HOMEWORK 1

>>Lane Schultz<<
>>907 929 6373<<

Problem 3

One single value of k does not do best for all training set sizes. I will refer to my learning_curve.pdf plot. For 20% of the training data used, the best k value is 4. However, that is no longer the case for 30% of training data used. Hence, different k values work better for different training set sizes.

Problem 4

One value of k in roc_curve.pdf does not seem to dominate over others. The best value of k is highly dependant on the application. For instance, we may want the false positive rate to be high while keeping the true positive rate high in exploring material properties for a composition space. This prevents missing interesting compounds at the cost of more experiments yielding undesirable results.