

ECE 1395 Problem Set 4

Bryan Hess: 4259226

ps4-1-b.)

Size of X:

ans =

1001

501

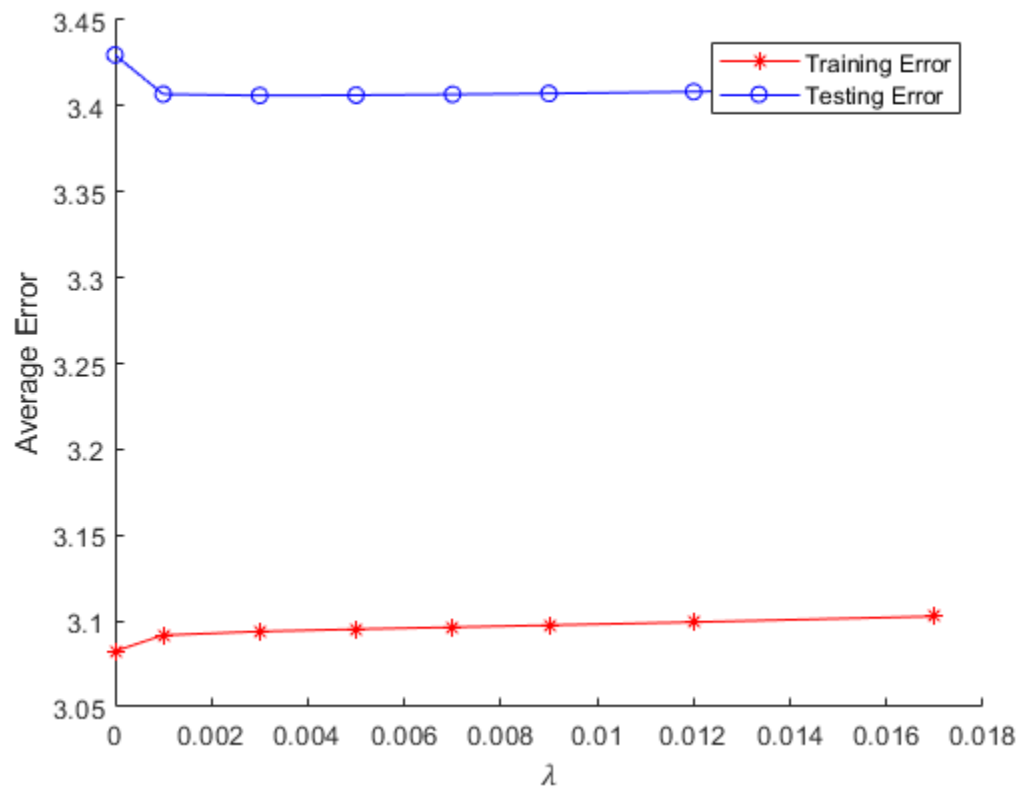
Size of y:

ans =

1001

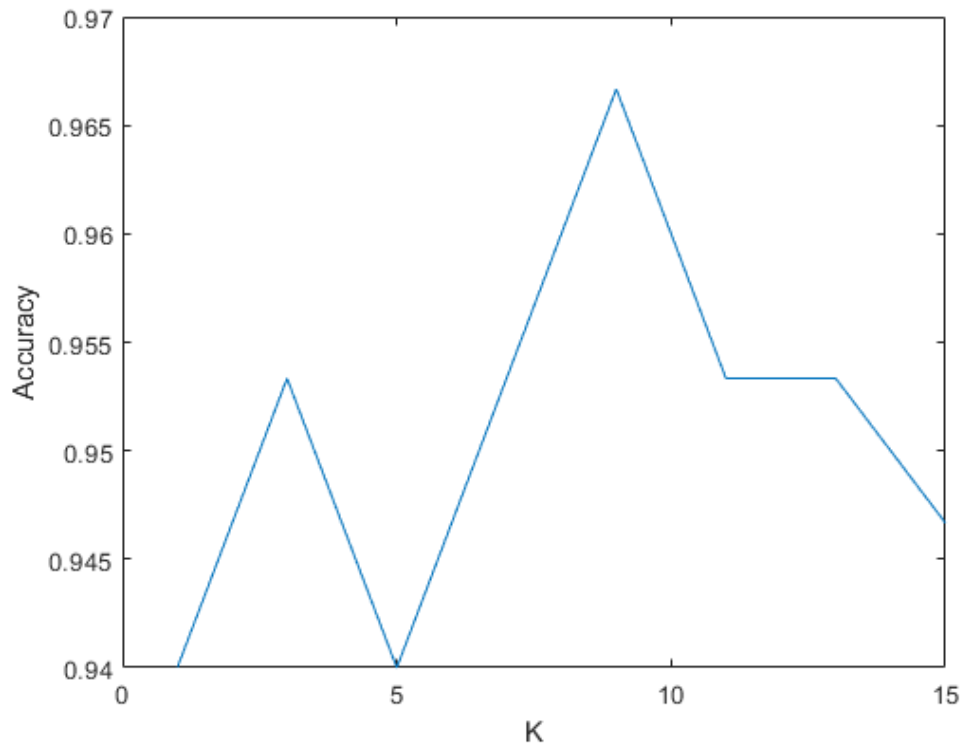
1

ps4-1-c.)



The lamda (λ) value should be used as the distance between the testing and the training error is minimized. For this data set that value would be 0.001. After this, the two errors diverge. While training error does go down, this only means we are over-relying on the specific testing data.

ps4-2-a.)



For this dataset I recommend using a K of 9, given it has the highest yielded accuracy. This value is not robust to other problems as KNN is very data specific. Some datasets might need a larger or smaller K depending on how the data is grouped.

ps4-2-b.)

ans =

6×2 [table](#)

| Sigma | Accuracy |
|--------------|-----------------|
| 0.01 | 0.52 |
| 0.1 | 0.92 |
| 0.5 | 0.92 |
| 1 | 0.92 |
| 3 | 0.72 |
| 5 | 0.72 |

The scaling in the distance metric affects the regional shapes. Too small a scaling value and the data is too broad, but too high and it's too specific. A sigma that adjusts the regions correctly should be chosen. For too small of a sigma, mathematically the weight would be 0 as it would be insignificant when taken as an exponential.