Contents

Q1. Beta-binomial Naive Bayes:

- Plots of training and test error rates versus α: (α = {0, 0.5, 1, 1.5, 2, · · · , 100})

- What do you observe about the training and test errors as α change?

From the above plot, I can observe that the training and test error both are increasing with the alpha value. However, the training error is more than the test error after a certain alpha value or the beta binomial parameter (a, b) (In our case after nearly after alpha=13). Further, it can be concluded that the test error is more than train error since the sample size of test is less than the training sample.

Accuracy: 86-88%

- Training and testing error rates for α = 1, 10 and 100:

alpha	Train error in %	Test error in %
1	11.549	11.328
10	12.006	11.914
100	14.192	13.671

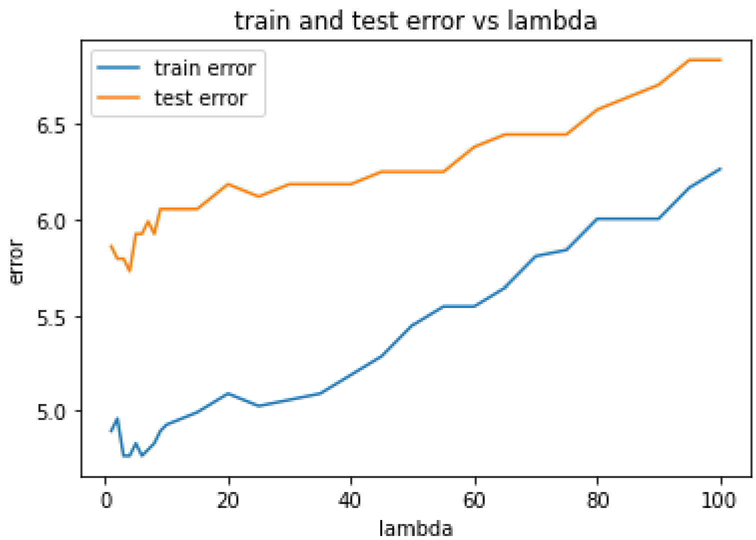
Q2. Gaussian Naive Bayes:

- Training and testing error rates for the log-transformed data.

Train error in %	16.57422
Test Error in %	16.01562

Q3. Logistic regression:

- Plots of training and test error rates versus λ: λ = {1, 2, · · · , 9, 10, 15, 20, · · · , 95, 100}



- What do you observe about the training and test errors as λ change?

From the observation of the train and test error for different lambda, I can state that the train and test error is rising as the lambda value is changing. Although the test error is always more than the train error. The error is constrained within 7% which means the accuracy is mostly in between 93 to 95%.

- Training and testing error rates for λ = 1, 10 and 100.

lambda	Train error in %	Test error in %
1	4.8939	5.8593
10	4.9265	6.0546
100	6.2642	6.8359

Q4. K-Nearest Neighbors

- Plots of training and test error rates versus K: k= {1, 2, · · · , 9, 10, 15, 20, · · · , 95, 100}

- What do you observe about the training and test errors as K change?

From the observation, it can be observed that the errors have an increasing nature then it's decreased for both train and test error. Later, the error was again increased for all the K values. However, for some values of K, train and test error is same.

Accuracy: 88-94%

- Training and testing error rates for K = 1, 10 and 100.

K-Values	Train error in %	Test error in %
1	6.4274	6.9661
10	6.0685	6.7057
100	9.2006	10.0911

Survey:

Estimated time:

Reading: 3 hours per week

Coding: 2 hours per day for 1 week

Report: 2 hours one day

The Knowledge I gathered from Professor along with TA was more than sufficient to understand the topics as well as the goal of the assignment.