P5-Randomness of a random forest

Jeetendra Gan

12/2/2019

What is random about a random forest?

- A random forest considers only a subset of features to create a node in one of its decision trees. Ideally, a value of $\sqrt{features}$ is selected.

- The node at each stage is selected randomly on purpose. The reason is to make the trees in a forest as different as possible from each other, i.e. to have as many uncorrelated trees in the data.

- The advantage of having multiple uncorrelated trees is that it helps the model to capture contributions of less significant variables. i.e. a random forest overcomes the disadvantage of bagging.

- i.e. if the search space for every tree is the same(even through the sample is bootstrapped), there might be a chance that an overpowering feature might be chosen almost all the time.