Uppsala University

Machine Learning

Mini-Project Instructions

Contents

1 Mini-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 to use it as an example of work you have done for future potential employers.

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

2 Project Group and Expected Workload

Having only one student in a group is possible, 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 40 hours per student in the group. Hence, a 3-group project should be the equivalent of a 120-hour project.

One student (first author) is the correspondning author. Thats the one that submit the paper on studium.

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

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)

4 Project Proposals

The project will be submitted multiple times during the course to even out the workload of the report and enable feedback.

4.1 Project Proposal Part 1

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].

Note! You should also follow the formating instructions in the ICML template, e.g. regarding Tables, Figures etc.

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. See the details on the different parts in Section 5.1.

1. Title
2. Abstract
3. Introduction
4. Data and Methods

In the Data and Methods Section, you only need to include the Data subsection and describe the data you will use.

4.2 Project Proposal Part 2

Part 2 contains additional information on the proposed method and some preliminary results. It should also fix comments to the previous proposal 1. Students must submit a project proposal part 2 and data description in the ICML paper format (see above).

The Project Proposal 2 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. See the details on the different parts in Section 5.1.

1. Title
2. Abstract
3. Introduction
4. Data and Methods
5. Results

In the Results section, you only need to include the preliminary result of one model/architecture.

5 Project Report

The requirements for the project report are:

1. 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]
2. You should follow the formating instructions in the ICML template, e.g. regarding Tables, Figures etc.
3. The project report should include all the group members' names but follow the LaTeX template, i.e. your names should be as authors in the template.
4. The paper should consist of *between three and a half (3.5) and four (4) pages*, excluding references, acknowledgements and eventual appendices.
5. The report and project proposals should follow the disposition in Section 5.1.
6. The audience is students in the machine learning course, i.e., there is no need to go deep into details of what has been presented during the course. Although, if models not introduced in the course are used, you need to explain these models in more detail.
7. All Figures using colour should have a colour-blind-friendly colour palette. See here and here.
8. All legends in figures and text in tables should be the same size as the body text in the report.
9. Before you submit 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 like Grammarly) will affect your grade.
10. The final report should be like a research paper, i.e., avoid bullet lists and get a good flow in the text.
11. This is an academic text. Hence, the claims you make should be backed by references.

5.1 Project Report Disposition and Content

The paper should include the following four sections and subsections:

1. Introduction
2. Data and Methods
   (a) Data
   (b) Methods
   (c) Evaluation
3. Results

4. Conclusions

In addition, the report should also include a title, abstract, and a bibliography. The report can also include Acknowledgements and Appendices, but this is optional. Below are further information and requirements for the different parts.

5.1.1 Title

The title should describe the problem and be like an actual article title, i.e., don't write "Mini-project:" or something similar in the title.

The project report and project proposals should have the same title.

5.1.2 Abstract

The abstract summarises your report in one paragraph of a maximum of 200 words. It should also be written in the present tense. We can write the abstract by writing one to two sentences on each of the following points:

1. An introduction to the topic.
2. Explanation of why the topic is important.
3. Your main research question, aim or problem.
4. Your research data, methods and models.
5. Your most important findings.

5.1.3 Introduction (0.5 page)

In the introduction, you should:

1. Give a sufficient background to understand the problem.
2. Description of the problem. The supervised problem should be well explained, including the target of the prediction (y) and the features (x) used. What variables are used here, and why is y relevant to predict using x?
3. Explanation of why the topic/prediction setting is important/relevant.

5.1.4 Data and Methods (1 page)

In the Data and Methods section, you should have (at least) the following three subsections:

1. Data
2. Methods
3. Evaluation

Data Describe the data you are using in your prediction model at a level so it is possible to replicate your analysis.

Methods The methods section should describe your methods so another student can replicate them in the course, independent of implementation language (e.g., R). The independence means that the hyperparameter settings should be explained and presented without reference to, for example, specific R function's default values.

If models with hyperparameters or architectures are used, there should be motivation. Why are they chosen, and how? You should describe the method used to set the hyperparameters.

Evaluation The evaluation should describe how you evaluate different models and what metrics you use. Hence, describe if and how you use training, validation and test sets and motivate the choices made. Then, present the metrics you use for evaluation and motivate these as well.

5.1.5 Results (roughly 1.5-2 pages)

Summarize the results of your models and compare your model results. Analyze the performance of your model and discuss the results. Try to summarize results in Tables and Figures to help the reader understand the results.

If you don't have perfect accuracy, which you probably dont have, you need to include an error analysis of your models in the report. You should then have at least one subsection on called "*Error analysis*" where you manually go through errors made by the model (or a random sample of errors) and analyze where (and maybe why) the model fails in the predictions.

5.1.6 Conclusion (roughly 0.5-1 pages)

Connect your results back to the introduction. Did the method work as expected? Are the results good or bad? Discussion of problems and potential improvements. Also, include a paragraph on potential ethical/fairness issues (in light of the guest lecture).

5.1.7 Bibliography

You should use the correct reference systems. A tip is to use
`citet`,
`citep`, and
`bibtex`. Using
`bibtex` will simplify your future thesis work.

5.1.8 Acknowledgements (optional)

If you can let me use your project report as an example in the next course, please state that in the Acknowledgement. Just add the following sentence: "We grant our consent

for the utilization of this project report as a reference material within the context of future editions of the course."

You can also add other people to thank.

5.1.9 Appendicies (optional)

Suppose there is additional material you want to include in the report that does not fit. You are then allowed to use appendices. Only include appendencies you also refer to from the main text. Also, the reader should not need to read the appendices to understand the main text.

6 Project Presentation

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 group members' names.
- 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.

To be able to get VG on the project the presentation needs to be turned in the day before the presentation at 23.59 the latest.

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 5.
2. Show basic knowledge and understanding of the course's core concepts by using concepts correctly.
3. Show an understanding of when certain methods should be used and how they should be used.
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 template guideline in Section 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 well 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.