Fianl Project, CPSC 8420, Fall 2020

The main goal of this course is to prepare you to apply machine learning algorithms to real-world tasks, or to leave you well-qualified to start machine learning or AI research. The final project is intended to start you in these directions.

Project Topics

Your first task is to pick a project topic. If you're looking for project ideas, please come to office hours, and we'd be happy to brainstorm and suggest some project ideas.

You can choose one of three kinds of projects:

1. **Application project**. Pick an application that interests you, and explore how best to apply learning algorithms to solve it.

2. **Paper Reading**. Pick a paper published recently (starting from 2015 appreared in ICML, NeurIPS, ICLR, JMLR) that interests you, investigate and explore how this paper made significant/incremental contribution to the field. If the paper contained experiments, you are require to implement it and compare with counterparts. Details including but not limited to derivative, experimental setting are required.

3. **Theoretical Project**. Prove some interesting/non-trivial properties of a new or an existing learning algorithm. (Usually this won't be easy.)

Projects can also combine elements of applications, algorithms and theory. Many fantastic class projects come from students picking either an application area that they're interested in, or picking some subfield of machine learning that they want to explore more. So, pick something that you can get excited and passionate about! Be brave rather than timid, and do feel free to propose ambitious things that you're excited about. (Just be sure to ask us for help if you're uncertain how to best get started.) Alternatively, if you're already working on a research or industry project that machine learning might apply to, then you may already have a great project idea. Once you have identified a topic of interest, it can be useful to look up existing research on relevant topics by searching related keywords on an academic search engine such as: `http://scholar.google.com`. Another important aspect of designing your project is to identify one or several datasets suitable for your topic of interest. If that data needs considerable pre-processing to suit your task, or that you intend to collect the needed data yourself, keep in mind that this is only one part of the expected project work, but can often take considerable time. We still expect a solid methodology and discussion of results, so pace your project accordingly.

Notes on a few specific types of projects:

1. **Deep learning projects**: Since our class discusses many other concepts besides deep learning, we ask that if you decide to work on a deep learning project, please make sure that you use other material you learned in the class as well. For example, you might set up logistic regression and SVM baselines, or do some data analysis using the unsupervised methods covered in class. We may grade these projects using different criteria to make sure that grading is fair for students who have not had exposure to DL before. Finally, training deep learning models can be very time consuming, so make sure you have the necessary computing resources.

2. **Datasets**: You can choose to collect the data if you want, though we don't want you to have to spend much time collecting raw data. The process of inspecting and visualizing the data, in addition to trying out different types of preprocessing and doing error analysis, is often important and tricky in machine learning. If you choose to use preprepared datasets (e.g. from Kaggle, the UCI machine learning repository, etc.), then you are encouraged to do some data exploration and analysis to get familiar with the problem.

3. **Replicating results**: You are expected to upload the codes to our course final project repository in Github, please make sure that your experiments and results are replicable.

Project Parts: Proposal, Milestone, Presentation, and Final Report

You should submit on Canvas as a group: that is, for each part, please make one submission for your entire project group and tag your team members.

Evaluation

Presentation and final report will combine to be the majority of the grade. Projects will be evaluated based on:

1. The technical quality of the work. (i.e., 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?)

2. 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?)

3. The novelty of the work. (Is this project applying a common technique to a well-studied problem, or is the problem or method relatively unexplored?)

It is important you present a solid discussion regarding the learnings from the development of your method, and summarizing how your work compares to existing approaches.

Project Proposals

In the project proposal, you'll pick a project idea to work on early and receive feedback from the TA. Your proposal should be a PDF document, giving the title of the project, the project category, the full names of all of your team members, the Clemson ID of your team members, and a 300-500 word description of what you plan to do.

Your project proposal should include the following information:

1. Motivation: What problem are you tackling? Is this an application or a theoretical result?

2. Method: What machine learning techniques are you planning to apply or improve upon?

3. Intended experiments: What experiments are you planning to run? How do you plan to evaluate your machine learning algorithm?

The project proposal is mainly intended to make sure you decide on a project topic and get feedback from TAs early. As long as your proposal follows the instructions above and the project seems to have been thought out with a reasonable plan, you should do well on the proposal.

Milestone

The milestone will help you make sure you're on track, and should describe what you've accomplished so far, and very briefly say what else you plan to do. You should write it as if it's an "early draft" of what will turn into your final project. You can write it as if you're writing the first few pages of your final project report, so that you can re-use most of the milestone text in your final report. Please write the milestone (and final report) keeping in mind that the intended audience are the instructors and the TA. Thus, for example, you should not spend two pages explaining what logistic regression is. Your milestone should include the full names of all your team members and state the full title of your project. Note: We will expect your final writeup to be on the same topic as your milestone.

Contributions Please include a section that describes what each team member worked on and contributed to the project. This is to make sure team members are carrying a fair share of the work for projects.

Grading The milestone is mostly intended to get feedback from TA to make sure you're making reasonable progress. As long as your milestone follows the instructions above and you seem to have tested any assumptions which might prevent your team from completing the project, you should do well on the milestone.

Format Your milestone should be at most 3 pages, excluding references. Similar to to the proposal, it should include:

1. Motivation: What problem are you tackling, and what's the setting you're considering?

2. Method: What machine learning techniques have you tried and why?

3. Preliminary experiments: Describe the experiments that you've run, the outcomes, and any error analysis that you've done. You should have tried at least one baseline.

4. Next steps: Given your preliminary results, what are the next steps that you're considering?

Presentation

Each team should present the work within 15 minutes in class. The grading will be based on peer review (40%), plus TA and instructor's review (60%).

Final Writeup

We know that most students work very hard on the final projects, and so we are extremely careful to give each writeup ample attention, and read and try very hard to understand everything you describe in it.

Format Final project writeups can be at most 5 pages long (including appendices and figures). We will allow for extra pages containing only 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 write-up must fully acknowledge their contributions. For shared projects, we also require that you submit the final report from the class you're sharing the project with.

Contributions Please include a section that describes what each team member worked on and contributed to the project. If you have any concerns working with one of your project teammates, please create a private Canvas post. We may reach out and factor in contributions and evaluations when assigning project grades.

Code Please include a link to a Github repository or zip file with the code for your final project. You do not have to include the data or additional libraries (so if you submit a zip file, it should not exceed 5MB).

Grading The final report will be judged based off of the clarity of the report, the relevance of the project to topics taught in this class, the novelty of the problem, and the technical quality and significance of the work.

After CPSC8420

After this class, if you want to submit your work to a machine learning conference, you are welcome to discuss with me.

Others

Template

Please strictly follow NeurIPS template and you can find here: `https://nips.cc/Conferences/2020/PaperInformation/StyleFiles`. Your final report should be up to 5 pages (including appendices and figures). We will allow for extra pages containing only references.

Methodology

We don't restrict you to only use methods/topics/problems taught in class. That said, you can always consult our TA you are unsure about any method or problem statement.

Dataset

We don't mind you using a dataset that is not public, as long as you have the required permissions to use it. We don't require you to share the dataset either as long as you can accurately describe it in the Final Report. Using datasets on Kaggle is allowed. However, the project need to focus on model performance and achieve a high leaderboard score to receive high grades. This is because a significant amount of work is needed to formulate the problem, obtain data and preprocessing data, which Kaggle challenges provide you well-defined problems and organized dataset at the start.

Team size

We recommend teams of 2 students, while teams sizes of 1 is also acceptable. The team size will be taken under consideration when evaluating the scope of the project in breadth and depth, meaning that a two-person team is expected to accomplish more than a one-person team would.

Fraction

The course project is 25% of the final grade.

Programming

Any programming language is Okay, and you can use any library for the project such as scikit-learn.

Timeline

1. Proposal: 10/12

2. Milestone: 10/31

3. Presentation: 11/17,19,24

4. Final Writeup: 12/3

Project Ideas

You may refer to: `http://cs229.stanford.edu/projects.html`
as well as: `https://people.eecs.berkeley.edu/~jrs/189/project.html`.

Computing Resource

You can use whatever you have, but it is not our responsibility to provide you GPUs stuff. If your project is mainly about deep learning, then Palmetto may be your option.