Technische Hochschule Rosenheim
Prof. Johannes Jurgovsky

Exercise 5

Learning Goals

- Decision Trees
- Impurity functions
- The ID3 algorithm

1. (a) What is overfitting?

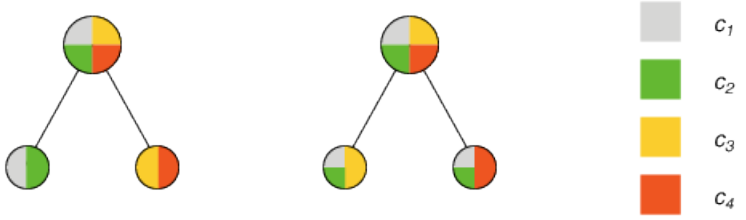
   (b) Check all correct statements.

   - ☐ A short training time leads to overfitting.
   - ☐ A smaller decision tree generalizes better to unseen examples than a larger decision tree.
   - ☐ The generalization capability of a decision tree depends on the training set.
   - ☐ By using entropy as an impurity measure, overfitting can no longer occur.

2. For each of the following boolean functions: Draw a decision tree that represents the boolean function. All variables take values in $\{0, 1\}$. (You do not need to apply any algorithm for this exercise)

   A. $x_1 \wedge \neg x_2$ B. x_1 XOR x_2 C. $x_1 \vee (x_2 \wedge x_3)$

3. Consider the following illustration of two possible splittings of an example set D with four classes $C = \{c_1, c_2, c_3, c_4\}$. Assume that in both splittings the resulting subsets each contain $\frac{1}{2}$ of the examples from D.



   (a) Compute the drop in impurity $\Delta\iota$ for the left split and the right split, respectively, using the misclassification rate $\iota_{misclass}$ as well as the Gini impurity ι_{Gini}. (Note: This will yield 4 values: one for every combination of "split" and "impurity measure".)

   (b) Which splitting is preferable under each impurity measure?

4. Implement the impurity functions in section 2.1 of the jupyter notebook and implement the impurity reduction calculation in section 2.2 of the jupyter notebook.