

EE5027 Programming Assignment Semester 1 2019/2020

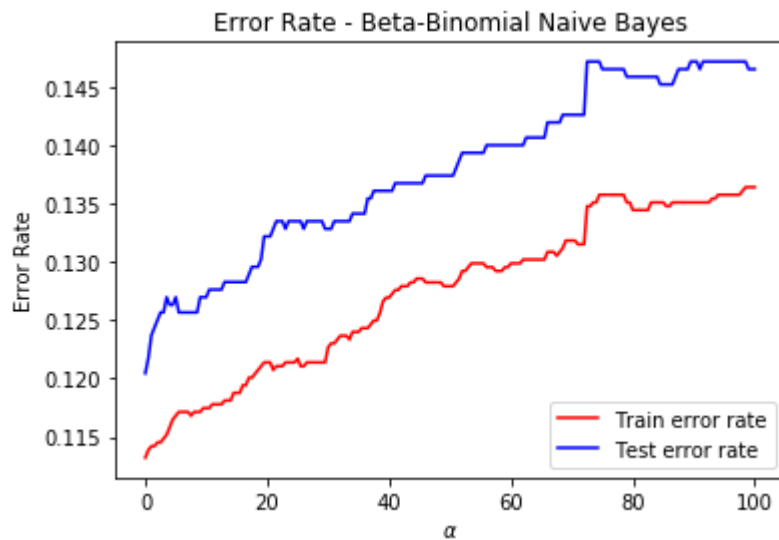
Name : Bain Zhen Jian Addison

Matric Number : A0072374E

Date : 27 September 2019

Q1. Beta-binomial Naive Bayes

(a) Plot of training and test error rates versus α



(b) As α increases, the training and test error rates also increase. Once α is more than 75, the increase in error rates appear to taper off. It is also noteworthy that the test errors are always slightly higher than the training errors.

(c)

α	Training Error Rate	Test Error Rate
1	0.1142	0.1237
10	0.1175	0.1270
100	0.1364	0.1465

(* Rounded to 4 decimal places)

Q2. Gaussian Naive Bayes

- Training Error Rate = 0.1680
- Testing Error Rate = 0.1634

(* Rounded to 4 decimal places)

Q3. Logistic Regression

(a) Plot of training and test error rates versus λ



(b) In general, as λ increases, the training and test error rates both increase. This is because more weight is given to the regularization, which means the model is not as “free” to fit the training data. When λ is large, it results in an increase in training error rate due to underfitting.

It was also noteworthy that when λ increased initially, the test error rate actually decreased as the model was more able to generalise for test data (marginally better). However, as λ continued to grow, the test error rate also increased due to underfitting.

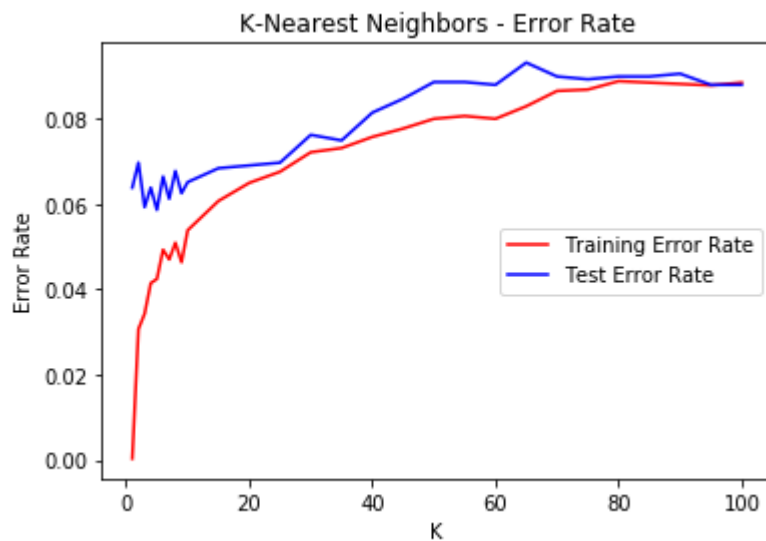
(c)

λ	Training Error Rate	Test Error Rate
1	0.0535	0.0605
10	0.0538	0.0599
100	0.0639	0.0664

(* Rounded to 4 decimal places)

Q4. K-Nearest Neighbors

(a) Plot of training and test error rates versus K



(b) As K increases, the training and test error rates will both increase, but will become closer. When K is too small, the model overfits to the training set.

(c)

K	Training Error Rate	Test Error Rate
1	0.0003	0.0638
10	0.0538	0.0651
100	0.0884	0.0879

Q5. Survey

I spent about 60 hours in total on this assignment. The bulk of the time was spent on Q1 and Q4, due to the computational complexity of the algorithm implemented. This made it challenging to debug as much time was spent to wait for the output to be generated to verify my code.