6.883 - Final Project Instructions

Spring 2019

Overview

Students taking the graduate version of the course (6.883) will be required to do a final project related to the formulation, design, implementation, and evaluation of machine learning solutions. Projects will be done **in teams**, and students will submit a write-up of their work (paper). Students are encouraged to undertake machine learning projects relevant to their areas of research.
Each team can be up to 4 people.

Topic

As a general guideline, the topic should be related to what has been taught in the course. This is not restricted to things we have already seen, but can be topics to be learned later in class like reinforcement learning or unsupervised deep learning. A natural choice is to find a topic that is relevant to your area of research.

Your first instinct should be to identify a learning problem (independent of the method) that is interesting, and think about several possible alternative formulations. Then find a method we have discussed in class (or will discuss), and try to adapt it in a manner that solves this task, and evaluate how well it actually does. You should identify key problems, issues, or advantages. One strategy for evaluation is to compare two or more methods (with different underlying assumptions) and try to understand differences in their performance. The project will be most interesting to you if you can make it relevant to your research or to a problem you know and care about.

A common hurdle is figuring out how to get the data you need, and this may limit your choices. You don't necessarily have to implement all (or even any) of the algorithms you use from scratch. There are several toolkits available with many learning algorithms already implemented in them. However, if you don't do any implementation yourself, we would expect something much deeper in the way of problem formulation or modeling.

Project Proposal

You are required to submit a project proposal. The proposal should be no longer than 1 page and should include:

- Problem definition and motivation.
- Short description of the data.
- Suggested approach(es).
- Evaluation (how will you evaluate your methods).
- Expected challenges.
- Team members and how do you plan to divide the work.

For the proposal — please submit only one document per team (i.e. only one of team members should make a submission).
The submission will be through the learning-modules web-site.

Report

The report will be a written document of length 4 pages (not counting references) in double-column, conference format (see for example CVPR format), including whatever graphs and tables are necessary to make your point. The 4 pages does not count references, and you have unlimited pages for references. Use 10-12 point type and 1.5 inch left and right margins.

The report is the means by which you communicate the process and results of your project, so it should be clear, coherent, and well written. Do not dump out large quantities of data or code or uninterpreted charts. Emulate the expositional style of a technical conference paper. [1] You do not need a detailed related work section, but be sure to cite and very quickly explain any technical work you referenced in formulating and carrying out your project.

You should describe and evaluate what you did in your project, which may not necessarily be what you hoped to do originally. A small result described and evaluated well will earn more credit than an ambitious result where no aspect was done well. Be accurate in describing the problem you tried to solve. Explain in detail your approach, and specify any simplifications or assumptions you have taken. Also demonstrate the limitations of your approach. When doesn't it work? Why? What steps would you have taken have you continued working on it? Make sure to add references to all related work you reviewed or used.

The main goals are to make clear what your findings are, why you think they came out the way they did, and why that might be important. Be precise enough to allow someone to replicate your experiments (or verify your proofs).

Submission

The submission will be through the learning-modules web-site. You should submit two files:

- `your_mit_user_name.pdf`
- `your_mit_user_name.zip`

The PDF file will be the report, as described above, and the ZIP file should include the code as well as any supplementary material.

Each student submits an individual copy. All the members of the group can share figures and text. However, your submission must clearly specify the **section** that you wrote and were responsible for. Since you are part of a team, your write-up can include shorter versions of sections written by other team members. But your section must be clearly identified and **describe the contributions you made.** Each report must include the names of all the collaborators.

Grading

The grade will reflect aspects of the work such as: modeling skill, originality, execution quality, and clarity of the report.

Resources

You are free to use any open source available Machine learning library (though if you choose one that is not common, you should explain your choice).

[1] Check this paper for an example of how to arrange the sections (yours will be shorter but you can follow a similar structure).

You are welcome to use any publicly available dataset or to construct your own one. If you do construct your own dataset, please verify in advance that you have all the required sources and know how to process them (when you submit the project proposal, you should have already processed at least some of the data). Some resources for data:

- https://nlpprogress.com/
- http://www.cvpapers.com/datasets.html
- https://github.com/jbhuang0604/awesome-computer-vision
- https://github.com/josephmisiti/awesome-machine-learning#python
- https://github.com/keon/awesome-nlp