CPSC 540 Project

Project Description

For 540 students, there is a project component to the course that will be worth 20% of your final grade. For 340 students, there is no requirement to do a project.

For 340 students, there is no requirement to do a project. However, we do want to allow and encourage undergraduates to do research projects (including this one), so you are able to do it. If you do, if (and only if) it helps your grade, we will swap in the grade on the research project in place of your two lowest-scoring assignments.[1]

These projects are done in groups of 2-3. In some circumstances we will allow groups of 4, but you cannot do an individual project so please find a group early. The final deliverable will be a 6-page report that is due near the end of the exam period. It is expected that this project will be a literature survey, but research projects are also ok. The project will be due close to the end of the exam period.

Project Proposal

If you are in 540, for the final part of Assignment 4 you must a submit a project proposal for your course project (no proposal is necessary for 340; if you want feedback you can ask in office hours). The proposal should be a maximum of 2 pages (and 1 page or half of a page is ok if you can describe your plan concisely). The proposal should be written for the instructors and the TAs, so you don't need to introduce any ML background but you will need to introduce non-ML topics.

You should submit this question as a group on Gradescope, separate from the other assignment questions.

There is quite a bit of flexibility in terms of the type of project you do, as I believe there are many ways that people can make valuable contributions to research. However, note that ultimately the final deliverable for the project will be a report that emphasizes a particular "contribution" (i.e., what doing the project has added to the world). The reason for this, even though it's strange for some possible projects, is that this is the standard way that results are communicated to the research community.

The three mains ingredients of the project proposal are:

1. What problem you are focusing on.
2. What you plan to do.
3. What will be the "contribution".

Also, for the course project note that negative results (i.e., we tried something that we thought would work in a particular setting but it didn't work) are acceptable (and often unavoidable).

I encourage you to follow the following default "template" for the project:

[1] The course is *not* graded on a curve, so 340 students are not hurt by choosing the to skip the project.

1. **Literature review**: you pick a specific topic in ML, read at least 10 papers on the topic, then write a report summarizing what has been done on the topic and what are the most promising directions of future work. In this case, the contribution would be your summary of the relationships between the existing works, and your insights about where the field is going.

The advantage of the above template is that the project will take a somewhat-predictable amount of time. If you want to explore a different style of project, here are some standard "templates":

2. **Application bake-off**: you pick a specific application (from your research, personal interests, or maybe from Kaggle) or a small number of related applications, and try out a bunch of techniques (e.g., random forests vs. logistic regression vs. generative models). In this case, the contribution would be showing that some methods work better than others for this specific application (or your contribution could be that everything works equally well/badly).
3. **New application**: you pick an application where people aren't using ML, and you test out whether ML methods are effective for the task. In this case, the contribution would be knowing whether ML is suitable for the task.
4. **Scaling up**: you pick a specific machine learning technique, and you try to figure out how to make it run faster or on larger datasets. In this case, the contribution would be the new technique and an evaluation of its performance, or could be a comparison of different ways to address the problem.
5. **Improving performance**: you pick a specific machine learning technique, and try to extend it in some way to improve its performance. In this case, the contribution would be the new technique and an evaluation of its performance.
6. **Generalization to new setting**: you pick a specific machine learning technique, and try to extend it to a new setting (for example, making a multi-label version of random forests). In this case, the contribution would be the new technique and an evaluation of its performance, or could be a comparison of different ways to address the problem.
7. **Coding project**: you pick a specific method or set of methods, and build an implementation of them. In this case, the contribution could be the implementation itself or a comparison of different ways to solve the problem.
8. **Theory**: you pick a theoretical topic (like the variance of cross-validation), read what has been done about it, and try to prove a new result (usually by relaxing existing assumptions or adding new assumptions). The contribution could be a new analysis of an existing method, or why some approaches to analyzing the method will not work.
9. **Reproduction**: you take a recent paper published in one of the top machine learning venue (either NeurIPS, ICML, ICLR, AI/Stats, or JMLR), and try to reproduce the results in the paper.

The above are just suggestions, and projects could mix several of these templates together, but if you are having trouble getting going then it's best to stick with one of the above templates (and again I recommend doing the literature review). Also note that the project can focus on topics not covered in the course (like GANs), so there is flexibility in the topic, but the topic should be closely-related to ML.

This proposal is mandatory but will not be formally marked: it is just a sanity check that you have at least one project idea that has an appropriate topic and scope, that you find a group early, and that you allocate some time to thinking about the project. Also, there is flexibility in the choice of project topics even after the proposal: if you want to explore different topics you can ultimately choose to do a project that is unrelated to the one in your proposal (and changing groups is ok too). If you aren't sure what to do, go bug the TAs in office hours (which is a good idea even if you are sure what you want to).

Project Submission Format

The deliverable for the final project is a PDF file. The PDF file should be made using LaTeX, using the NeurIPS 2023 style file:
`https://neurips.cc/Conferences/2023/PaperInformation/StyleFiles`.

Use the "final" option and use the names and student numbers of your group members in the author field.

The main text of the project should be restricted to 6 pages in this format. However, following the NeurIPS format note that references and appendices do not count against the page limit. Appendices are optional but can include additional figures, tables, pseudocode, and proofs. So you can submit a 10-page PDF file, but the last 4 pages can only contain references and appendices. There is no need to fill up the page limit. or have a long appendix. If your project fits in 4 pages of text, then go for it. Short papers are more likely to be read carefully in academia.

Marking Scheme

There are three components to the marking scheme. These components and their weight are:

- 50%: Paper organization.
- 20%: Writing style.
- 30%: Contribution.

More details on each of these items are given in the following sub-sections.

Contribution

The first 70% of the mark above is based on writing a clear paper that follows the structure and addresses the relevant issues in a generic scientific/engineering publication. That part is largely independent of the project you choose. The final 30% percent of the mark is based on the "contribution" of the project, which is what doing this project adds to the world. Note that there is no expectation that the course project yields a novel or high-impact result, but it should still have some contribution. Also, note that negative results are ok; for example, "existing supervised learning methods are not well-suited for this task" is an ok contribution. Here are some ways to argue that the paper has made a contribution:

- The paper provides a summary of how existing works on a particular topic are related to each other, and gives insights about where work on this topic is going [**Recommended**].
- The paper shows that some methods work better than others for a specific application (e.g., random forests work better than logistic regression).
- The paper shows that machine learning is well-suited for a particular type of prediction problem, or that the methods you tried do not work on that type of problem.
- The paper shows how you can solve something much more efficiently or much more accurately than other reasonable approaches.
- The paper shows a new theoretical result, or argues that an existing theoretical technique can not be applied in a straightforward way to a problem.
- The paper could/did make a software package that many people might use. Or it added components to an existing package.

You should explicitly state your contribution in the introduction of the paper, and the paper should argue that the course project does make a contribution. (One of the worst things you can do in your academic writing is leave the reader unclear about what your contribution is.)

Paper organization

Half of the mark for the course project will be based on whether your report satisfies a set of standard criteria for a scientific/engineering publication. If you are doing a literature review, then I *strongly* suggest the following organization:

1. Introduction: Clearly state the problem being addressed. Explain why it is an important problem. At a high level, briefly summarize the history of the works that will be discussed in the project.
2. Review: Go through the different works in some logical order, such as chronologically or by going from simple to complex models. Do not just list the methods, but say how they relate to each other (going through the strengths/weaknesses of the different methods, both in comparison to each other and compared to an ideal method that solves the problem).
3. Discussion: Discuss the trends that have occured over time. Speculate about where the next steps in the trend could lead. Point out issues that are not properly addressed by existing methods. State some interesting directions to explore, or opprotunities to use existing tools in new applications.

For most other projects, I *strongly* suggest using the following traditional outline:

1. Introduction: Clearly state the problem being addressed. **Explain why it is an important problem to work on**. At a high level, briefly summarize what the limitations of existing approaches that your work will be addressing. At a high level, summarize what the contribution of your project is.
2. Related Work: Identify at least three publications on related topics; usually these will be papers that have worked on slightly different problems or papers that have proposed an approach to your problem that is not fully satisfactory. For each paper, briefly say either how the problem addressed is related and/or different (if they address a different problem) or why it doesn't solve the problem you are working on (if it addresses the same problem). **You do not want the reader to be confused about whether the problem you are addressing is already solved by related work**.
3. Description and justification of what you did (divided into sub-sections): There is a lot of flexibility here, and it will depend on the type of project you are doing. For example, if you are applying standard machine learning methods to a new dataset or doing a Kaggle competition, you should have one sub-section describing the dataset and why you think machine learning could help, and one sub-section stating the methods you will try and why you think these are appropriate methods (you do not have to go into detail describing the methods if they are covered in class). If you are extending an existing technique, you should have one sub-section describing the existing technique, and one sub-section for each of the extensions you explored. If your project has a theoretical component, you should have one sub-section discussing the assumptions, one sub-section describing the results, and one sub-section describing implications.
4. Experiments and/or analysis (if you have an experimental component): Describe each experiment that you did. Say what each experiment is trying to test. Ideally, each experiment should only try to test one thing and you should control for as many other factors as possible. Subsequently, summarize the result of your experiment, in both the text and in a nice visual form such as a figure; in most cases, **a table with a huge list of numbers is not a nice way to summarize information**.
5. Discussion and future work: State the main conclusions that are obtained from this course project. List at least one strength and one weakness of your contribution. Briefly state what you would do with more time.

50% of your mark will be based on whether the sections of your project report achieve the goals described above for each section. For example, for the introduction: (i) is the problem clearly stated? (ii) is it explained why the problem is important? (iii) are the limitations of existing approaches briefly summarized (or history for literature reviews)?, and (iv) is the contribution of the project briefly summarized?

Other structures would be more appropriate for different types of projects, but this portion of the mark will be based on whether the project is logically structured: there should be some introductory material saying *why* this project is interesting and what the contribution is, some discussion of the literature and possibly presenting a new idea/approach, and then a discussion of where this fits into a larger view of science/engineering. If you are not sure how to structure your paper for the particular type of course project you are doing, message me on Piazza and I can give you some guidance.

Writing Style

20% percent of the mark is based on writing style. I put some notes on writing clearly and some common writing conventions here:
`https://www.cs.ubc.ca/~schmidtm/Courses/Notes/writing.pdf`

Project Repository

Some previous projects have been posted to Piazza as examples.