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 3 · Due at 2016-12-08

Association Rule Mining with R

Execute the following tasks with R[1]:

1. Load the libraries `arules` and `arulesViz`. You may have to install them. Hint: `library` and `install.packages` are the commands for loading and installing.
2. Load the example data using the command `associationData <- read.transactions("http://user.informatik.uni-goettingen.de/~sherbold/AssociationRules.csv",rm.duplicates = FALSE,format="basket",sep=" ")`
3. Train association rules with the `apriori` command.
4. Visualize the results with the `plot` command.
   - Use the parameter `interactive=TRUE` with the plot command and see what happens.
5. While doing this, think about reasonable values for support and confidence. See what happens when you choose different values.

Logistic Regression with R

Execute the following tasks with R:

1. Load the example data using the command `cuse <- read.table("http://data.princeton.edu/wws509/datasets/cuse.dat", header=TRUE)`
2. Train a logistic regression model over the formula `cbind(using, notUsing) ~ age + education + wantsMore`
   - The command for logistic regression models is `glm` and the family is `binomial(logit)`.
3. Print a summary of the trained regression model. Interpret the coefficients. Which features are important? Why?

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