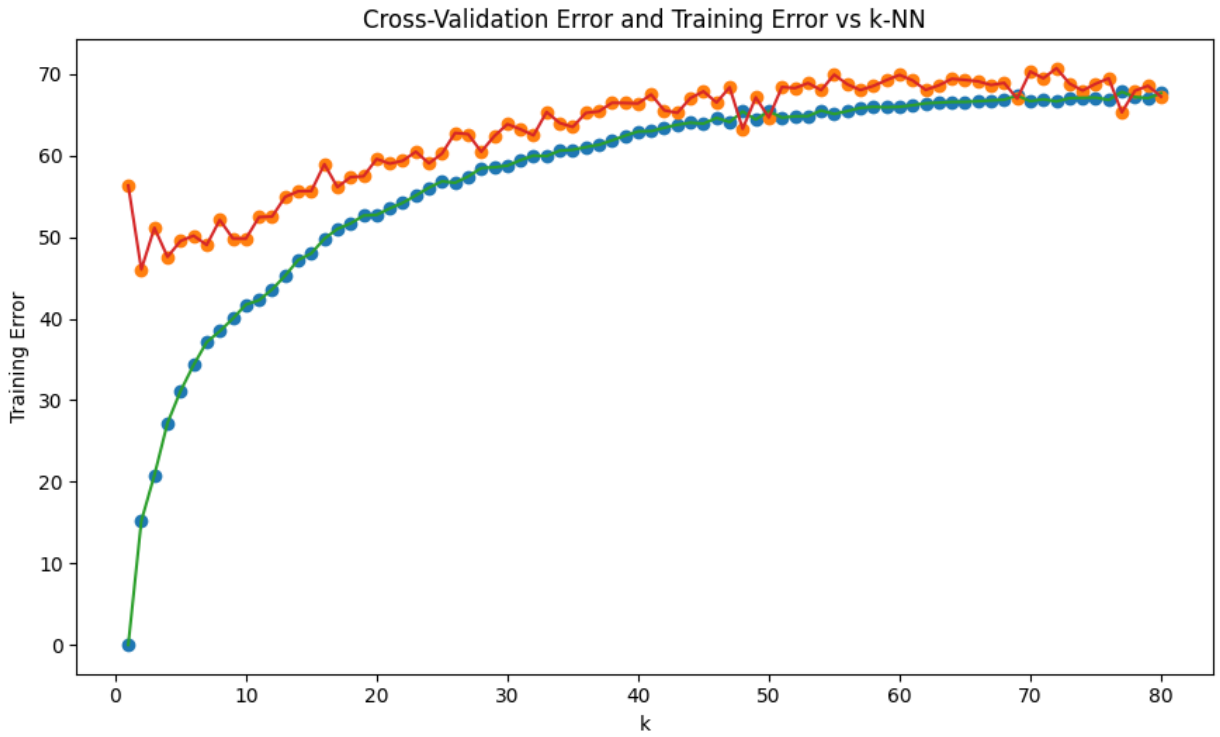
LAB 3 4SL4
PARTHAV PATEL
400251342
Patep62

Random Seed/Random state used: 1342

5-Fold Cross-Validation error and Training Error vs k from 1 to 80.
The data was initially split 80/20 into training and test sets respectively. The cross-validation was then performed on the resulting training set by first shuffling examples then splitting data into 5 folds.



We can see that when k = 1, our training error is 0. This is expected as the nearest neighbor will always be the same as the target, giving us an exact prediction. However, the validation error is high at this point, due to overfitting. The model is adapting to the noise in the training sets and throwing off the predictions.

However, when k becomes too large (greater than 10), the validation error and training error begin to rise. This depicts underfitting, as we are now considering too many neighbors.

According to the results, the lowest recorded mean squared error is 46.05636875. This value was produced with k = 2. Using this as our optimal k, we find a test error of 39.94985294117647 However, any k value in the range of 2 to 10 will provide a similarly low cross-validation error, and a good trade-off between variance and bias.