CS 534 Machine Learning Homework 4
Ruixiang Qi

Readme:

Two python files are submitted.

Among the two files, Adaboost.py solves problem 1a, 1b , randomForest.py solves 2a,2b.

Since I used the package sklearn to generate the weak learners, you need make sure that the package sklearn has been installed in your python environment. What's more, numpy is also required. Finally, since I used matplotlib to plot graphs, you also need to install matplotlib.

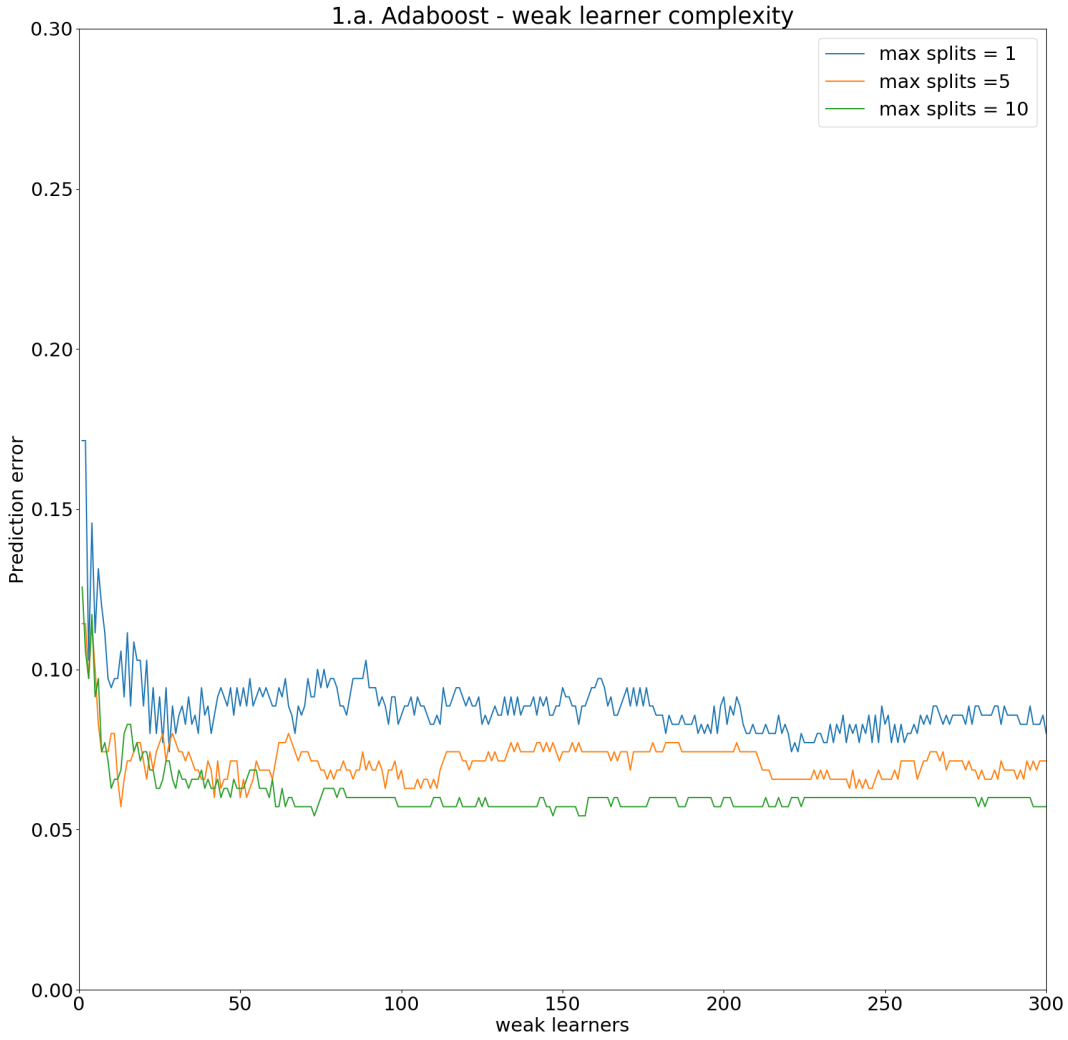
In conclusion, the required packages are : sklearn, numpy , matplotlib

After installing the above packages in your python environment, you could run the code in terminal by typing:

python Adaboost.py
python randomForest.py

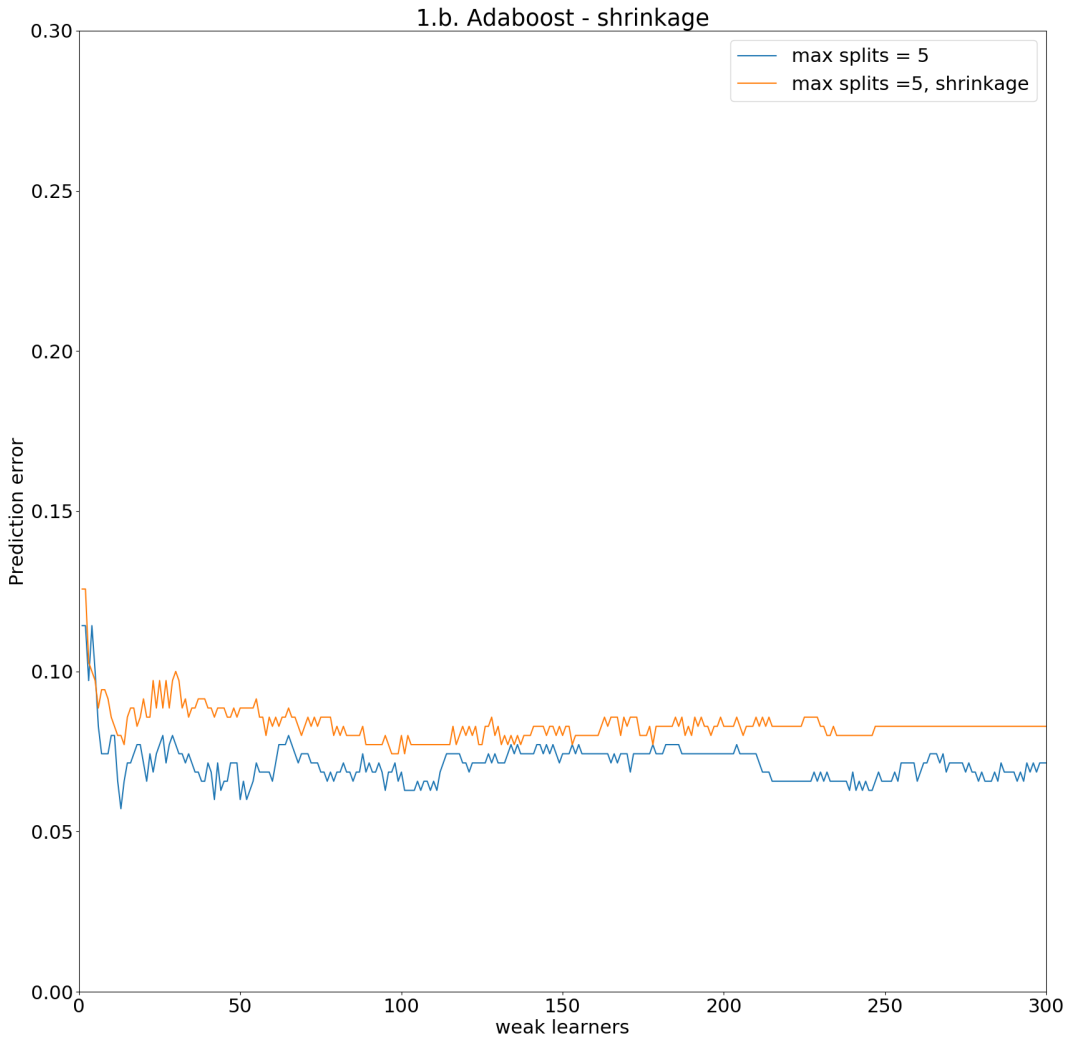
Result and analysis are on the following pages

1.a.



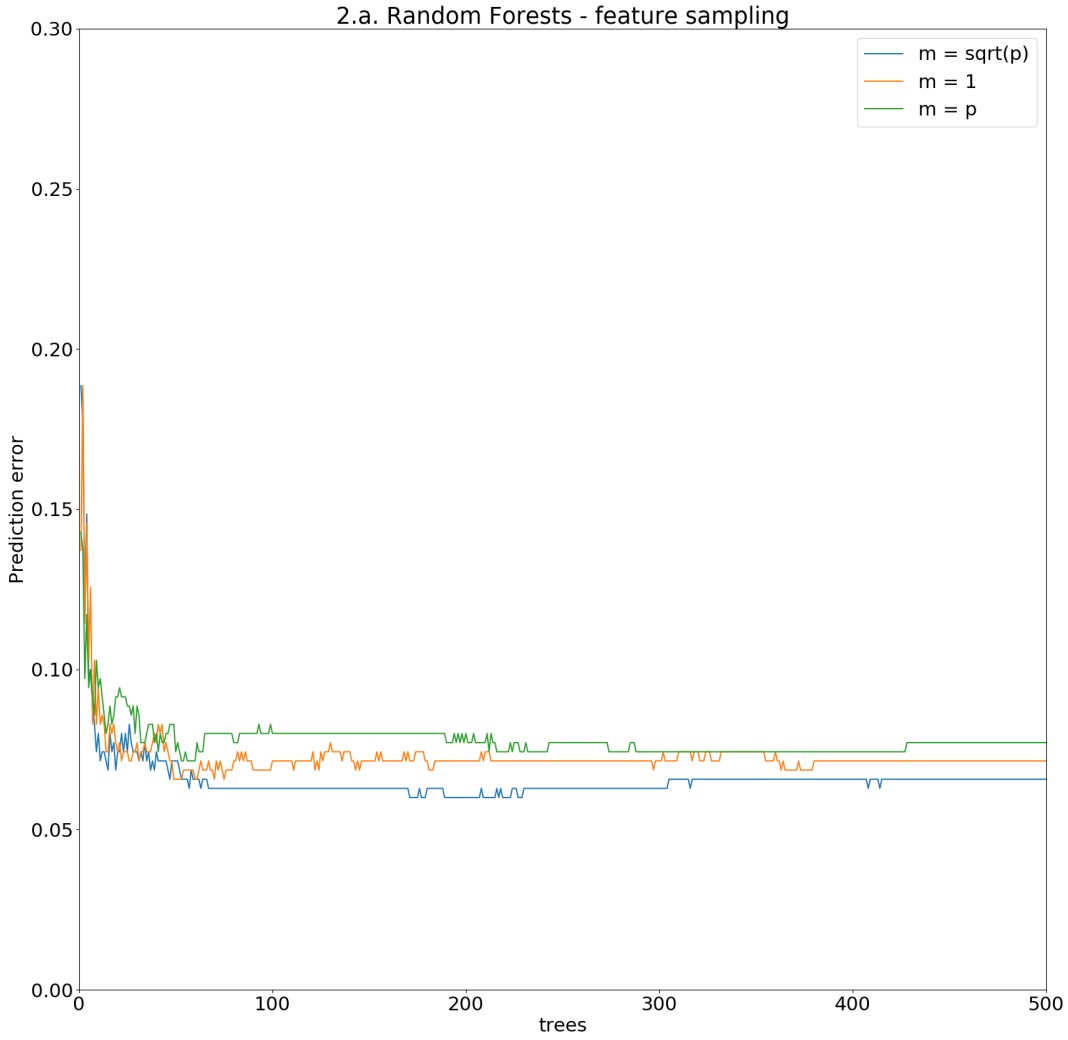
As we can see, the model's accuracy increases with number of tree nodes. The accuracy also increases with number of weak learners

1.b.



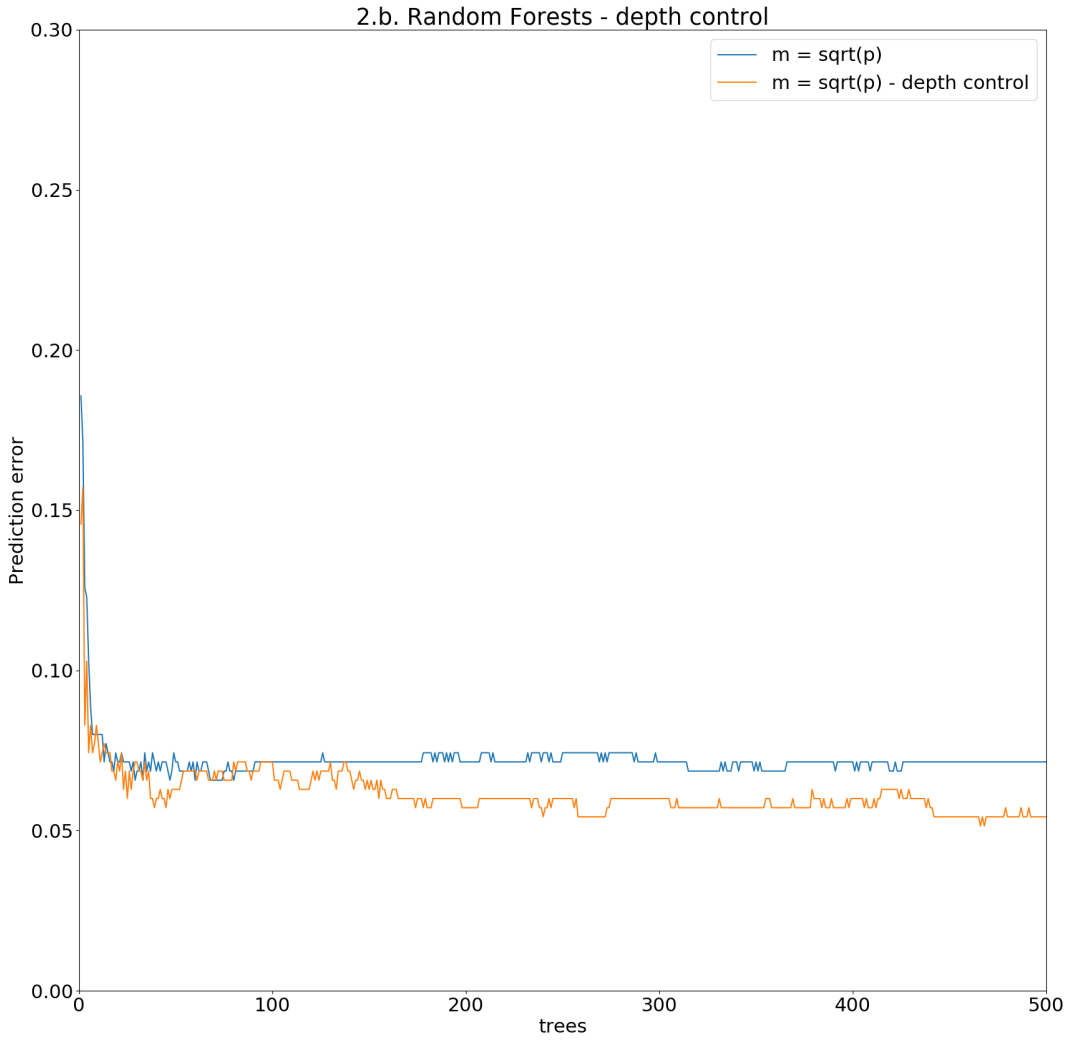
As we can see, adding an appropriate shrinkage does increase the model's accuracy.

2.a.



As we can see, the model's accuracy increases with number of trees. What's more, considering sqrt(p) features has the best accuracy while considering all features has the worst accuracy.

2.b.



As we can see, adding an appropriate depth control to the random forest trees does reduce the algorithm's error when the number of trees is large.