CS 669 Assignment 1

Rohit Patiyal
Devang Bacharwar
Group 6

September 14, 2015

Contents

1 Objective

To build Bayes and Naive-Bayes classifiers for different types of data sets :

1.1 2-D artificial Data of 3 or 4 classes

1. Linearly separable data set
2. Nonlinearly separable data sets (3 Data sets)
3. Overlapping data set

1.2 Real World data set

2 Procedure

1. Data for each class is partitioned into 75 % for training and 25 % for testing
2. Mean and Covariances are calculated for each class using the training data.
3. For points in a grid, likelihood is calculated for each class and is labeled as of the class with the maximum likelihood probability.
4. For bayes classifier, the likelihood is assumed to be a multivariate gaussian distribution
5. These labelled points are plotted with different colors to visualize the different regions separated by the decision boundaries.
6. The testing data is also plotted over the regions, and observations are made.

3 Observations

3.1 Bayes Classifier

3.1.1 Linearly separable data set

The decision boundary clearly separates the testing data according to the estimated classes, as the data forms widely separated clusters. Results are similar even when the covariance is taken as the average of individual class covariances and when the covariance is calculated using all classes' data together.

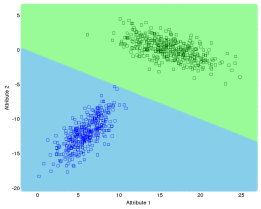
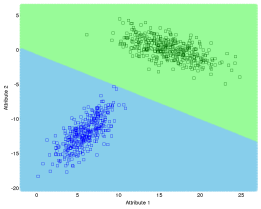
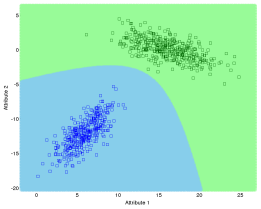
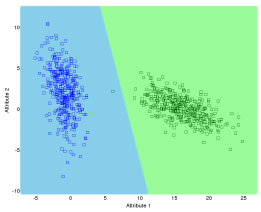
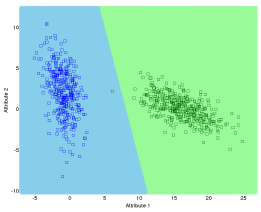
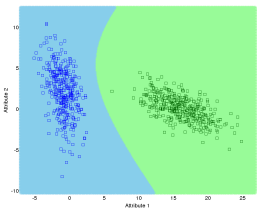
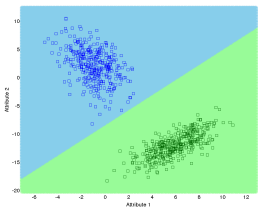
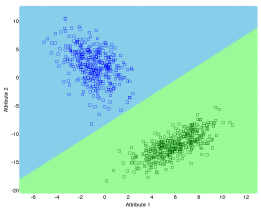
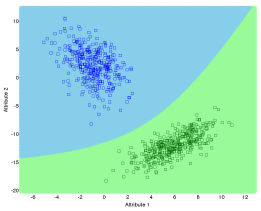
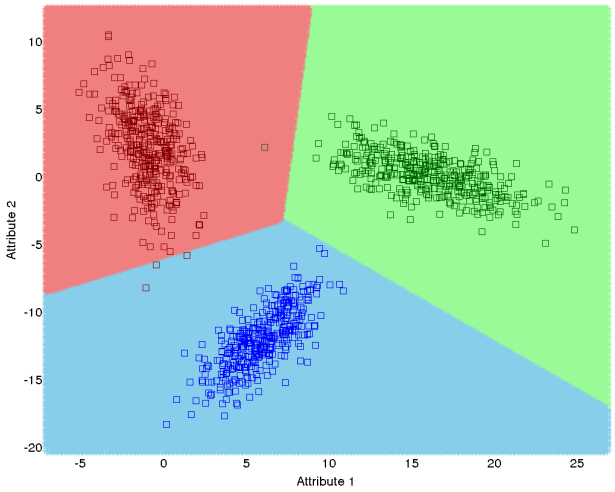
Pair/Cov	Same(Alltogther)	Same(Average)C	Different
1 and 2			
1 and 3			
2 and 3			

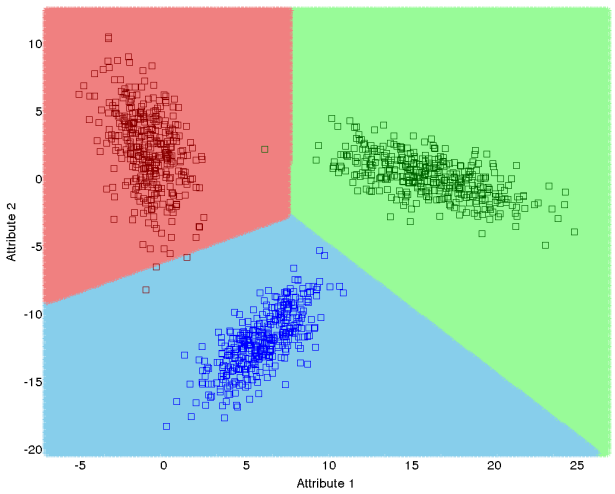
Figure 1: Decision region plot for every pair of classes



Correct : 374
Incorrect : 1
Accuracy : 99.733

		Predicted		
		Class 1	Class 2	Class 3
Act.	Class 1	125	0	0
	Class 2	0	125	0
	Class 3	0	1	124

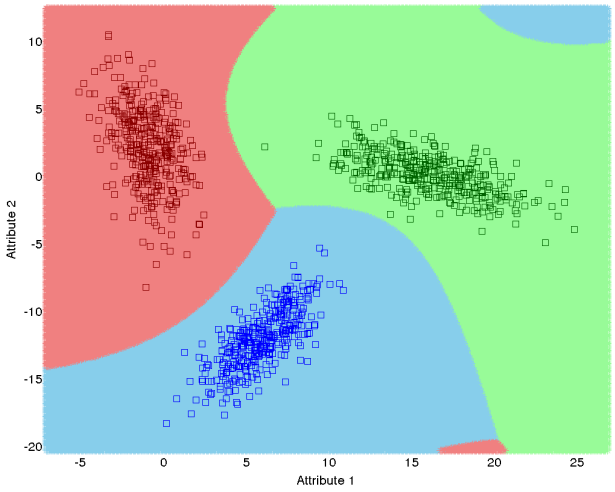
Figure 2: Decision region plot for all the classes together with the training data superposed with alltogether covariance



Correct : 374
Incorrect : 1
Accuracy : 99.733

		Predicted		
		Class 1	Class 2	Class 3
Act.	Class 1	125	0	0
	Class 2	0	125	0
	Class 3	0	1	124

Figure 3: Decision region plot for all the classes together with the training data superposed with average covariance



Correct : 375
Incorrect : 0
Accuracy : 100

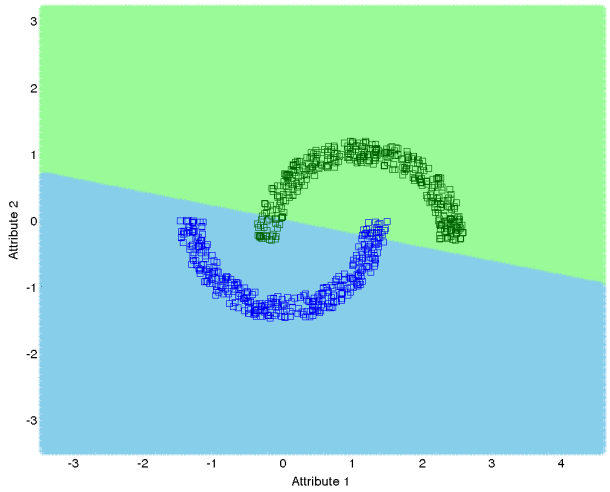
		Predicted		
		Class 1	Class 2	Class 3
Act.	Class 1	125	0	0
	Class 2	0	125	0
	Class 3	0	0	125

Figure 4: Decision region plot for all the classes together with the training data superposed with different covariance

Accuracy comes out to be 100 percentage when variance is taken differently for different classes. In this case the curvature of the decision boundary ensures that the points are accurately clasified.

3.1.2 Non-Linearly separable data set

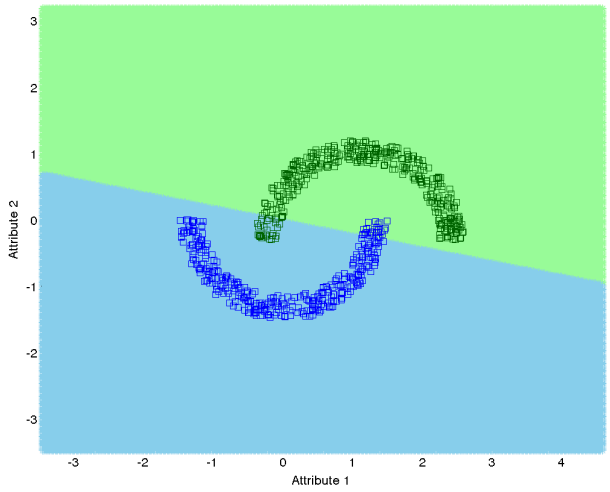
3.1.2.1 Data of Interlocking Classes As per the decision plots, it can be observed that the spiral data is not easily classifiable even in the case of different covariance(the curvature of the decision boundary does not seem to change much.)



Correct : 239
Incorrect : 11
Accuracy : 95.6000

		Predicted	
		Class 1	Class 2
Act.	Class 1	118	7
	Class 2	4	121

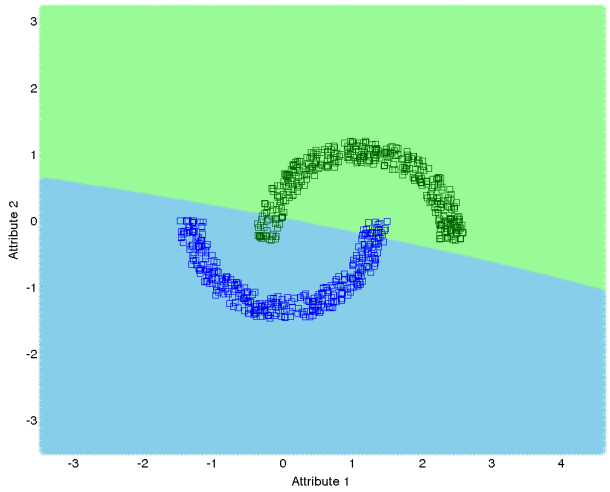
Figure 5: Decision region plot for all the classes together with the training data superposed with alltogether covariance



Correct : 239
Incorrect : 11
Accuracy : 95.6000

		Predicted	
		Class 1	Class 2
Act.	Class 1	118	7
	Class 2	4	121

Figure 6: Decision region plot for all the classes together with the training data superposed with average covariance

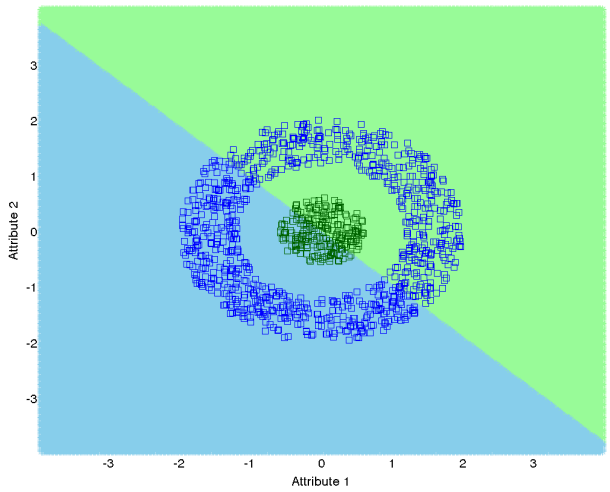


Correct : 240
Incorrect : 10
Accuracy : 96

		Predicted	
		Class 1	Class 2
Act.	Class 1	118	7
	Class 2	3	122

Figure 7: Decision region plot for all the classes together with the training data superposed with different covariance

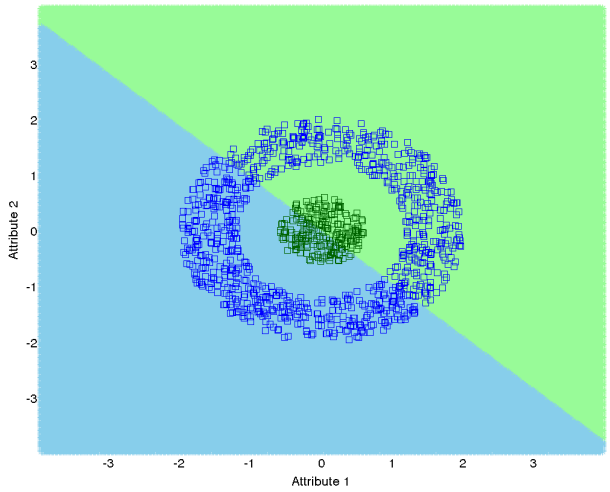
3.1.2.2 A ring with a central mass The decision boundary remains a straight line for the same covariance cases and are very inaccurate for the ring data.
But, the accuracy significantly improves when the covariance is introduced of each of the classes, leading to an accuracy of 100%.



Correct : 188
Incorrect : 187
Accuracy : 50.133

		Predicted	
		Class 1	Class 2
Act.	Class 1	46	29
	Class 2	158	142

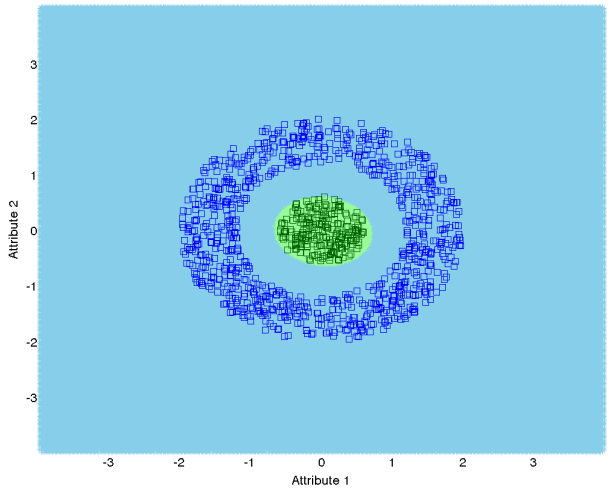
Figure 8: Decision region plot for all the classes together with the training data superposed with alltogether covariance



Correct : 188
Incorrect : 187
Accuracy : 50.133

		Predicted	
		Class 1	Class 2
Act.	Class 1	46	29
	Class 2	158	142

Figure 9: Decision region plot for all the classes together with the training data superposed with average covariance

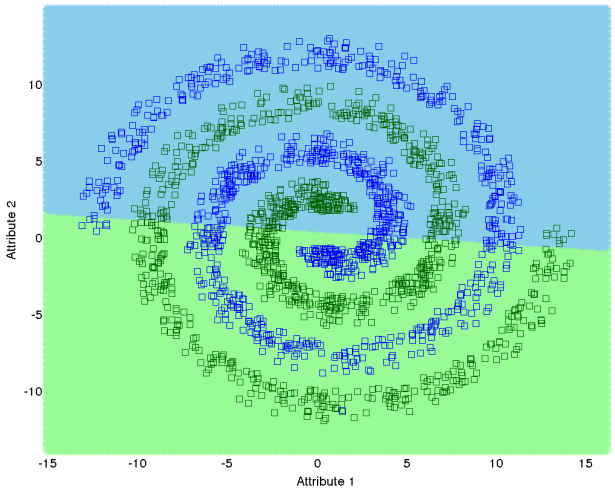


Correct : 375
Incorrect : 0
Accuracy : 100

		Predicted	
		Class 1	Class 2
Act.	Class 1	75	0
	Class 2	0	300

Figure 10: Decision region plot for all the classes together with the training data superposed with different covariance

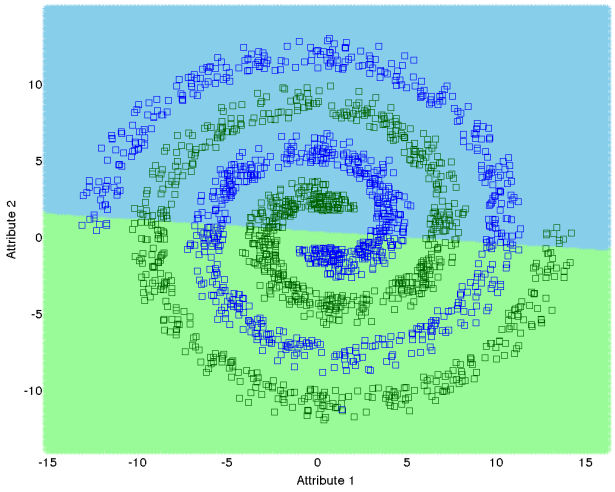
3.1.2.3 Spiral Dataset The bayesian classifier is in all cases unable to classify the spiral data accurately resulting in accuracies less than 55 %.



Correct : 350
Incorrect : 302
Accuracy : 53.68

		Predicted	
		Class 1	Class 2
Act.	Class 1	175	151
	Class 2	151	175

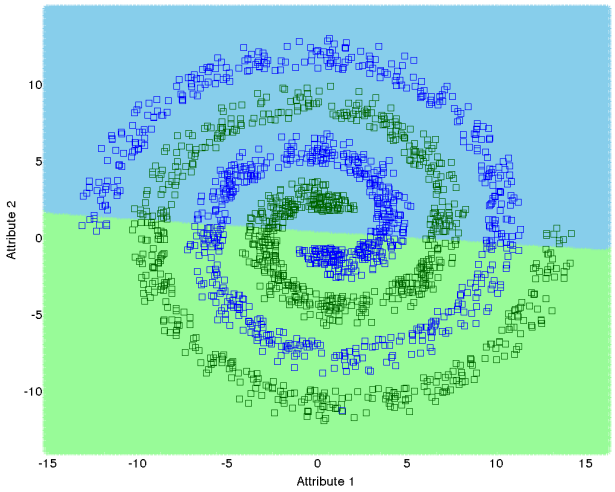
Figure 11: Decision region plot for all the classes together with the training data superposed with alltogether covariance



Correct : 350
Incorrect : 302
Accuracy : 53.68

		Predicted	
		Class 1	Class 2
Act.	Class 1	175	151
	Class 2	151	175

Figure 12: Decision region plot for all the classes together with the training data superposed with average covariance



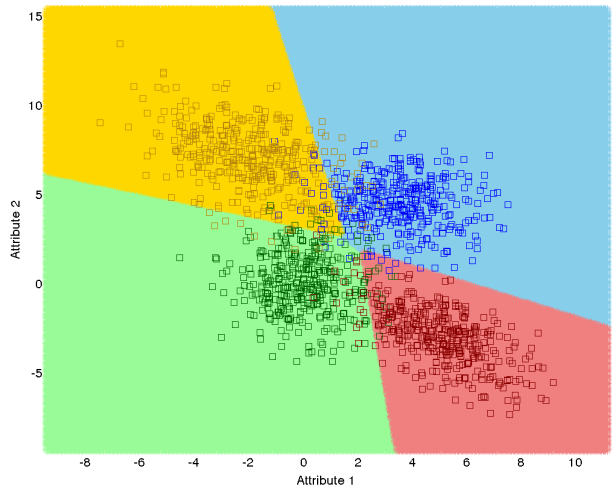
Correct : 350
Incorrect : 302
Accuracy : 53.68

		Predicted	
		Class 1	Class 2
Act.	Class 1	175	151
	Class 2	151	175

Figure 13: Decision region plot for all the classes together with the training data superposed with different covariance

3.1.3 Overlapping data set

Overlapping data is classified to a satisfactory level by the bayesian classifier, clearly, the more overlapped the data is more are the chances of incorrect labelling leading to less than best accuracies.



Correct : 450
Incorrect : 50
Accuracy : 90.000

		Predicted			
		Class 1	Class 2	Class 3	Class 4
Act.	Class 1	111	4	4	6
	Class 2	1	116	0	8
	Class 3	9	0	116	0
	Class 4	6	12	0	107

Figure 15: Decision region plot for all the classes together with the training data superposed with alltogether covariance

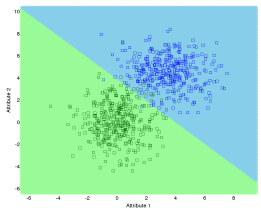
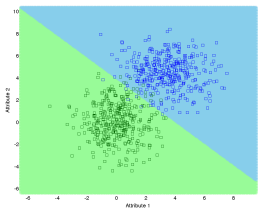
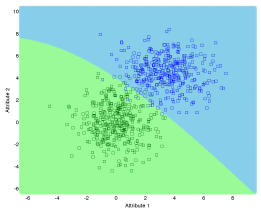
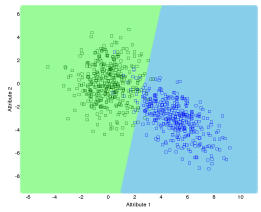
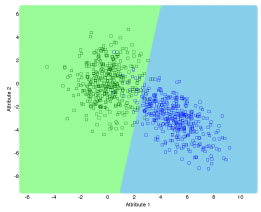
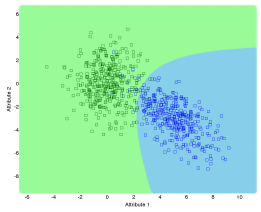
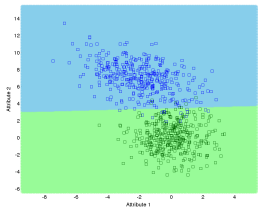
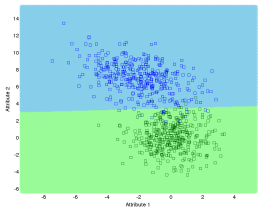
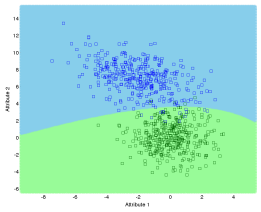
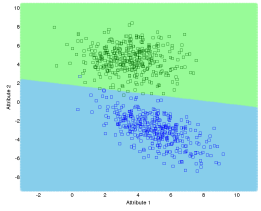
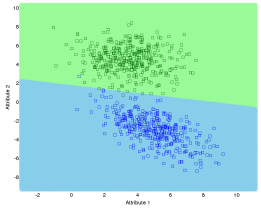
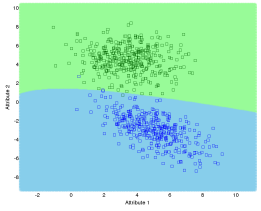
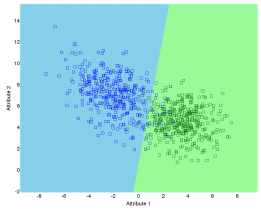
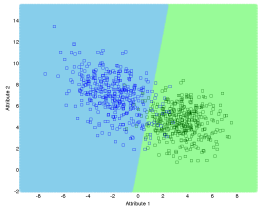
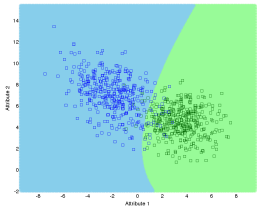
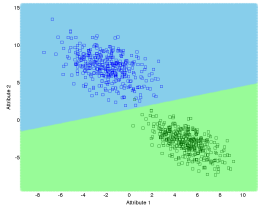
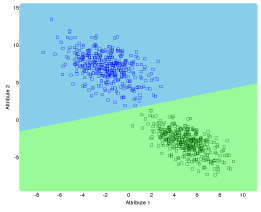
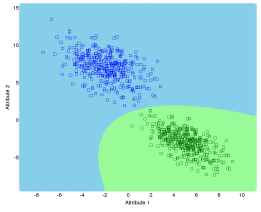
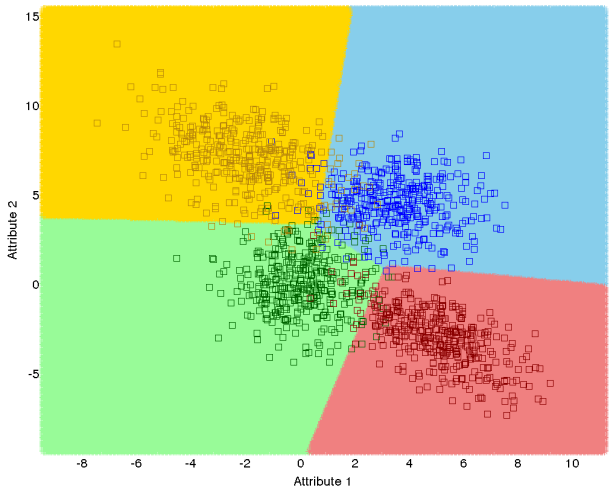
Pair/Cov	Same(Alltogther)	Same(Average)C	Different
1 and 2			
1 and 3			
1 and 4			
2 and 3			
2 and 4			
3 and 4			

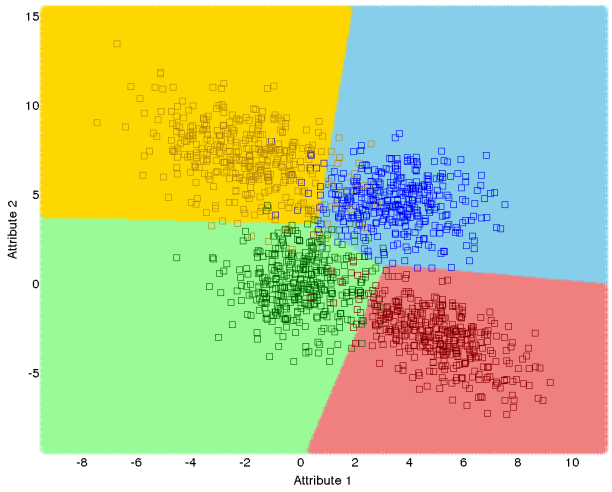
Figure 14: Decision region plot for every pair of classes



Correct : 453
Incorrect : 47
Accuracy : 90.600

		Predicted			
		Class 1	Class 2	Class 3	Class 4
Act.	Class 1	111	6	4	4
	Class 2	2	118	0	5
	Class 3	9	0	116	0
	Class 4	5	12	0	108

Figure 16: Decision region plot for all the classes together with the training data superposed with average covariance



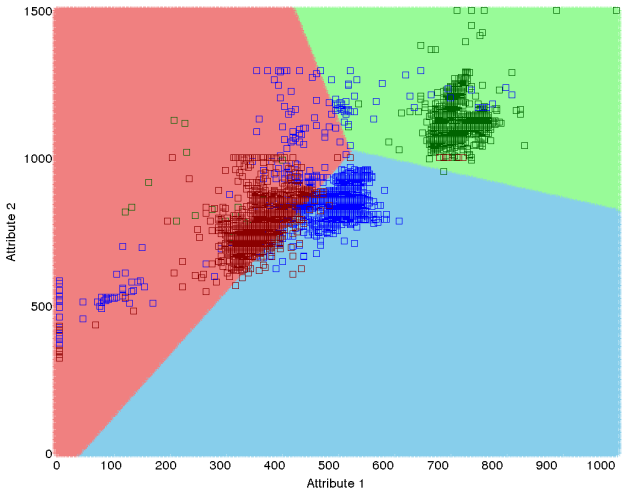
Correct : 452
Incorrect : 48
Accuracy : 90.400

		Predicted			
		Class 1	Class 2	Class 3	Class 4
Act.	Class 1	113	4	4	4
	Class 2	2	118	0	5
	Class 3	12	0	113	0
	Class 4	5	12	0	108

Figure 17: Decision region plot for all the classes together with the training data superposed with different covariance

3.1.4 Real world data set

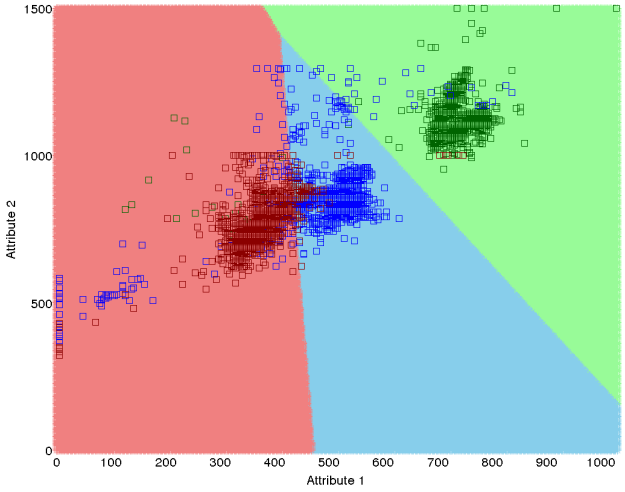
In case of real world data, bayesian classifier is put to a test where the accuracies drop to 80s, not significantly different for the three cases. When covariance matrix is taken as same, the decision boundaries are straight lines, but when different covariances are considered, elliptical regions are observed.



Correct : 1386
Incorrect : 391
Accuracy : 77.9966

		Predicted		
		Class 1	Class 2	Class 3
Act.	Class 1	510	1	30
	Class 2	21	332	261
	Class 3	7	71	544

Figure 18: Decision region plot for all the classes together with the training data superposed with alltogether covariance



Correct : 1509
Incorrect : 268
Accuracy : 84.9184

		Predicted		
		Class 1	Class 2	Class 3
Act.	Class 1	510	5	26
	Class 2	20	397	197
	Class 3	7	13	602

Figure 19: Decision region plot for all the classes together with the training data superposed with average covariance

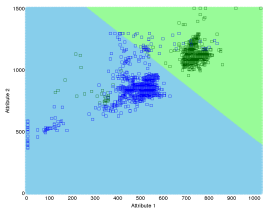
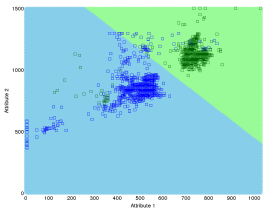
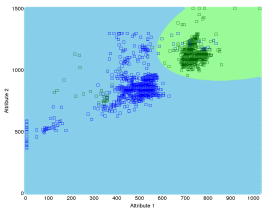
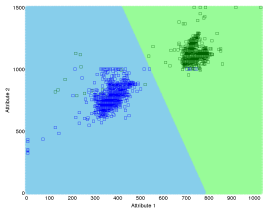
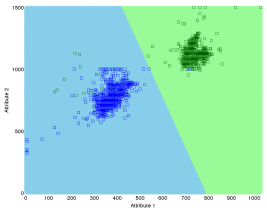
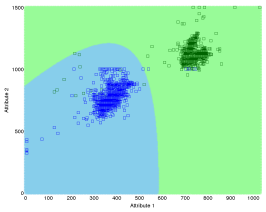
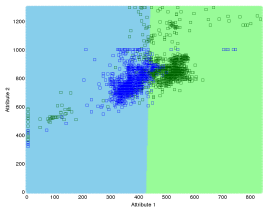
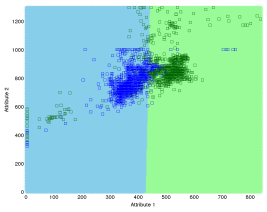
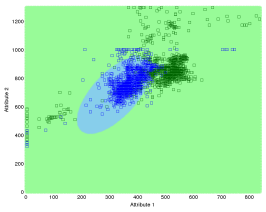
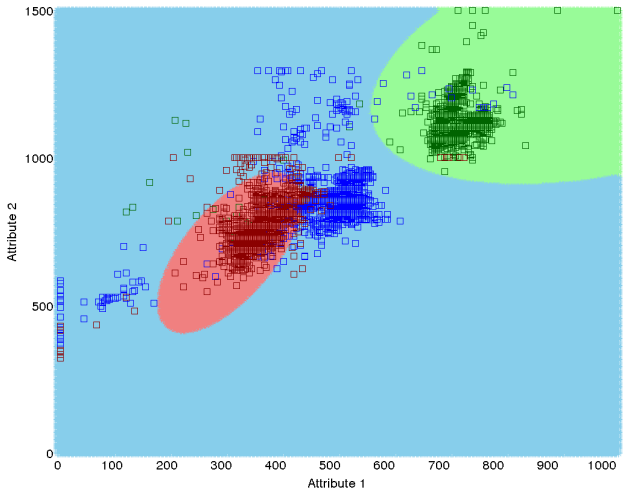
Pair/Cov	Same(Alltogther)	Same(Average)C	Different
1 and 2			
1 and 3			
2 and 3			

Figure 21: Decision region plot for every pair of classes



Correct : 1433
Incorrect : 344
Accuracy : 80.6415

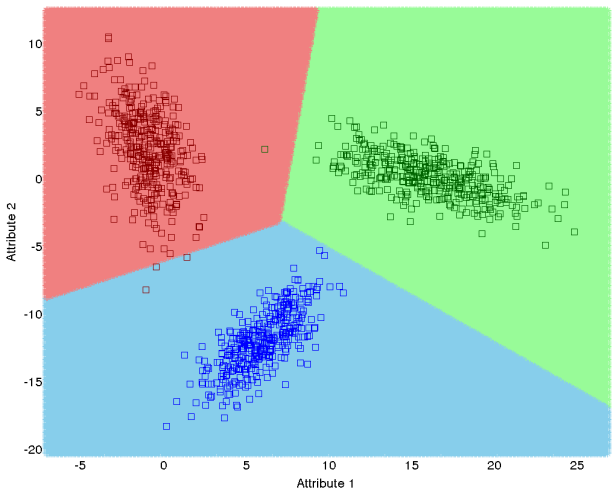
		Predicted		
		Class 1	Class 2	Class 3
Act.	Class 1	509	19	13
	Class 2	5	392	217
	Class 3	7	83	532

Figure 20: Decision region plot for all the classes together with the training data superposed with different covariance

3.2 Naive-Bayes classifier

3.2.1 Linearly separable data set

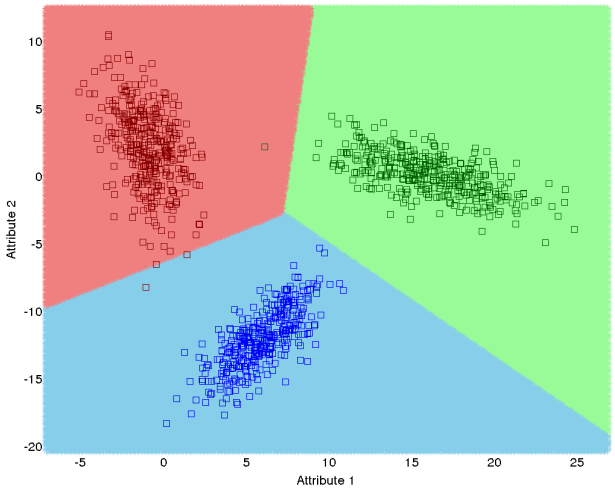
The decision boundaries are very similar to the bayes classifier, where most of test data fit in the estimated class regions for the linearly separable data set.



Correct : 374
Incorrect : 1
Accuracy : 99.733

		Predicted		
		Class 1	Class 2	Class 3
Act.	Class 1	125	0	0
	Class 2	0	125	0
	Class 3	0	1	124

Figure 22: Decision region plot for all the classes together with the training data superposed with Same(σ^2I) covariance



Correct : 374
Incorrect : 1
Accuracy : 99.733

		Predicted		
		Class 1	Class 2	Class 3
Act.	Class 1	125	0	0
	Class 2	0	125	0
	Class 3	0	1	124

Figure 23: Decision region plot for all the classes together with the training data superposed with average covariance

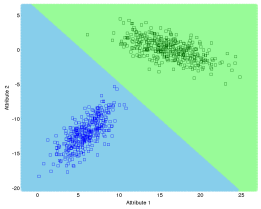
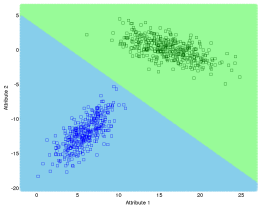
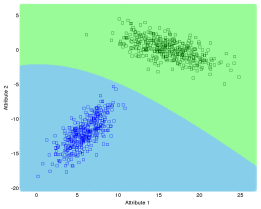
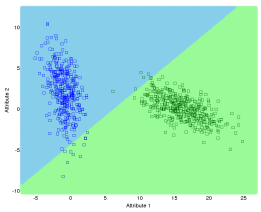
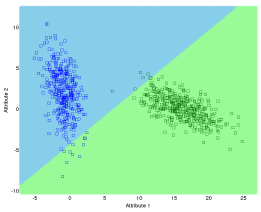
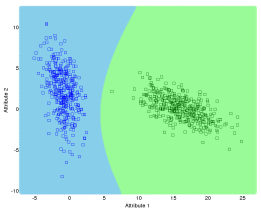
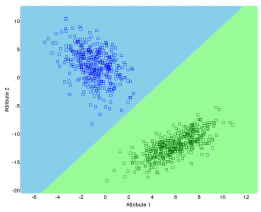
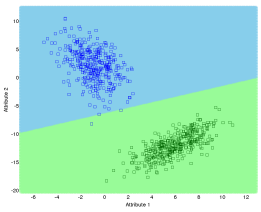
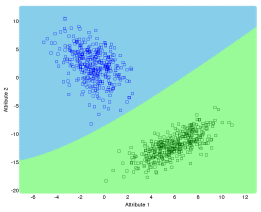
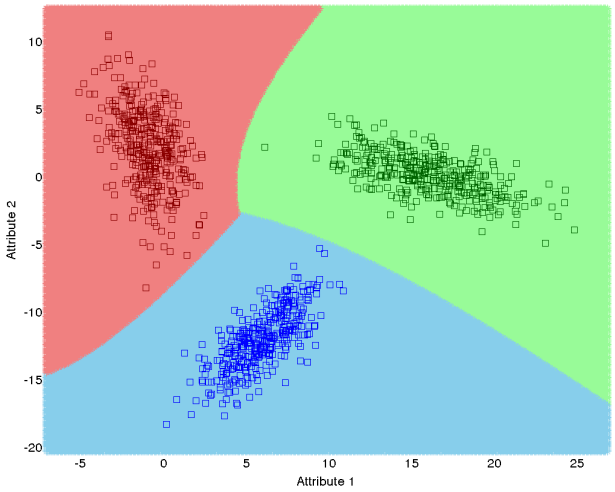
Pair/Cov	Same(σ^2)I	Same(Average)C	Different
1 and 2			
1 and 3			
2 and 3			

Figure 25: Decision region plot for every pair of classes



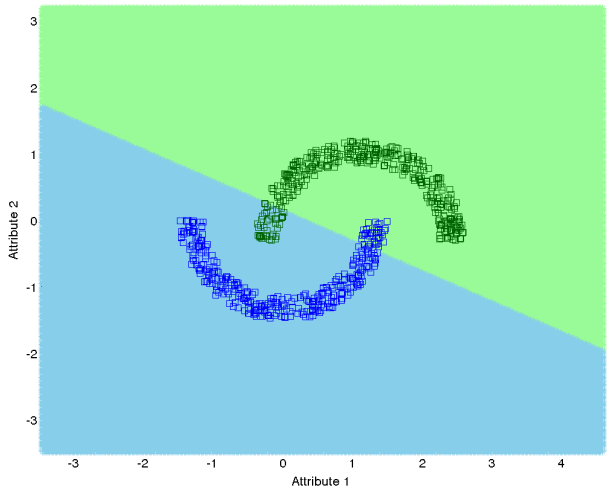
Correct : 374
Incorrect : 1
Accuracy : 99.7333

		Predicted		
		Class 1	Class 2	Class 3
Act.	Class 1	125	0	0
	Class 2	0	125	0
	Class 3	0	1	124

Figure 24: Decision region plot for all the classes together with the training data superposed with different covariance

3.2.2 Non-Linearly separable data set

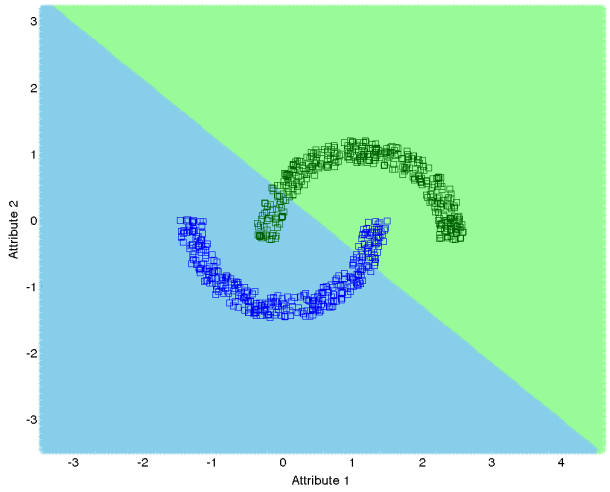
3.2.2.1 Data of Interlocking Classes The decision region plots show that the naive bayes performes worse than bayes, leading to drop in accuracies, although there is not much difference in their shape.



Correct : 231
Incorrect : 19
Accuracy : 92.400

		Predicted	
		Class 1	Class 2
Act.	Class 1	112	13
	Class 2	6	119

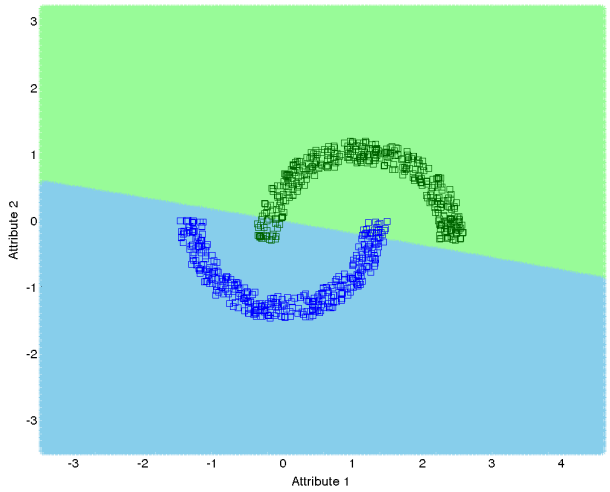
Figure 26: Decision region plot for all the classes together with the training data superposed with Same(σ^2I) covariance



Correct : 219
Incorrect : 31
Accuracy : 87.6000

		Predicted	
		Class 1	Class 2
Act.	Class 1	106	19
	Class 2	12	113

Figure 27: Decision region plot for all the classes together with the training data superposed with average covariance

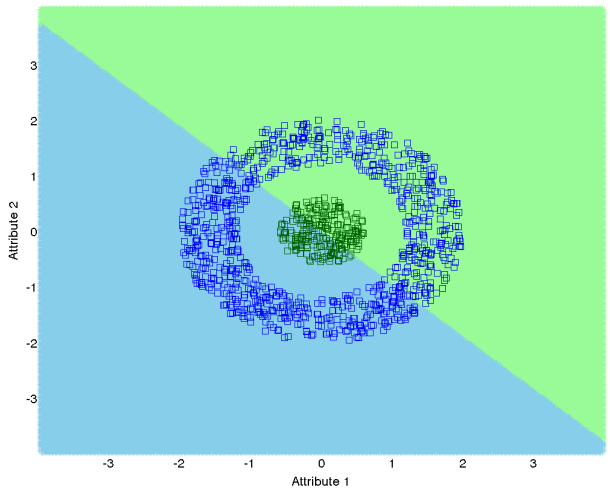


Correct : 240
Incorrect : 10
Accuracy : 96.000

		Predicted	
		Class 1	Class 2
Act.	Class 1	118	7
	Class 2	3	122

Figure 28: Decision region plot for all the classes together with the training data superposed with different covariance

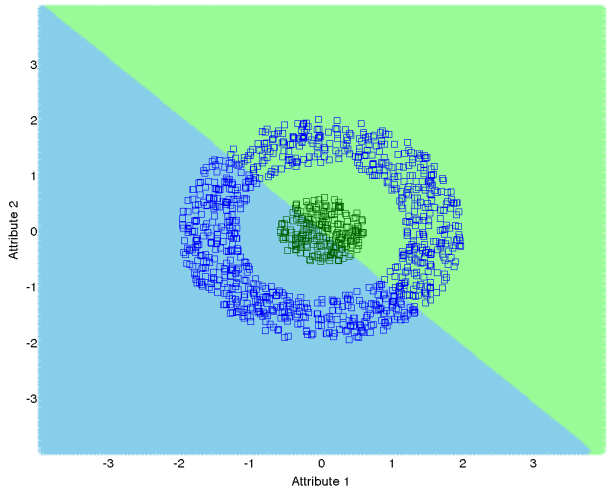
3.2.2.2 A ring with a central mass Again the plots shows similar behaviour as the bayes case, but a slight drop in accuracy is observed. This can be reasoned with the help of fact that naive bayes assumed independence between the attributes.



Correct : 188
Incorrect : 187
Accuracy : 50.133

		Predicted	
		Class 1	Class 2
Act.	Class 1	46	29
	Class 2	158	142

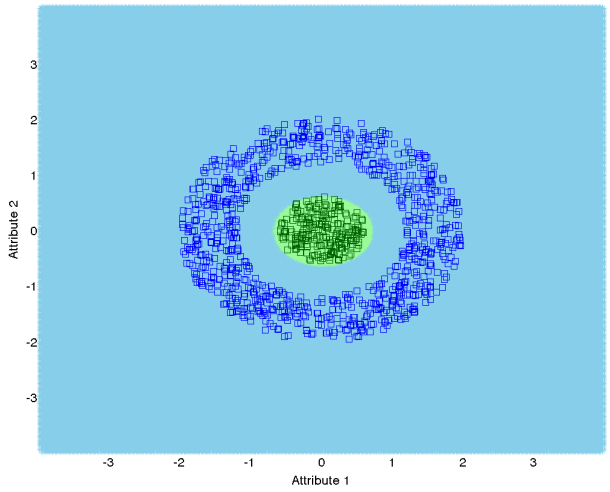
Figure 29: Decision region plot for all the classes together with the training data superposed with alltogether covariance



Correct : 192
Incorrect : 183
Accuracy : 51.2000

		Predicted	
		Class 1	Class 2
Act.	Class 1	47	28
	Class 2	155	145

Figure 30: Decision region plot for all the classes together with the training data superposed with average covariance

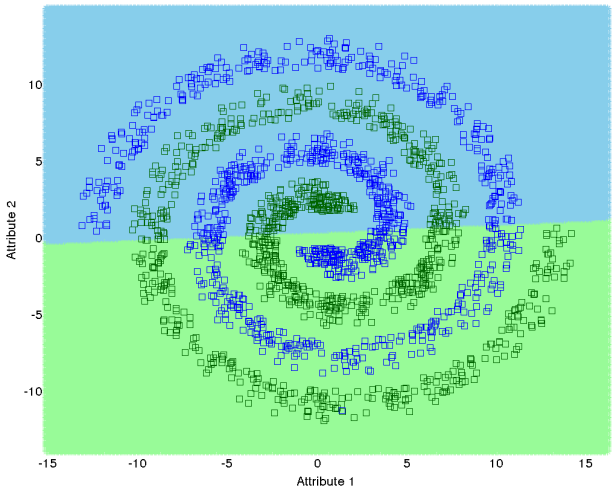


Correct : 375
Incorrect : 0
Accuracy : 100

		Predicted	
		Class 1	Class 2
Act.	Class 1	75	0
	Class 2	0	300

Figure 31: Decision region plot for all the classes together with the training data superposed with different covariance

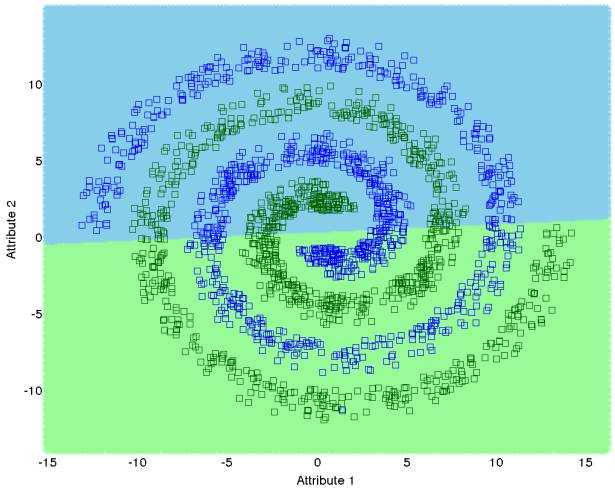
3.2.2.3 Spiral Dataset Similar to the bayes classifier with inaccuracies which are worst as compare to other data sets.



Correct : 354
Incorrect : 298
Accuracy : 54.2945

		Predicted	
		Class 1	Class 2
Act.	Class 1	117	149
	Class 2	149	177

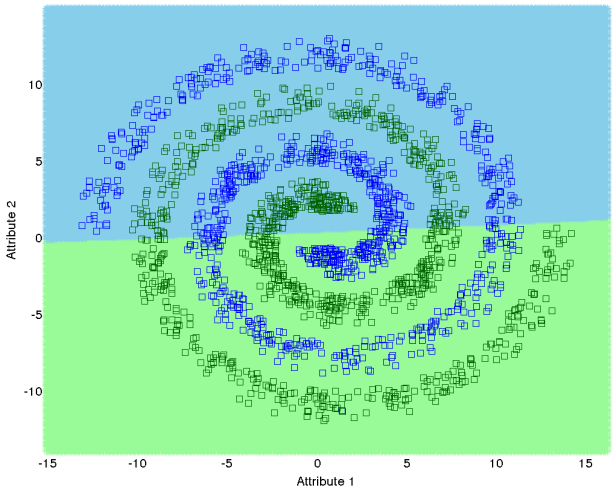
Figure 32: Decision region plot for all the classes together with the training data superposed with Same(σ^2I) covariance



Correct : 353
Incorrect : 299
Accuracy : 54.1411

		Predicted	
		Class 1	Class 2
Act.	Class 1	176	150
	Class 2	149	177

Figure 33: Decision region plot for all the classes together with the training data superposed with average covariance



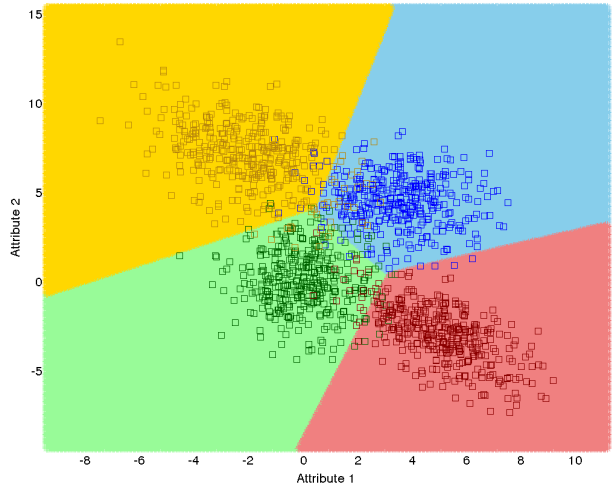
Correct : 354
Incorrect : 298
Accuracy : 54.2945

		Predicted	
		Class 1	Class 2
Act.	Class 1	177	149
	Class 2	149	177

Figure 34: Decision region plot for all the classes together with the training data superposed with different covariance

3.2.3 Overlapping data set

There is a drop in accuracies as compared to the bayes plots. The boundaries are similar to those generated by the bayesian classifier.



Correct : 449
Incorrect : 51
Accuracy : 89.800

		Predicted			
		Class 1	Class 2	Class 3	Class 4
Act.	Class 1	112	5	4	4
	Class 2	2	117	1	5
	Class 3	9	2	114	0
	Class 4	6	13	0	106

Figure 36: Decision region plot for all the classes together with the training data superposed with Same(σ^2I) covariance

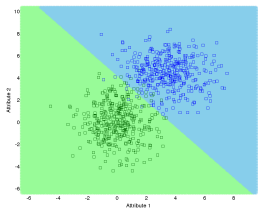
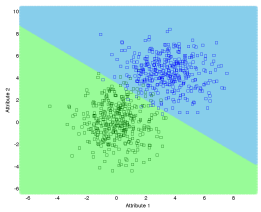
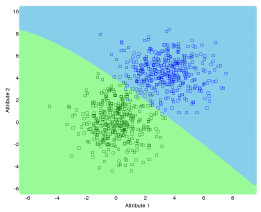
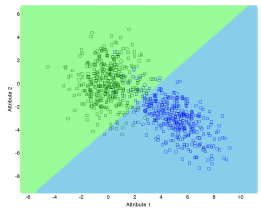
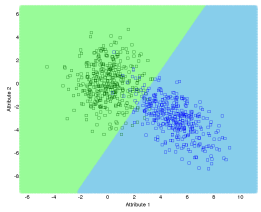
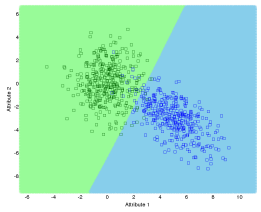
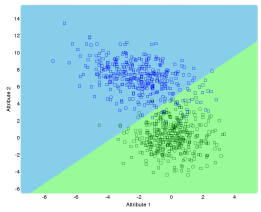
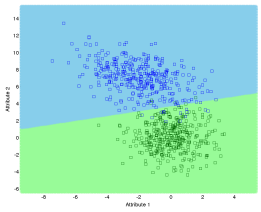
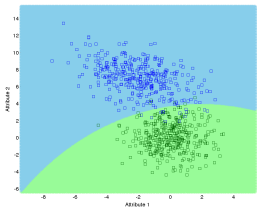
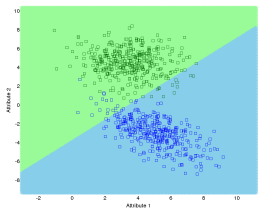
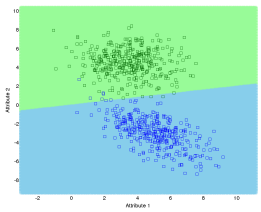
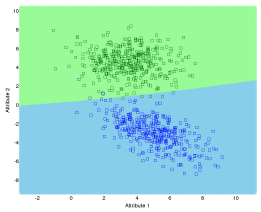
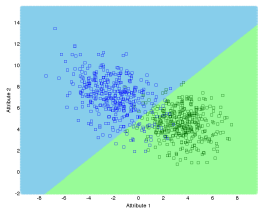
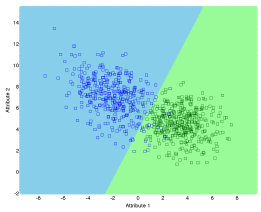
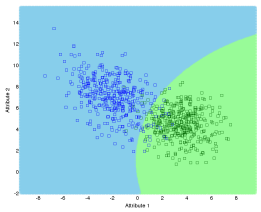
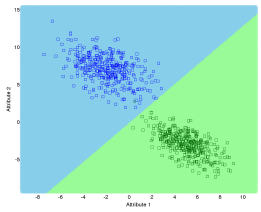
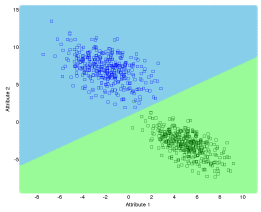
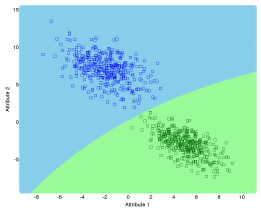
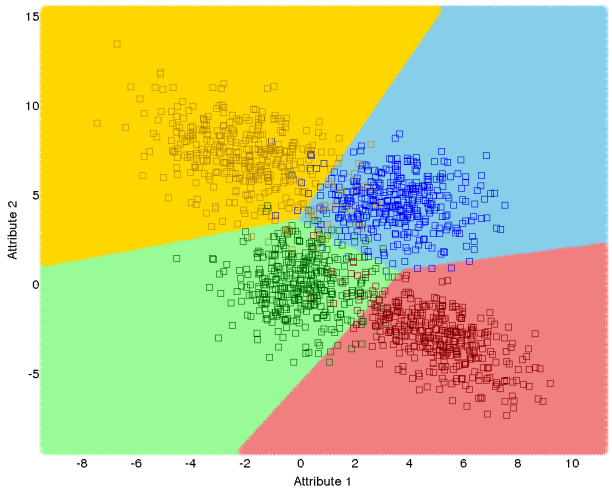
Pair/Cov	Same(σ^2)I	Same(Average)C	Different
1 and 2			
1 and 3			
1 and 4			
2 and 3			
2 and 4			
3 and 4			

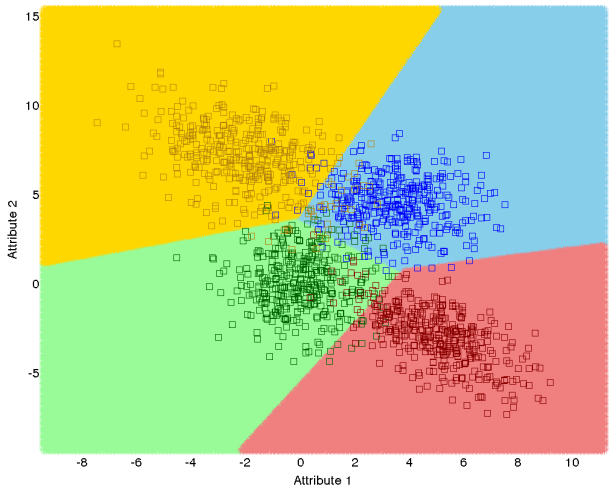
Figure 35: Decision region plot for every pair of classes



Correct : 445
Incorrect : 55
Accuracy : 89.000

		Predicted			
		Class 1	Class 2	Class 3	Class 4
Act.	Class 1	111	6	4	4
	Class 2	2	117	1	5
	Class 3	13	0	112	0
	Class 4	5	15	0	105

Figure 37: Decision region plot for all the classes together with the training data superposed with average covariance



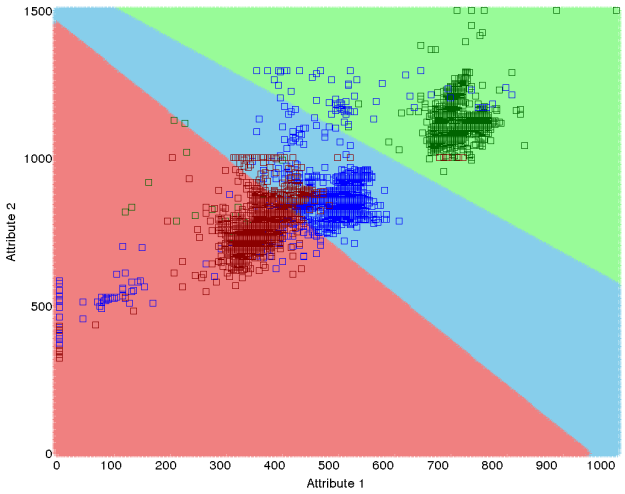
Correct : 445
Incorrect : 55
Accuracy : 89.000

		Predicted			
		Class 1	Class 2	Class 3	Class 4
Act.	Class 1	110	6	5	4
	Class 2	2	115	1	7
	Class 3	11	0	114	0
	Class 4	5	14	0	106

Figure 38: Decision region plot for all the classes together with the training data superposed with different covariance

3.2.4 Real world data set

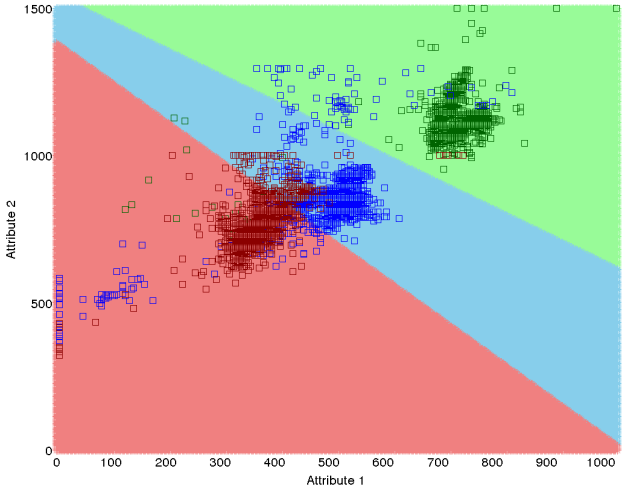
The results are comparable to the results obtained in the bayesian classifier with minute ups and downs in different cases of covariance matrices.



Correct : 1487
Incorrect : 290
Accuracy : 83.6804

		Predicted		
		Class 1	Class 2	Class 3
Act.	Class 1	512	5	24
	Class 2	43	400	171
	Class 3	7	40	575

Figure 39: Decision region plot for all the classes together with the training data superposed with Same(σ^2)I covariance



Correct : 1482
Incorrect : 295
Accuracy : 83.3990

		Predicted		
		Class 1	Class 2	Class 3
Act.	Class 1	512	5	24
	Class 2	44	398	172
	Class 3	7	43	572

Figure 40: Decision region plot for all the classes together with the training data superposed with average covariance

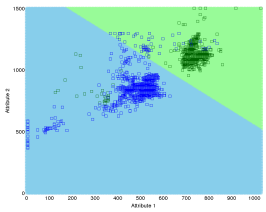
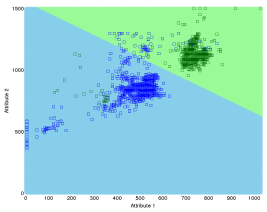
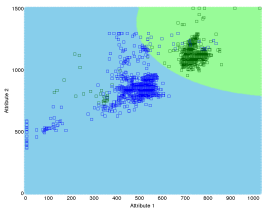
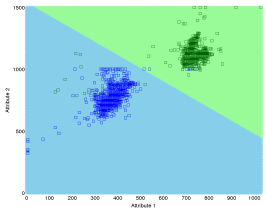
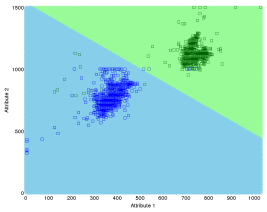
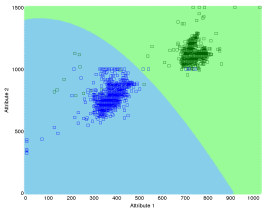
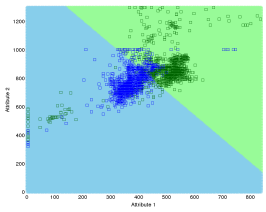
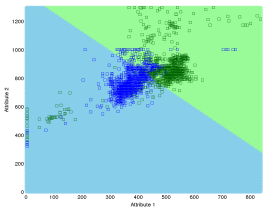
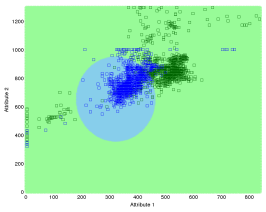
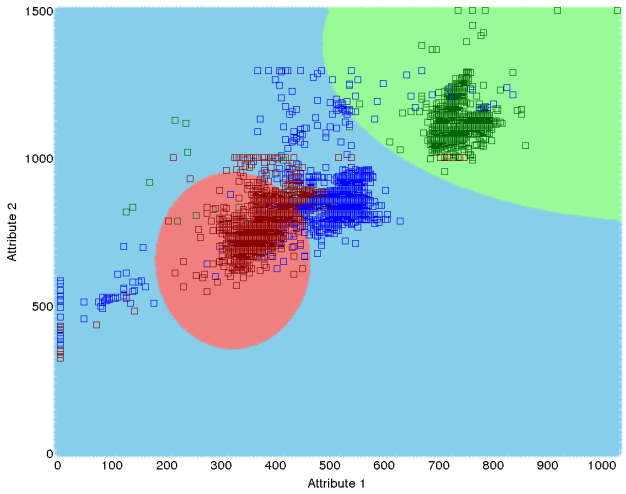
Pair/Cov	Same(σ^2)I	Same(Average)C	Different
1 and 2			
1 and 3			
2 and 3			

Figure 42: Decision region plot for every pair of classes



Correct : 1466
Incorrect : 311
Accuracy : 82.4986

		Predicted		
		Class 1	Class 2	Class 3
Act.	Class 1	510	16	15
	Class 2	19	410	185
	Class 3	7	69	546

Figure 41: Decision region plot for all the classes together with the training data superposed with different covariance

4 Conclusion

As per the observations, we can make the following conclusions :

1. The Decision Boundaries are more accurate in the case of different covariance for different classes as compared to the other cases.
2. The curvature of the decision boundaries is due to the covariance term in the likelihood probabilty which makes the surface quadratic.

3. The Decision Boundaries are better in cases where data is not overlapping and is separable either linearly or non linearly.

4. In case of real data, the data is more overlapping and non linear, resulting in lesser accuracy of the testing data.