

## **Chapter 6**

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

A well-balanced binary tree with  $m$  leaves has a depth equal to  $\log_2(m)^2$ , rounded up. If trained without limits, a binary Decision Tree (one that makes only binary decisions, as all trees in Scikit-Learn do) would end up more or less well balanced at the conclusion of training, with one leaf per training instance. The Decision Tree will have a depth of  $\log_2(10^6) \approx 20$  if the training set comprises one million occurrences (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 effort.