ID3 - Implementation for given test data

Johann Wagner, 212276
Johannes Wünsche, 211720
Marten Wallewein-Eising, 212277
Paul Stang, 204561

16. November 2017

Overview

This submission includes a python 3.6 script for creation of a ID3-decision-tree, as well as the needed data, an output file created by the script in case of any incompatability or other obstacles restricting the usage.

Code

Most things in need of explanation is commented in the code, but the base ideas are listed here anyways:

- *function id3* contains implementation of the ID3-algorithm

- *function printTree* contains output into the decision_tree xml file

- *class Node* used for storing information about single node with reference to following ones

- *function readData* parses the plain text data file

- *informationGain* calculates the informationGain value for given attribute

- *collectInformation* used for informationGain to calculate entropy

Result

The result is a large and probably overfitted decision tree build with the ID3-algorithm.
While many leaves of the tree only contain a single sample from the given data this is traceable to the algorithm itself and can be undone by simply unifying leaves which have the same parent node and point to the same class value.