CS/CNS/EE/IDS 165: Foundations of Machine Learning Project Guidelines

Prof. Anandkumar

January 8, 2019

1 Overview

One of the main goals of this course is for you to explore an interesting machine learning problem of your choice in the context of a real-world data set. Best idea would be to combine machine learning with problems you may have faced in your research area. Projects can be done in teams of two students. Group members are responsible for dividing up the work equally and making sure that each member contributes. If you have a particularly ambitious project idea that cannot be completed by a team of two people, you may propose a team of three students, but you must have a strong justification for such a larger team.

1.1 Deadlines

- Proposal Report Deadline: Due February 2

- Proposal Presentation Sessions: February 5 and February 7

- Final Report Deadline: Due March 3

- Final Presentation Sessions: March 7 and Mach 12

1.2 Project Grading Policy

A major part of your final grade is based on this project. This is how your project grade is distributed among different parts of the project.

- Proposal Report: 20%

- Proposal Presentation: 10%

- Final Report: Due 50%

- Final Presentation: 20%

1.3   Project Evaluation

Evaluation is based on the following factors:

1. **Reproducibility**[1]:
   For all the **Algorithms** presented, check if you include:

   - A clear description of the algorithm.

   - An analysis of the complexity (time, space, sample size) of the algorithm.

   - A link to a downloadable source code, including all dependencies.

   For any **theoretical claim**, check if you include:

   - A statement of the result.

   - A clear explanation of any assumptions.

   - A complete proof of the claim.

   For all **figures** and **tables** that present empirical results, check if you include:

   - A complete description of the data collection process, including sample size.

   - A link to downloadable version of the dataset or simulation environment.

   - An explanation of how samples were allocated for training / validation / testing.

   - An explanation of any data that were excluded.

   - The range of hyper-parameters considered, method to select the best hyper-parameter configuration, and specification of all hyper-parameters used to generate results.

   - The exact number of evaluation runs.

   - A description of how experiments were run.

   - A clear definition of the specific measure or statistics used to report results.

   - Clearly defined error bars.

   - A description of results with *central tendency* (e.g. mean) & *variation* (e.g. stddev).

   - A description of the computing infrastructure used.

2. **Technical quality**:
   Does the technical material make sense? Are the things tried reasonable? Are the proposed algorithms or applications clever and interesting? Do the authors convey novel insight about the problem and/or algorithms?

3. **Significance**:
   Did the authors choose an interesting or a "real" problem to work on, or only a small "toy" problem? Is this work likely to be useful and/or have impact?

4. **Novelty of the work**:
   Is the proposed application and approach novel or especially innovative?

5. **Clarity of presentation**:
   Is the presentation clear? Could we reconstruct the method entirely from the report?

---

[1]Reproduced from: www.cs.mcgill.ca/ jpineau/ReproducibilityChecklist.pdf

1.4 Choosing a Project

There are many avenues that you may pursue for this project and we encourage you to be creative even if you don't think you'll necessarily get "great" results.

After you come up with a project idea, you may consult with the TA's to confirm that the project is sufficiently significant. **Please get your project approved by at least one of the TA's before starting to work on it.**

For instance, the followings doesn't count as a project; Running a standard benchmark/Kaggle challenge/new dataset without any detailed ablation and exploratory studies, or Reviewing a paper and just reproducing original experiments. you need to do a critical study. Present new experiments or find shortcomings in theory and ideally try to resolve that. In the interest of diversity in projects, we will maintain a link on the Piazza, displaying a list of people who have already (tentatively) decided on a project topic. In order that we all learn maximum from the course, we should try to avoid duplicate projects.

2 Project Proposal

Your proposal consists of two parts. A proposal report and a short 5 minutes presentation. One submission of behalf of the whole group will be sufficient.

Proposal report is a short report that includes the following information: project title, who is in your group, describing the problem you've chosen, sketching the general approach you intend to take, and stating the kinds of data you're using. More importantly, you should mention which team member is expected to do what and how tasks will be shared. The proposal mainly serves to give us a chance to make sure you're on a good path and to help us get a sense of who is doing what.

You should write a compelling proposal that describes your project in detail and demonstrates that you have the understanding and ability to complete it. Your proposal should also discuss sources of real-world data for your chosen application or how you plan to obtain real-world data. Since you may wish to use machine learning methods that we have not yet covered, you may need to read ahead. Do not worry if there are particular aspects of the project that you can't answer currently (such as which ML method is best); this is a proposal for future work, after all. However, your plan should demonstrate that you've started to think through the various issues involved with your project and present a compelling argument in support of it.

Proposal presentation session will be held during the second week of February. Each team will make a short presentation (5 minutes for the presentation, 2 minutes for the questions) describing their preliminary results. An important skill in research is to be able to tell in a week or two whether your ideas are basically going to work, well before you've fully done all engineering and experiments. The order of the presentations will be randomized.

3 Final Submission

Your final submission will consist of three deliveries: (1) a group final report, (2) an individual final report and (3) a final presentation.

3.1 Final Reports

Group final report can be at most 5 pages long (include all text, appendices, figures, and anything else), with 1 additional page that can contain nothing but references. If you did this work in collaboration with someone else, or if someone else (such as another professor) had advised you on this work, your report must fully acknowledge their contributions. At a minimum your final report must describe the problem/application and motivation, survey related work, discuss your approach, and describe your results/conclusions/impact of your project. It should include enough detail such that someone else can reproduce your approach and results. You will likely end up with a better report if you start by writing a long report and then edit it down to 5 pages of well-written and concise prose.

We know that most students work very hard on the final projects, and so we are careful to give each report sufficient attention. After the class, we are also considering posting the final reports online so that you can read about each others work. If are okay with having your final report posted online, be sure to give us explicit permission when you submit. Under normal circumstances, all group members will receive the same grade for the final project.

You shouldn't worry about getting "great" results. The idea and your understanding of the machine learning issues involved are much more important than getting "great" results.

    Individual final report is at most a one page report written separately by each one of the team members. It includes the collaboration process such as what you did personally in the project and what you think others did. We want to make sure that each team member has a fair contribution in the project.

3.2 Final Presentation

Final presentation session will be held during last week of classes. Presentations should be 10 minutes long fully explaining your results. You may use any format you wish for the slides. The goal is not to cram as much as possible in the slides, but to provide a clear and concise presentation of the main points of your project.Consider is as a conference presentation when you'd like to present your submitted paper (in this case your final report).