Project

Machine Learning II
2021-2022
UMONS
Souhaib Ben Taieb

1 Overview

The goal of this project is to study and experiment with a new machine learning model. The first task is to choose a machine learning model used to solve any learning problem (regression, classification, clustering, etc). You can choose any machine learning model, but bonus points will be given for originality, including if your topic has not been covered in previous courses. Examples include XGBoost, differential privacy, generative adversarial networks, Gaussian processes, density-based clustering (DBSCAN) and energy-based models.

The second task is to find and read some of the relevant references. For example, you might look at some machine learning research papers. Two of the main machine learning conferences are ICML[1] and NeurIPS[2].

The third task is to run some experiments and compare your algorithm/method to at least one alternative on a dataset that you have chosen.

You will write a report describing your algorithm/method, explaining the experiments you have performed and summarizing the key results. Bonus points will also be given if you include theoretical properties of your machine learning algorithm. You will also provide the (commented) Python code used for your experiments. It should run without errors.

We ask the students to form a group of two people, and each group will present a different topic. Note that the topic will be assigned to each group based on first-come, first serve basis. Each group will make a presentation of their work for the class.

2 Project report

The report can be a maximum of **10 pages**, and must abide by the section structure described below.

1. **Section 1: Models**. This section presents the learning problem and models you have considered. Some of the questions you could answer include "what is the hypothesis set?", "what are the parameters/hyperparameters?", "how do you optimize these parameters?", "explain how these parameters control the complexity/flexibility of the algorithm", "does your method scale well with large data sets?", "What are the theoretical properties?", etc.

2. **Section 2: Experiments**. This section describes the data set you have chosen and the experiment you have performed, including a justification of your choices (evaluation metric, optimization procedure, etc).

3. **Section 3: Results and Discussion**. This includes for example graphs and tables, as well as a discussion of the results. You should also present your model fitting, diagnostics, etc.

4. **Section 4: Conclusion**. This includes summary of the findings.

3 Project presentation

You will prepare a maximum of **15 slides** for your project presentation. Each group will be given **20 minutes** for presentation and **5 minutes** for questions. Each member of the team should participate by speaking in the presentation for **10 minutes**.

---

[1] https://icml.cc/Conferences/2021/Schedule
[2] https://neurips.cc/Conferences/2021/Schedule

4   Grading

Overall, you will be graded based on your report, code and presentation. The report will be graded based on clarity of writing, quality of presentation, level of machine learning content, and technical communication of main ideas. Your report should clearly explain what you have done, using figures to supplement your explanation. Your figures must be of proper size with labeled, readable axes. In general, you should try to make your report readable and clear. You code should be reproducible. Your presentation should be clear and understandable to everyone. We would like to remind the students that **plagiarism** will be taken very seriously.

5   Deliverable

- Choose a group of two and submit one topic per group.

- Submit one project report per group.

- Submit one presentation per group.

- Present you work.

6   Deadlines

- **April 1, 11:59 pm**: Submit the names of each member of the team and proposed topic on Moodle.

- **May 8, 11:59 pm**: Upload to Moodle your project **report** and **code**, one per group.

- **May 11, 11:59 pm**: Upload to Moodle your **slides**, one per group. Note that you will not be able to change it after submission.

- **May 12, 10:30 am**: Present your work.

Late submissions will not be allowed.