Decision Tree

* Using the decision algorithm, we start at the tree root and split the data on the feature that results in the largest **Information Gain** (**IG**).
* In practice, this can result in a very deep tree with many nodes, which can easily lead to overfitting. Thus, we typically want to **prune** the tree by setting a limit for the maximal depth of the tree.

**Maximizing information gain**

In order to split the nodes at the most informative features, we need to define an objective function that we want to optimize via the tree learning algorithm. Here, our objective function is to maximize the information gain at each split, which we define as follows: