Statistical Deep Learning (MT7042) - Project 4

Task: Pick up a dataset from a databases/search engines, e.g. those listed in the course page, then analyze it using an appropriate deep learning method with regularization. You can use any program package in the analyses. This project should be solved by teamwork with 2 students in each team. Since we may not have even number of students, it is also permitted if 3 students as a team in carrying out the project. Please be aware that some datasets (e.g., images, videos) could be very big. To avoid turning the project into purely a computational problem, please either downsample your data or choose a dataset with workable size.

Guidelines for the written report:

1) Give the student names of the group.
2) State which data-set 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 deep 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 setting, etc.) and the analysis results in terms of graphs, table, etc. Do your graphs effectively present the message you want to convey?
6) **Most importantly: Discuss the implications and interpretations from the training network and 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 oral presentation, i.e., at **noon on Mar 7, 2023**. Any source code associated with the analyses should be attached as appendices in the report with sufficient documentations. **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.