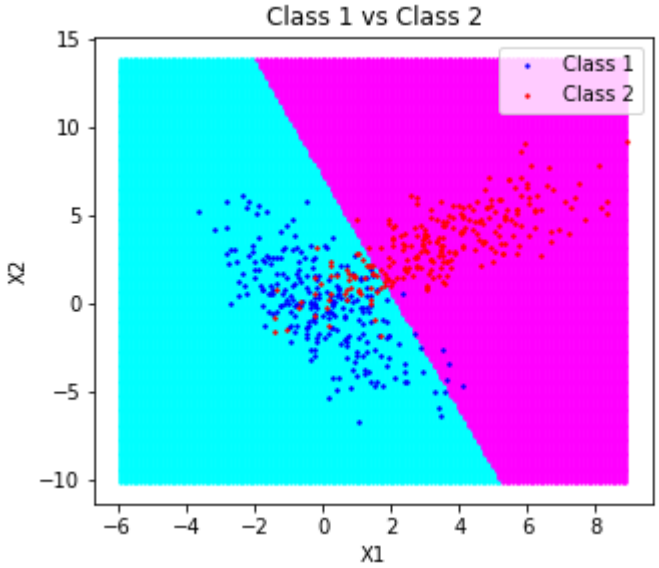
1.3.2.2 Classification accuracy, precision for every class, mean precision, recall for every class, mean recall, F-measure for every class and mean F-measure on test data

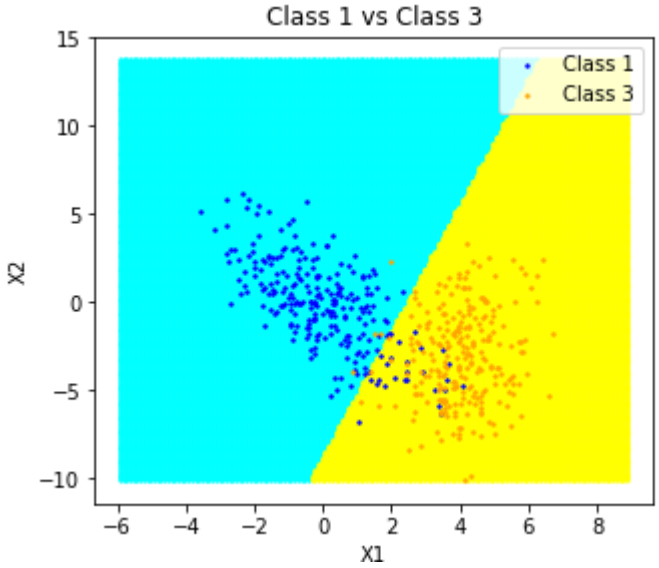
Accuracy: [79.0, 80.0, 90.0, 98.0] Mean Accuracy: 86.75
Precision: [0.79, 0.8791208791208791, 0.9, 0.8990825688073395]
Mean Precision: 0.8670508619820546
Recall: [0.79, 0.8, 0.9, 0.98] Mean Recall: 0.8675
F_measure: [0.79, 0.837696335078534, 0.9, 0.937799043062201]
Mean F_measure: 0.8663738445351837

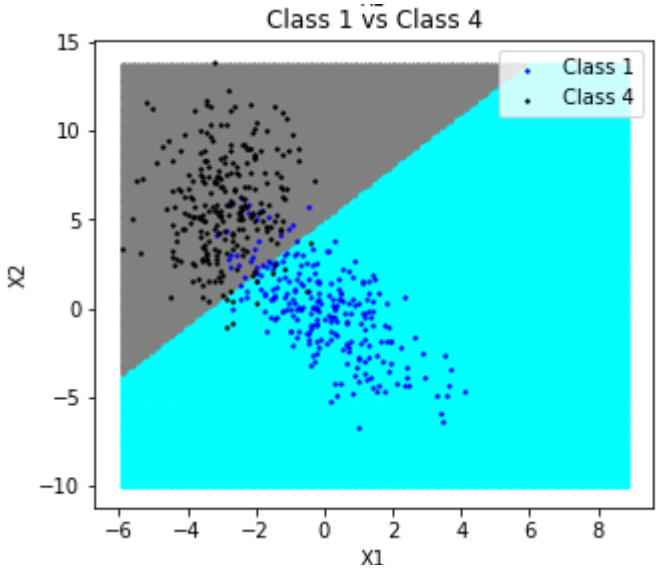
1.3.2.3 Confusion matrix based on the performance for test data. The entries in confusion matrix must be made in percentage.

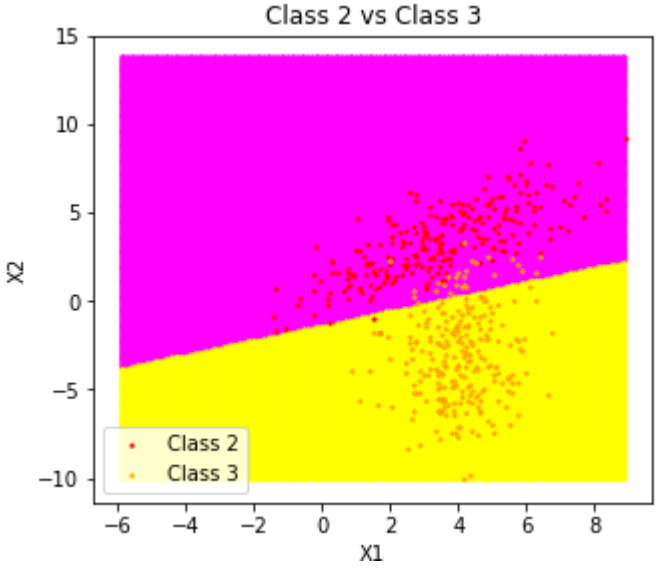
	True 1	True 2	True 3	True 4
Predicted 1	79.0	19.0	0.0	2.0
Predicted 2	1.0	80.0	10.0	0.0
Predicted 3	9.0	1.0	90.0	0.0
Predicted 4	11.0	0.0	0.0	98.0

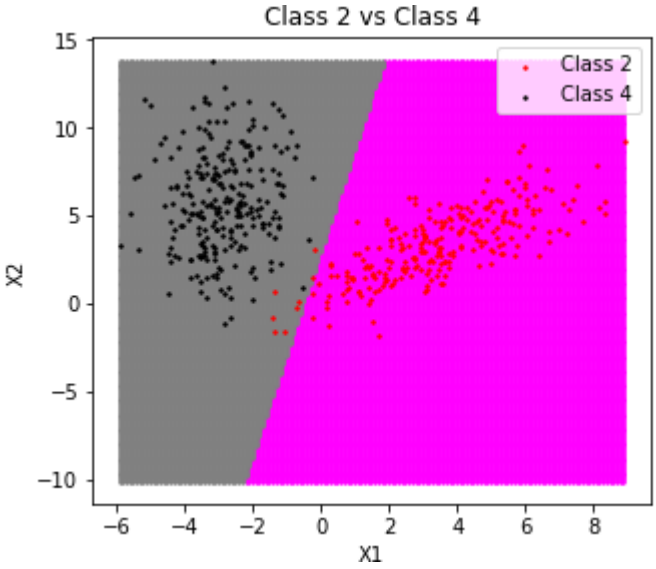
1.3.2.4 Decision region plot for every pair of classes with the respective training data superimposed

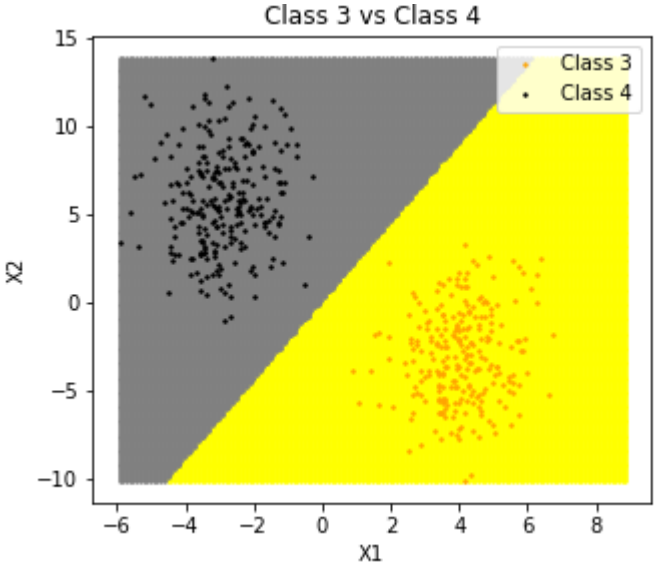




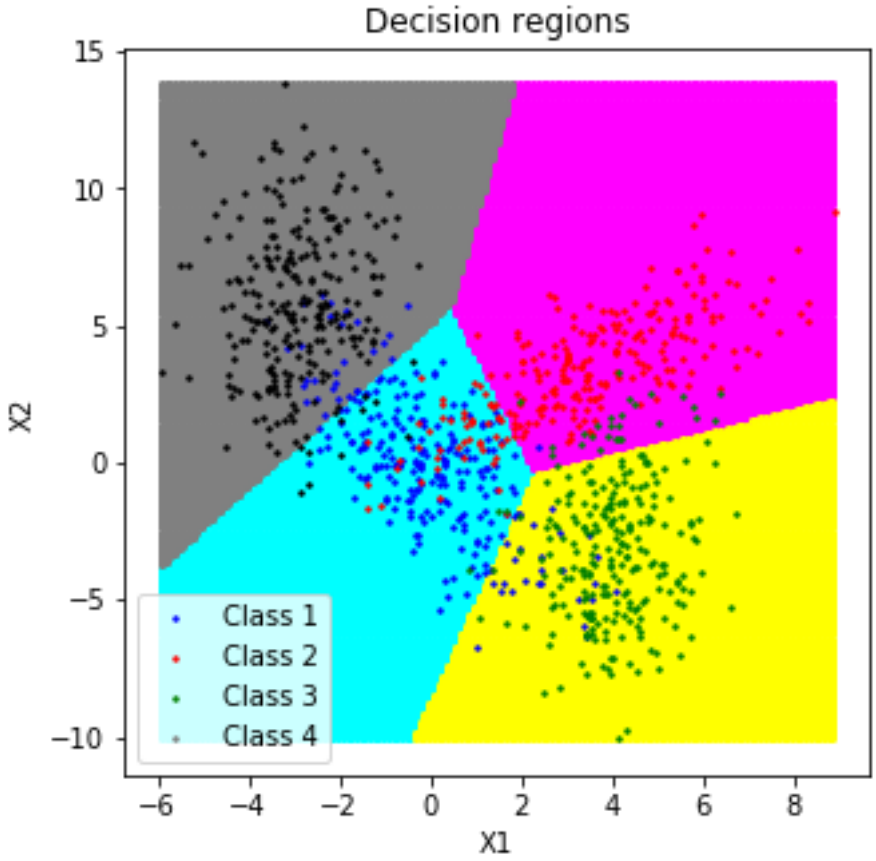



Class 2 vs Class 3
Class 2
Class 3
X1
X2

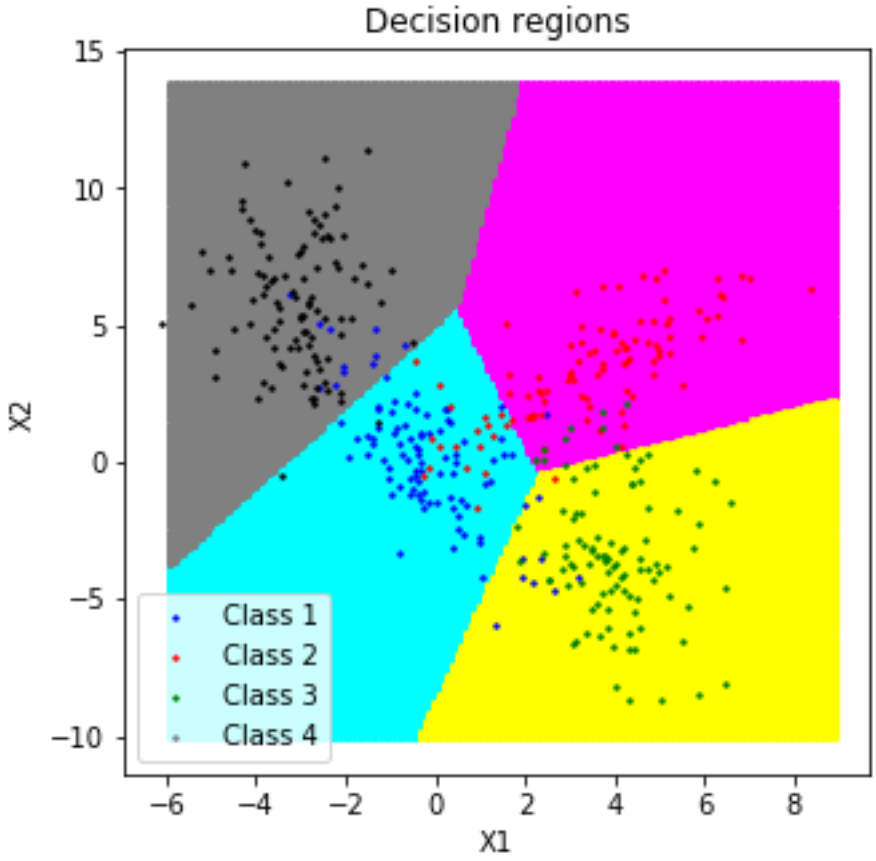

Class 2 vs Class 4
Class 2
Class 4
X1
X2


Class 3 vs Class 4
Class 3
Class 4
X1
X2

1.3.2.5 Decision region plot for all the classes together with the training data superposed

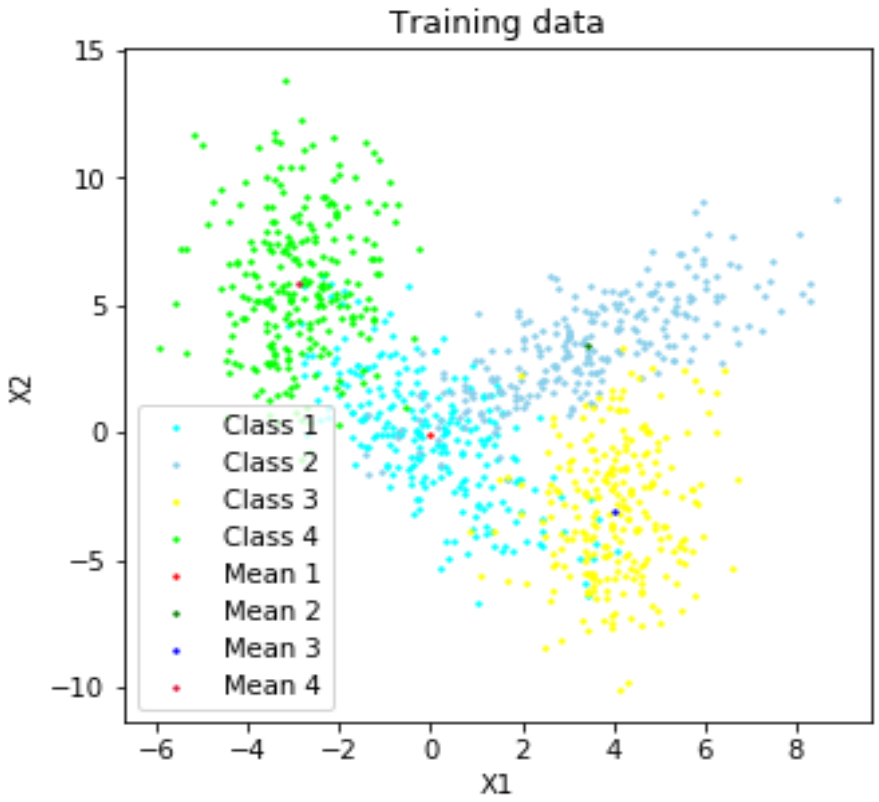


1.3.2.6 Decision region plot for all the classes together with the testing data superimposed



1.3.3 Bayes Classifier with Different Diagonal Covariance Matrix

1.3.3.1 Plot of training data with mean displayed in different colour



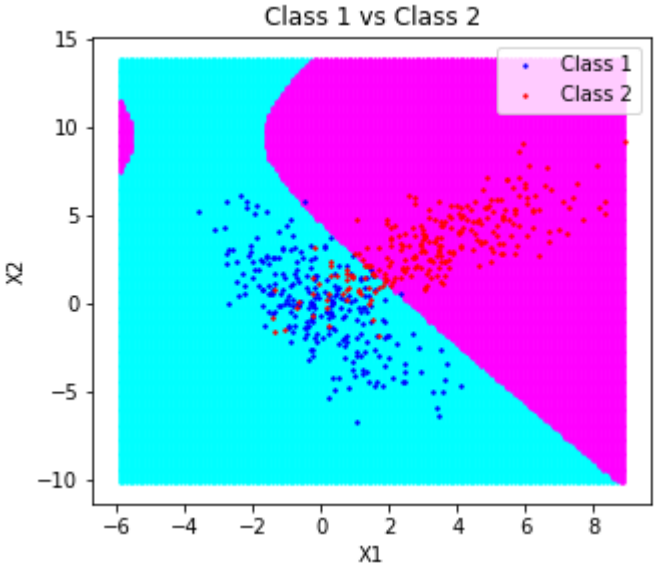
1.3.3.2 Classification accuracy, precision for every class, mean precision, recall for every class, mean recall, F-measure for every class and mean F-measure on test data

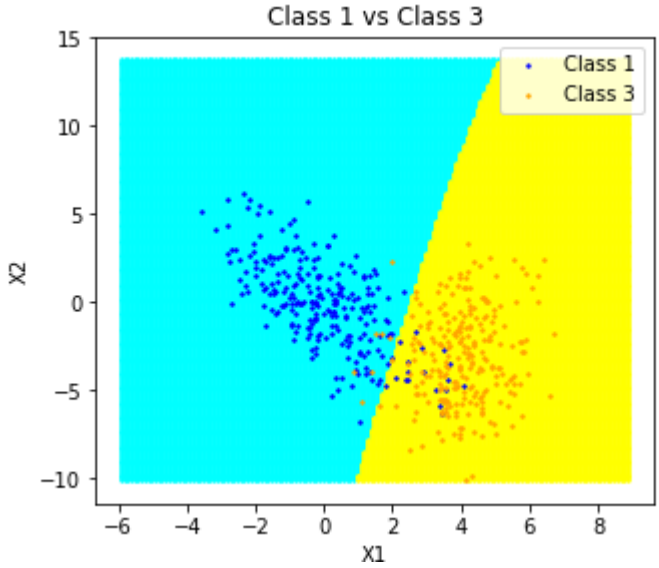
Accuracy: [83.0, 80.0, 90.0, 98.0] Mean Accuracy: 87.75
Precision: [0.7757009345794392, 0.9090909090909091, 0.9278350515463918, 0.9074074074074074]
Mean Precision: 0.880008575656037
Recall: [0.83, 0.8, 0.9, 0.98] Mean Recall: 0.8775
F_measure: [0.8019323671497584, 0.8510638297872342, 0.9137055837563451, 0.9423076923076924]
Mean F_measure: 0.8772523682502575

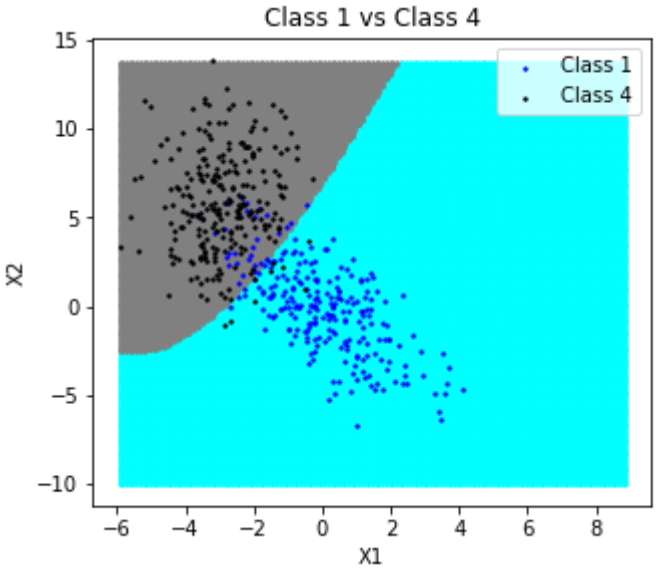
1.3.3.3 Confusion matrix based on the performance for test data. The entries in confusion matrix must be made in percentage.

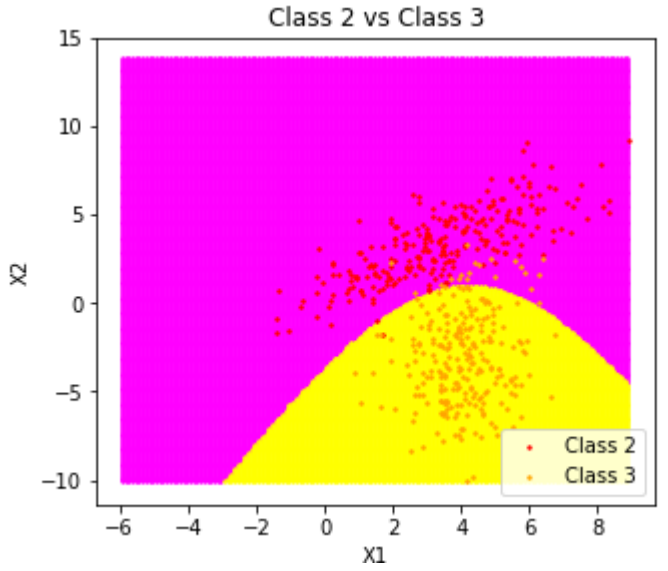
	True 1	True 2	True 3	True 4
Predicted 1	83.0	18.0	4.0	2.0
Predicted 2	2.0	80.0	6.0	0.0
Predicted 3	5.0	2.0	90.0	0.0
Predicted 4	10.0	0.0	0.0	98.0

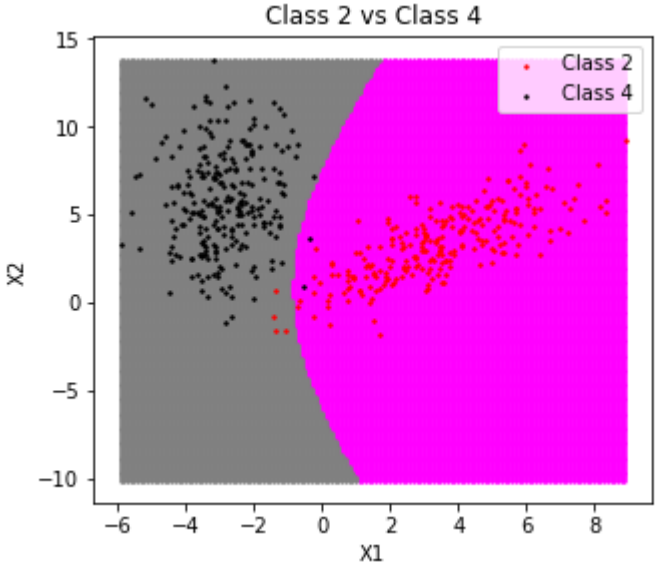
1.3.3.4 Decision region plot for every pair of classes with the respective training data superimposed

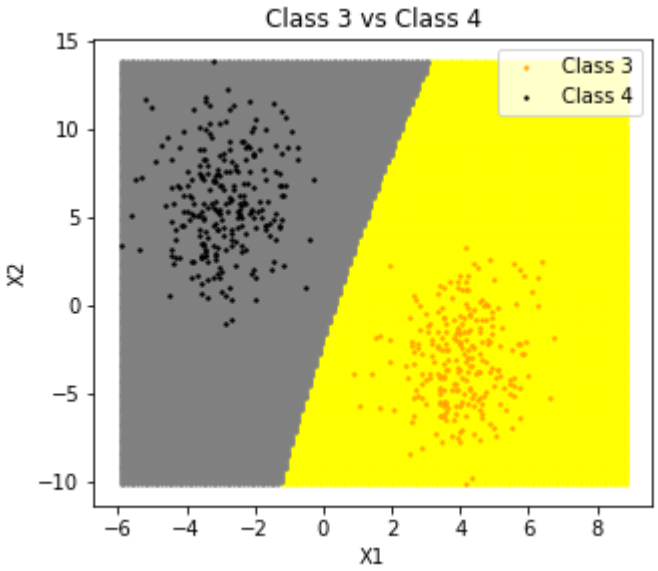




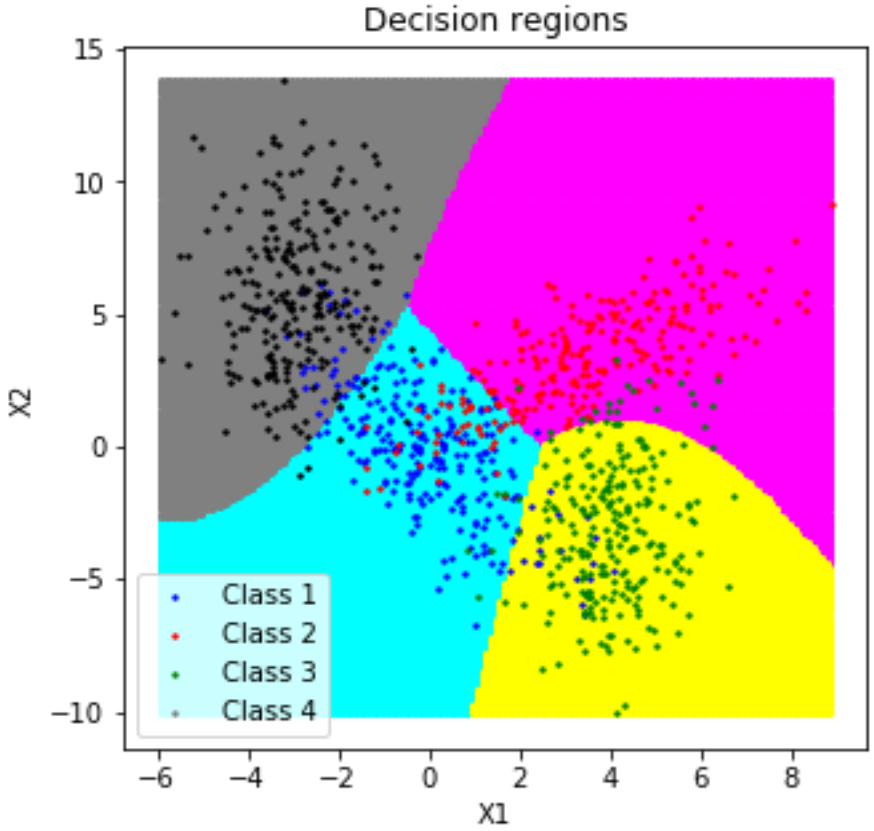



Class 2 vs Class 3
X1
X2
Class 2
Class 3

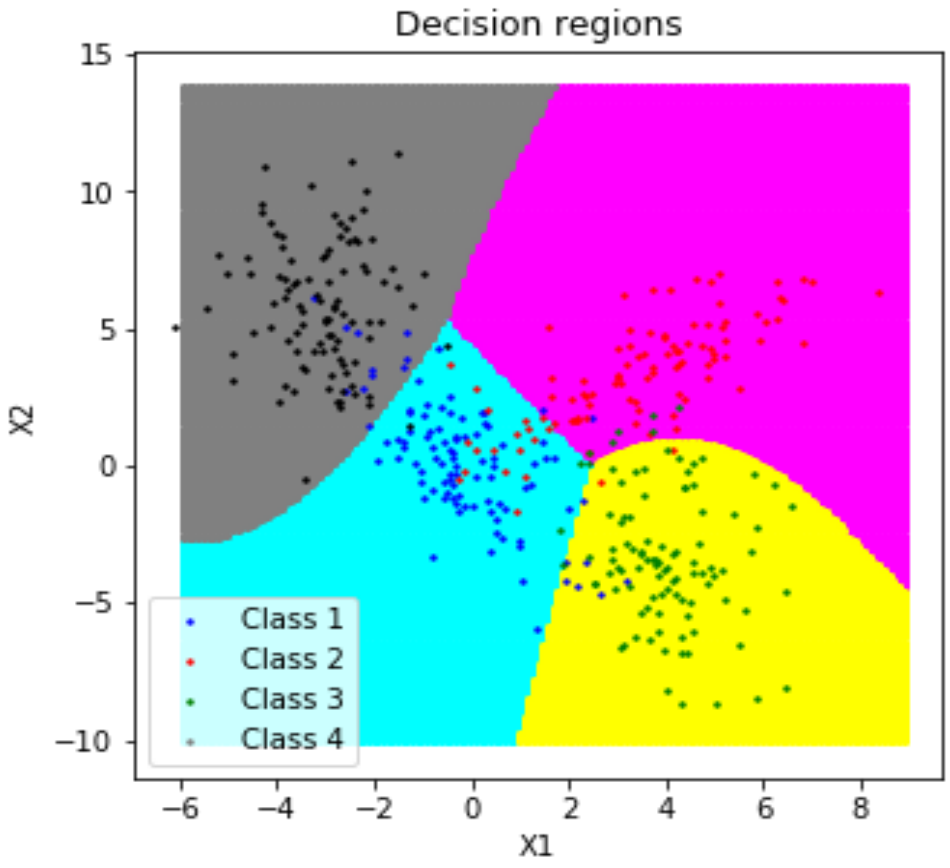

Class 2 vs Class 4
X1
X2
Class 2
Class 4


Class 3 vs Class 4
X1
X2
Class 3
Class 4

1.3.3.5 Decision region plot for all the classes together with the training data superposed

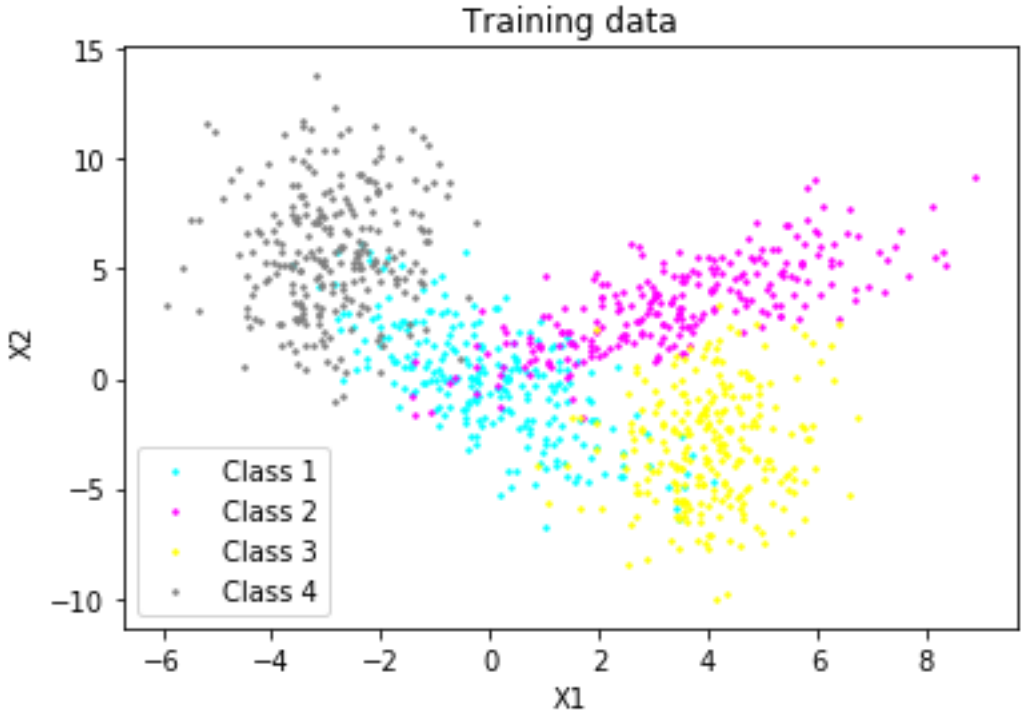


1.3.3.6 Decision region plot for all the classes together with the testing data superimposed



1.3.4 Bayes Classifier with Different Full Covariance Matrix

1.3.4.1 Plot of training data with mean displayed in different colour



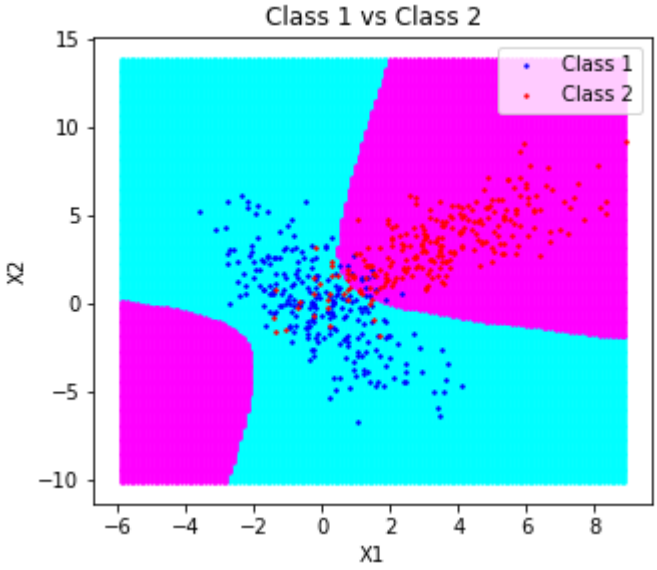
1.3.4.2 Classification accuracy, precision for every class, mean precision, recall for every class, mean recall, F-measure for every class and mean F-measure on test data

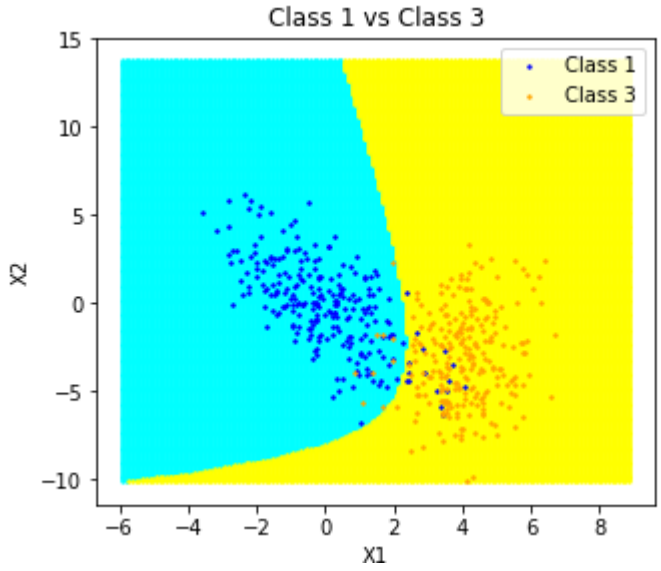
Accuracy: [86.0, 84.0, 91.0, 96.0] Mean Accuracy: 89.25
Precision: [0.8269230769230769, 0.8842105263157894, 0.9381443298969072, 0.9230769230769231]
Mean Precision: 0.8930887140531742
Recall: [0.86, 0.84, 0.91, 0.96] Mean Recall: 0.8925
F_measure: [0.8431372549019608, 0.8615384615384616, 0.9238578680203046, 0.9411764705882353]
Mean F_measure: 0.8924275137622406

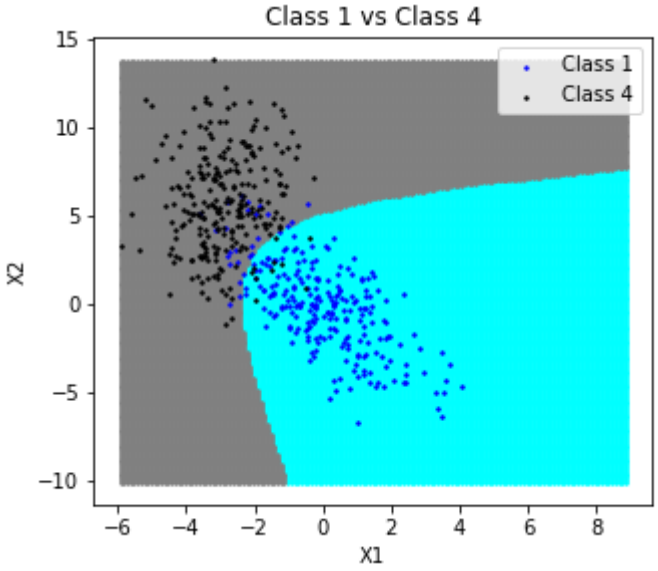
1.3.4.3 Confusion matrix based on the performance for test data. The entries in confusion matrix must be made in percentage.

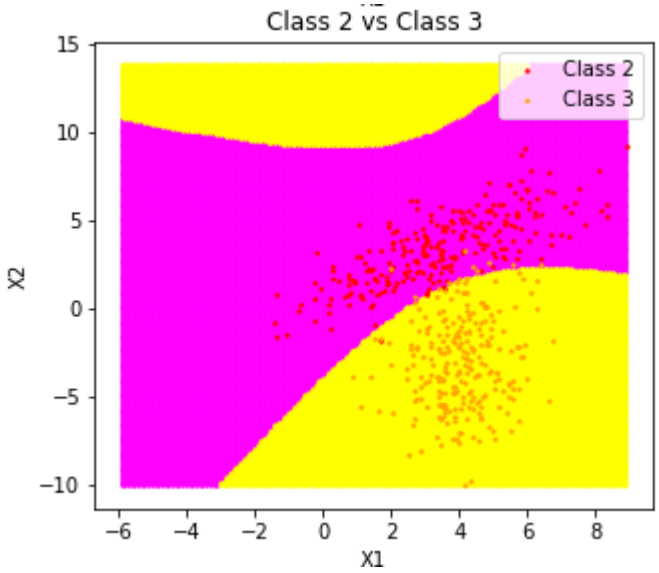
	True 1	True 2	True 3	True 4
Predicted 1	86.0	12.0	2.0	4.0
Predicted 2	4.0	84.0	7.0	0.0
Predicted 3	2.0	4.0	91.0	0.0
Predicted 4	8.0	0.0	0.0	96.0

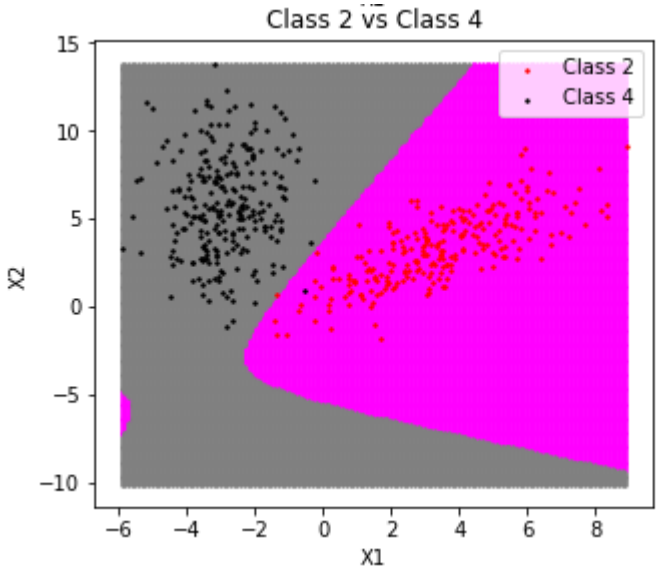
1.3.4.4 Decision region plot for every pair of classes with the respective training data superimposed

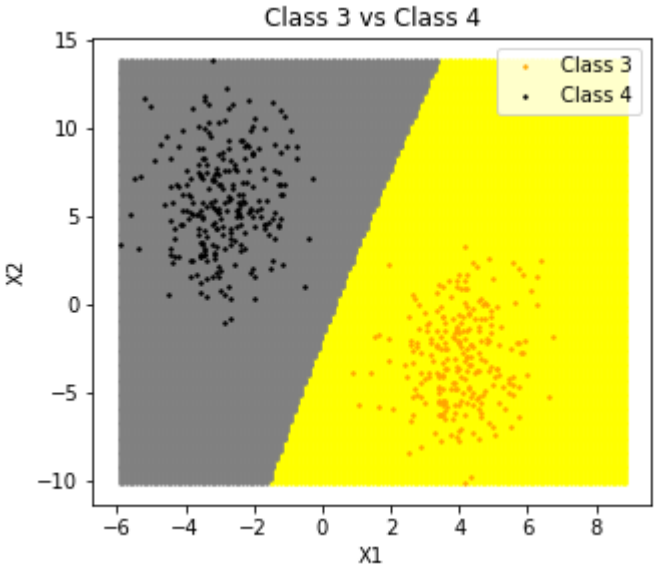




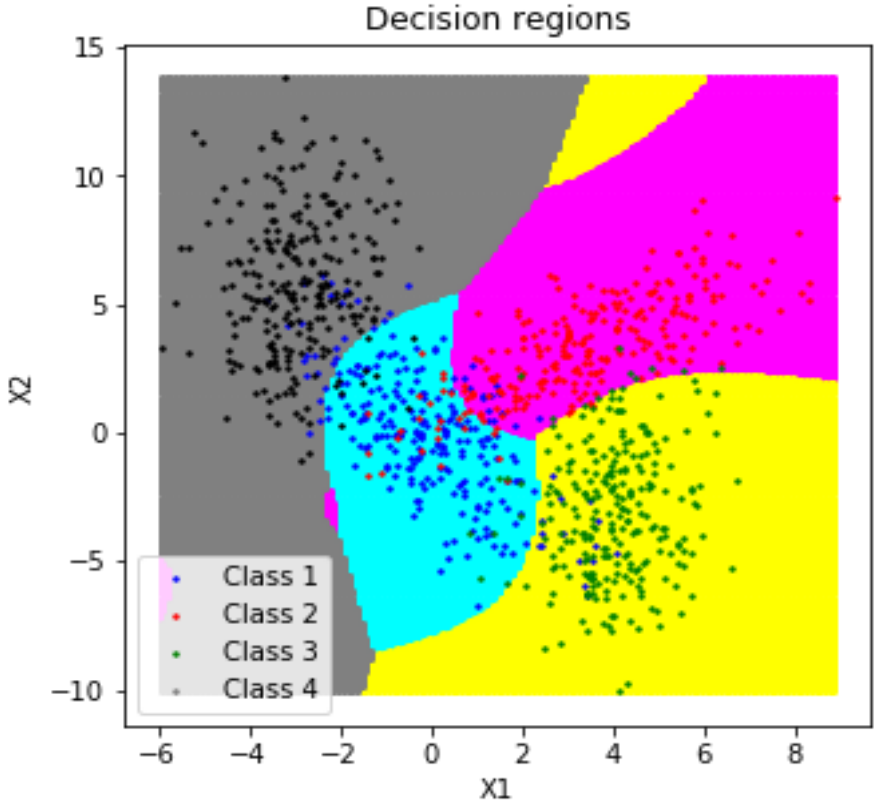



Class 2 vs Class 3
Class 2
Class 3
X1
X2

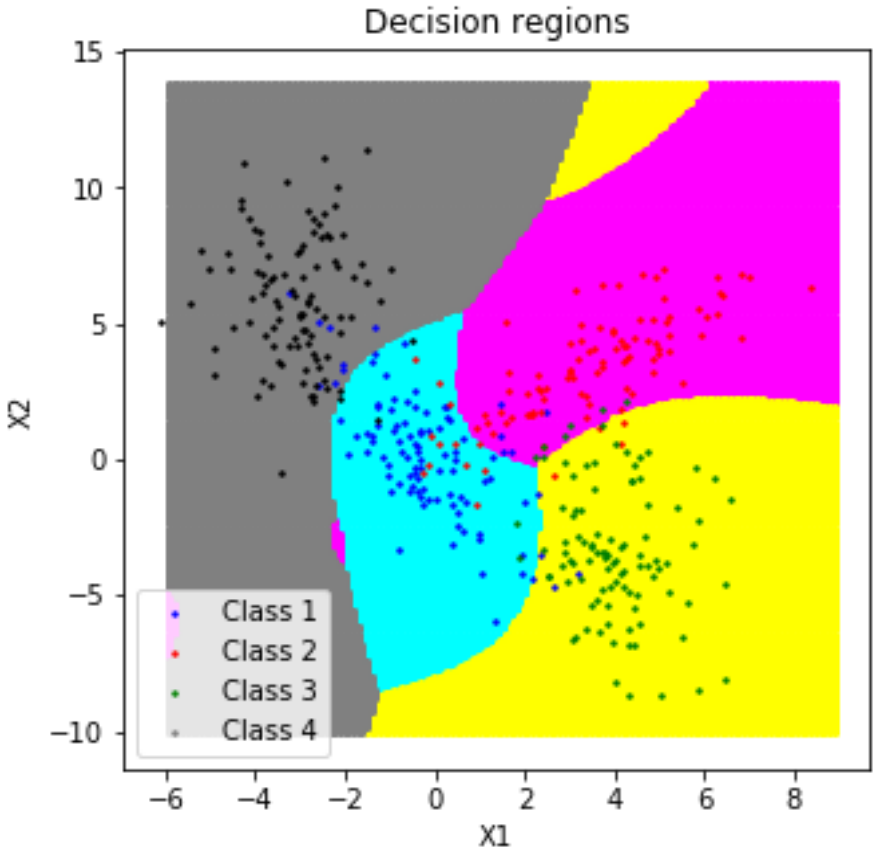

Class 2 vs Class 4
Class 2
Class 4
X1
X2


Class 3 vs Class 4
Class 3
Class 4
X1
X2

1.3.4.5 Decision region plot for all the classes together with the training data superposed



1.3.4.6 Decision region plot for all the classes together with the testing data superimposed



Conclusion

2.1 Linear Separable Classes

Since, the classes are linearly separable Bayes classifier achieves 100% accuracy in all cases.

2.2 Non-Linearly Separable Classes

The non-linearly separable data of the Bayes Classifier the data cannot be classified when we have same covariance matrix since linear boundary cannot classify concentric circular data and when we have different covariance matrix the mean accuracy is 86% & 83%. (diagonal different & full different covariance matrix). The decision boundary is concentric circles for last two cases.

2.3 Overlapping Classes

As the dataset given here is of overlapping nature, all four cases of Bayes classifier again perform poorly but when the covariance matrix changes from same covariance to different covariance matrix the average accuracy is increasing (84.25% to 89.25%), so Bayes Classifier with Different Full Covariance Matrix performs the best.