Machine Learning Project Proposals

Instructions

The project can be done independently or in pairs. If you decide to do the project individually, it should take at most 25 hours to complete. If you decide to do the project in a group, it should take at most 50 hours, equally subdivided. Developing a bigger project has no impact on your grade.

Your project can be selected from the proposal list, down below. The proposals are presented as a general guide, with a few examples: in fact, we invite you to research and propose a personal topic, meeting the Professor to evaluate it. The Professor will guide you to understand if the project is feasible, suggesting possible paths and assessing the overall value of the proposal.

It is also possible to propose bigger projects, to be later expanded to a Master's degree thesis under the supervision of the Professor.

The project is not mandatory and will give you a bonus grade that will be added to your final grade. The grade for the project is usually in the range [0, 8], reserving higher scores only for outstanding projects. Smaller projects will be evaluated (<25h), albeit giving a smaller bonus grade. Bonus grade obtained by submitting a project **substitute any bonus grade given during the course activities** posted on Moodle: hence, you should submit a project only if you are not satisfied with the bonus grade given to your submitted activities of the previous months, and you are willing to give up your current bonus grade.

Project proposal

Once you choose your project, you need to send an email to aiolli@math.unipd.it, containing the project proposal. You also need to send it as a CC to bergamin@math.unipd.it and tcarraro@fbk.eu.

Your proposal should contain:

- References to the dataset chosen
- A brief description of the task chosen
- A summary of the tools (e.g., libraries, frameworks) and methods (e.g., SVM, neural networks) you want to use
- Any relevant reference to prior work you will use to develop the project (e.g., papers, tutorials, videos, etc.)
- Your current level of expertise with the tools you chose

Output

- You need to produce a PDF or a well-documented Colab document. The report has no constraint in size, nor in the software used to compile it (Word, Latex, etc., are allowed).
- You also need to attach any code produced, preferably using Jupyter Notebooks. The code should be executable, with clear instructions (ideally, executing all code cells in the notebook should give no errors).
- You need to publish your work in the Moodle forum.

Deadlines

It is needed to submit your project at least 15 days before taking the written test.

Evaluation

The project will be evaluated considering the following factors:

- Clarity: the work should be clearly organized, stating a research question and your objectives, followed by a recap of the necessary concepts needed to fully appreciate your work, and finally the design of your experiment.
- Critical evaluation: the work should have a critical evaluation of your results, showing a good awareness of machine learning subjects stated in the course.
- Literature: the work should have reputable sources cited.
- Experimentation: the work should discuss a non-trivial experiment.

The following will be positively considered as a bonus (but are not necessary):

- Novelty, e.g. discussing state-of-the-art techniques or proposing novel ideas.
- Interdisciplinarity, e.g. proposing projects spanning material covered by different courses.
- Focus on interpretability, e.g. works that give insight into the inner workings of ML models, good data visualization, etc.
- Theoretical discussion, e.g. studying and explaining the foundation of ML models covered or cited in the course. Note that experimentation should still be a relevant part of the project.

The following is not considered a bonus:

- Model performance: your work should focus first on a solid foundation, showing a good comprehension of the models used.

Proposals (not exhaustive)

1. Study a Machine Learning library/framework and perform some experiments with it, e.g. training a model. Examples of frameworks are Scikit-Learn, PyTorch, and Tensorflow. You can also experiment with libraries for some specific tasks (for any relevant domain, e.g., signal processing, computer vision, NLP)
2. Take a Kaggle competition and see if you are able to place your work in a good position on the ranking. Alternatively, you could take a dataset or model from Kaggle and perform some experiments on it, possibly reporting on the forum discussions about other people's work.
3. Read a reputable survey and do some experiments in the selected research field. (e.g., recommender systems, anomaly detection)
4. Try to reproduce the experiments of a reputable paper/tutorial about models seen during the course. See if you are able to get the same results using the parameters that have been specified in the paper, possibly exploring different datasets or tasks. (e.g. feature visualization, variational autoencoder)
5. Take a dataset and analyze it, following the full machine learning pipeline discussed in class.