Uppsala University

Machine Learning

Mini-Project Instructions

Contents

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). Choose a dataset you find interesting!
- Minimally, two methods (or two different neural net architectures) should be compared and evaluated.

Tip! Write the report so you can add your final report as an example of work you have done to future potential employers.

For PhD students: You can choose to do 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 project's requirements.

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.

Hint! If you choose to work alone, try to choose a subject related to your master thesis 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]

You should not use the following data sets in the project:

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

1.4 Project Proposal

Students must turn in a half-page project and data description in the ICML paper format that can be found [here]. The ICML format is also available in overleaf here: [here].

The project report should include all the group members' names following the LaTeX template, i.e. your names should be as authors in the template.

The project proposal must include the following pieces:

1. Title
    - The title should describe the problem and be like a real article title, i.e. don't write "Mini-project:" or similar in the title.
2. Abstract (1 paragraph)
3. Introduction (roughly 0.5 pages)
    - Description of the problem.
4. Data and Methods (roughly 0.5 pages)
    - Description of the data.
    - A first idea on potential methods and evaluation.

1.5 Project Report

The requirements for the projects are:

- 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 include all the group members' names, but following the LaTeX template. I.e. your names should be as authors in the template.
- The supervised problem should be well explained. This includes the target of the prediction y and the features x. What variables are used here, and why is y relevant to predict using x?
- If models with hyperparameters are used, there should be motivation. Why are they chosen and how? You should describe the method used to set the hyperparameters.

The paper should consist of *between three and a half (3.5) and four (4) pages*, excluding references and eventual appendices. The paper should include the following four (4) sections:

1. Title
    - The title should describe the problem and be like a real article title, i.e. don't write "Mini-project:" or similar in the title.
2. Abstract
3. Introduction (roughly 0.5 pages)
    - Description of the problem.
4. Data and Methods (roughly one page)
    - Description of the data.
    - Description and motivation of the method
    - Description of how you conducted the evaluation
5. Results (roughly 1.5-2 pages)
    - Results of the proposed methods
    - Evaluation of the proposed method
6. 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/fairness issues (in light of the guest lecture).
7. Acknowledgements (optional)
    - If you are willing to let me use your project report as an example in the next course, please state that in the Acknowledgements. Just add the following sentence: "We hereby grant our consent for the utilization of this project report as a reference material within the context of future editions of the course."
    - Other people you might want to thank.

The project report should have the same title as the project proposal.

Additional requirements and hints for the report:

1. All Figures using colour should have a colour-blind-friendly colour palette. See here and here.
2. Before you turn in the project, do a language check with a tool such as Grammarly. A project with poor English (errors you would have spotted with a tool such as Grammarly) will affect your grade.
3. The final report should be like a research paper, i.e. try to avoid bullet lists and get a good flow in the text.
4. You should use the correct reference systems. A tip is to use
   `citet`,
   `citep`, and
   `bibtex`. Using
   `bibtex` will simplify your future thesis work.

1.6 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.

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). You should use
- Big enough font size for text and figure labels 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 the group members' names.

Missing the project presentation

Suppose you cannot attend the presentations due to sickness or similar. In that case, you will have to turn in a video presentation where you present the whole project (i.e. the entire presentation). The teacher then grades this presentation.

1.7 Project Grading

Below are the criteria 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), you should fulfil the following criteria:

1. Turned in a correctly formatted report that follows 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 of when certain methods should be used or not and how.
4. Use at least two (2) different methods (or architectures) and correctly compare them.
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 additional criteria also apply:

1. Show deep 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 master's program, i.e. not just repeat what has been done in previous labs.