

ST790 HW5

Steven Xu

October 22, 2019

1.

Table 1: Training Error Scenario 1

	False Negative	False Positive	Classification Error
Bayes	0.19	0.25	0.22
Linear	0.2	0.26	0.23
LDA	0.2	0.26	0.23
Logistic	0.19	0.24	0.215

Table 2: Testing Error Scenario 1

	False Negative	False Positive	Classification Error
Bayes	0.228	0.252	0.24
Linear	0.242	0.248	0.245
LDA	0.242	0.248	0.245
Logistic	0.246	0.246	0.246

We can see that the LDA and linear regression are identical. Although logistic regression fits the training data better, Bayes classifier performs best in the testing data and is thus optimal.

2.

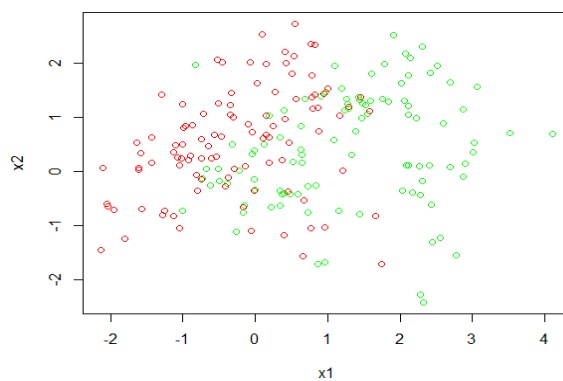


Figure 1: Scatter plot for scenario 2

3.

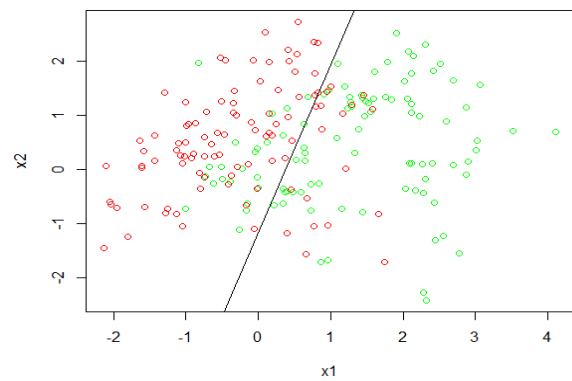


Figure 2: Linear classifier scenario 2

Table 3: Training Error Scenario 2

	False Negative	False Positive	Classification Error
Linear	0.18	0.23	0.205
LDA	0.18	0.23	0.205
QDA	0.17	0.23	0.20

Table 4: Testing Error Scenario 2

	False Negative	False Positive	Classification Error
Linear	0.132	0.314	0.223
LDA	0.132	0.314	0.223
QDA	0.118	0.32	0.219

3.

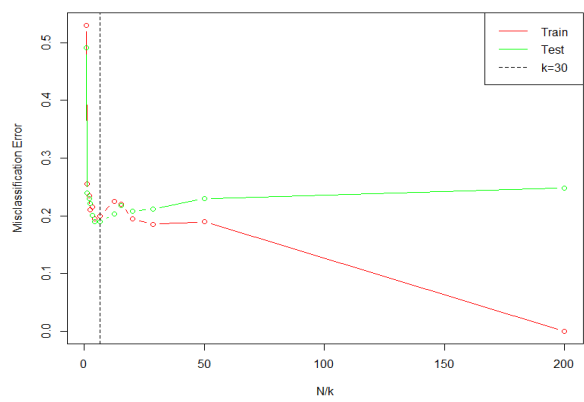


Figure 3: kNN Scenario 2

Since testing error is a better estimate of the true error, we should choose $k=30$ to minimize the testing error.

5

Table 5: kNN ZIP data

k	Training Error	Testing Error
1	0.000	0.022
3	0.004	0.024
5	0.005	0.024
7	0.006	0.027
15	0.010	0.032

The training error and testing error for LDA is 0.004 and 0.033 respectively. QDA cannot be fitted since the number of constant columns in each group is too large, making the group covariance matrices singular.