Project 1 Write Up

WU1

The boolean statement inside the mean function returns 1 when prediction matches training label and 0 otherwise. It essentially tallies up how many correct predictions the classifier made. Taking the mean then normalizes the tally, i.e. generates a number between 0 and 1 that represents classification accuracy. For example, if m out of n training predictions are correct, then that's the sum of m 1's and n-m 0's over a total of n, i.e. m/n

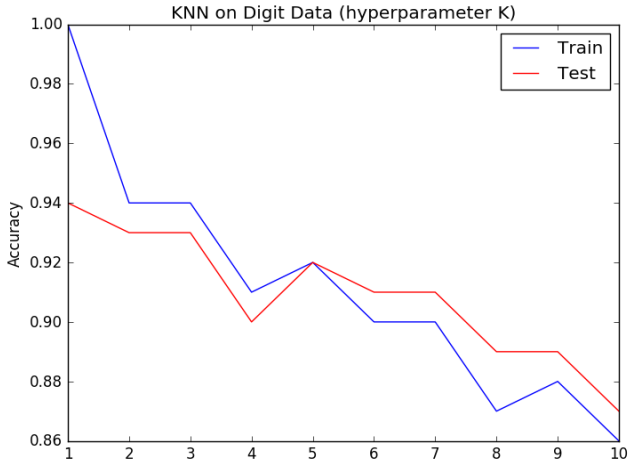
WU2

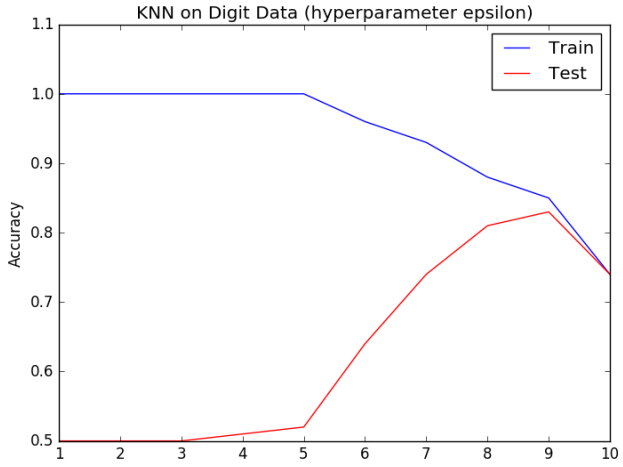
As the number of training examples increases, the probability of finding a decision tree that can correctly classify a set percentage of the examples decreases. 8/10 examples correct vs 800/1000 examples correct is a huge difference in the raw number of examples to correctly classify. Naturally, training accuracy will tend to decrease as a function of number of examples. Test accuracy is not monotonically increasing because for smaller sample sizes, the training process might not learn enough about the overall data generating distribution, so while the training accuracy is high, the learned tree could be very off about data it hasn't seen before. This could explain the jagged part of the curve for test accuracy.

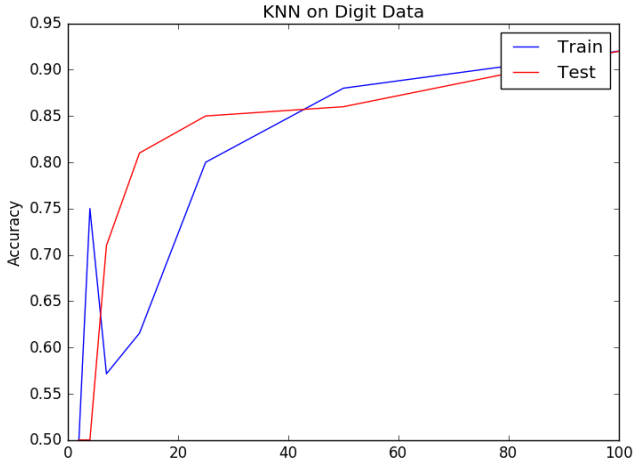
WU3

Training accuracy monotonically increasing is guaranteed to happen, while the hill in test accuracy is expected to happen. As max depth increases, the algorithm is able to use more features to more accurately fit the data. However, it might (not always) start to overfit at some point and therefore be less accurate on test data, resulting in a hill shaped curve.

WU4



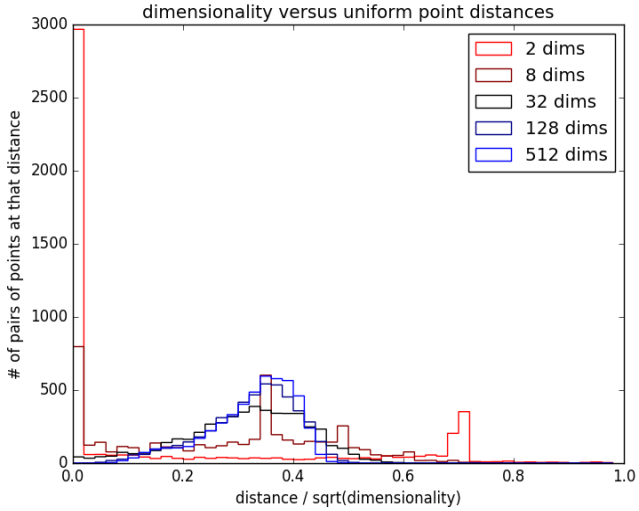




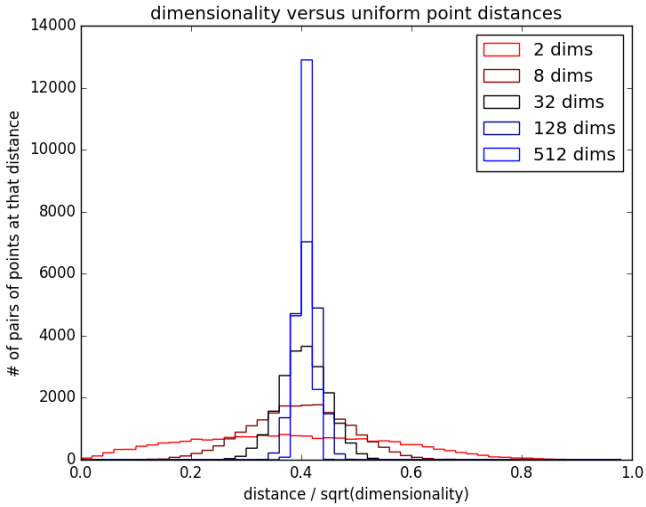
There appears to be underfitting for small epsilon, like for epsilon<5 it's trivially easy to get 100% training accuracy since it's basically the same as using k=1

As for the k hyperparameter, the general decreasing trend in both accuracies doesn't provide evidence for overfitting or underfitting. There's a sweet spot at k=5.

WU5



Digit data

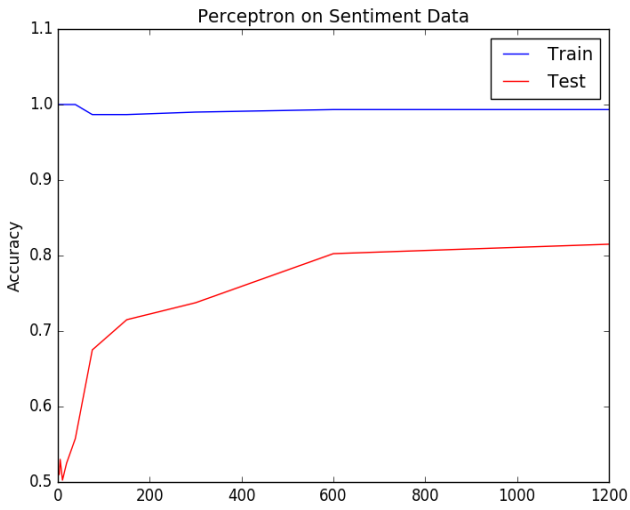


Random data

Random data shows a decidedly normal distribution whose variance decreases as the number of dimensions increase. On the other hand, the digit data appears to have skewed and multimodal distributions depending on the number of dimensions; fewer dimensions create multimodal distributions, while more dimensions give left skewed distributions.

WU6

a) Learning curve



b) Number of epochs vs accuracy