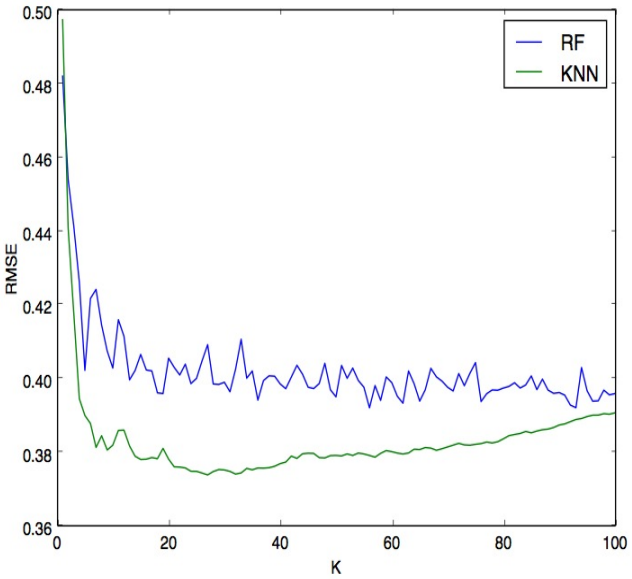
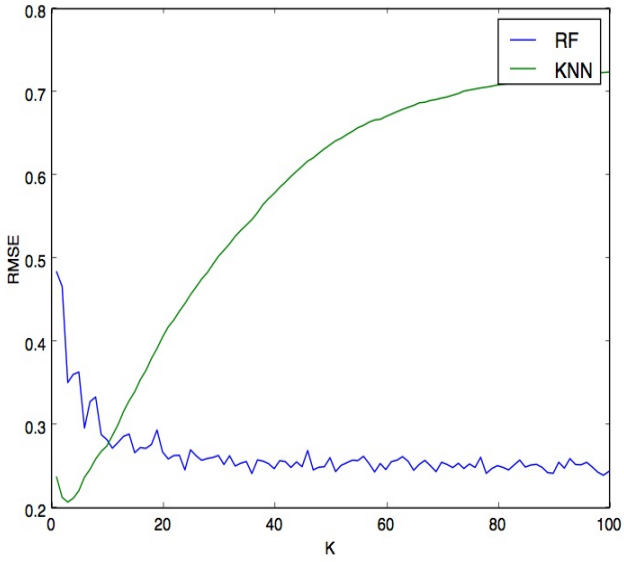
Machine Learning for Trading
Project 2 – Random Forest Learner
Sriram Madapusi Vasudevan
GT ID: 902916994

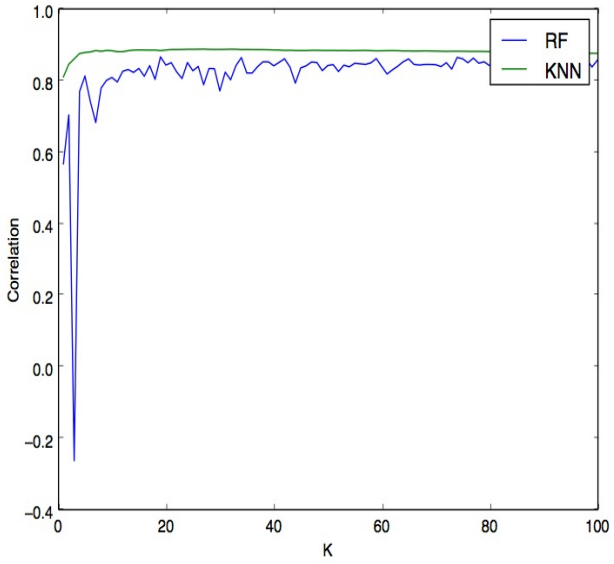
Graphs:

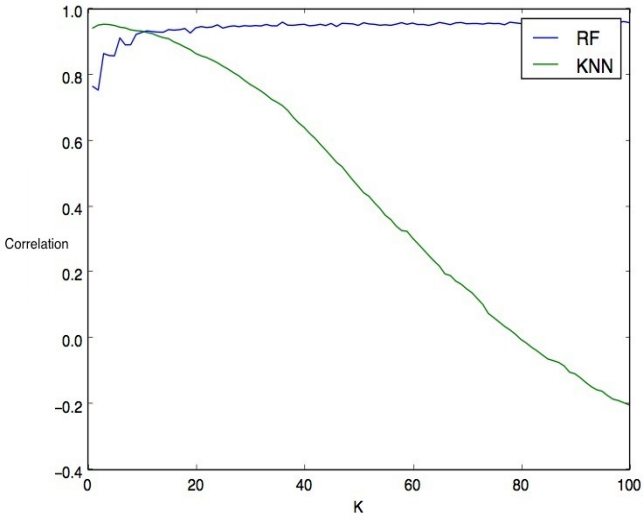
RMSE vs K : dataclassification (left) ripple(right)





Correlation vs K : dataclassification(left) ripple (right)





Did you see improved performance using more trees in one data set or the other (or both)?

Both datasets showed improved performance on increasing the number of trees. RMSE error went down, while correlation increased.

If there was a difference, explain why you think the improvement is better for one data set.

It is to be noted here that the ripple dataset provided the most improvement because of having far more unique values of Y to train on, but in the case of the classification data set the the Y values were just either 1,-1 or 0. So operation such as taking mean of values obtained from multiple trees suffer from loss in quality of predicted output.

Now that you have compared KNN, linear regression and Random Forests, which approach do you think is best, and why?

The Random forests as the name suggests involves random creation of trees. So the results obtained from it are usually not repeatable. But rather they form a trend on how the predicted results could be. Usually with increase in number of trees, better results are formed due to increase in the number ways of a tree can be built. This is in turn leads to more values, which on taking mean would give a much better and close result to what is expected.

Linear Regression on the other hand will suffer from poor predictive performance if the relationships that are portrayed in the dataset are non linear. So Linear regression is best only if the relationships modeled are linear.

K Nearest Neighbors approach again only works when the dataset is densely populated which will allow more accuracy while considering more number of neighbors.

So each of these approaches is great under some condition of the training set, But as seen in the graphs for this dataset, Random forest work out to be better.