732A99/732A68 Machine Learning
Division of Statistics and Machine Learning
Department of Computer and Information Science

Computer lab block 2

Instructions

The purpose of this laboratory work is that you use the knowledge of methods of tools that you have learned in the course (from Block 1 or 2) to perform a mini-project in the groups of 2 persons. The groups can be decided by students themselves.

Find a dataset with at least 1000 observations and at least 10 features from internet and state a research question (perhaps devoted to classification, regression or clustering) that is related to this data set. The two students are supposed to use two different kinds of models to address this research question. Use your imagination and knowledge from the course and squeeze out the most you can from the model you are using to achieve the best possible model: try different data partitioning, test wide ranges of hyperparameter values, do feature/model selection, test various cost functions, distributions, regularizations etc. What and how much you do depends a lot on your ambition level and interest of the subject.

After the two group members are done with their individual work, they need to make a careful comparative analysis of their models and results.

Format of the report

The report should include the following sections:

- Data description and research objective (0.5 page max)
- Descriptions of models used and experimental design (1 page max)
- Results (4 pages max)
    - Should include your results, your analysis of these results and comparative analysis of the two models used.
- Discussion and conclusion (1 page max)

Evaluation criteria

The laboratory work is passed if all of these criteria are fulfilled:

1. Clarity: the report can be understood with a minor effort
2. There are few or no grammar errors or sentences having bad structure
3. The research objective is clearly stated and is relevant for the data used.
4. Data description is complete enough with respect to the research objective stated

5. Description of methods is correct to a large extent and is complete enough to understand the results of the work
6. Models and the design of the experiments are appropriate for the research objective stated
7. Results demonstrate that models and methods are appropriately applied, results to a large extent answer the stated research objective
8. Results: comparison of the two models used is correct to a large extent.
9. Conclusions are appropriate, discussions include reasoning about limitations of the study performed
10. You have submitted the report as PDF, your data and the codes. After running the codes, the results can be reproduced.