Random Forests

Random Forests are built using bagging, which relies on bootstrapping

A bootstrapped sample of a dataset of size n is obtained by drawing n samples independently with replacement. Read about bagging in ISLR, pages 315-321.

What is the idea of bagging?

What is the trade off in choosing the number m of features to sample in every node?

What’s Out Of Bag Sampling? What is the size of the Out Of Bag set when the dataset is large?

How can one measure feature importance in random forest?

Now read about random forest in chapter 15 of ESL, pages 587-596. Skip 15.3.3 (Proximity Plots).

Solve exercise 15.1. What is its implication on random forest?

Demonstrate in code that when the number of variables is large, but the fraction of relevant variables is small, random forests are likely to perform poorly with small m.