

Unsupervised Learning (MT7050) - Project 4

Task: Choose a dataset from the list of databases/search engines stated in the course page, then analyze it using an appropriate unsupervised learning method with validation. You can use any programming language in the analyses. The project should be solved with 3 students in each group.

Guidelines for the written report:

- 1) Give the student names of the group.
- 2) State which dataset is picked and why it is picked
- 3) Describe what the data is about and the data properties with suitable exploratory analyses.
- 4) Explain which unsupervised learning and validation methods are used, and justify why the methods are suitable for the chosen data.
- 5) Concisely state how the analysis is performed (e.g. which functions and packages are used, parameter settings, etc.) and present the analysis results using graphs, tables, etc. Do your graphs effectively present the message you want to convey?
- 6) **Most importantly: Explain and discuss the implications from the analysis results, what you learn from the analysis, what are the possible improvements, etc.**
- 7) Give appropriate references.
- 8) Your performance of the project will be judged on how clear, logical and organized your report and talk are given, and on whether the suitable methods are employed and correctly implemented for the analyses. Please include enough explanations and discussion and avoid writing a technical manual only on how things are done. Moreover, a report size (not including the appendices) of around 5 pages is preferred.
- 9) The written report as a single .pdf file should be submitted to the moodle system before the start of the oral presentation (date and time will be announced on the course page). Any source code associated with the analyses should be attached as appendices in the report with sufficient documentation. **All members of the same group should submit the same report to the moodle system.**

Guideline for the oral presentation:

- 1) The presentation is ~ 20min for each group plus ~ 3min for questions. (Talk duration may be changed later depending on the number of talks)
- 2) The talk should cover concisely items 1 to 6 above in the written report.
- 3) It is up to you how to distribute the talk time to each member of the group.