

# INFO411

## Laboratory exercises on Classification and Prediction

### Overview:

This lab is a non-assessed group exercise. Multi-Layer Perceptron (MLP) networks are popularly used for classification and regression in Data Mining and Big Data applications. They are often trained for the purpose of predicting the classification of unknown input patterns.

In this laboratory class we learn how to use MLP for predicting the credit worthiness of bank customers. **This will prepare you for the second part of assignment 1.** The same dataset as for assignment 1 will be used in today's lab.

### What you need:

1. R software package (already installed on the lab computers)
2. The file "Lab\_Classification.zip" on Moodle.

### Preparation:

1. Work in a group of size three to four (no less than 3 but no more than 4 students are to work together).
2. Boot the computer into Windows mode.
3. Download Lab\_Classification.zip then save to an arbitrary folder, say "C:\Users\yourname\Desktop"
4. Uncompress Lab\_Classification.zip into this folder
5. Start "R"
6. Change the working directory by entering: `setwd("C:/Users/yourname/Desktop")`  
(Note that R expects forward slashes rather than backwards slashes as used by Windows.)

### Your task:

You are to submit a PDF document which contains your answers to the questions in this laboratory exercise. One document is to be submitted by each group (thus only one member from each group submits the answer to the group exercise). The header of the document must list the name and student number of all students in the group. Clearly indicate which question you have answered.

The following link provides a documentation of the MLP module (called RSNNS) which we will use in R. The link can help you with finding the right answers in today's lab and for the assignment:

<http://cran.r-project.org/web/packages/RSNNS/RSNNS.pdf>

### Work through the following step and answer given questions:

Step1: Open file Classification.R by using a text editor (i.e. Notepad)

Copy one command-line at a time from this file and paste it into the R-command window. Observe what the command does and then develop an understanding of the copied commands (do not just blindly copy without understanding each of the command lines). You may have to consult the "help" function of R or refer to documentation on the Internet to develop the required understanding.

**Question 1:** What does the function `mlp()` do in this script?

**Question 2:** What does the function `predict()` do?

**Question 3:** Explain the result shown in the first two confusion matrices. Give an interpretation of the results.

**Question 4:** Explain the four plots that will be produced by this script. Give an interpretation of the results.

**Question 5:** Explain the difference of the third confusion matrix (the one that uses the 402040) method when compared to the first of the confusion matrices.

**Question 6:** Inspect the weights of the MLP. Explain why the weights can become larger than 1 (or smaller than -1) and explain problems that can arise out of having very large weights.

Write up all your answers, then submit your answer as a PDF document via the link provided on MOODLE. One submission per group!