EE657 : Pattern Recognition and Machine Learning

Assignment 1 Report
Name : Oksiya Ayush Puranlal
Roll No : 150121028

Q1) Problem: Learn the parameters μ5 , μ6 ,Σ_5, Σ_6 and π by maximizing the likelihood,
$\pi = Pr(r = C_5)$ $\qquad Pr(x|C_5) = N(x|\mu_5, P_5)$
$1 - \pi = Pr(r = C_6)$ $\qquad Pr(x|C_6) = N(x|\mu_6, P_6)$

Use Bayes decision criterion to classify the test data. Estimate the misclassification rates of both classes and populate the 2x2 confusion matrix.

Dataset : Modified version of the *Optical Recognition of Handwritten Digits Dataset* from the *UCI repository*.The original dataset consists of normalized bitmaps of handwritten digits (0 − 9). 32x32 bitmaps are divided into non-overlapping blocks of 4x4 and the number of ON pixels are counted in each block.This generates an input matrix of 8x8 where each element is an integer in the range 0 − 16. This reduces dimensionality and gives invariance to small distortions.

Training data : 'P1_data_train.csv' consisting of 777 instances(rows) of 64 attributes(cols) corresponding to the handwritten digit value(5 or 6) given in 'P1_labels_train.csv'.

Test data : 'P1_data_test.csv' consisting of 333 instances(rows) of 64 attributes(cols) corresponding to the handwritten digit value(5 or 6) given in ' P1_labels_test.csv'.

Findings :

The Estimates Mean, Covariance, and Probabilities for each class(i.e. For character to be 5 & 6) and overall Mean & Covariance was calculated using Bayes Decision Criterion.

Using this estimates and probabilities the predictions were made on the test data for different possible covariance matrices for both the classes(i.e. For character to be 5 & 6).

Case 1 : When we use the Covariance Matrix obtained from training data for both the classes. (Here both will have different independent Covariance Matrices).

Misclassification rate for class 1 (Character == 5) : 31.6129 %
Misclassification rate for class 2 (Character == 6) : 15.1685 %

Confusion Matrix :

106	49
27	151

Case 2 : When we use same covariance matrix for both classes

Case 2(a) : Using Covariance Matrix obtained for entire training data as covariance matrix for both class 1(Character == 5) and class 2(Character == 6).

Misclassification rate for class 1 (Character == 5) : 12.2580 %
Misclassification rate for class 2 (Character == 6) : 15.7303 %
Confusion Matrix :

136	19
28	150

Case 2(b) : Using weighted average of covariance matrix obtained for each class in training data as covariance matrix for both class 1(Character == 5) and class 2(Character == 6).

Misclassification rate for class 1 (Character == 5) : 13.5483 %
Misclassification rate for class 2 (Character == 6) : 15.1685 %

Confusion Matrix :

134	21
27	151

Case 2(c) : Using Identity Matrix as Covariance Matrix for Both class 1(Character == 5) and class 2(Character == 6).

Misclassification rate for class 1 (Character == 5) : 14.1935 %
Misclassification rate for class 2 (Character == 6) : 22.4719 %

Confusion Matrix :

133	22
40	138

Q2) Problem: Learn a binary classifier for the given data taking class conditional densities as normal density. Estimate the misclassification rates of both classes, plot the discriminant function and iso-probability contours for the following cases:

(a) Equal diagonals Σ_s of equal variances along both dimensions,
(b) Equal diagonal Σ_s with unequal variances along different dimensions,
(c) Arbitrary Σ_s but shared by both classes,
(d) Different arbitrary Σ_s for the two classes.

Dataset :

Training data :'P2_train.csv' consisting of 310 instances, 2 attributes +1 class label.
Test data : ' P2_test.csv' consisting of 90 instances, 2 attributes +1 class label.

Findings :

The Estimates Mean, Covariance, and Probabilities for each class(i.e. For 0 & 1) and overall Mean & Covariance was calculated using Bayes Decision Criterion.

Using this estimates and probabilities the predictions were made on the test data for different possible covariance matrices for both the classes(i.e. For 0 & 1).

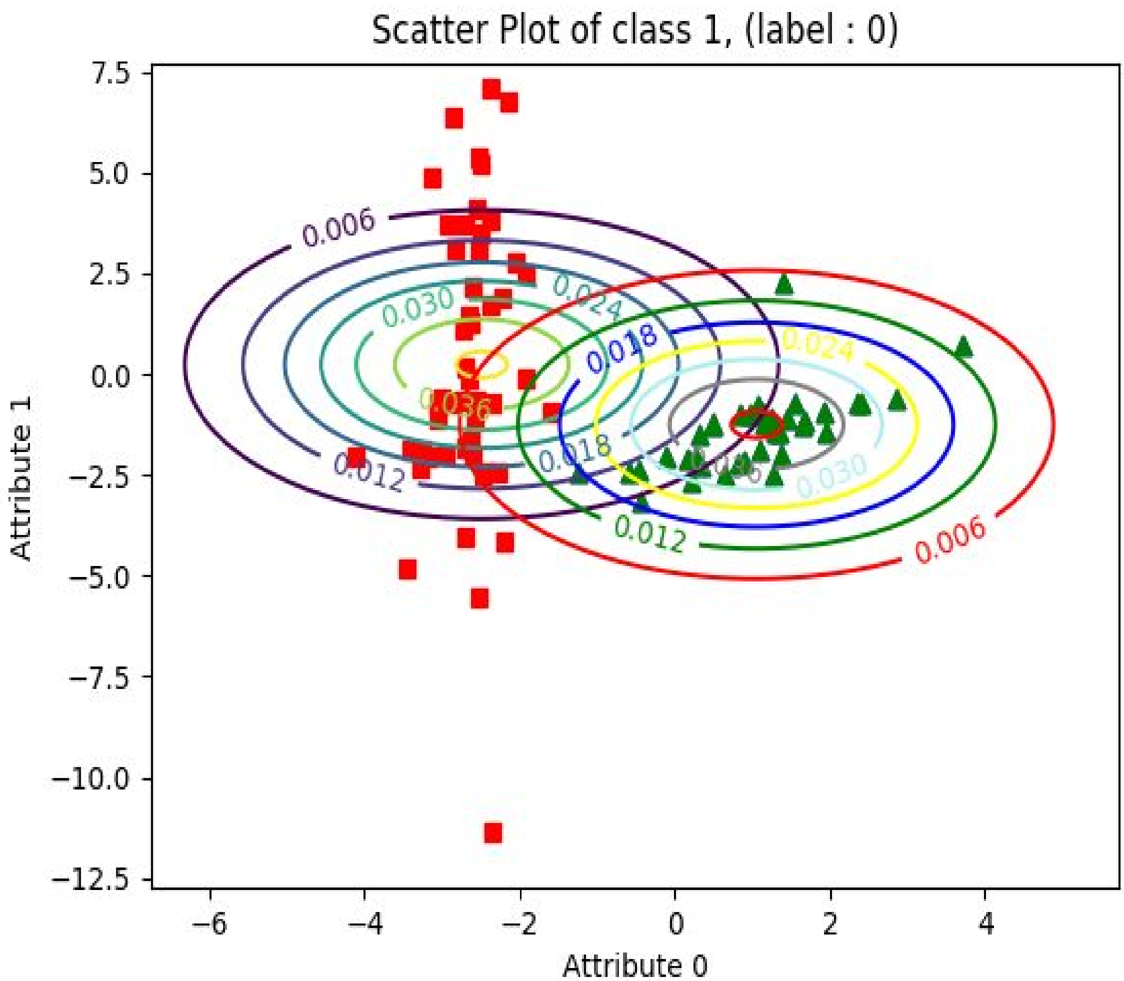
Case 1 : Equal diagonals Σ_s of equal variances along both dimension.
Here I have used the diagonals with the value of first element of the covariance (estimate) of whole training data.

Misclassification rate for class 1 (For label : 0) : 6.0 %
Misclassification rate for class 2 (For label : 1) : 0.0 %

Confusion Matrix :

47	3
0	40

Plot :



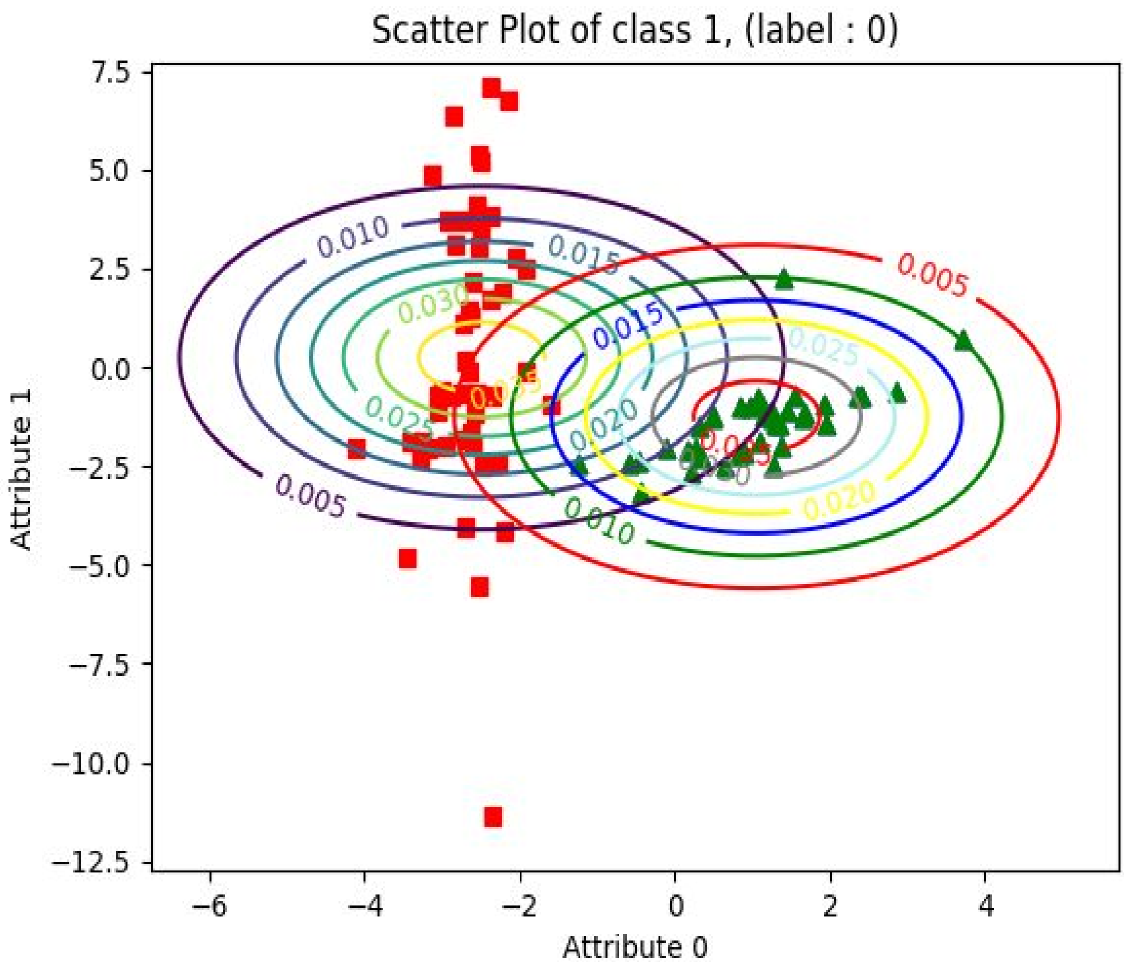
Case 2 : Equal diagonal Σ_s with unequal variances along different dimensions.
Here I have used the diagonal elements of covariance matrix (estimate) of whole training data. And resultant matrix to be the covariance in each class.

Misclassification rate for class 1 (For label : 0) : 2.0 %
Misclassification rate for class 2 (For label : 1) : 0.0 %

Confusion Matrix :

49	1
0	40

Plot :



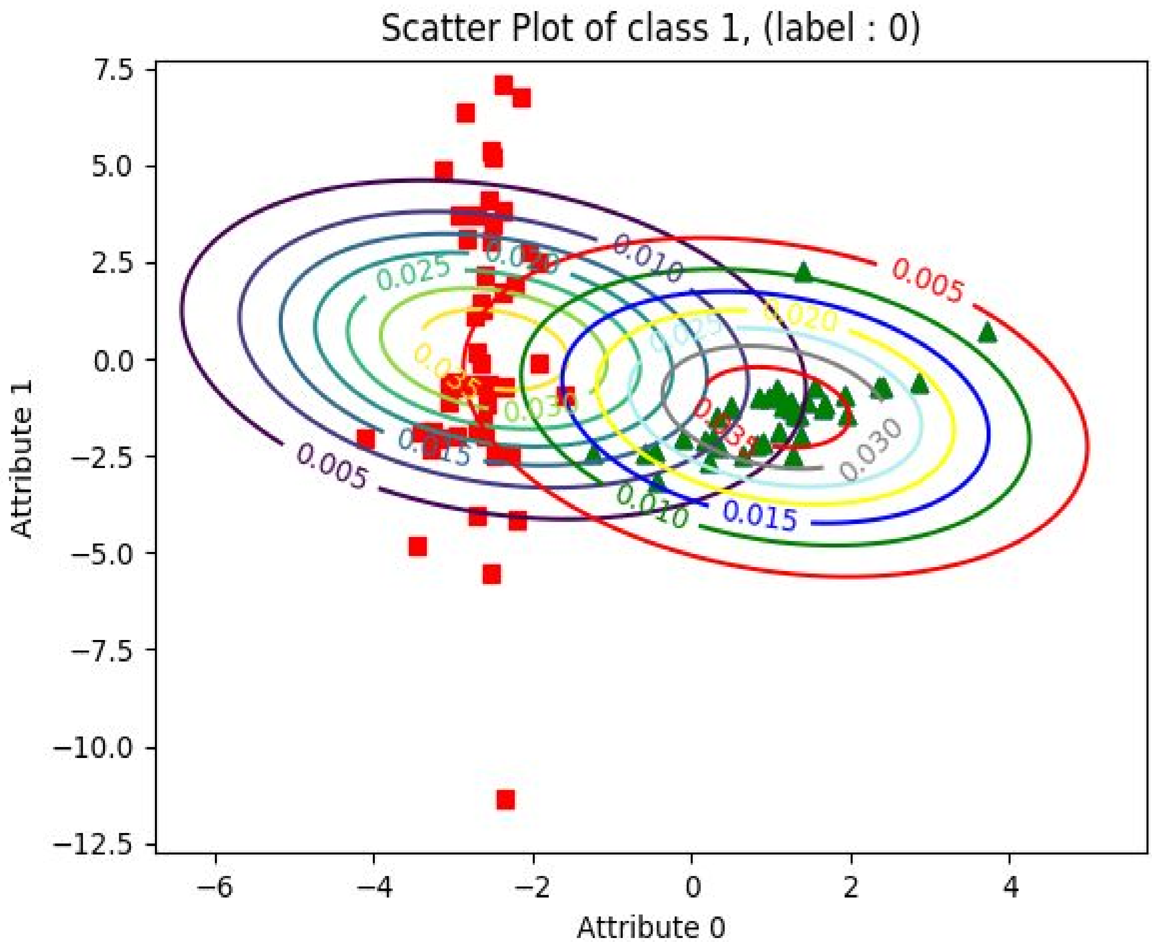
Case 3 : Arbitrary Σ_s but shared by both classes.
Here I have used the covariance matrix(estimate) of entire training data to be the covariance in each class (as it can be arbitrary).

Misclassification rate for class 1 (For label : 0) : 0.0 %
Misclassification rate for class 2 (For label : 1) : 2.5 %

Confusion Matrix :

50	0
1	39

Plot :



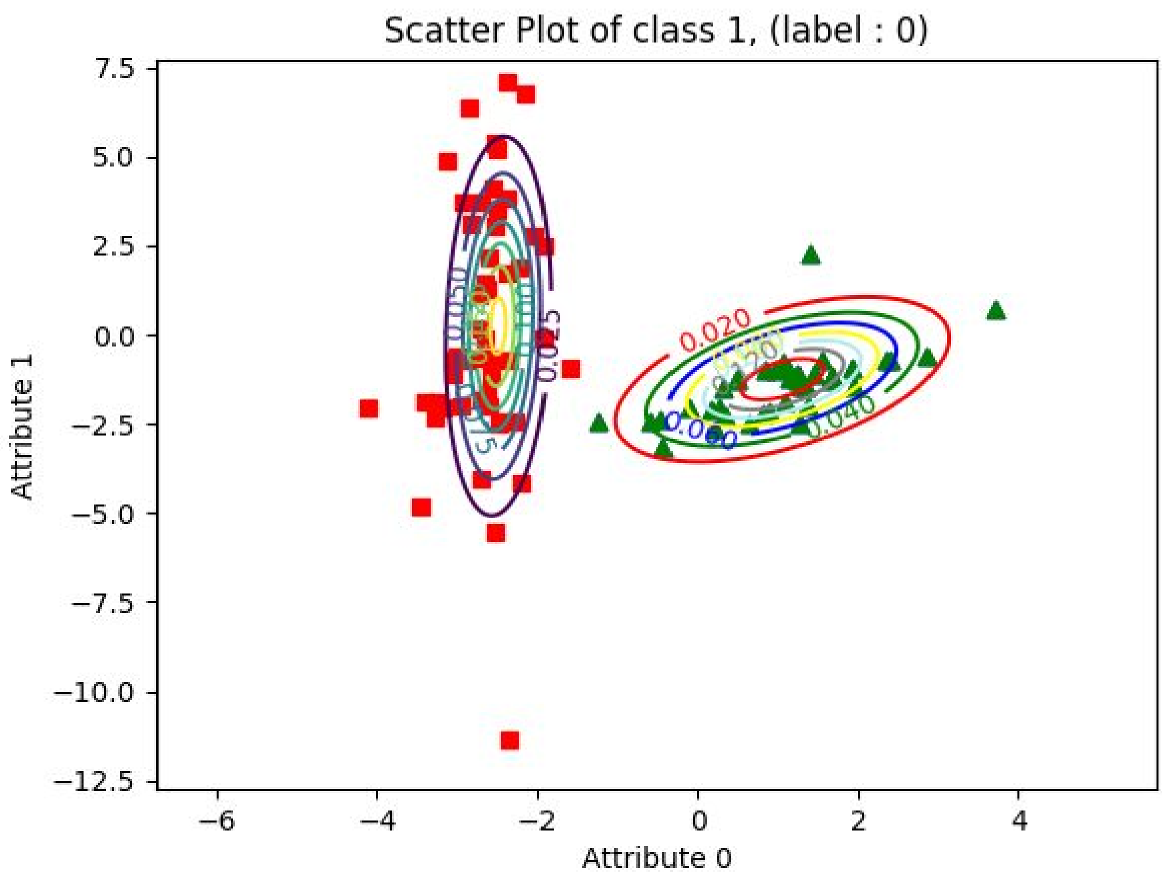
Case 4 : Different arbitrary Σ_s for the two classes.
Here I have used the covariance extracted from training data for each class using bayes decision to be the covariance in that particular class.

Misclassification rate for class 1 (For label : 0) : 0.0 %
Misclassification rate for class 2 (For label : 1) : 0.0 %

Confusion Matrix :

50	0
0	40

Plot :



As we can see for different arbitrary covariance matrices the misclassification rate for both classes are 0.0 %, for arbitrary shared covariance matrices it is 0.0 & 2.5 %, for equal diagonal and different variance covariance it is 2.0 & 0.0 % and for equal diagonal and equal variance covariance it is 6.0 & 0.0 %.

Q3) Problem: We wish to understand the association between an employees age and education, as well as the calendar year, on his wage. Perform polynomial regression on age vs wage, year vs wage, plot education vs wage. Provide description of your observations on the

variation of wage as a function of each these attributes. Can we get an accurate prediction of a particular man 0 s wage from one of these 3 attributes alone?
Experiment with your own ideas, related to those discussed in the class and make a brief report of your findings.

Dataset : *Wage dataset* contains the income survey information for a group of males from Atlantic region of the United States.

Data : ‘Wage_dataset.csv’ has the numerical data from ‘Wage_original.csv’ extracted (except the 1st column). The data consist of 3000 instances, 9 attributes (including the age and education and the calendar year)+ 2 columns giving the natural log of the wage, wage Respectively.

Findings :

For Wage Vs Year : Best Order For Polynomial regression : 3
For Wage Vs Age : Best Order for polynomial regression : 9 (I checked till 50 order)