ULID: C00535934

Name: Rajesh Awal

Department: Informatics

1. What is the approximate depth of a Decision Tree trained (without restrictions) on a training set with 1 million instances?

Decision Tree splits the sample in two at every layer for balanced tree. Approximate depth will be function of log with base two. For one million instances, the depth will be log2(1,000,000) which is slightly less than 20. With round up the answer is 20.

3. If a Decision Tree is overfitting the training set, is it a good idea to try decreasing max_depth?

Yes, it is good idea to decrease max_depth. Instances and features with high correlations causes over fit. Decreasing the max_depth regularizes a model. As a result model will be constrained.

4. If a Decision Tree is underfitting the training set, is it a good idea to try scaling the input features?.

No, Scaling the training set will not work if Decision Tree model is underfitting. Whether the data is centered or scale, it will not matter if you are using decision tree.

7. Solution is attached in jupyter notebook file.