

Solution 1:

- 1) Bagging means drawing n observations with replacement from the same training data M times, fitting a model with the same inducer on each of the M data sets, and average their predictions to obtain a final model.
- 2) Random forests optimize performance by
 - randomly drawing $mtry \leq p$ candidate features at each node of each tree to consider for splitting. This leads to less correlated trees within the forest, which improves performance.
 - fully expanding trees (without aggressive stopping criteria or pruning). This leads to more variable trees within the forest, which improves performance.
- 3) Example explanations:
 - **Variable importance measure based on improvement in split criterion**
The variable importance of a feature is the average decrease in the splitting criterion (e.g. accuracy) achieved in all nodes of all trees that use the feature for splitting.
 - **Variable importance measure based on permutations of OOB observations**
The variable importance of a feature is calculated by randomly shuffling around the values of a feature within the data set (this is called permutation). Then, OOB observations are funneled through the random forests, and an error estimate is calculated. The difference between this error estimate and the error estimate without permutation is the variable importance of the feature.