GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN
INSTITUTE OF COMPUTER SCIENCE
Software Engineering for Distributed Systems
http://www.swe.informatik.uni-goettingen.de

Data Science and Big Data Analytics
WS 2016/2017
Dr. Steffen Herbold

Exercise 4 · Due at 2017-01-05

Classification with Naive Bayes and Decision Trees

Execute the following tasks with R[1]:

1. Load the libraries `e1071` and `party`. You may have to install them.
2. Create a training set and a test set from the `iris` data set. The training set shall contain 100 data points and the test set shall contain 50 data points. The training set and the test set shall be disjunctive.
3. Train a naive bayes classifier for the species of the iris using Sepal.Length, Sepal.Width, Petal.Length, and Petal.Width with the training data (Hint: `naiveBayes`).
4. Train a decision tree with the command `ctree` for the species of the iris using Sepal.Length, Sepal.Width, Petal.Length, and Petal.Width with the training data.
5. Plot the decision tree.
6. Evaluate the results of both classifiers on both the training and the test data using the `predict` function.

[1] You can start RStudio typing rstudio into the bash in the CIP pool.