Name: Alvee Chowdhury

Answer 1

	train error rate	test error rate
Classifier 1	0.468	0.5160
Classifier 2	0.216	0.2240
Classifier 3	0.096	0.1270
Classifier 4	0.050	0.0665

Classifier 4 perform the best among all these 4 classifiers since it has the least test error rate. Classifier 1 performs the worst forllowed by classifier 2.

Answer:2

Summary of the test and error rates of the models

Model	Train Error Rate	Test Error Rate
model 1	0.334	0.3305
model 2	0.334	0.3305
model 3	0.320	0.3500
model 4	0.080	0.1110

Here, We notice that Model 1 and Model 2 have identical train and test error rates. This may because of having similar intercept. We notice that the test error rates reduces after model 3. Model 4 has the the least test error rates. From Model 2 to Model 3 we notice an increase in the test error rates. But then it decreases as we move to model 4 from model 3. As the model complexity increases, bias should increase. For, train error rate, we see a steady decrease from Model 1 to Model 4 (as complexity increases).

Answer: 3

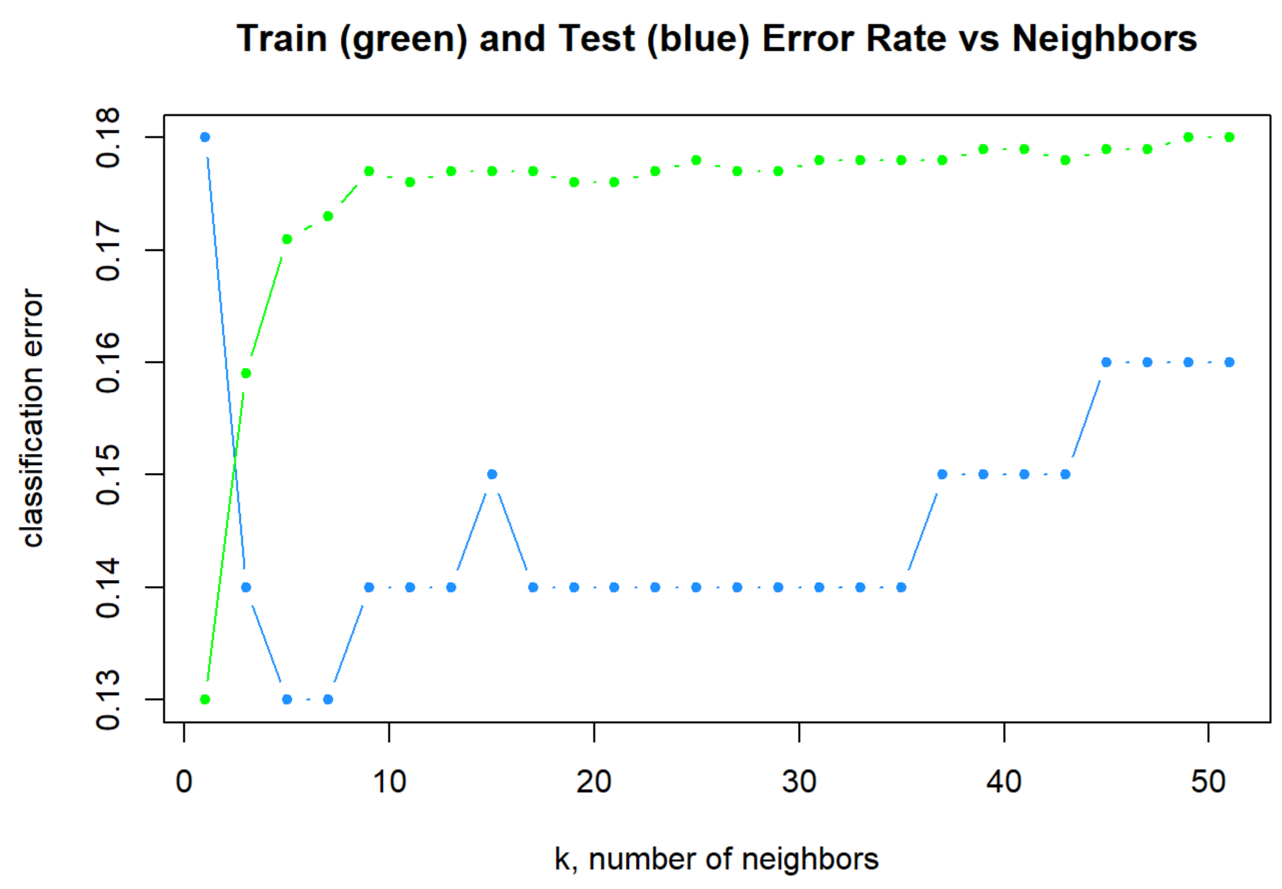
Summary of the Squared Bias, MSE and Variance of the logistic models

Logistic Regression Model	Mean Squared Error	Bias Squared	Variance
Intercept Only	0.05821	0.04889	0.00932
Additive	0.01001	0.00002	0.00999
Second Order	0.02575	0.00043	0.02533

The Sqaured should decrease as model complexity increase. However, here we see the opposite happening. As we move from the additive model to the second order model, the bias increases. The variance increases as the model complexity increases. Since Mean squared error is equal to the sum of Bias squared and variance, it initially decreases as we move from the intercept model to the additive model but then increases as we move from the additive model to the second order model. This is the Bias varinace tradeoff. The only problem here is that the sqaured bias should not increase as we move from the additive to the second order model. There may be some error in fitting the model.

Answer 4

The train and test error vs k on a single plot is as follows:



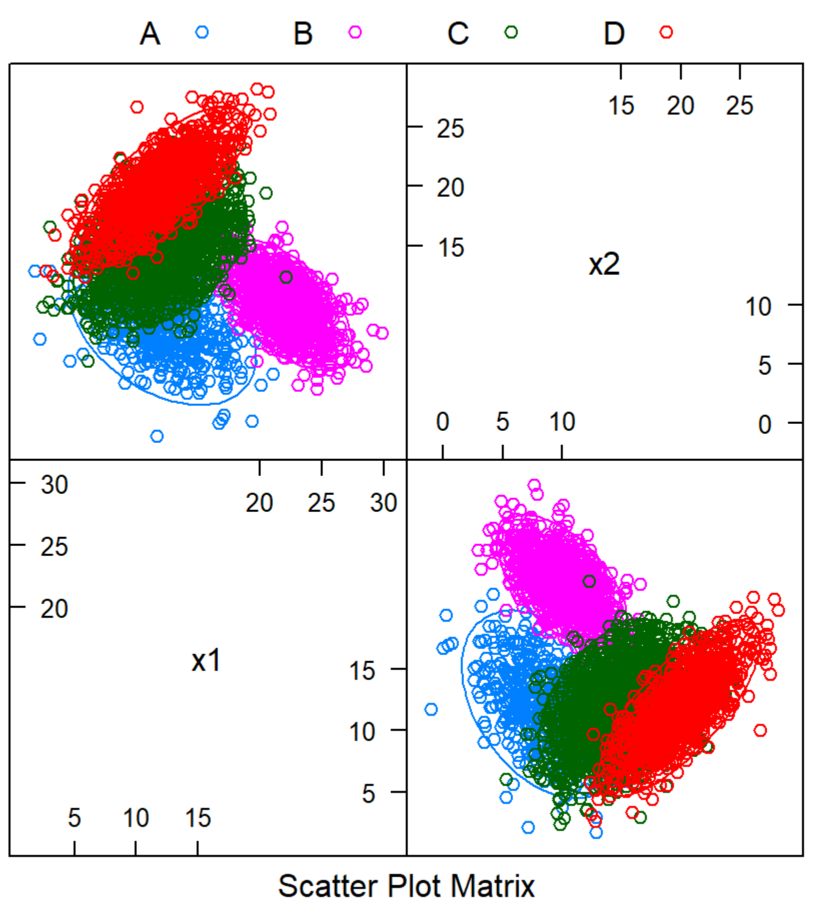
Summary of the Test Accuracy, Test Sensitivity and Test Specificity of the logistic models

cut-off for predicted probability	Accuracy	Sensitivity	Specificity
0.1	0.8017058	0.9651163	0.7070707
0.5	0.8912580	0.8255814	0.9292929
0.9	0.8017058	0.6104651	0.9966330

Sensitivity decreases as the cut-off is increased. Conversely, specificity increases as the cut-off increases. This is useful if we are more interested in a particular error, instead of giving them equal weight. The 0.5 cut-off has the best accuracy.

Answer:5

The pairs plot with ellipses for the training data is as follows:



Summary of the Test and Train error rates of the models

Model	Train Error Rate	Test Error Rate
Additive Logistic Model	1.0000000	1.00000
LDA (with Priors estimated from data)	0.1620000	0.19825
LDA with Flat Prior	0.1906667	0.16875
QDA (with Priors estimated from data)	0.1477778	0.16925
QDA with Flat Prior	0.1791111	0.14000
Naive Bayes (with Priors estimated from data)	0.1733333	0.20000

From the above table, Additive Logistic regression performs the wrost followed by Naive Bayes. Since QDA (with flat priors) has the least test error rate, we can conclude that it performs the best.