



University of Louisiana at Lafayette

School of Computing and
Informatics

INFX-598 Machine Learning Applications

Homework # 06

Submitted By

Navid Yousuf

ULID: C00419219

Fall 2021

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

The depth of a well-balanced binary tree containing m leaves is equal to $\log_2(m)$, rounded up. A binary Decision Tree (one that makes only binary decisions, as is the case of all trees in Scikit-Learn) will end up more or less well balanced at the end of training, with one leaf per training instance if it is trained without restrictions. Thus, if the training set contains one million instances, the Decision Tree will have a depth of $\log_2(10^6) \approx 20$ (actually a bit more since the tree will generally not be perfectly well balanced).

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

If a Decision Tree is overfitting the training set, lowering max depth may be a good option, since this will limit the model and make it more regular.

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

Scaling or centering the training data is irrelevant to Decision Trees; scaling the input features is a waste of time.