Uppsala University

Introduction to Machine Learning, Big Data, and AI

Project Instructions

1 Project Instructions

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

Requirements for the projects are:

- Your project should be a supervised learning project.

- Real data should be used (see below for details).

- Minimally two methods (or two different neural nets) should be compared and evaluated.

For PhD students: You can choose to make a small project related to your research interest instead. Although, it should still be a 4 page paper output.

1.1 Suggested Reading

The project will be a small practical exercise in supervised machine learning. Suggested reading (before starting the project) is:

- Chapter 11 in Ian Goodfellow, Yoshua Bengio, and Aaron Courville. 2016. *Deep Learning*. The MIT Press.

1.2 Project Group and Expected Workload

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.

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.

1.3 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 (and problems) at:

- The UCI Machine Learning repository: [here]

- The machine learning competition site Kaggle: [here]

The following data sets should not be used in the project:

- Titanic (R data set)

- mtcars (R data set)

1.4   Project Proposal

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 proposal must include the following pieces.

The project proposal should include all the group members names!

- Description of the problem.

- Description of the data.

- A first idea on potential methods and evaluation.

To minimize the project work's total workload, I suggest you see the proposal as a first draft of Sections 1 and 2 of the project report (see below). Then you will already have written the first part of your project report.

1.5   Project Report

The Project outcome is a report in the ICML paper format that can be found [here]. The ICML format is also available in overleaf here: [here]

The project report should be anonymous!

The paper should consist of *four (4) pages*, excluding references and eventual appendices. The paper should include the following four (4) sections:

1. Abstract

2. Introduction (roughly 0.5 pages)

   - Description of the problem.

3. Data and Methods (roughly 1 pages)

   - Description of the data.
   - Description and motivation of the method
   - Description of how the evaluation was conducted

4. Results (roughly 1.5-2 pages)

   - Results of the proposed methods
   - Evaluation of the proposed method

5. Conclusions (roughly 0.5-1 pages)

   - Conclusions from the results. Did the method work as expected?
   - Discussion of problems and potential improvements
   - Discussion of potential ethical issues (in light of the guest lecture).

Additional requirements and hints for the report:

1. All Figures using color should have a color-blind friendly color palette. See here and here.

2. Befor you turn in the project, do a language check with a tool such as Grammarly. A project will poor english (errors that would have been spotted with a tool such as Grammarly) will affect your grade downwards.

3. The final report should like like a research paper, i.e. try to avoid bullet list and get a good flow in the text.

4. You should use correct reference systems. A tip is to use
`citet`,
`citep`, and
`bibtex`. This will also simplyfy your future thesis work.

1.6 Project Review

Each student will review one other group's project report. The project review should consist of a written report of 1-2 pages using the ICML paper format that can be found [here]. The review *have to* contain the following sections with roughly one to two paragraphs per section.

1. Summary and contributions:
   Briefly summarize the project report and its contributions in one paragraph.

2. Strengths:
   Describe the strengths of the work. Examples can be a clever use of the method. An interesting and non-trivial problem.Comparisons of more than one method.

3. Weaknesses:
   Describe the weaknesses of the work. Examples can be a lack of reproducibility, poorly chosen method, or evaluation.

4. Correctness:
   Are the claims and methods correct? Is the empirical methodology correct?

5. Clarity:
   Is the paper well written? Is it easy to follow or understand the report?

6. Formatting:
   Is the paper correctly formated, and are the required content included in the project report?

7. Ethical considerations:
   Does the report raise potential ethical concerns? Does methods, applications, or data create or reinforce unfair bias or have a primary purpose of harm or injury. If so, please explain briefly.

Note Reviewing others' work is an essential part of the scientific process. These reviewer guidelines have been in large part been taken from the NeurIPS reviewer guidelines that you can find here: NeurIPS reviewer guidelines

1.7 Project Presentations

Presentation details:

- Each project needs to be presented in addition to submitting the mini-project report

- 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.

- Reviewers are anonymous so as a reviewer you are *not* expected to do any additional questions.

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 for the audience to read slides.

- A good rule of thumb is to expect one slide to take 2 minutes to present.

- The last/final slide needs to include your conclusion and names of the group members.

1.8 Project Grading

Below are the criterias used when grading the mini-projects. Some general comments on grading are:

1. The more students the higher the quality expected of the project, i.e. a better report is expected from a three-student report than a two-student report.

To pass the report (G), the following criterias should be fulfilled:

1. The report should be turned in and follow the general outline of Section 1.5.

2. show basic knowledge and understanding of the core concepts of the course by using concepts correctly

3. show an understanding in when certain methods should be used or not, and how

4. use at least two (2) differnt methods (or architectures) and compare them in a correct way

5. state what has been done in the report with clarity, good english and rigour so it is easy for a reader to understand and follow the paper.

6. correctly use references in the report following the guideline of the template in Section 1.5

To pass the mini-project with distinction (VG), the following criterias also apply in addition to the criteria for passing the report above:

1. show deep knowledge knowledge and understanding of the core concepts and how to adapt them in a good way to a new situation

2. connect the analysis in the report with other areas in statistics or machine learning or previous courses taken in the masters program, i.e. not just repeat what has been done in previous labs.