Cross Validated

Gradient Boosting Tree vs Random Forest

Asked 4 years ago Active 1 year, 2 months ago Viewed 115k times

110

89

Gradient tree boosting as proposed by Friedman uses decision trees as base learners. I'm wondering if we should make the base decision tree as complex as possible (fully grown) or simpler? Is there any explanation for the choice?

Random Forest is another ensemble method using decision trees as base learners. Based on my understanding, we generally use the almost fully grown decision trees in each iteration. Am I right?

machine-learning random-forest cart boosting ensemble

edited Jul 30 '18 at 14:42

Ferdi

4,021 5 28 56

asked Sep 20 '15 at 20:44

FihopZz

718 2 7 8

mlibby Jun 5 at 11:32

2 Answers

148

$$\text{error} = \text{bias} + \text{variance}$$

- Boosting is based on **weak** learners (high bias, low variance). In terms of decision trees, weak learners are shallow trees, sometimes even as small as decision stumps (trees with two leaves). Boosting reduces error mainly by reducing bias (and also to some extent variance, by aggregating the output from many models).
- On the other hand, Random Forest uses as you said **fully grown decision trees** (low bias, high variance). It tackles the error reduction task in the opposite way: by reducing variance. The trees are made uncorrelated to maximize the decrease in variance, but the algorithm cannot reduce bias (which is slightly higher than the bias of an individual tree in the forest). Hence the need for large, unpruned trees, so that the bias is initially as low as possible.

Please note that unlike Boosting (which is sequential), RF grows trees in **parallel.** The term `iterative` that you used is thus inappropriate.

edited Feb 5 '18 at 18:29
gung ♦
114k 34 284 557

answered Sep 24 '15 at 17:09

Antoine
3,730 5 22 47

1 "The trees are made uncorrelated to maximize the decrease in variance, but the algorithm cannot reduce bias (which is slightly higher than the bias of an individual tree in the forest)" -- the part about "slightly higher than the bias of an individual tree in the forest" seems incorrect. See web.stanford.edu/~hastie/Papers/ESLII.pdf section 15.4.2: "As in bagging, the bias of a random forest is the same as the bias of any of the individual sampled trees." Maybe you mean "slighly higher than the bias of a single fully-grown tree fit to the original data"? – Adrian Sep 17 '17 at 5:24

1 @gung I think there is a key question unanswered in OP, which is: why not use a fully grown tree at the 1st step of GBM? Why use a sequence of weak learner is better than one single fully grown tree? I am curious about that – ftxx Sep 13 '18 at 5:45

54

This question is addressed in this very nice post. Please take a look at it and the references therein. http://fastml.com/what-is-better-gradient-boosted-trees-or-random-forest/

Notice in the article that the speaks about calibration, and links to another (nice) blog post about it. Still, I find that the paper Obtaining Calibrated Probabilities from Boosting gives you a better understanding of what calibration in the context of boosted classifiers is, and what are standard methods to perform it.

And finally one aspect missing (a bit more theoretical). Both RF and GBM are ensemble methods, meaning you build a classifier out a big number of smaller classifiers. Now the fundamental difference lies on the method used:

1. RF uses decision trees, which are very prone to overfitting. In order to achieve higher accuracy, RF decides to create a large number of them based on bagging. The basic idea is to resample the data over and over and for each sample train a new classifier. Different

2. GBM is a boosting method, which builds on weak classifiers. The idea is to add a classifier at a time, so that the next classifier is trained to improve the already trained ensemble. Notice that for RF each iteration the classifier is trained independently from the rest.

edited Apr 13 '17 at 12:44
Community ♦
1

answered Feb 13 '16 at 14:46
jpmuc
10.8k 22 55

3 Would it be a fair conclusion from your answer that RF overfits more than GBM? – 8forty May 8 '18 at 22:42

4 @8forty I wouldn't draw that conclusion - while a single tree in RF will overfit more than a single tree in GBM (because these are much smaller), in RF these overfit will be averaged out when employing a lot of trees, while in GBM the more trees you add, the higher the risk of overfitting. In short, as N (number of trees used) goes to infinity, I expect RF to overfit much less than GBM – Ant Jun 28 '18 at 15:49

protected by Sycorax Jul 14 '18 at 22:23

Thank you for your interest in this question. Because it has attracted low-quality or spam answers that had to be removed, posting an answer now requires 10 reputation on this site (the association bonus does not count).

Would you like to answer one of these unanswered questions instead?