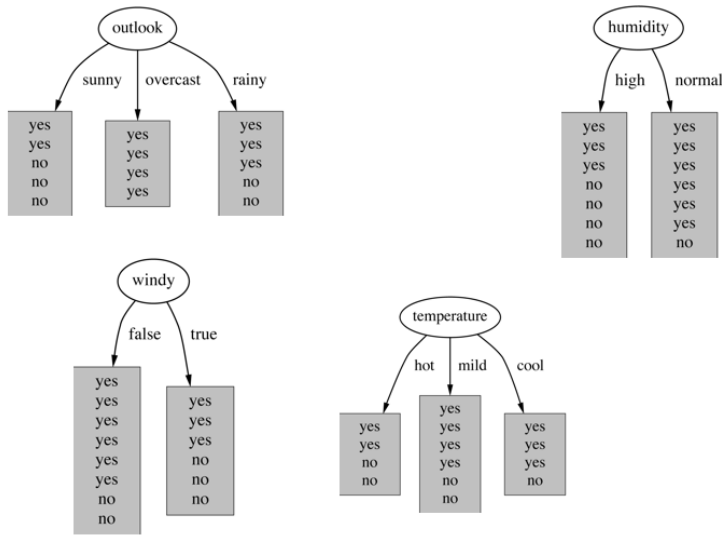


Data Mining, CSCI 347, Fall 2019
Decision Trees, Sept. 18

A simple decision tree algorithm is to calculate the information gained when splitting on each attribute, split on the attribute that yields the greatest gain, and recursively repeat this process for each branch.

Consider the weather dataset and splitting on the outlook attribute.



1. Calculate the purity of the top node.
2. Give the formula to calculate the information gained by splitting on Outlook.
3. Calculate the purity of each of the three branches.

4. Calculate the weighted average information value for splitting on Outlook

5. Information gain from splitting on Outlook.

6. Calculate the intrinsic information value of the split

7. Calculate the ratio by dividing the attribute gain by the intrinsic information value