



ADVERTISEMENT



Professional email.
For the business you built.



Gmail for Work

What is the fundamental difference between RandomForest and Gradient Boosting algorithms?

machine_learning, gradient_boosting, random_forest

mukesh

Jul '15

Hello,

While reading about the gradient boosting algorithm, I read that Gradient boosting is a machine learning technique for regression and classification problems, which produces a prediction model in the form of an ensemble of weak prediction models, typically decision trees.

This is essentially what RandomForests do too. Then how are both of these algorithms different from each other?

Thanks.

manizoya_1

Jul '15

There are two parts of the model building story. They are

1. How to do training and testing?
2. Which algorithm to use for prediction?

The difference between Random Forest and Boosting can be understood easily by understanding the above two questions.

Random Forest use bootstrapping method for training/testing (Q1 above) and decision trees for prediction (Q2 above) . Bootstrapping simply means generating random samples from the dataset with replacement. Every bootstrapped sample has a corresponding hold out or 'out-of-bag' sample which is used to test. This means that if you generate 100 bootstrapped samples, each time you will get a set of predictions. Final prediction is simply the average of all 100 predictions.

When using a boosting technique, we are essentially concerned about only Q1 above. We are free to use any algorithm ... it could be decision tree, or NN or KNN or SVM. Now lets look at how is training/testing done here. Boosting methods also produce multiple random samples, but it is done more thoughtfully. The subsequent samples depend on weights given to records in the previous sample which did not predict correctly - hence called weak learners. The final prediction is also not a simple average

of all 100 predictions, but a weighted average.

Hope this helps, random forest and boosting are two powerful advanced methods which are difficult to grasp quickly. Cheers!

ADVERTISEMENT



© Copyright 2015. Analytics Vidhya