

2nd year Ensimag ISI, MMIS / M1 MSIAM

Statistical Analysis and Data Mining

Instructions

Duration: ISI/MMIS: 3 hours, MSIAM: 2 hours. MSIAM students should not address part 3, unless they satisfactorily addressed all questions of parts 1 and 2.

All handwritten documents allowed, as well as the electronic documents on your computer. No calculator or mobile phone allowed.

No internet connexion is available from your computer.

To achieve this practical, you have to use RStudio. Open each .Rmd file located in the \$HOME/exam/ directory

Answer the questions directly in each .Rmd file, ensure that you can knit it into an html file, and save the .Rmd file. The different .Rmd files correspond to three independent exercises.

Please answer in English only.

Some questions can be multiple, do not forget to read and answer all of these.

Any answer not contained into the .Rmd file will not be taken into account, unless instructed explicitly by you or the teacher into the .Rmd file (you can state into a .Rmd file: 'Please see figure myfig.png' for example).

The contents of the \$HOME/exam/ directory are

- Part1.Rmd, Part2.Rmd and Part3.Rmd (the files you have to complete and save)
- DigestASM_Data_Mining.pdf (booklet of the course)
- instructions.pdf (this document)
- rmarkdown-cheatsheet.pdf (basic RMarkdown commands)
- short-math-guide.pdf (to help you to write \LaTeX -like formulas in RStudio)
- *.csv: data sets
- classification_slides.pdf (slides on document representation and classification)
- GraphCourse: slides of the course on graph clustering, including some .Rmd to be taken as an example on dummy data.

Please check the presence of some `whoami.txt` file. This file is automatically generated at logon and should already contain your first and last names, login and computer identifier. If the file is missing, please create and fill it.

Do not hesitate to ask the teacher if you need help on RStudio, rmarkdown or R.

Important – You can (and you should!) save the current state of your whole home directory using the icons on your desktop.

Before that, save your currently open files using your editor(s) – particularly the .Rmd and .html files, which you should try to knit.

These icons allow you to either:

1. save your current work and continue your work;
2. save your current work and stop the session. You have to perform this operation before leaving the room.

The backup operation is based on rsync, and does not store every intermediate saved version. If your PC crashes, your last backup can be recovered with the help of the teaching staff.