CS 669 Assignment 1

Rohit Patiyal
Devang Bacharwar

September 17, 2015

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

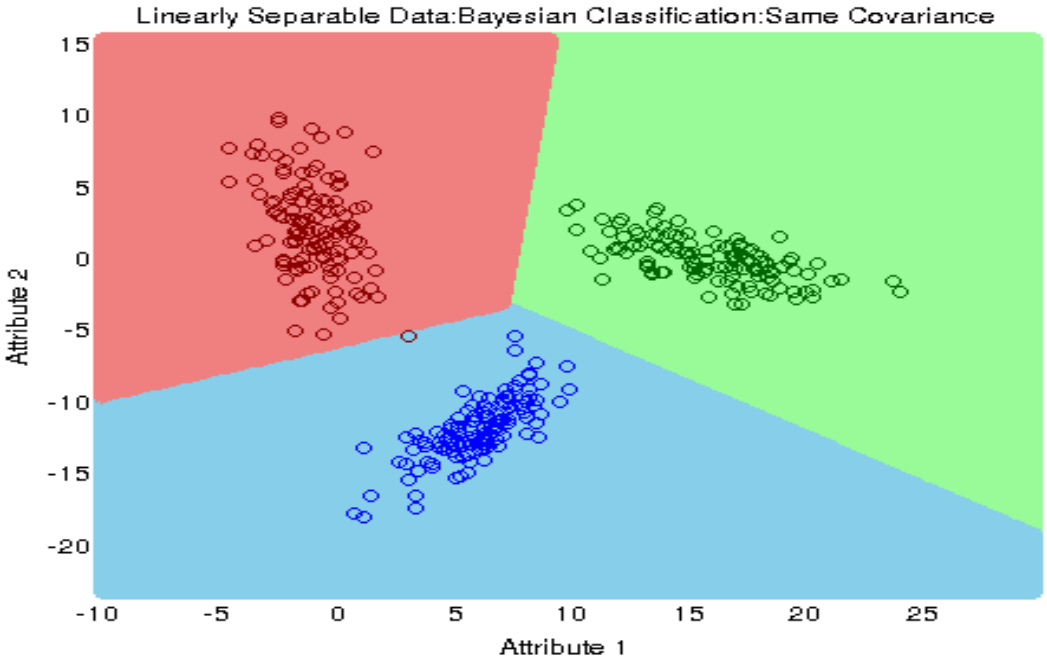
1. Data for each class is partitioned into 75 % for training and 25 % for testing
2. Mean and Covariances are calculated for each class using the training .
3. For points in a grid, likelihood is calculated for each class and is labeled as of the class with the maximum likelihood probability.

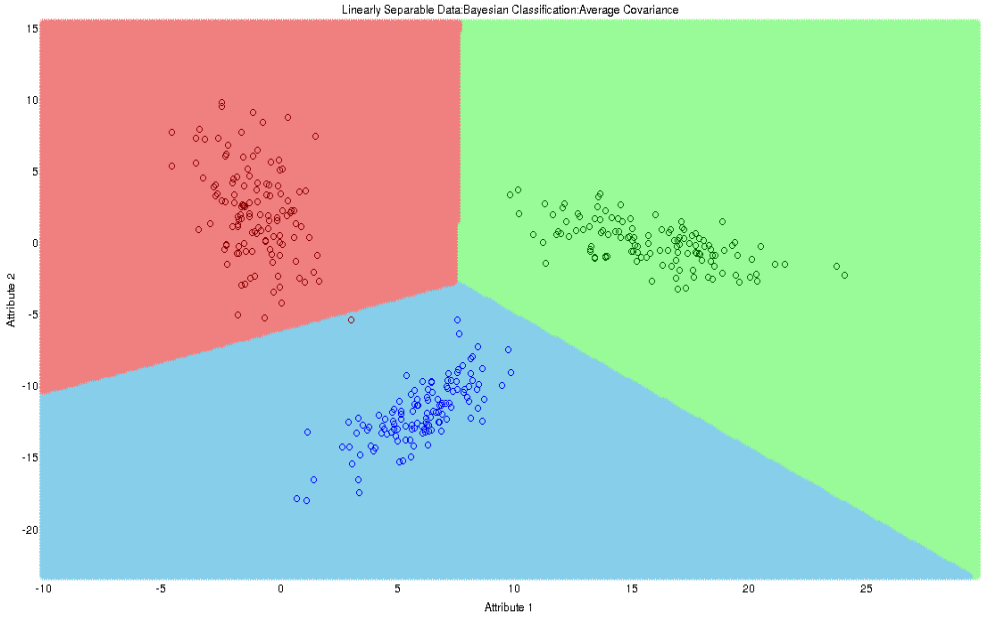
   For bayes classifier, the likelihood is assumed to be a multivariate gaussian distribution
4. These labelled points are plotted with different colors to see the different regions separated by the decision boundaries.
5. The testing data is also plotted over the regions, and observations a re made.

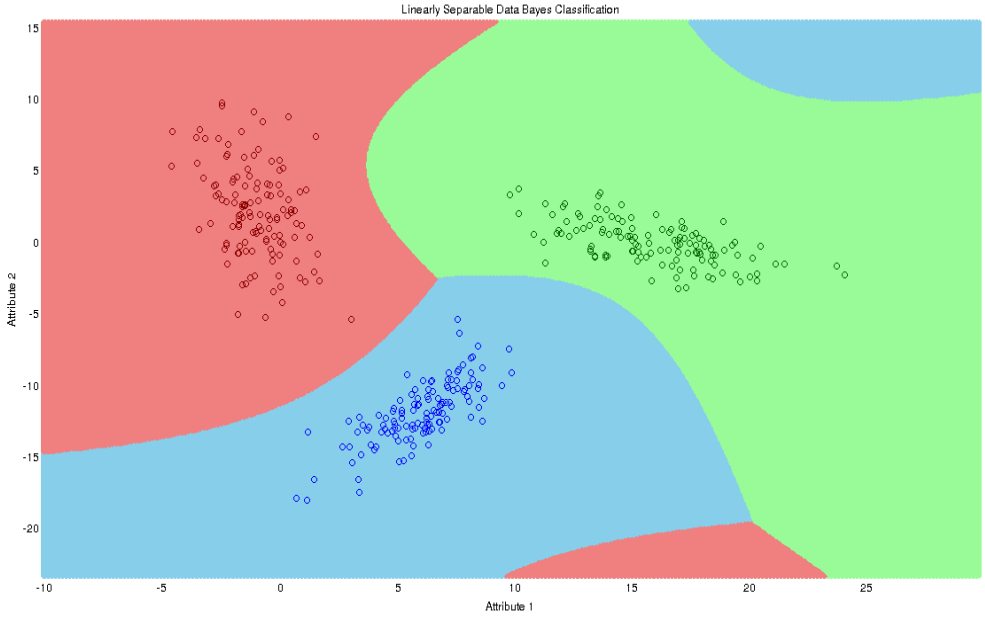
3 Observations

3.1 Bayes Classifier

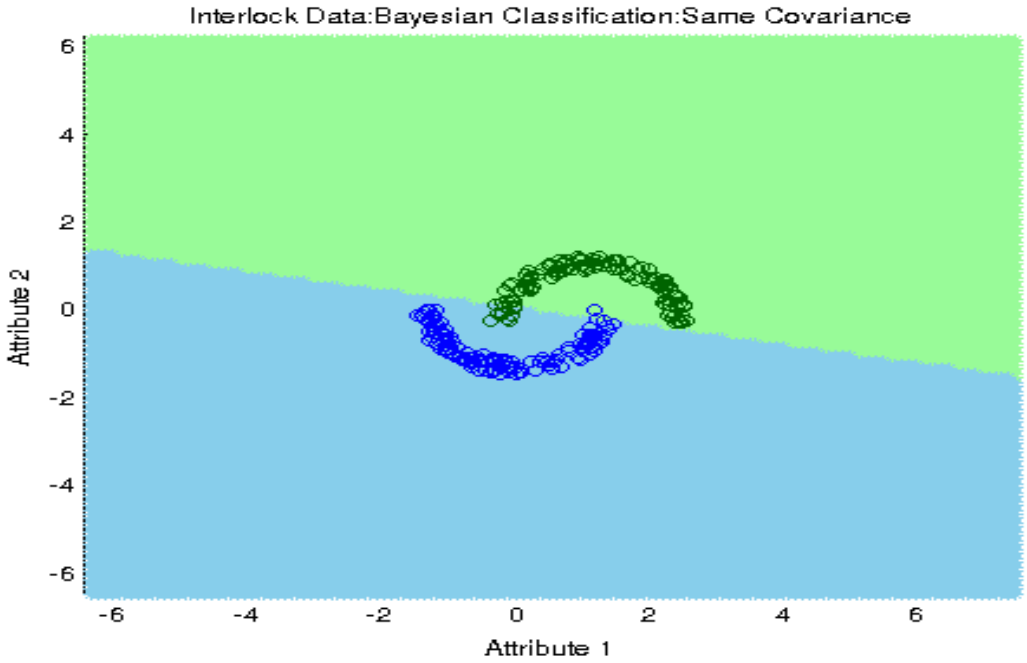
3.1.1 Linearly separable data set

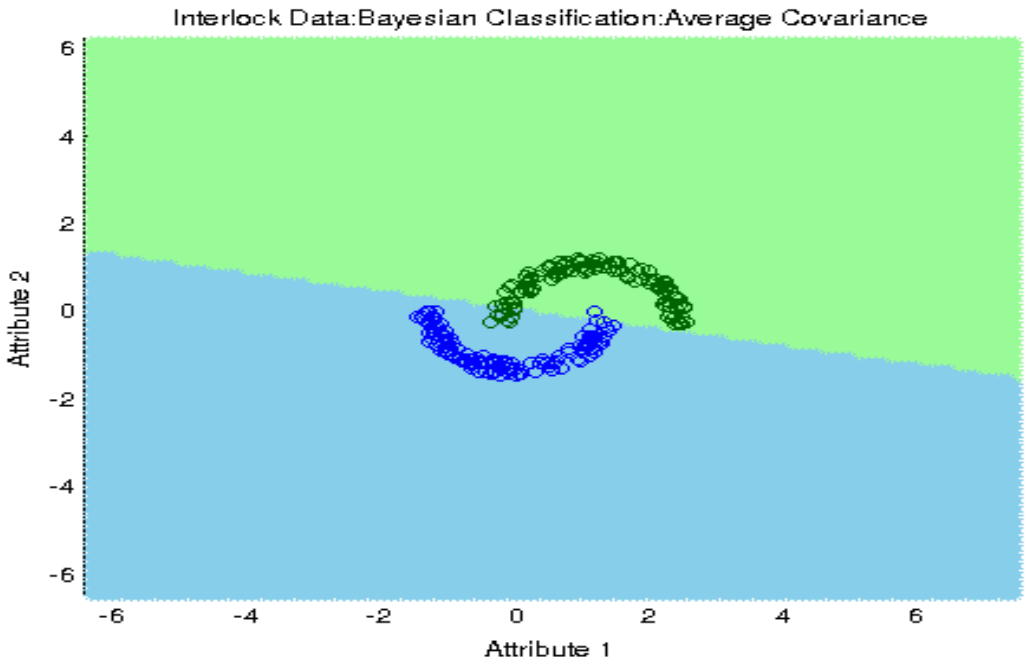


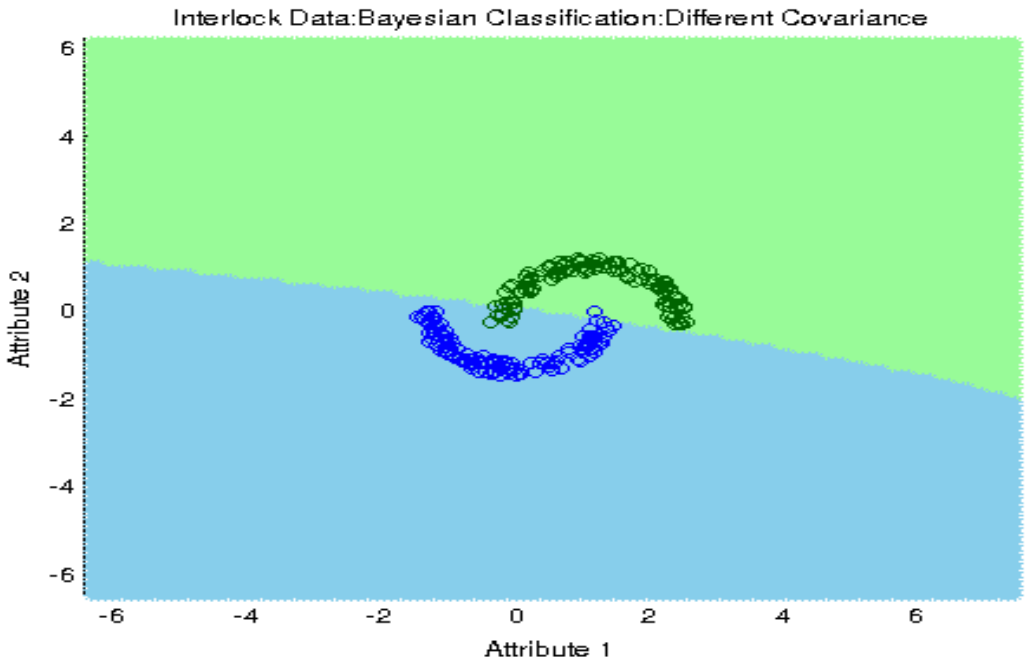




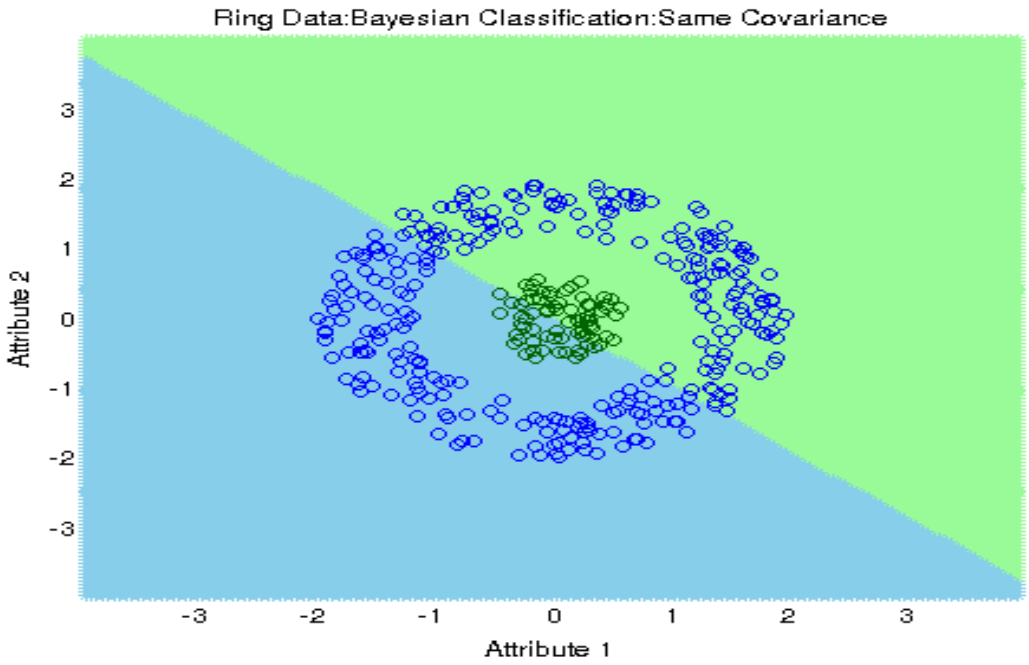
3.1.2 Non-Linearly separable data set

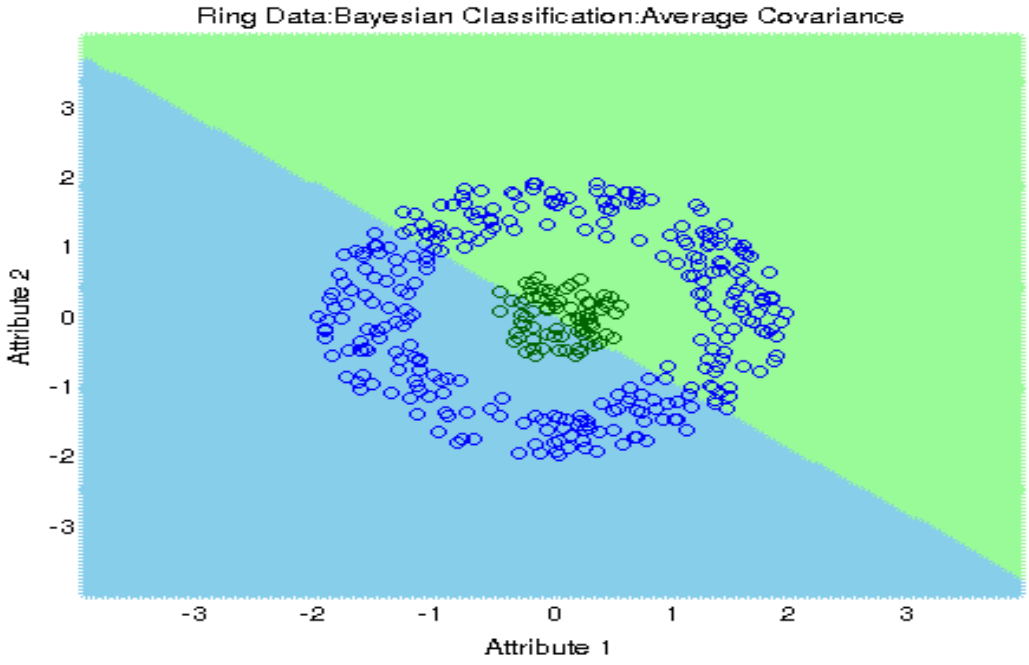


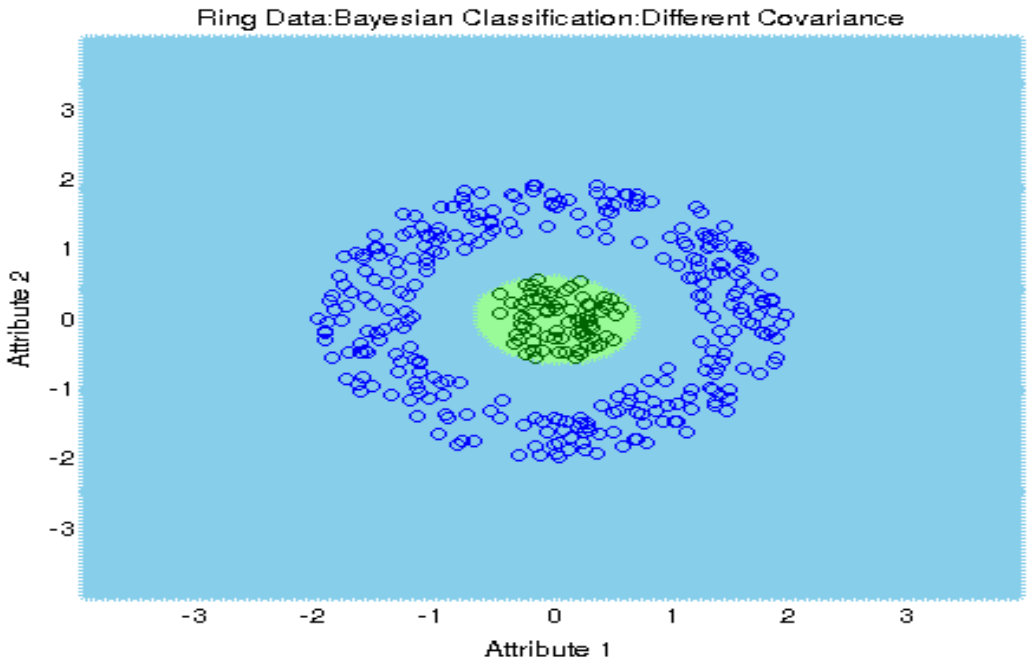




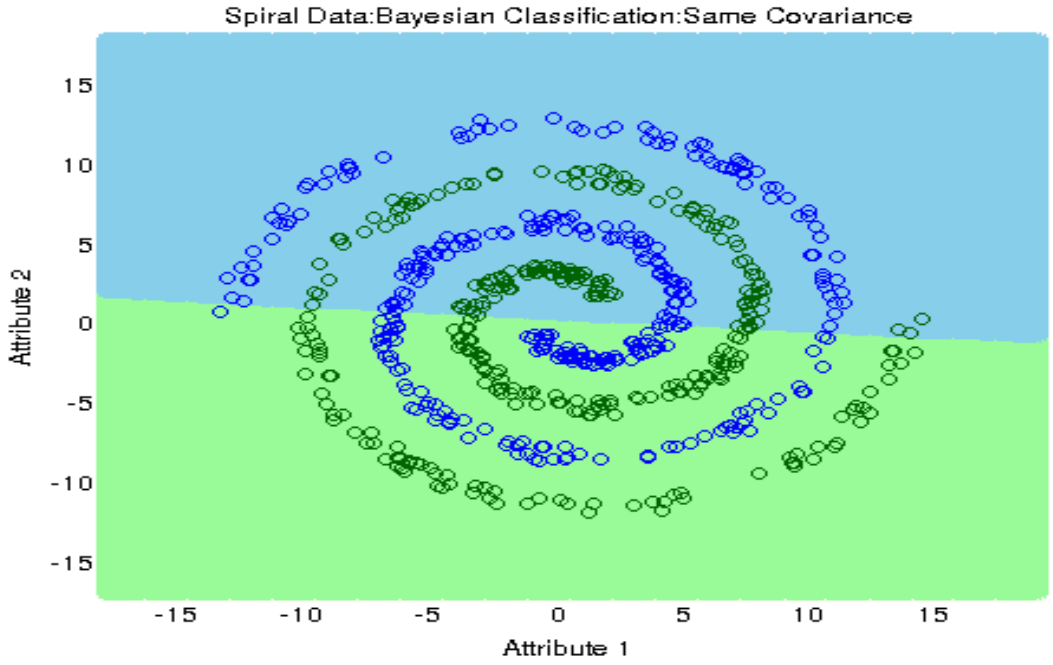
3.1.2.1 Data of Interlocking Classes

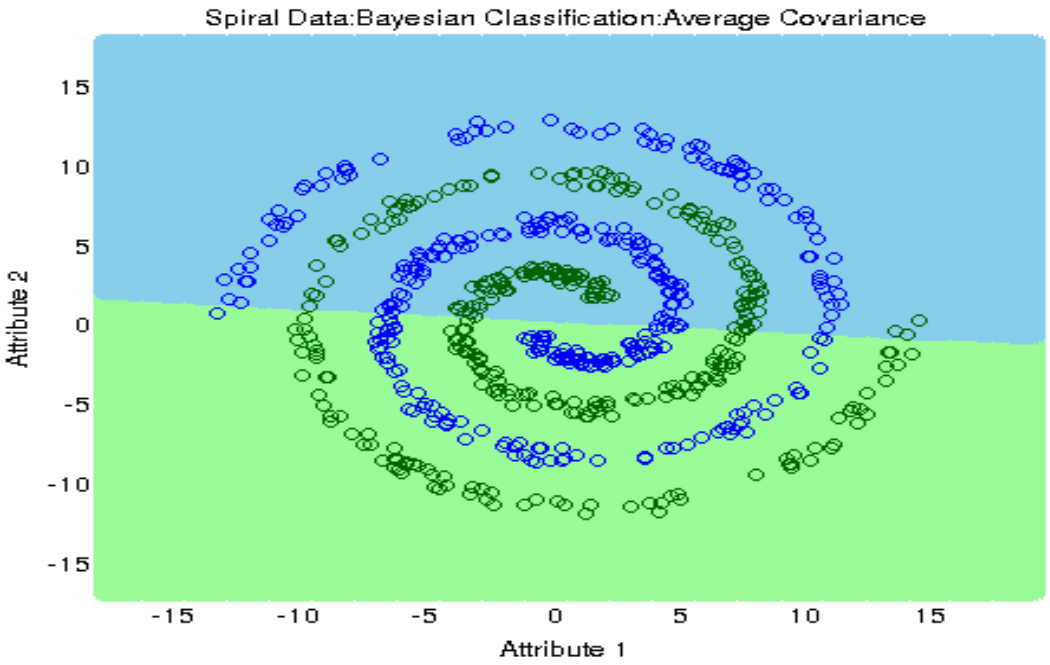

Ring Data:Bayesian Classification:Same Covariance
Attribute 2
Attribute 1
3
2
1
0
-1
-2
-3
-3
-2
-1
0
1
2
3

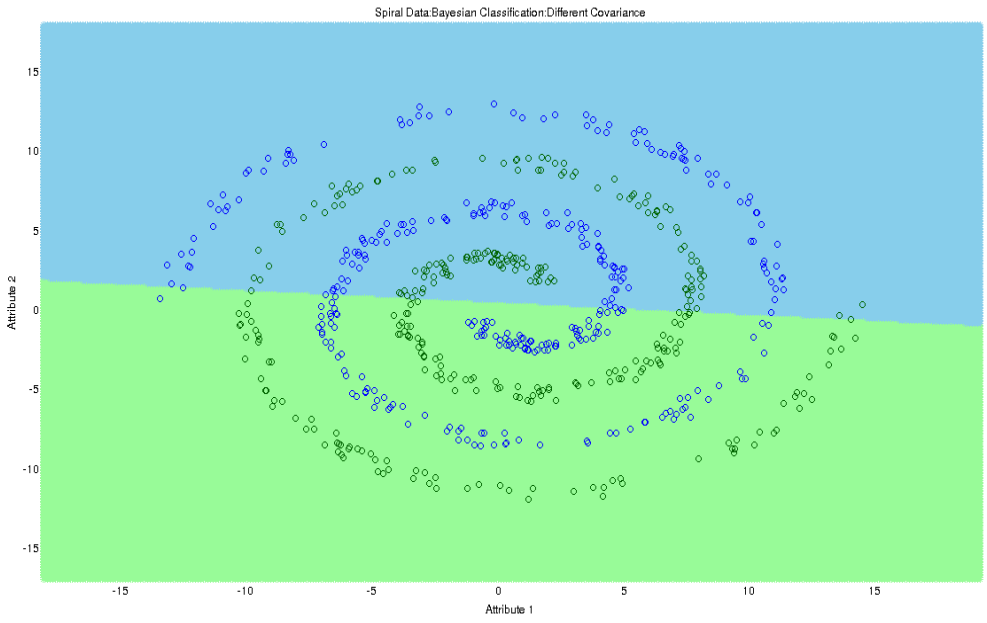

Ring Data:Bayesian Classification:Average Covariance
Attribute 2
Attribute 1
3
2
1
0
-1
-2
-3
-3
-2
-1
0
1
2
3



3.1.2.2 A ring with a central mass

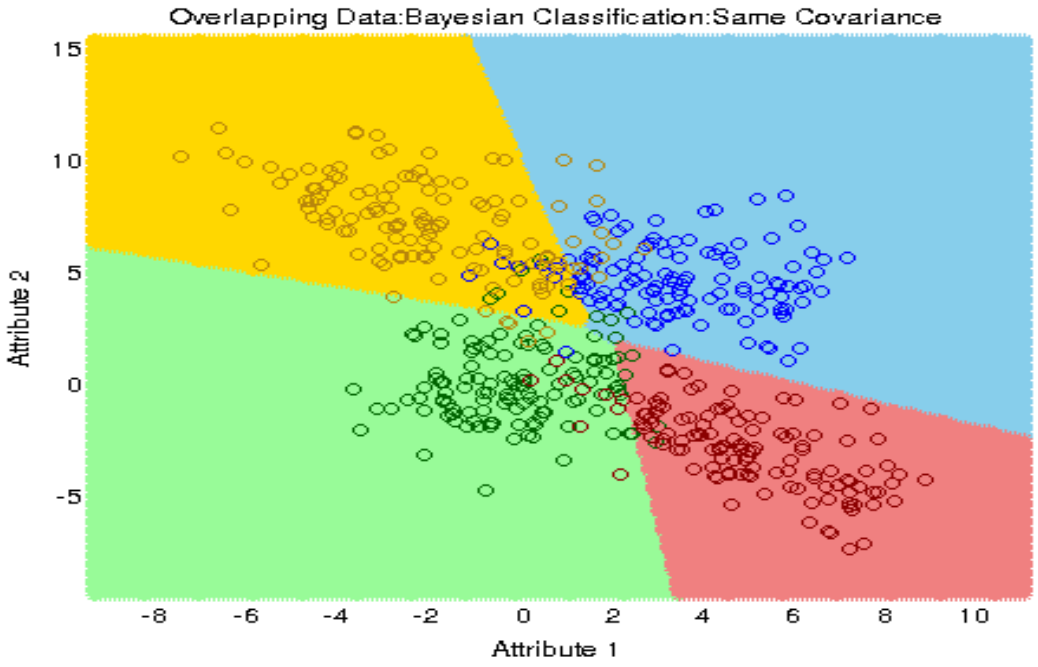


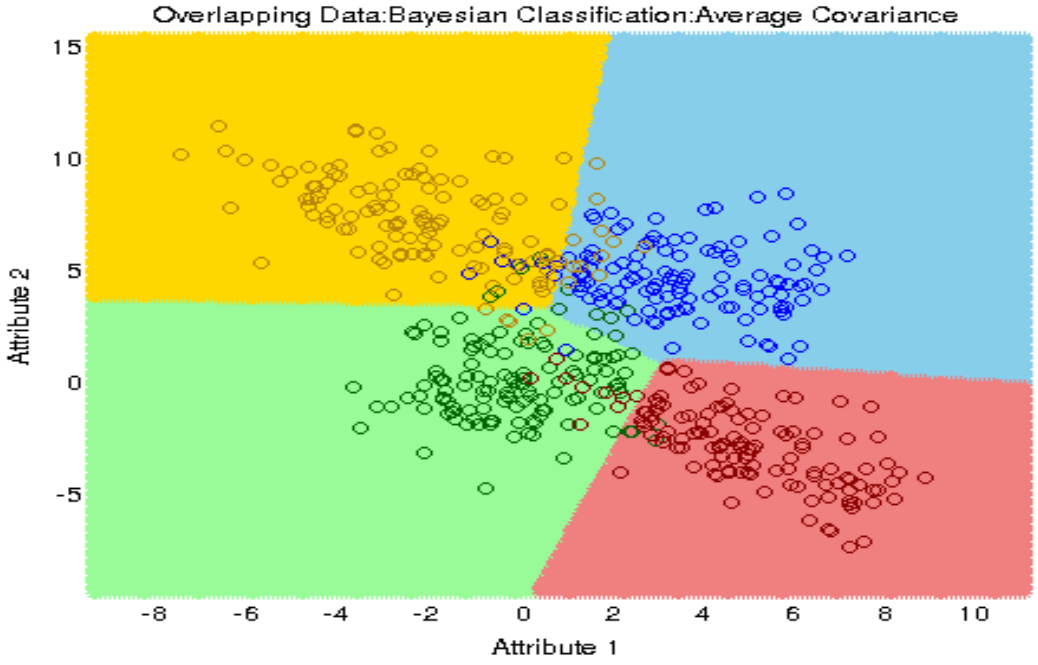


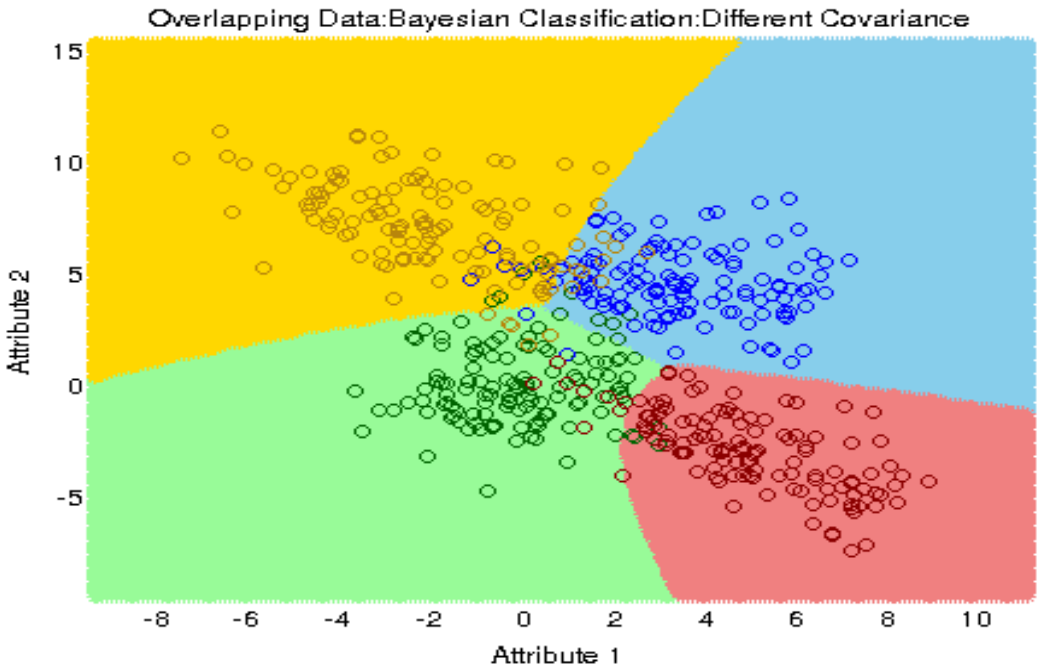


3.1.2.3 Spiral Dataset

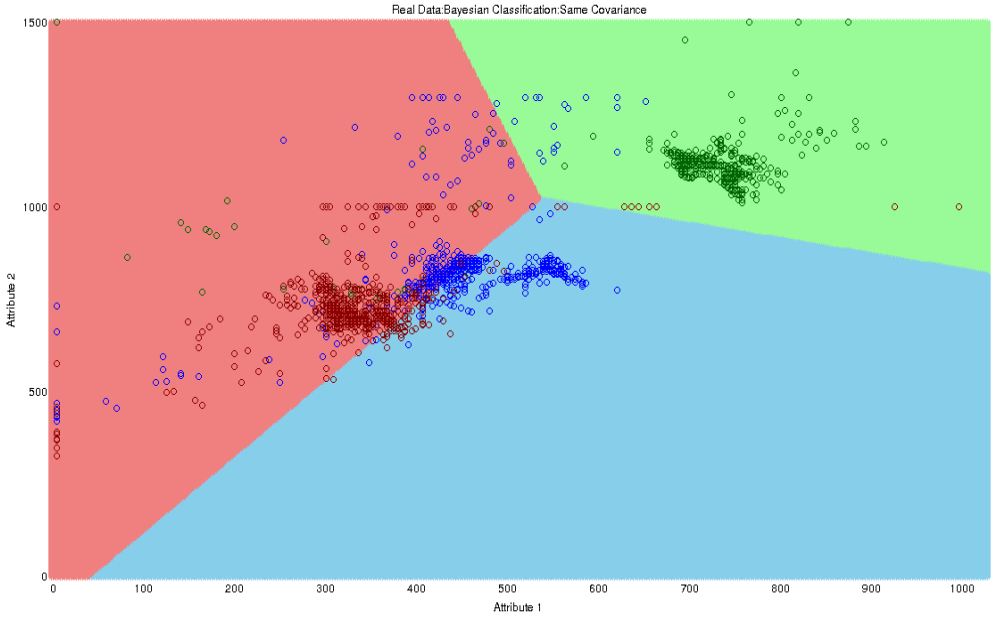
3.1.3 Overlapping data set

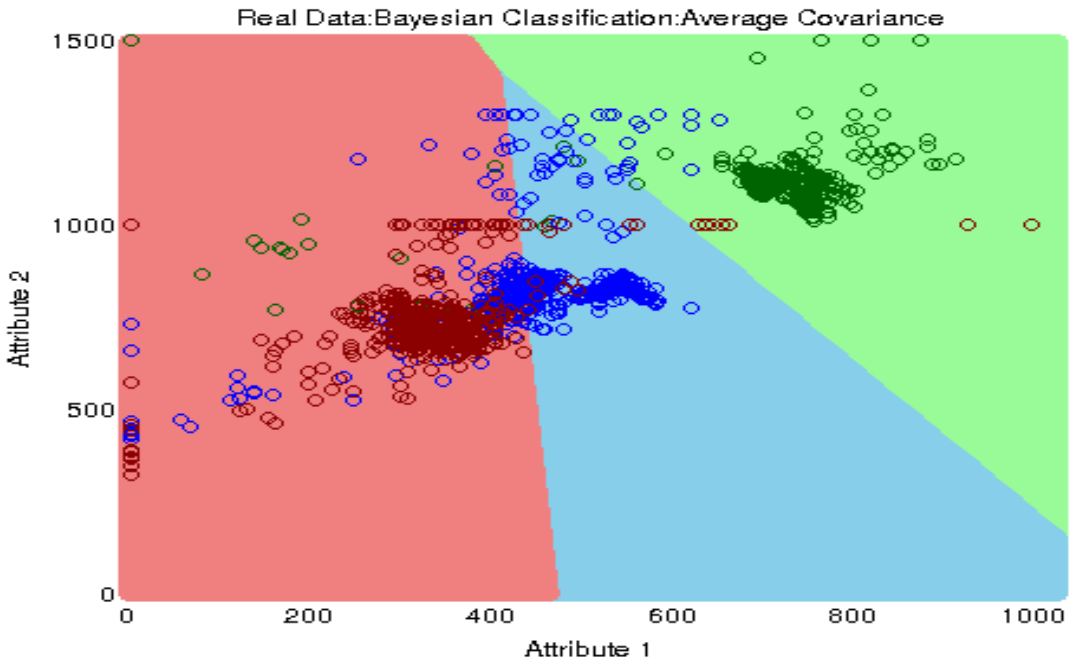






3.1.4 Real world data set





plots/bayes/real/diff_cov.png

3.2 Naive-Bayes classifier

3.2.1 Linearly separable data set

3.2.2 Non-Linearly separable data set

3.2.2.1 Data of Interlocking Classes

3.2.2.2 A ring with a central mass

3.2.2.3 Spiral Dataset

3.2.3 Overlapping data set

3.2.4 Real world data set

4 Conclusion

```
> 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)')
```