Uppsala University

Introduction to Machine Learning, Big Data, and AI

Project Instructions

1 Project Instructions

The last two weeks will be focused on a course project where a group of 2-3 students choose data and create a supervised machine learning predictor for a real-world dataset.

It is possible to have only one student in a group, although this is not recommended. One student group will, in practice, mean additional work due to the requirements of the project.

Students need to turn in a half-page project and data description by the end of block six and get approval for the proposed project.

The project is expected to take 40h per student in the group. Hence a 3 group project should be the equivalent of a 120h project.

1.1 Data Sets and Methods Recommendations

We recommend that you find a dataset you are interested in using yourself. If you have a hard time finding a dataset to use, there are a lot of available datasets at the UCI Machine Learning repository: https://archive.ics.uci.edu/ml/index.php

Some data sets should not be used in the project:

- Titanic (R data set)
- mtcars (R data set)

Modeling requirements and recommendations:

- Your project should be a supervised learning project.

1.2 Project Report

The Project outcome is an R or Python notebook/markdown report. Both the R-markdown and the final PDF should be supplied.

The submitted notebooks need to illustrate the knowledge of the Bayesian workflow. It has to include:

- Description of the problem.
- Description of the data.
- Description of the method used and motivation.
- Detailed description of how the evaluation was conducted and metrics that were computed.
- Discussion of problems and potential improvements.
- Discussion of potential ethical problems (in light of the guest lecture).

1.3 Project Presentations

Presentation details:

- Each project needs to be presented in addition to submitting the notebook
- The presentation should be high level, but sufficiently detailed information should be readily available to facilitate answering questions from the audience
- Within each session, about four groups, will be presenting
- For 1-2 person groups, the presentation should be 10 minutes
- For three-person groups, the presentation should be 15 minutes
- Afterwards, questions will be asked first by other students and then by attending teachers.
- Each group will be responsible for (critically) discussing one other project report.
- Grading of the presentation will be done by the attending teachers using standardized grading instructions 8see below).
- Presenters' ID cards will be checked to ensure the right persons are presenting

Specific presentation recommendations:

- The first slide needs to include the project title and names of the group members.
- The chosen methods(s) should be explained and justified (you are *not* holding this presentation for a hypothetical customer who doesn't care about the details of your methods).
- Big enough font size for text and figure labels should be used to make it easy to read the slides for the audience.
- The last slide needs to include to conclusion and names of the group members.
- The best presentations we have seen were groups who discussed with teachers and showed intermediate results to get feedback on improving. So we recommend you visiting the computer labs.