

Classification Trees

Data Mining 2015: assignment 1

Sebastiaan Jong (5546303) & Bas Geerts (5568978)

This report is written for the first assignment of the Data Mining (2015) course at Utrecht University. The goal of this assignment was to write a function in the R programming language that constructs a classification tree on a certain dataset, and to figure out efficient parameters for this tree.

For this assignment we used the Heart Disease dataset from the University of California, Irvine machine learning repository. The preprocessed version for this assignment contains 297 instances and 14 attributes. A brief description of the used attributes:

1. Age, numeric attribute.
2. Sex, categorical attribute.
3. ChestPain, categorical attribute.
4. RestBP, numeric attribute, resting bloodpressure.
5. Chol, numeric attribute, serum cholesterol.
6. Fbs, categorical attribute, fasting bloodsugar test result.
7. RestECG, categorical attribute, resting electrocardiographic results.
8. MaxHR, numeric attribute, maximum heart rate achieved.
9. ExAng, categorical attribute, exercise induced angina.
10. Oldpeak, numeric attribute, exercise induced ST-depression.
11. Slope, categorical attribute, slope of the peak exercise ST segment.
12. Ca, numeric attribute, number of major vessels colored by flourosopy.
13. Thal, categorical attribute, thallium heart scan.
14. AHD, the class label for this assignment, heart disease diagnosis.