EEL5840 Elements of Machine Intelligence - HW 4

Madhusudan Govindraju 39267182

The document explains all the assumptions and the steps undertaken to complete the assignment with the results attached.

1

1. The Formula for Naive Bayes classifier is

$$\boxed{P(\omega_j / x) = \frac{p(x/\omega_j)P(\omega_j)}{p(x)} = \frac{likelihood \times prior}{evidence}}$$

2. Based on the instructions in the question we assume the sample distribution to be a gaussian distribution and the formula for that is as available in the question booklet.

$$g_j(x) = -\frac{1}{2}(x-\mu_j)^\top \Sigma_j^{-1}(x-\mu_j) - \frac{d}{2}\log(2\pi) - \frac{1}{2}\log(\det(\Sigma_i)) + \log(p(j)).$$

3. In this equation the Σ_i is taken as the covariance of the entire data set and the Σ_j is taken as the covariance matrix for that particular class for which we are computing the posterior probability.
4. This is also the discriminant function for this case. So we tag the sample to the class which gives the highest output from the discriminant function.
5. We first separate the data set into the training and testing data set with a ratio of 70% 30% with 140 samples for training data and 60 samples for testing data.
6. The covariance matrices, class prior and the class mean is calculated from the training data.
7. We predict the classes using the above mentioned discriminant function and plot the confusion matrix and give the error Rate.
8. The time to give the predictions is also noted in each case.
9. the output with the error rate and the execution times are available below

```
Calculating the Covariance Matrix for Class 0
Calculating the Covariance Matrix for Class 1
Calculating the Covariance Matrix for EntireDataSet
Elapsed time is 0.015846 seconds.

-------Testing Data -------

The Error Rate for Testing Data is 0.016667
Elapsed time is 0.912333 seconds.

----- Training Data --------

The Error Rate for Training Data is 0.007143
Elapsed time is 0.582898 seconds.

------Complete Dataset ------

The Error Rate for Complete dataset is 0.010000
Elapsed time is 0.639021 seconds.
```

10. The time taken is directly proportional to the number of samples in the dataset.
11. the figures 2, 1 & 3 are the confusion matrices obtained for the 7:3 training:testing ratio

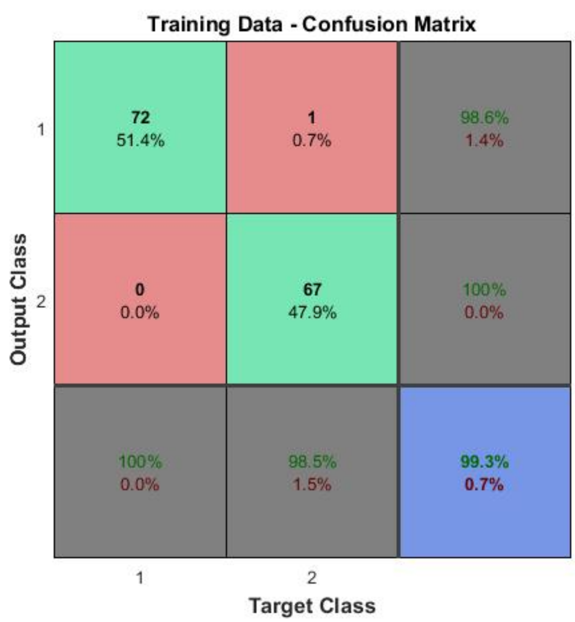

Figure 1: Confusion Matrix for Training Data

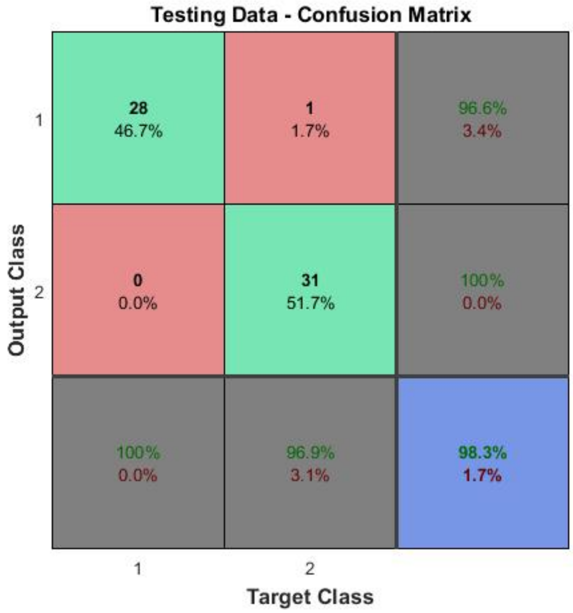

Figure 2: Confusion matrix for Testing Data

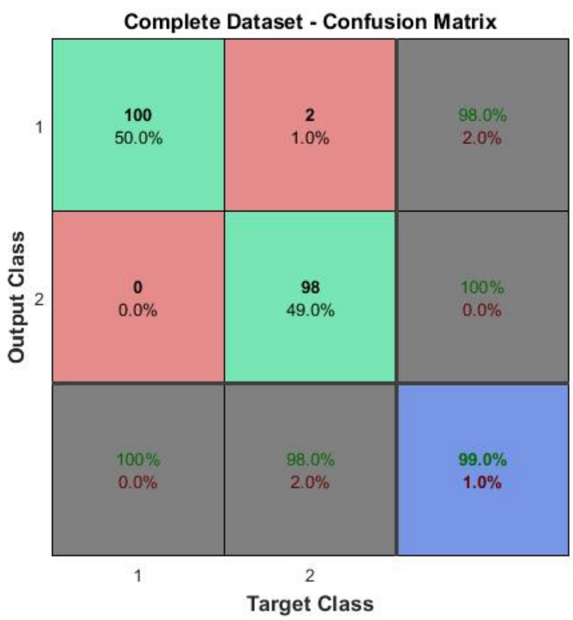

Figure 3: Confusion Matrix for Complete Dataset

2

1. Given the distribution to be a gaussian distribution , using the LDA approach for bayesian probability estimation we have the following formula.

2. This equation is also the discriminant function in our case. The greater than symbol explains that this is an example of Dichotomizer. Where we have 2 classes. If the $g_1(x) - g_2(x) > threshold$ we can tag the sample as class one else class two. Here $g_i(x)$ is discriminant function.

$$g(x):\ (\Sigma_0 + \Sigma_1)^{-1}(\mu_1 - \mu_0) \cdot x > \frac{1}{2}(-\mu_0^\top \Sigma_0^{-1} \mu_0 + \mu_1 \Sigma_1^{-1} \mu_1).$$

3. In our case the the threshold is assumed to be zero and the constant values in the discriminant function is moved to the right hand side to obtain the above equation

4. the LDA based approach assumes that the class covariances are equal to achieve the equation above, where the threshold is 0. So this causes a small drop in accuracy.

5. the output with the error rate and the execution times are available below

```
Calculating the Covariance for Class 0
Calculating the covariance matrix for Class 1
Calculating the covariance matrix for EntireDataSet

Elapsed time is 0.028948 seconds.
----------- Testing Data --------------
CM =

    28     0
     2    30

The Error Rate for Testing Data is 0.033333
Elapsed time is 0.860959 seconds.

-------------Training Data ----------------
CM =

    72     0
     5    63
The Error Rate for Training Data is 0.035714
Elapsed time is 0.553176 seconds.

--------- For Complete Dataset ----------
CM =

   100     0
     7    93
The Error Rate for Complete Data is 0.035000
Elapsed time is 0.563648 seconds.
```

6. The time required for the prediction is lesser than the previous set because we don't need to calculate the covariance of the entire sample dataset and the prior of the dataset. Also the time consumption is going to be directly proportional to the number of samples in the dataset. The LDA based approach only requires the mean and the class covariances from the training dataset.

7. The confusion matrices for the testing training and complete dataset are available in figures 4 5 6

8. we can see that the accuracy drops for the LDA based approach because we assume that the class covariances are almost equal so that costs us in accuracy a little. In our case the accuracy drops by approximately 3%

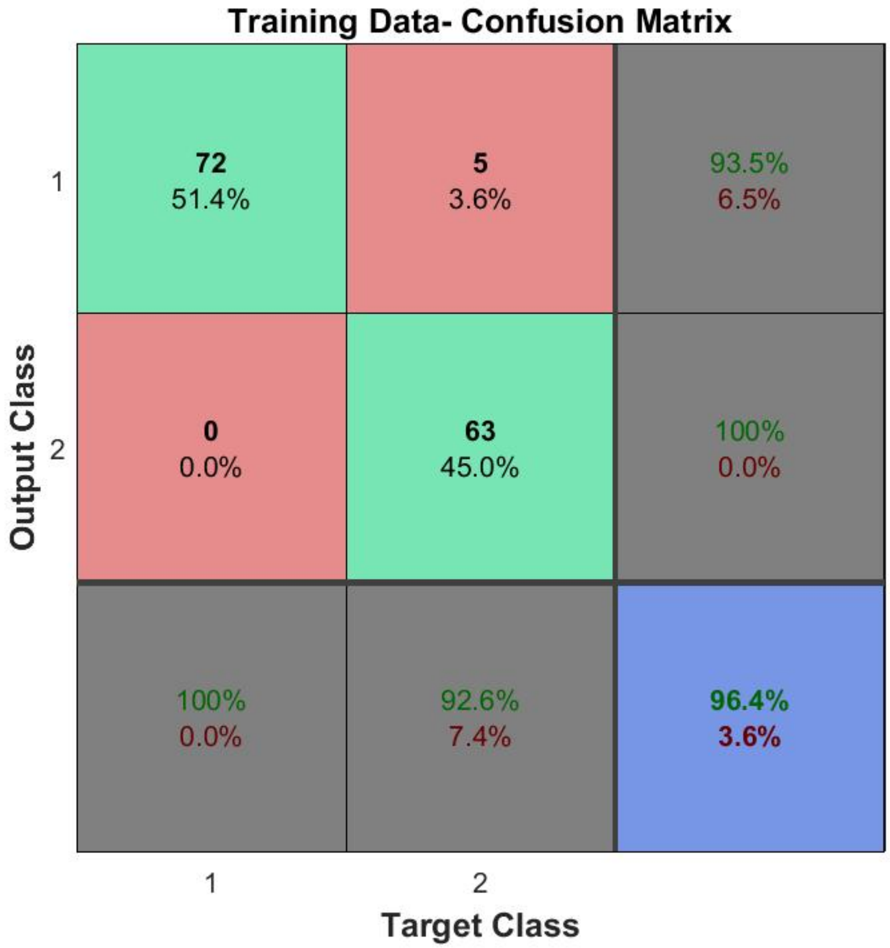

Figure 4: Confusion Matrix for Training Data : Set 2

9. false positives, false negatives, true positives, true negatives, accuracy and recall Explain all therse

3 Conclusion

1. we can see that the accuracy drops for the LDA based approach because we assume that the class covariances are almost equal so that costs us in accuracy a little. In our case the accuracy drops by approximately 3%
2. The samples which are classified into the wrong class are referred to as true negatives.
3. The accuracy for the normal naive bayes classifier is approximately 99% with one to two smaples being classified worng(True Negatives). The number of True Negatives are 1,1 and 2 for class 2 for training testing and complete dataset respectively. This can be checked in the confusionmatrix.
4. The accuracy for the LDA based approach gives us approximately 96% accuracy and the approximately two to 7 seven smaples being classified wrongly(True Negatives). The number of True Negatives are 5,2,7 for class 2 in the training, testing and the complete dataset respectively. This can also be checked in the confusion matrix.

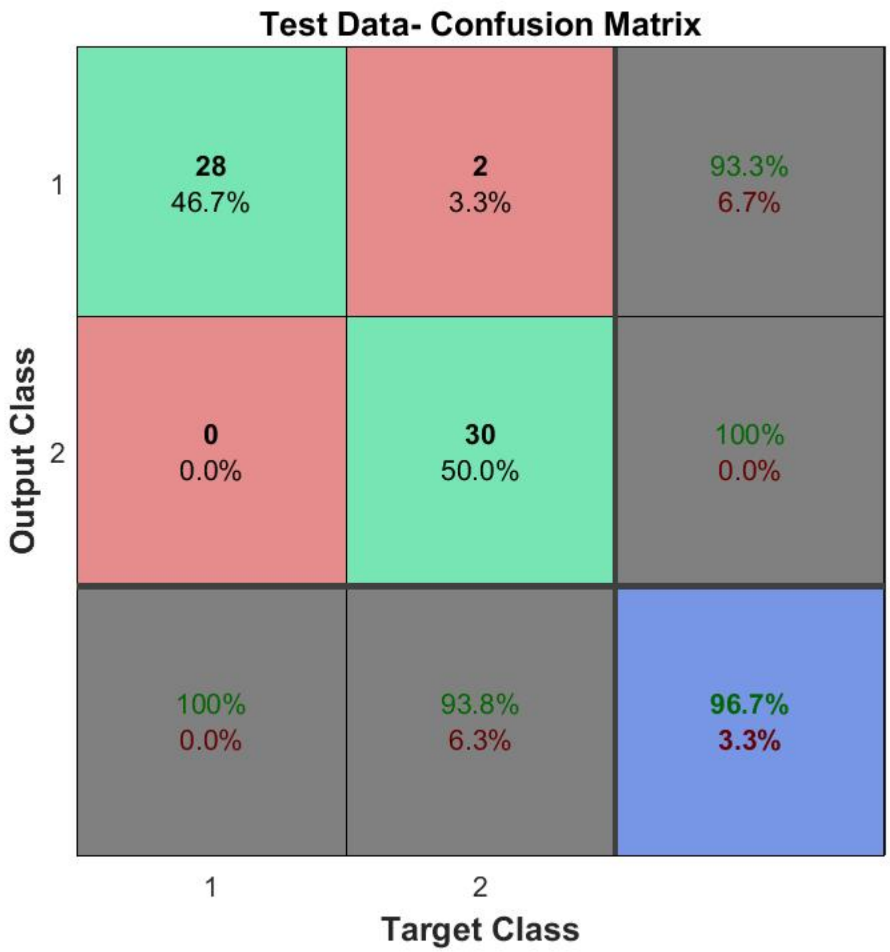

Figure 5: Confusion matrix for Testing Data : Set 2

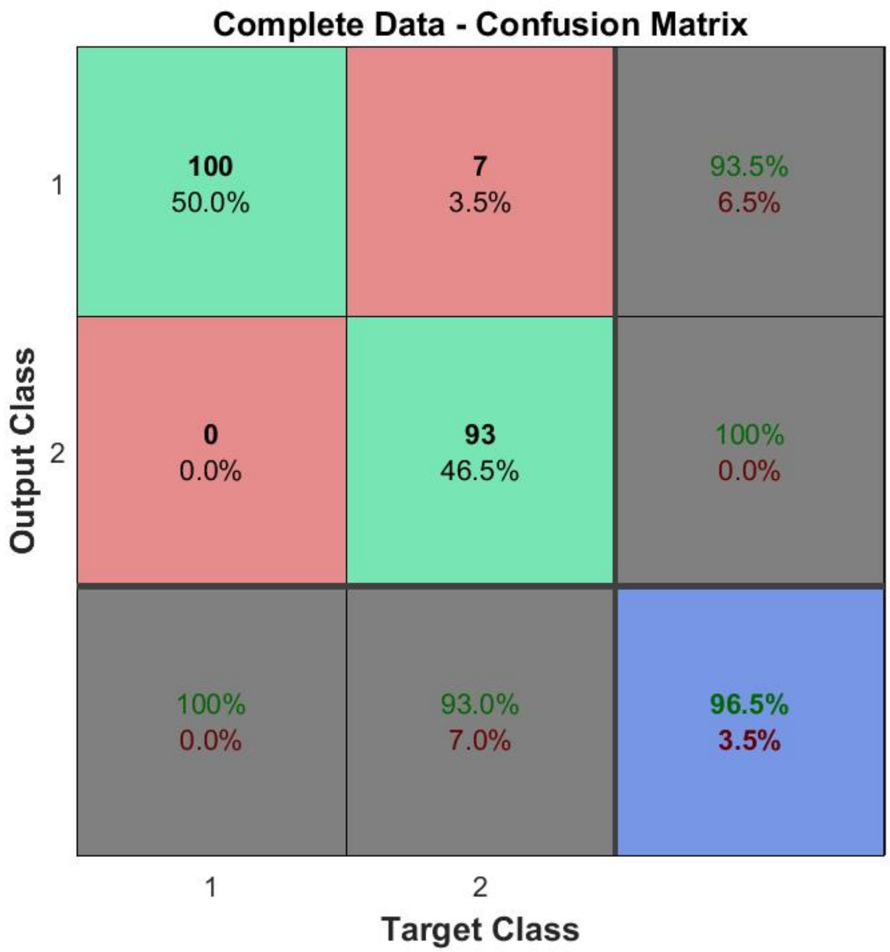

Figure 6: Confusion Matrix for Complete Dataset : Set 2