Independent Research Homework 1

The purpose of this homework is to test your ability to (1) manipulate data, (2) run experiments, and (3) visualize the effects of dimensionality on accuracy.

1. Iris is a great dataset because it is simple and the 3 classes are mostly distinct. This is not indicative of most natural data sets which tend to be messy and tend to have more overlap between classes. For this problem, you are going to make Iris messy.
   1. Add a new feature to the current data set. The feature values should be random and will range between 0 and 1.
   2. Split your new dataset into a training set and test set.
   3. use the nearest neighbor classifier on your new data.
   4. Record the accuracy and number of features.
   5. Repeat steps 1a-1e until you reach 100 features.
   6. Graph accuracy vs number of features in a line graph. Try to create the line graph using python. Use excel to graph the results if you need to.
2. Repeat the exercise above using the random forest classifier instead of the nearest neighbor classifier.