

Random Forest Classifier

Eli

Random Forest...and more...

Random forest, like its name implies, consists of a large number of individual decision trees that operate as an ensemble. Each individual tree in the random forest spits out a class prediction and the class with the most votes becomes our model's prediction.

Random Forest...and more...

The fundamental concept behind random forest is a simple but powerful one — **the wisdom of crowds**. In data science speak, the reason that the random forest model works so well is:

A large number of relatively uncorrelated models (trees) operating as a committee will outperform any of the individual constituent models.

Random Forest...and more...

The low correlation between models is the key. Just like how investments with low correlations (like stocks and bonds) come together to form a portfolio that is greater than the sum of its parts, uncorrelated models can produce ensemble predictions that are more accurate than any of the individual predictions.

The reason for this wonderful effect is that the **trees protect each other from their individual errors** (as long as they don't constantly all err in the same direction).

Random Forest...and more...

While some trees may be wrong, many other trees will be right, so as a group the trees are able to move in the correct direction. **So the prerequisites for random forest to perform well are:**

Random Forest...and more...

1. There needs to be some actual signal in our features so that models built using those features do better than random guessing.
2. The predictions (and therefore the errors) made by the individual trees need to have low correlations with each other.

Random Forest...and more...

<https://www.activestate.com/blog/comparing-decision-tree-algorithms-random-forest-vs-xgboost/>

<https://towardsdatascience.com/basic-ensemble-learning-random-forest-adaboost-gradient-boosting-step-by-step-explained-95d49d1e2725>

Thank You