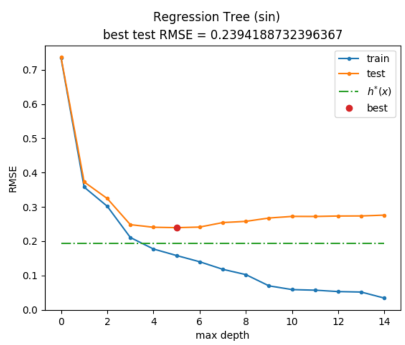
Ensembling

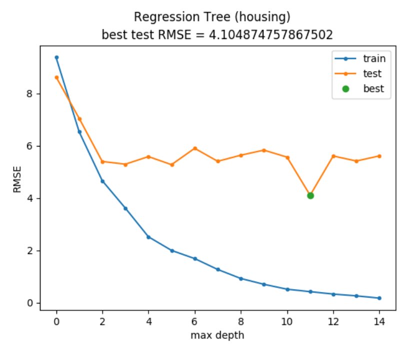
Andrii Zakharchenko

January 2020

1 Assignment 1



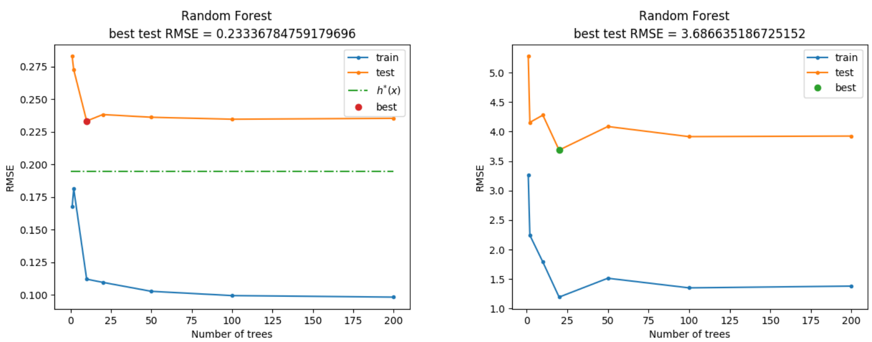
(a) Regression tree on Sin dataset



(b) Regression tree on Boston dataset

Figures above describe effect of regression tree depth on RMSE. The key takeaway from these figures is that with growing depth train error decreases to zero, while test error stalls at some point or even begin to grow. This is due to overfitting. With growing depth we get more leafs and decision nodes resulting in a higher variance of the model.

2 Assignment 2



(a) Random forest on Sin dataset (b) Random forest on Boston dataset

Figures above describe effect of number of tress on RMSE of random forest algorithm. We can see that there is no sign of overfitting with growing number of trees. Since we didn't limit the max depth of a tree in random forest each tree there has a low bias, but high variance. But since we average results of all those trees this reduces the variance of the model. However, important point to actually reduce variance is to train independent base models, and since regression trees have high sensitivity to changes in dataset this is why we used dataset bootstraping.

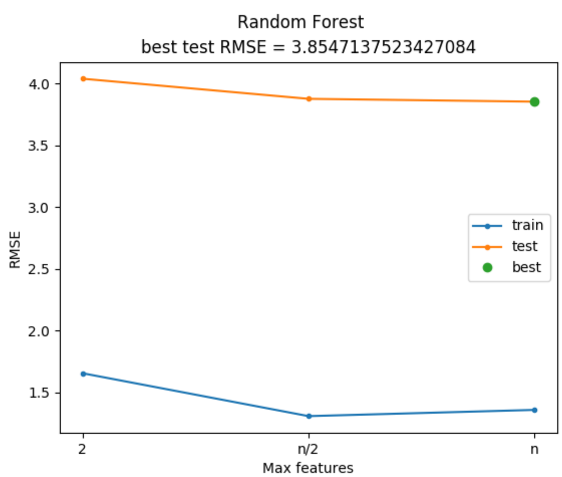
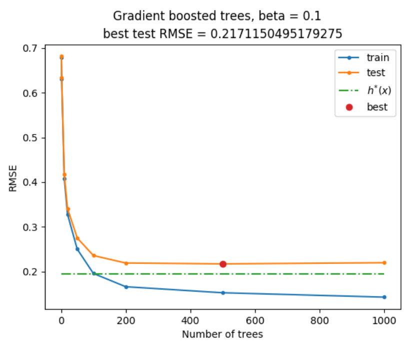


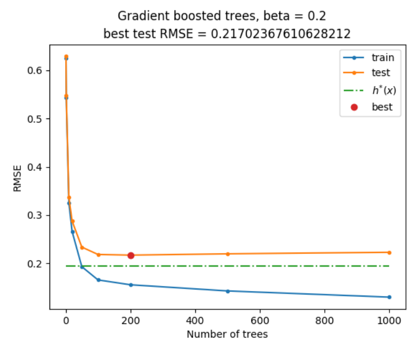
Figure 3: Random forest by max features

The figure above shows effect of number of features on RMSE of random forest. We can see that the difference in RMSE is not that big for different number of max features. This is because each base model was learnt on some subset of features. If some base models learned on unimportant features, other base model will compensate for this.

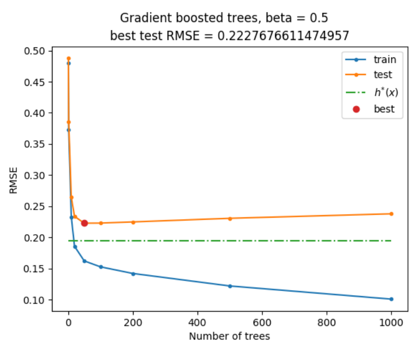
3 Assignment 3



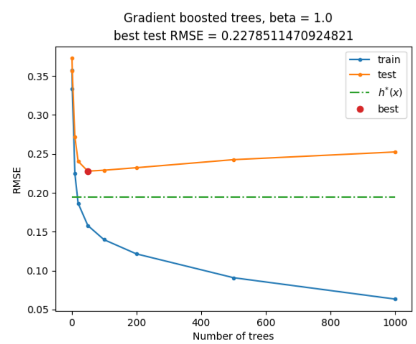
(a) GBT on Sin dataset, with $\beta = 0.1$



(b) GBT on Sin dataset, with $\beta = 0.2$



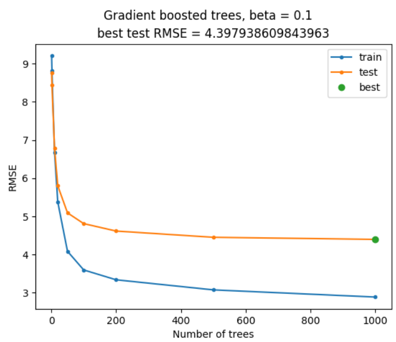
(a) GBT on Sin dataset, with $\beta = 0.5$



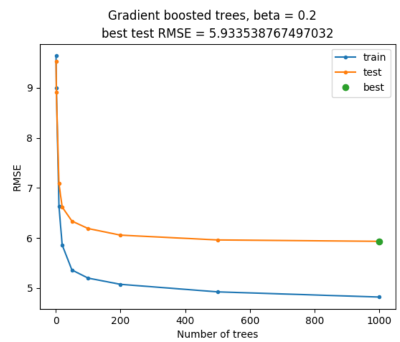
(b) GBT on Sin dataset, with $\beta = 1$

On a sin dataset we can clearly see two things, firstly lower learning rate requires more iterations (number of trees) to fit a better model, secondly with increasing number of trees the model begins to overfit.

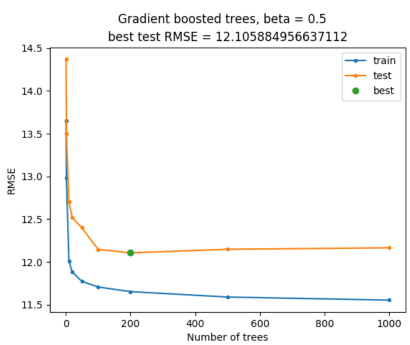
On the Boston dataset we see that model required to have more trees to produce a better fit, since it's a bit more complex dataset and each tree in the GBT is of a depth 1, which means that a single base model has high bias and we would require more base models. Also we can see that the higher learning rate produces worse RMSE.



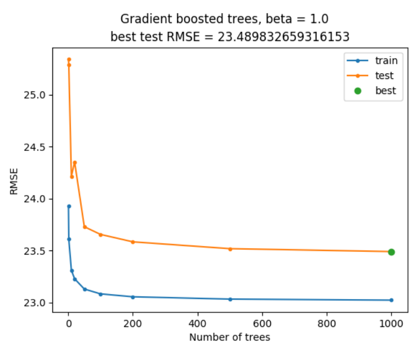
(a) GBT on Boston dataset, with $\beta = 0.1$



(b) GBT on Boston dataset, with $\beta = 0.2$



(a) GBT on Boston dataset, with $\beta = 0.5$



(b) GBT on Boston dataset, with $\beta = 1$