

INFO411

Laboratory exercises on Clustering and Visualization

Important notes:

This lab is a group exercise. All students must work as part of a group. Student must be a member of one group (and one group only) and group sizes must be 3 or 4 (no less than 3, no more than 4 students in each group). You can chose your own group, or you can ask the tutor during the lab to be allocated to a group. Each group makes one submission (one member of each group submits).

This submission will not be assessed.

Overview:

Self-Organising Maps (SOMs) are an unsupervised data visualisation technique that can be used to visualise high-dimensional data sets in lower (typically 2) dimensional representations. In this laboratory class we will mine census data from Ireland using R:

1. Visualize population properties
2. Aquire some knowledge on the difficulty of data segmentation of a given set of data
3. Obtain geographically dispersed clusters

What you need:

1. R software package (already installed on the lab computers)
2. The file "laboratory_week3.zip" on Moodle.

Preparation:

1. Work in groups. Minimum group size is 3, maximum size of a group is 4.
2. Boot computer into Windows mode.
3. Download laboratory_week3.zip then save to an arbitrary folder, say
"C:\Users\yourname\Desktop"
4. Uncompress laboratory_week3.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 backward slashes as used by Windows.)

Your task:

Your group is 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 of students (only one student in each group needs to submit). 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 can help you with finding the answers:

<http://www.r-bloggers.com/self-organising-maps-for-customer-segmentation-using-r/>

Note that this link does not offer any interpretation of the results. It is your task to interpret the results (i.e. clearly explain the results), what knowledge can be extracted from the results, and explain why the results are of value.

Work through the following steps and answer given questions:

Step1: Open file step1_preprocessing by using a text editor (i.e. Notepad)

Copy each line in this file and paste it into the R-command window. Try to understand as many of the copied commands as possible (do not just blindly copy)

Step 2: Open file step2_SOMtraining by using a text editor.

Copy each line in this file and paste it into the R-command window. Try to understand all of the copied commands.

Question 1: Explain the purpose of the som_grid and the som function in this set of commands.

Step 3: Open the file step3-visualization

Copy each line in this file (one line at a time) into the R-command window. Try to understand as many of the copied commands as possible.

Question2: This step will create eight different plots. Explain what can be seen in each of these plots. It is important that you give an interpretation of the plotted results, what knowledge can be extracted from these plots. Explain the potential value of the knowledge extracted.

Step4: Open the file step4-clustering

Copy each line in this file (one line at a time) into the R-command window. Try to understand as many of the copied commands as possible.

Question3: Explain the plots shown. Again, it is important that you give an interpretation of the results, knowledge that can be extracted, and the value.

Step5: Open the file step5-mapping

Copy each line in this file and paste it into the R-command window.

Question4: A plot is generated. Explain what can be seen in this plot.

Step6: Open file step2_SOMtraining by using a text editor.

Change the SOM so that it is double in size, and trained for 2000 iterations. Then execute all the commands of step2 through to step5.

Question5: Did any of the interpretations of results (in step3, step4,step5) change? If so, what has changed?

Step7: A SOM can be useful for finding out how difficult a learning problem is.

Question 6: Explain why.

Write up all your answers, add at the top of the document the name and student number of each member in the group, then one member of each group is to submit your answer as a PDF document via the submission link provided for this lab on MOODLE.