CS 669 Assignment 1

Rohit Patiyal
Devang Bacharwar

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 when the covariance is either taken as the average of individual class covariances and when the covariance is calculated using all classes' data together.

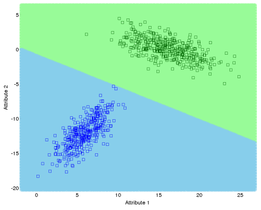
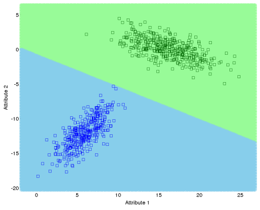
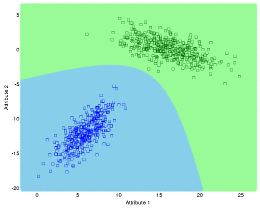
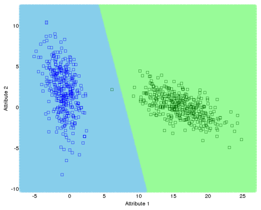
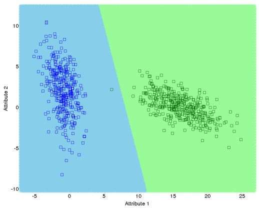
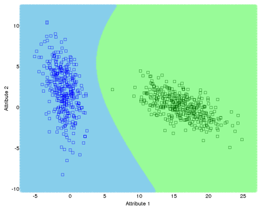
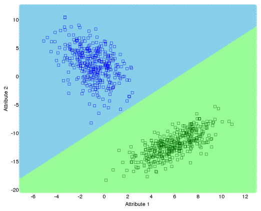
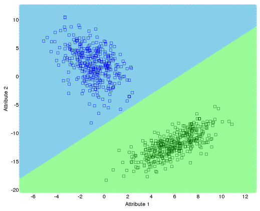
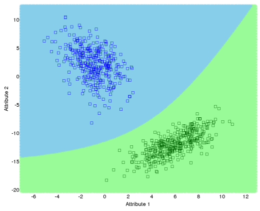
Pair/Cov	Alltogther	Average	Different
1 and 2			
1 and 3			
2 and 3			

Figure 1: Decision region plot for every pair of classes

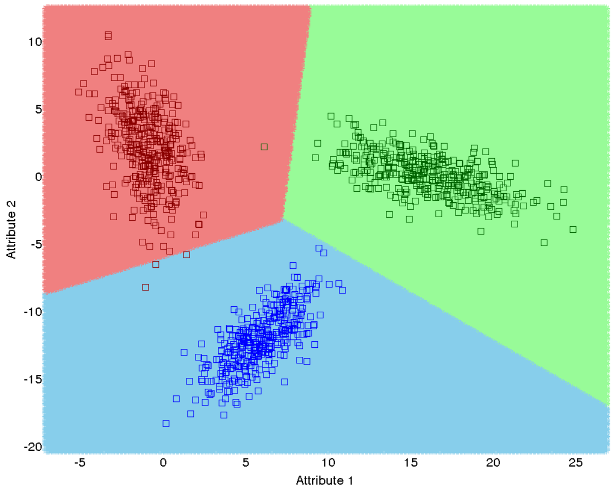
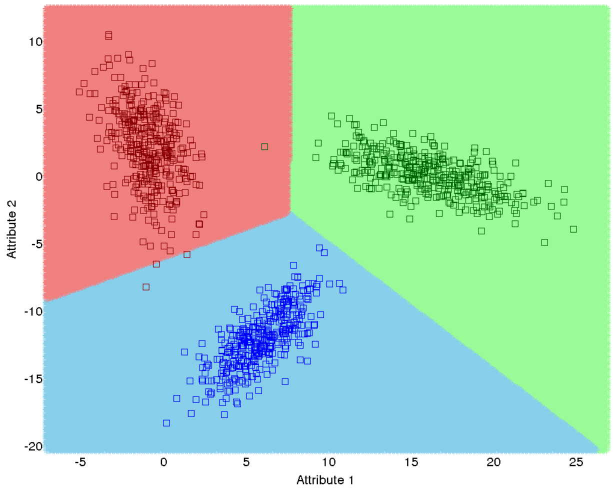


Figure 2: Decision region plot for all the classes together with the training data superposed with alltogether covariance

Correct : 374
Incorrect : 1
Acurracy : 99.733

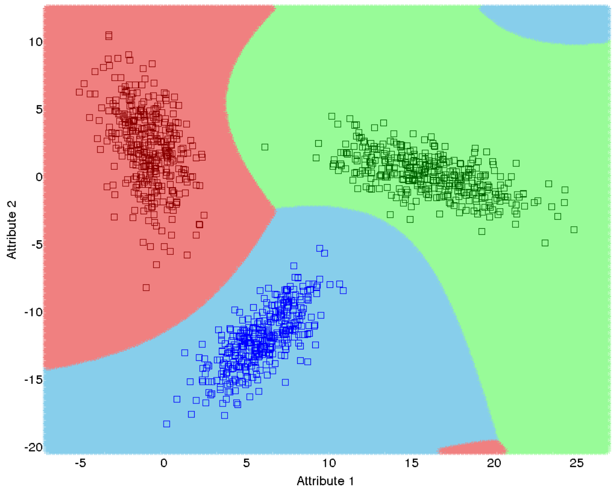
		Predicted		
		Class 1	Class 2	Class 3
Act.	Class 1	125	0	0
	Class 2	0	125	0
	Class 3	0	1	124



Correct : 374
Incorrect : 1
Acurracy : 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



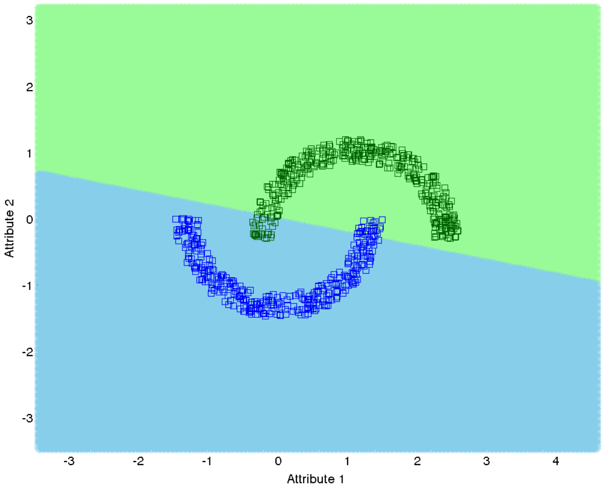
Correct : 375
Incorrect : 0
Acurracy : 100

		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

3.1.2 Non-Linearly separable data set

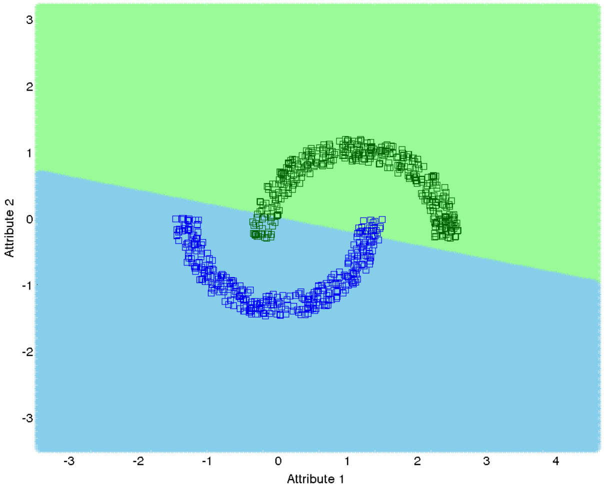
3.1.2.1 Data of Interlocking Classes



Correct : 188
Incorrect : 187
Acurracy : 50.133

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

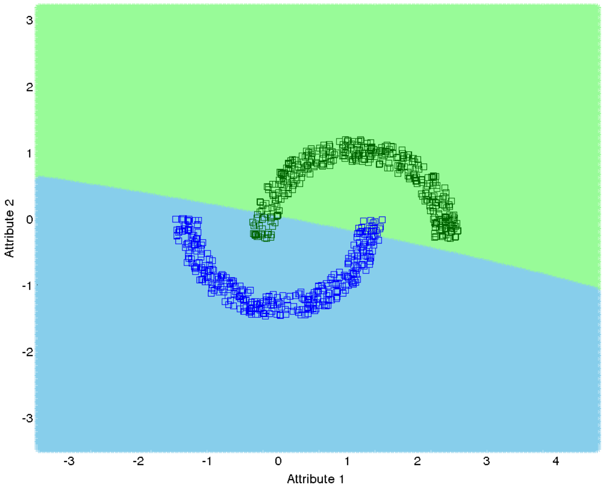
Figure 5: Decision region plot for all the classes together with the training data superposed with alltogether covariance



Correct : 188
Incorrect : 187
Acurracy : 50.133

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

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

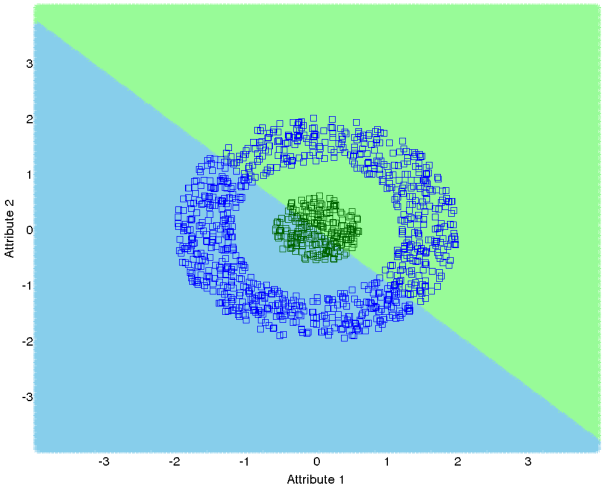


Correct : 375
Incorrect : 0
Acurracy : 100

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

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

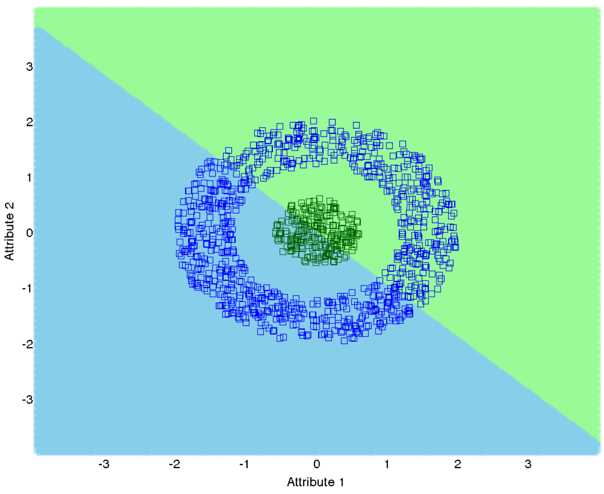
3.1.2.2 A ring with a central mass



Correct : 188
Incorrect : 187
Acurracy : 50.133

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

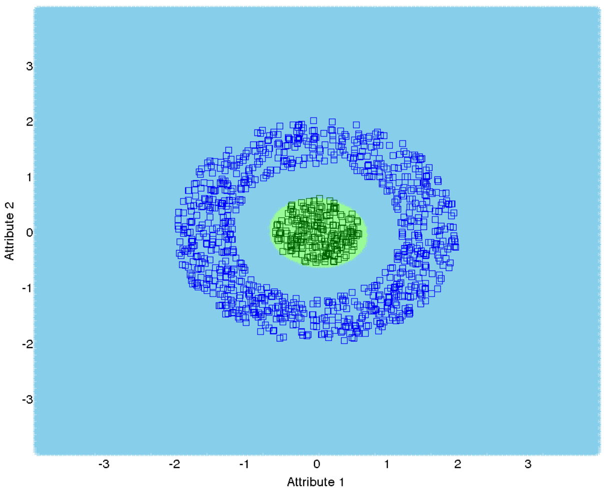
Figure 8: Decision region plot for all the classes together with the training data superposed with alltogether covariance



Correct : 188
Incorrect : 187
Acurracy : 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
Acurracy : 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

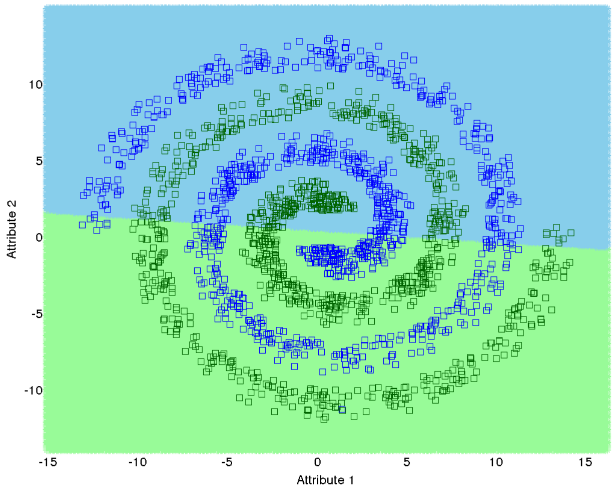


Figure 11: Decision region plot for all the classes together with the training data superposed with alltogether covariance

Correct : 188
Incorrect : 187
Acurracy : 50.133

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

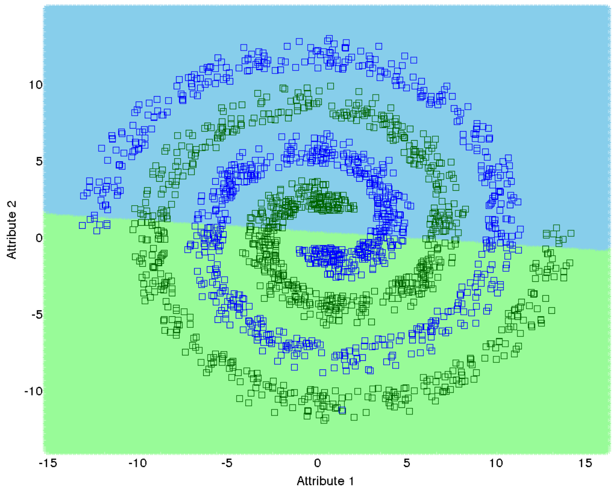
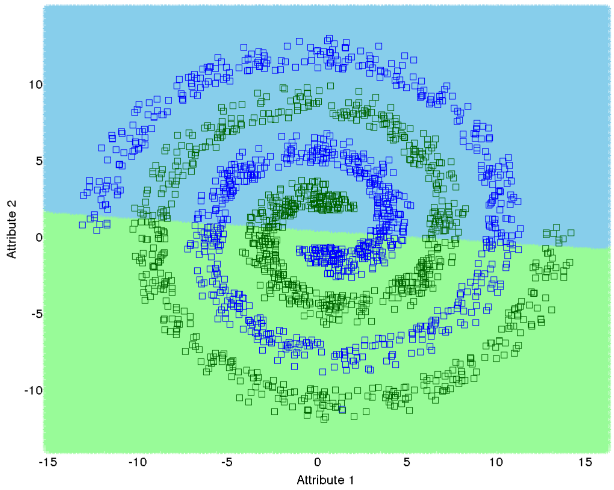


Figure 12: Decision region plot for all the classes together with the training data superposed with average covariance

Correct : 188
Incorrect : 187
Acurracy : 50.133

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

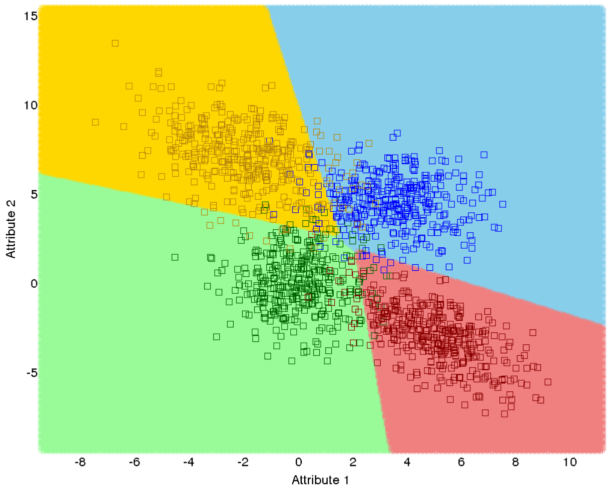


Correct : 375
Incorrect : 0
Acurracy : 100

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

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

3.1.3 Overlapping data set



Correct : 450
Incorrect : 50
Acurracy : 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

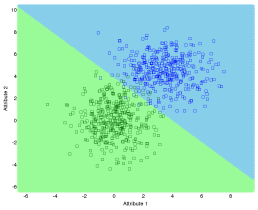
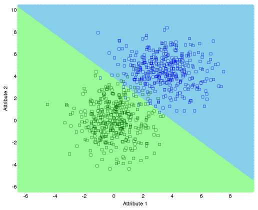
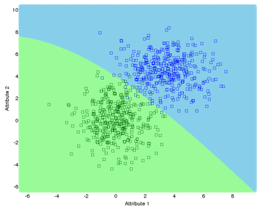
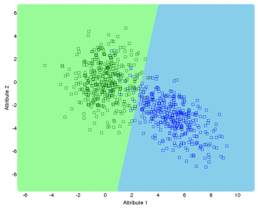
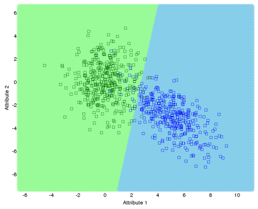
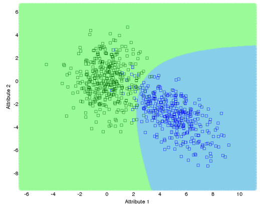
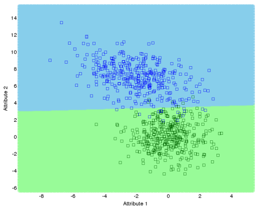
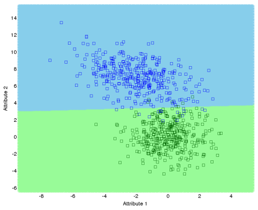
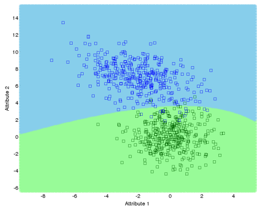
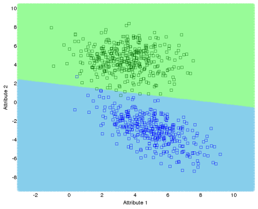
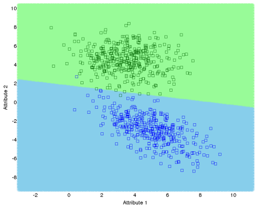
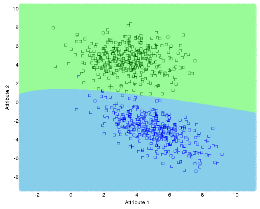
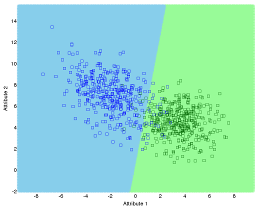
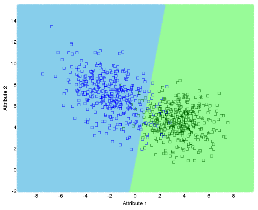
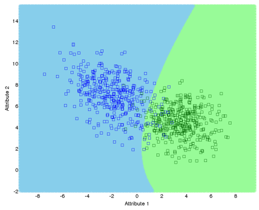
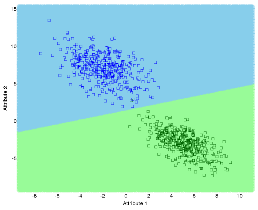
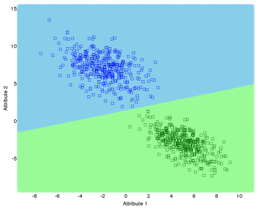
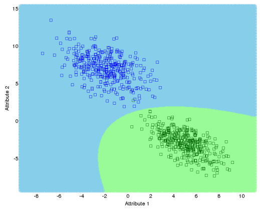
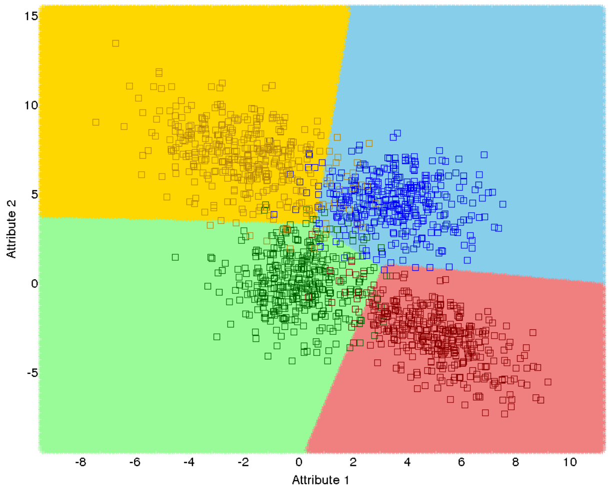
Pair/Cov	Alltogther	Average	Different
1 and 2			
1 and 3			
1 and 4			
2 and 3			
2 and 4			
3 and 4			

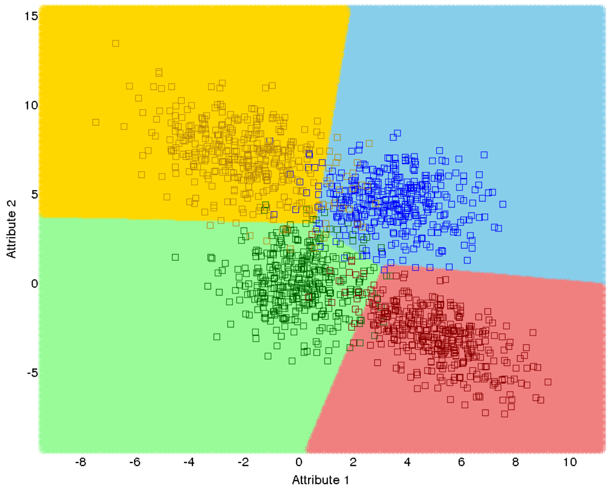
Figure 14: Decision region plot for every pair of classes



Correct : 453
Incorrect : 47
Acurracy : 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

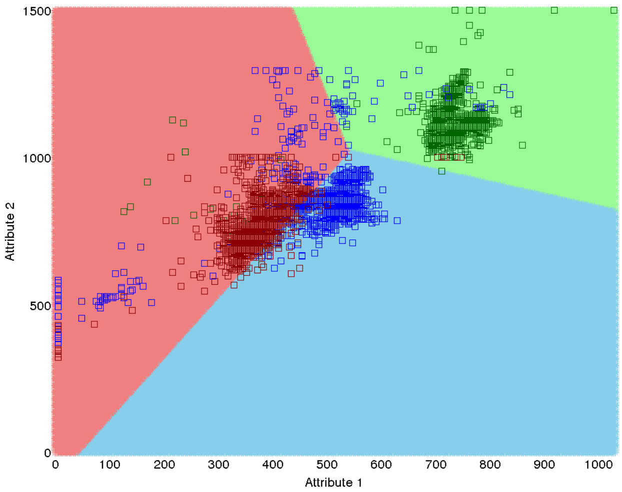


Correct : 452
Incorrect : 48
Acurracy : 90.400

		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

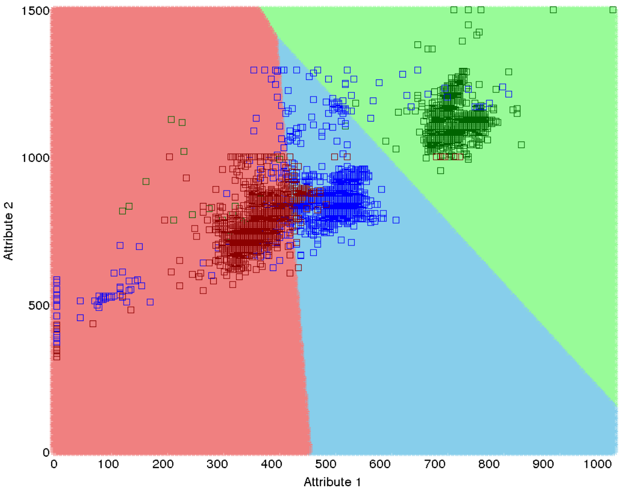
3.1.4 Real world data set



Correct : 374
Incorrect : 1
Acurracy : 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 18: Decision region plot for all the classes together with the training data superposed with alltogether covariance



Correct : 374
Incorrect : 1
Acurracy : 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 19: Decision region plot for all the classes together with the training data superposed with average covariance

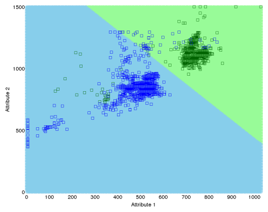
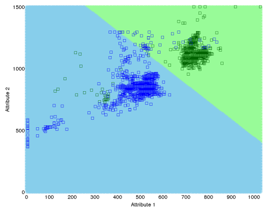
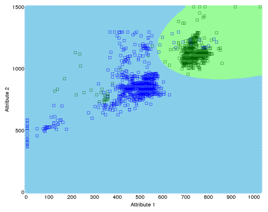
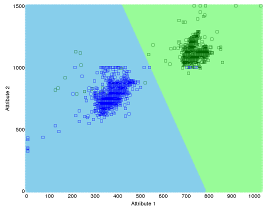
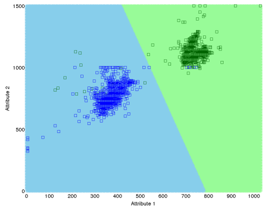
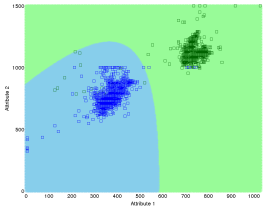
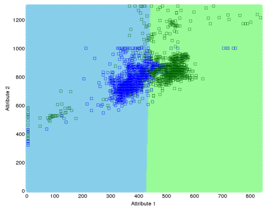
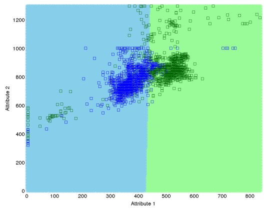
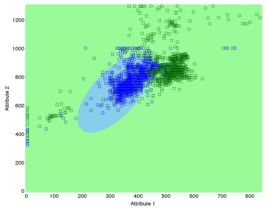
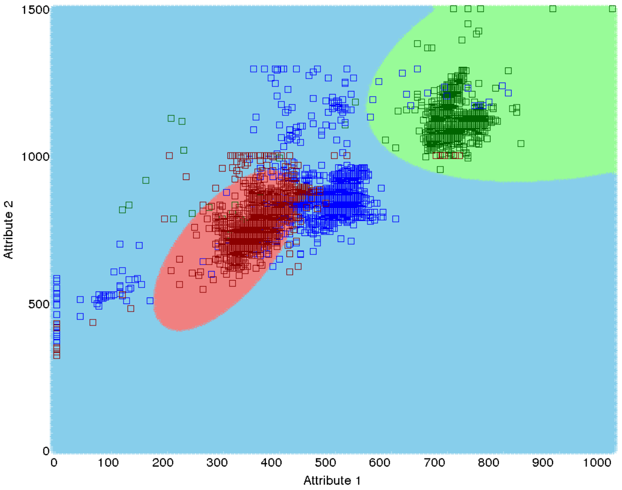
Pair/Cov	Alltogther	Average	Different
1 and 2			
1 and 3			
2 and 3			

Figure 21: Decision region plot for every pair of classes



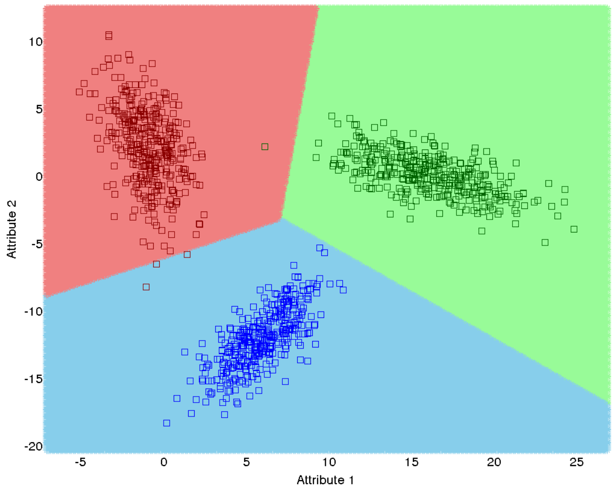
Correct : 375
Incorrect : 0
Acurracy : 100

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

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

3.2 Naive-Bayes classifier

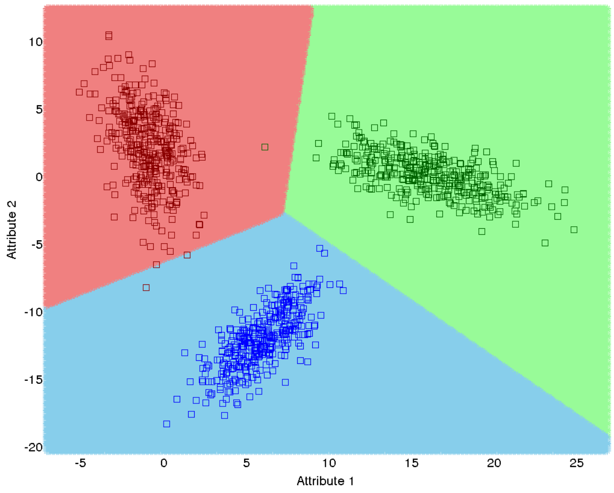
3.2.1 Linearly separable data set



Correct : 374
Incorrect : 1
Acurracy : 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 alltogether covariance



Correct : 374
Incorrect : 1
Acurracy : 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

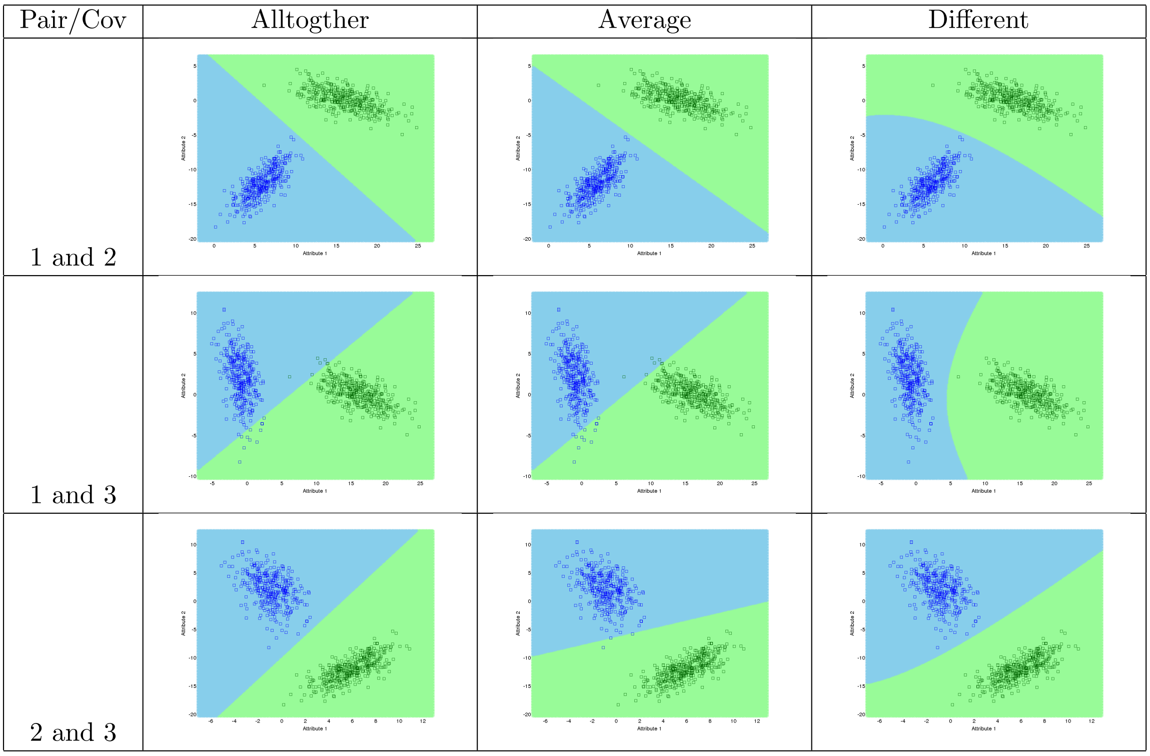
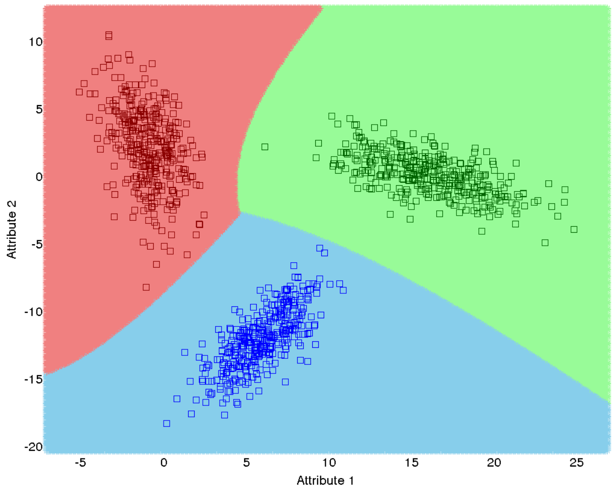


Figure 25: Decision region plot for every pair of classes



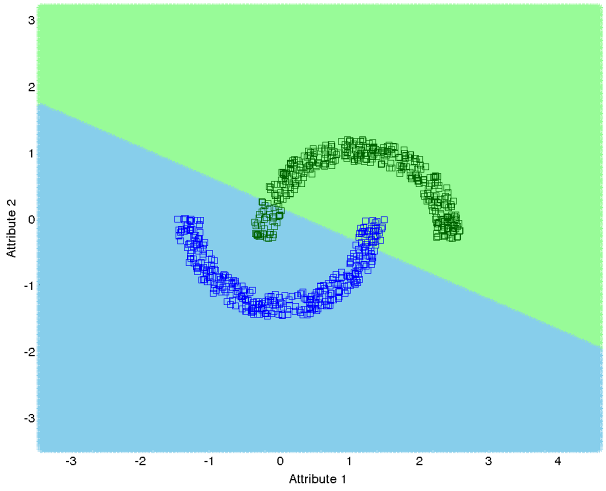
Correct : 375
Incorrect : 0
Acurracy : 100

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

Figure 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

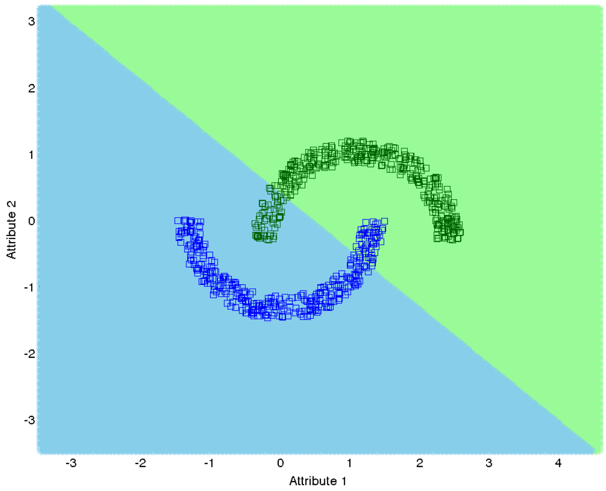
3.2.2.1 Data of Interlocking Classes



Correct : 188
Incorrect : 187
Acurracy : 50.133

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

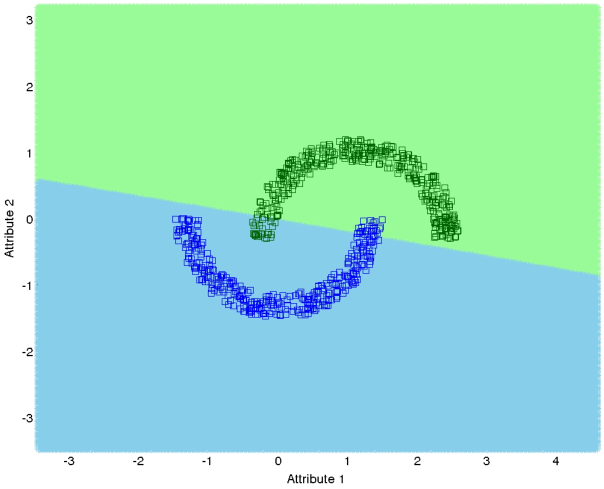
Figure 26: Decision region plot for all the classes together with the training data superposed with alltogether covariance



Correct : 188
Incorrect : 187
Acurracy : 50.133

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

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

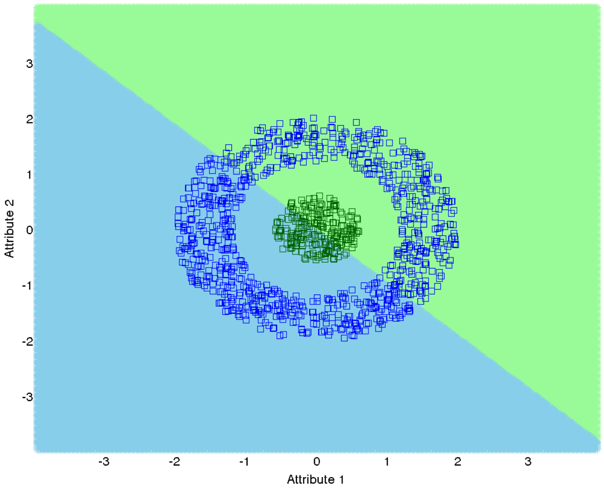


Correct : 375
Incorrect : 0
Acurracy : 100

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

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

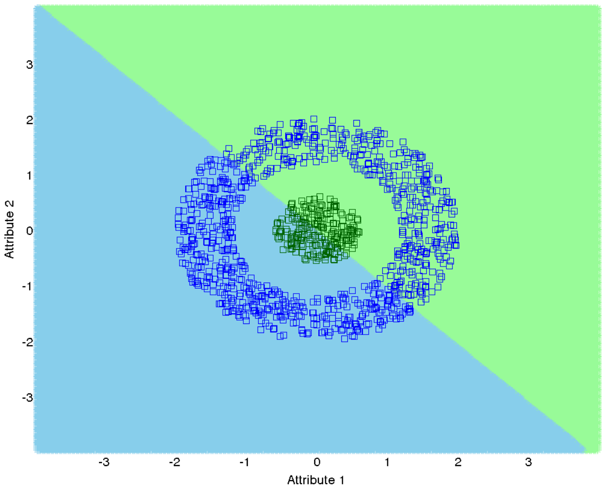
3.2.2.2 A ring with a central mass



Correct : 188
Incorrect : 187
Acurracy : 50.133

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

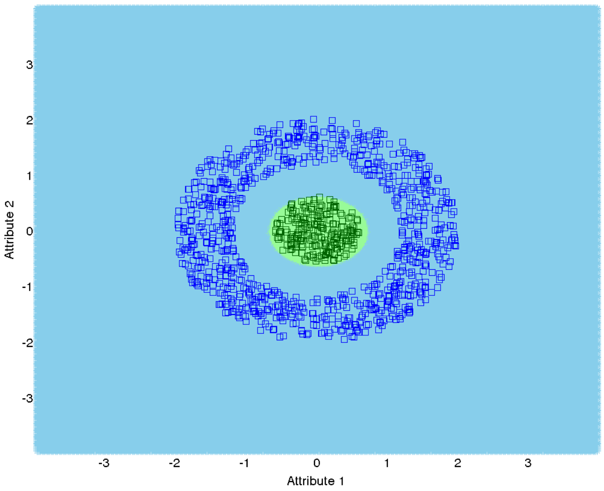
Figure 29: Decision region plot for all the classes together with the training data superposed with alltogether covariance



Correct : 188
Incorrect : 187
Acurracy : 50.133

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

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

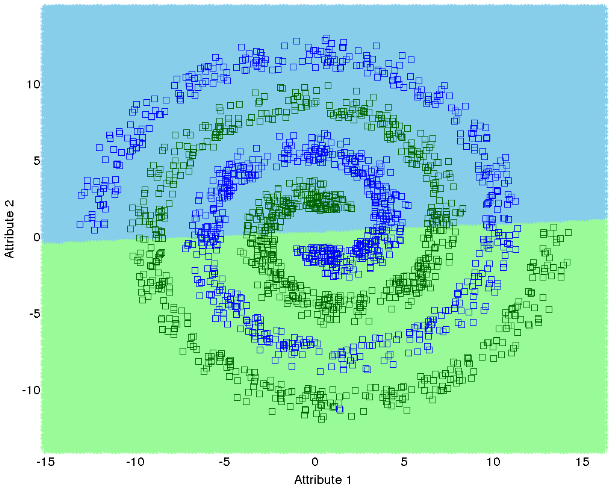


Correct : 375
Incorrect : 0
Acurracy : 100

		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

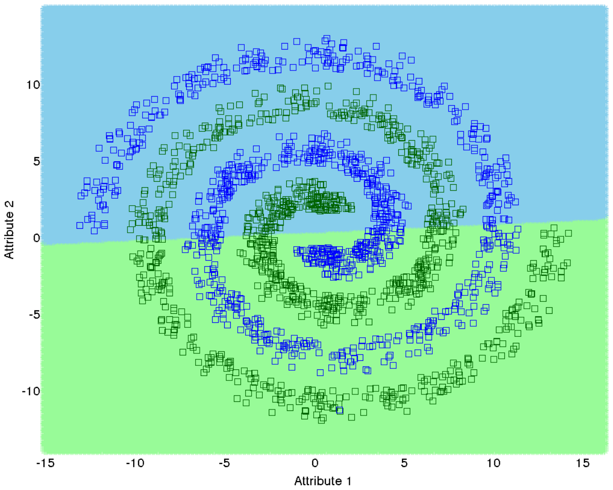
3.2.2.3 Spiral Dataset



Correct : 188
Incorrect : 187
Acurracy : 50.133

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

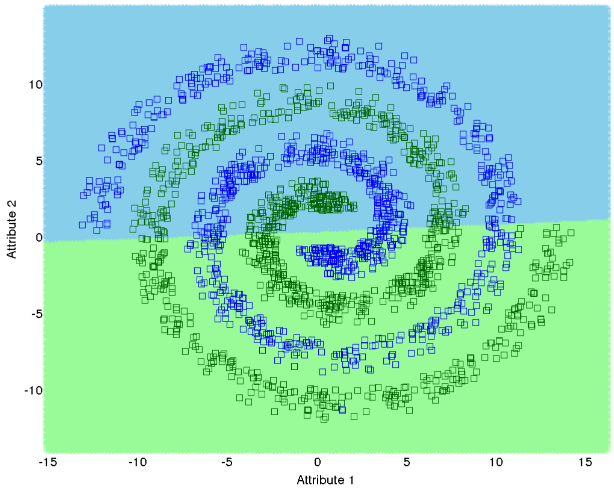
Figure 32: Decision region plot for all the classes together with the training data superposed with alltogether covariance



Correct : 188
Incorrect : 187
Acurracy : 50.133

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

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

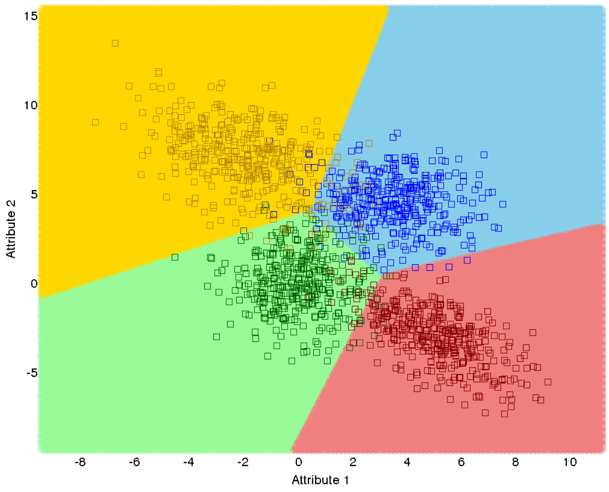


Correct : 375
Incorrect : 0
Acurracy : 100

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

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

3.2.3 Overlapping data set



Correct : 450
Incorrect : 50
Acurracy : 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 36: Decision region plot for all the classes together with the training data superposed with alltogether covariance

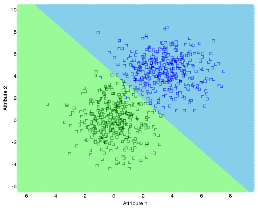
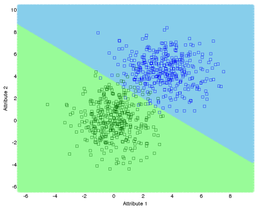
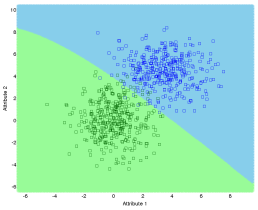
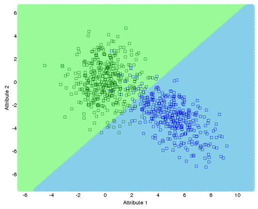
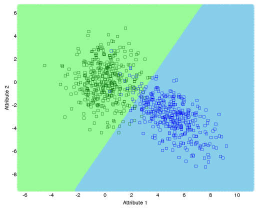
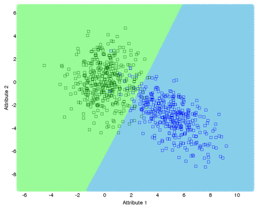
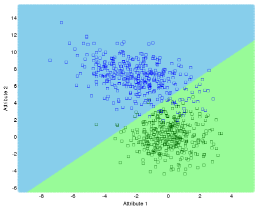
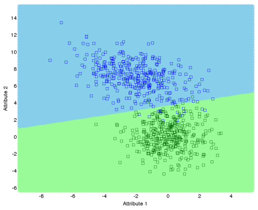
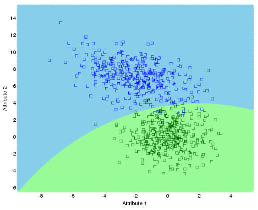
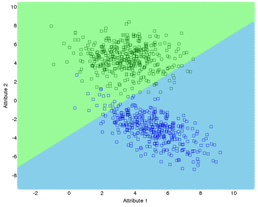
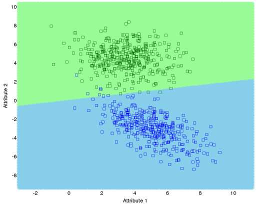
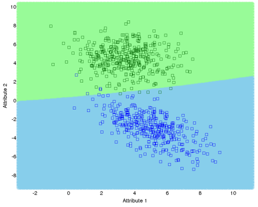
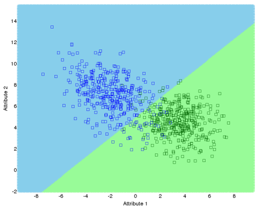
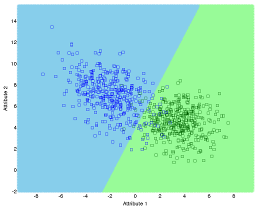
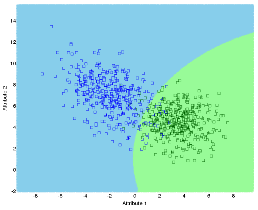
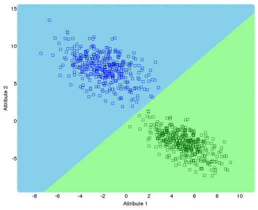
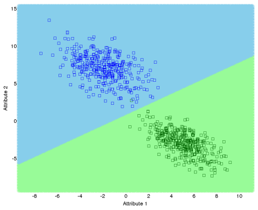
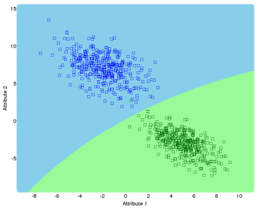
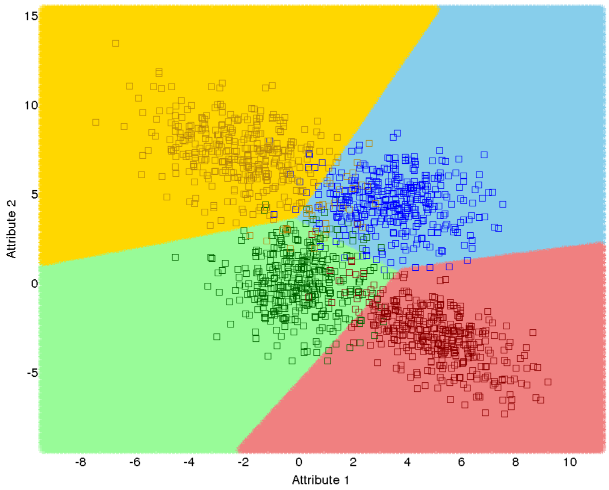
Pair/Cov	Alltogther	Average	Different
1 and 2			
1 and 3			
1 and 4			
2 and 3			
2 and 4			
3 and 4			

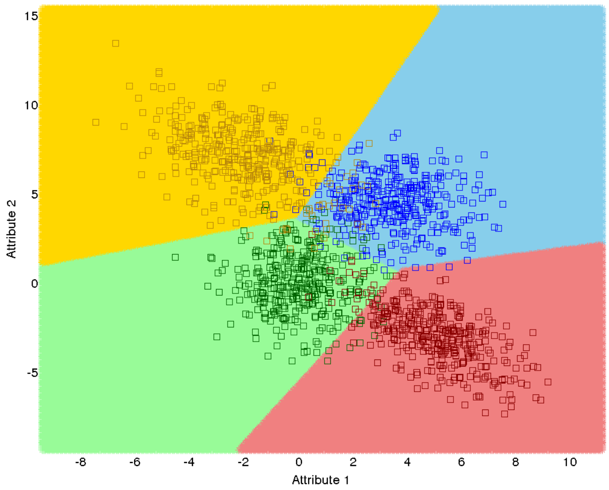
Figure 35: Decision region plot for every pair of classes



Correct : 453
Incorrect : 47
Acurracy : 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 37: Decision region plot for all the classes together with the training data superposed with average covariance

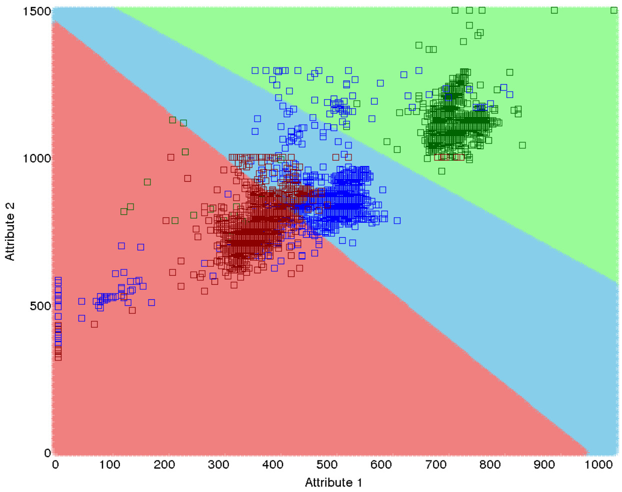


Correct : 452
Incorrect : 48
Acurracy : 90.400

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

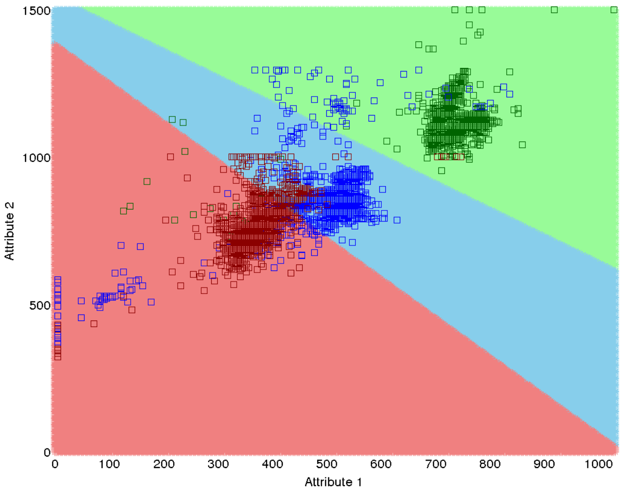
3.2.4 Real world data set



Correct : 374
Incorrect : 1
Acurracy : 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 39: Decision region plot for all the classes together with the training data superposed with alltogether covariance



Correct : 374
Incorrect : 1
Acuracy : 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 40: Decision region plot for all the classes together with the training data superposed with average covariance

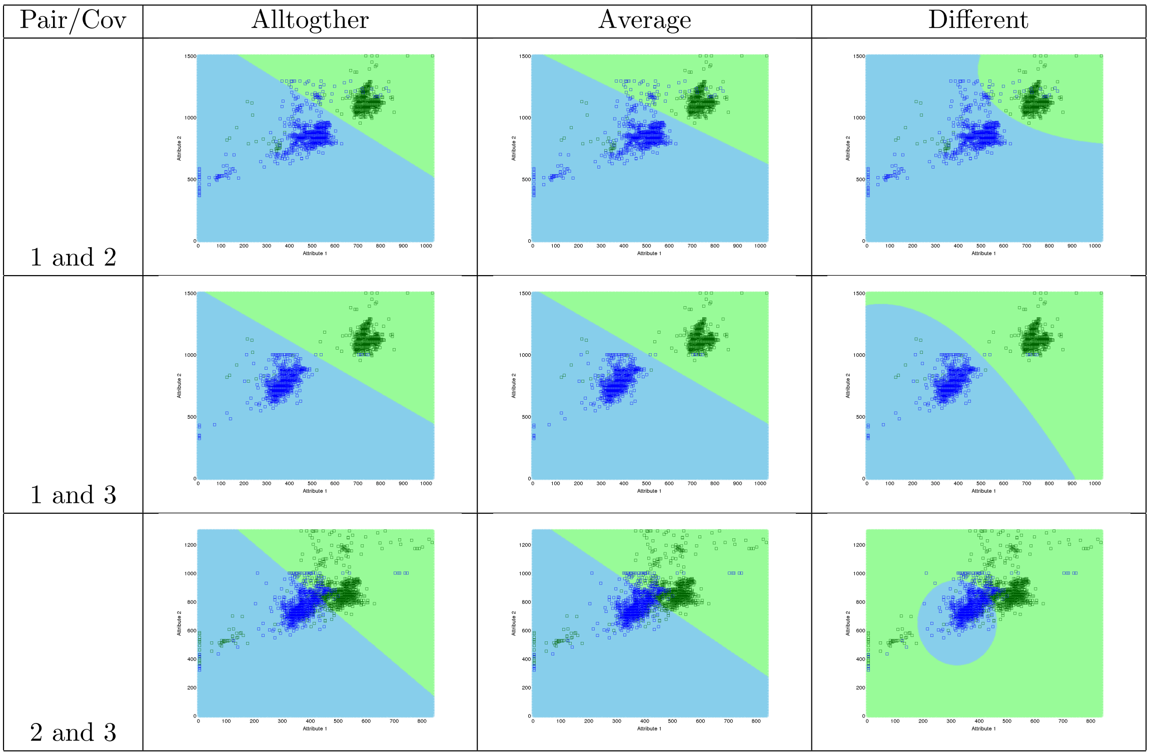
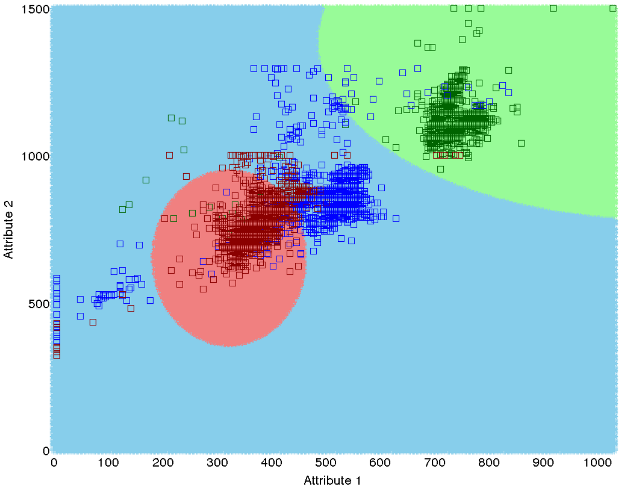


Figure 42: Decision region plot for every pair of classes



Correct : 375
Incorrect : 0
Acurracy : 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 41: Decision region plot for all the classes together with the training data superposed with different covariance

3.2.5 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.

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.

```
> data=read.table("hw2_chol.txt")
> hist(data$V1,xlab='Cholesterol (mg/dL)',main='Histogram of Total Cholesterol')
> boxplot(data$V1,main='Total Cholesterol',ylab='Cholesterol (mg/dL)')
```