**Ender F.Morgul Assignment 7**

I used cv2 package in python which offers several classification methods. I chose four methods for the 10-fold cross-validation:

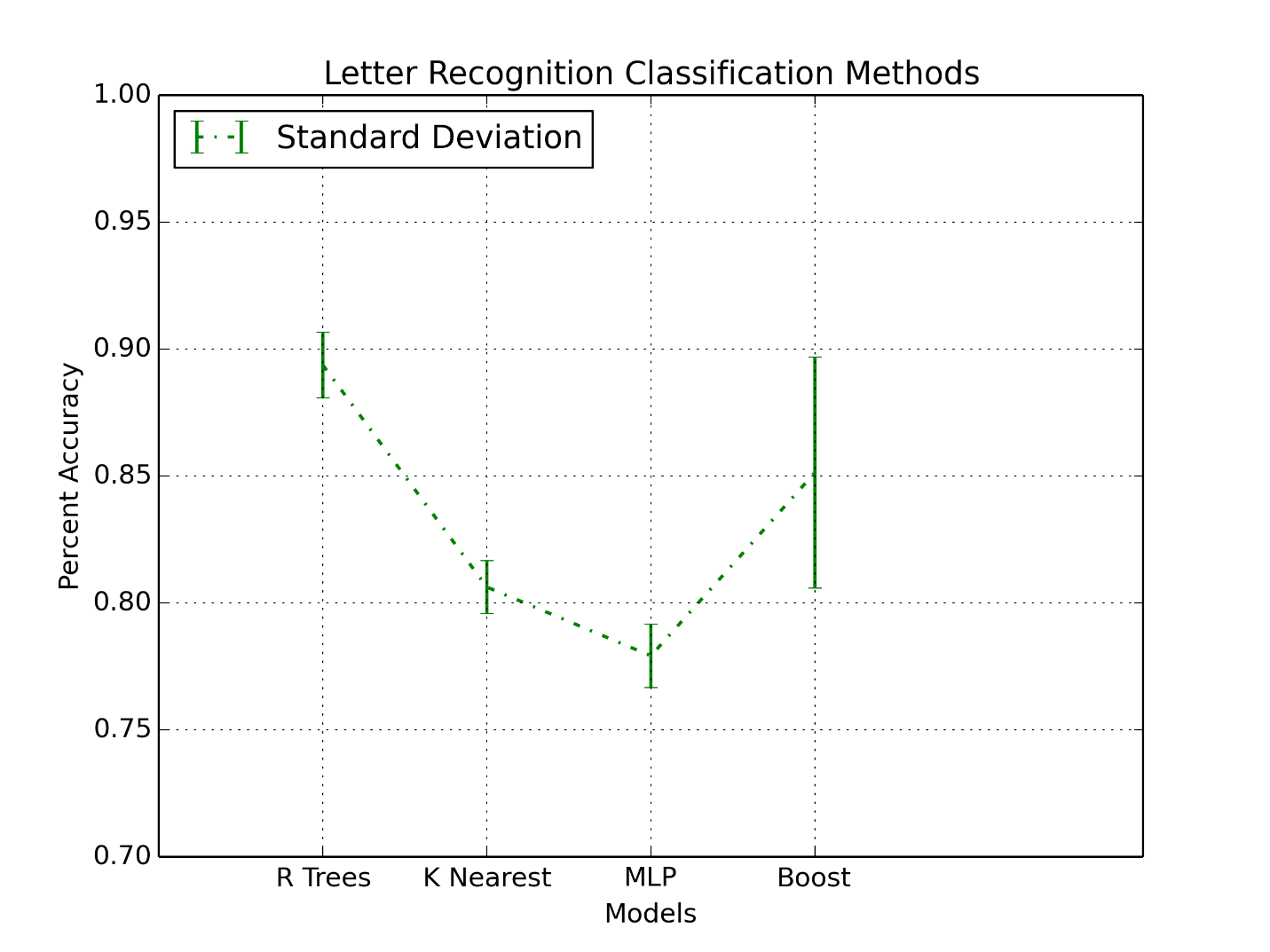
-R Tree

-K-Nearest Neighbor

-Boosting

-Multilayer Perceptron-Neural Networks

10 fold cv results using the first 16k sample are written in separate csv files for each method. Measured accuracies are depicted in below graph.



According to these results R Trees method has the highest accuracy with an average of approximately 89% for the 10 training folds. Then comes Boosting method and followed by K-Nearest and MLP respectively. The execution times are also recorded. According to the measurements K-nearest is by far the fastest method. It is followed by R Trees, Boosting and MLP respectively. Therefore these results are in favor of R Trees method for the final model.

Next I trained the whole 16k sample using R Trees method and tested the model on the remaining 4k sample. I obtained the following accuracies:

**Training sample accuracy rate: 83.29%; Test sample accuracy rate: 79.33%**

(Code: Full\_model\_RTrees.py)

These results are not satisfactory and lower than the cross validation estimates, so I tried the other models for the full dataset. K-nearest method gave the following results:

**Training sample accuracy rate: 96.09%; Test sample accuracy rate: 94.18%**

(Code: Full\_model\_KNearest.py)

The results are much better than the cv average. This might be the effect of sample size on the performance of the method.

For MLP, I obtained:

**Training sample accuracy rate: 95.36%; Test sample accuracy rate: 93.7%**

(Code: Full\_model\_MLP.py)

Finally for Boosting:

**Training sample accuracy rate: 79.09%; Test sample accuracy rate: 76.9%**

(Code: Full\_model\_Boost.py)

Therefore full model results show that **K-nearest neighbor method** gives the **best** accuracy.

I subset the Letter ‘A’ from the full dataset and measured the accuracy and F 1 score. Since I converted the letters to digits while reading the dataset, letter A corresponds to zero.

I used two different codes to calculate the F 1 score, both gave a value of 0.667.

The accuracy for Letter A is 94.18%.

All the codes and calculations are included in Full\_model\_Knearest.py.