CPSC 440/550 Machine Learning (Jan-Apr 2024)
Project Proposal – due Friday March 29 at **11:59pm**

As discussed in the syllabus, 40% of the final grade for the course is from:

1. CPSC 440: the best of the final exam or the research project.
2. CPSC 550: the average of the final exam and the research project.

If you're a 440 student and confident you want to only do the final, you don't need to hand in a proposal.

If you choose to do the research project, 10% of the corresponding part of the final grade is for this proposal. That is, if you're a 440 student doing only the research project, the research project proposal will count for 4% of your final course grade, and the research project writeup will be 36% of your final grade. If you're a 550 student, this proposal will be worth 2% of your final grade.

Projects are **strongly encouraged** to be done in groups of 2-3. If you really want to do it alone, that's allowed, but **expectations will not be decreased** so it'll be harder, and it will probably be a worse learning experience for you as well – ML research is almost always collaborative.

These proposals are short and lightly graded: if you put some effort in, you'll get full proposal marks. The point is so that (a) you form your project groups now, (b) you think about what you want to do, and (c) we check that the scope and topic of the project is suitable for the course.

It's okay with me for your project to overlap with a project in another simultaneous course, as long as you check with the other instructor (though do an appropriate writeup for each course). Having some overlap with your ongoing thesis research/similar is also fine, but make sure this is a relatively discrete project with a clear scope and will be self-contained enough that the course staff can read and understand it.

The proposal is not necessarily "binding" – research projects very often shift from the idea you started with, and it's even okay to do something totally unrelated to what you proposed. If so, though, you probably want to check in with me or a TA that the new project is in scope: if you hand in a final project we think is very out of scope for the course without having checked with us, you'll get a bad grade on it.

The proposal can be done in any format you'd like; default LaTeX style with `\usepackage{fullpage}` is fine.

The proposal should be a maximum of 2 pages; it's okay to be shorter if you can describe the plan concisely. The proposal should be written for the instructor and the TAs, so you don't need to introduce ML background that's covered in the course or that you would reasonably expect the TAs (graduate students working in ML) to know, but you should introduce any required background for non-ML topics.

There is quite a bit of flexibility in terms of the type of project you can do, since there are many ways that people can make valuable contributions to research. The final deliverable will be a written report consisting of at most 6 pages (in a LaTeX format to be provided), with unlimited additional space for references and possible appendices (which, as with e.g. NeurIPS reviewers, you shouldn't count on the graders reading). That report should emphasize one particular "contribution": what has doing this project added to the world?

The three main questions your project proposal needs to answer are:

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

For the course project, negative results are acceptable (and often unavoidable). In that case, the paper should probably include something like "here's why we thought this thing we tried would work in this setting, here's us convincing you that it didn't, and here's our best understanding at why we think it failed."

Here are some standard project "templates"; you might want to roughly follow one of these templates, but it's fine if your project mixes and matches between these project types, or does something else entirely.

Some of the examples below include topics not covered in the course (like random forests): there is flexibility in the topic, but it should be closely related to ML, and ideally use tools covered in this course.

- **Application bake-off**: you pick a specific application (from your research, personal interests, 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 classifiers). In this case, the contribution could be showing that some methods work better than others for this specific application, and hopefully some idea why – or your contribution could be that everything works equally well/badly.
- **New application**: you pick an application where 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 (and perhaps how to prepare the data and constructed features).
- **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. (For example: how do we apply kernel methods when n is very large?) Your improvements might be a new approximation, a distributed version, a smarter implementation, or so on. 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.
- **Improving performance**: you pick a specific machine learning technique, and try to extend it in some way to improve its performance (for example, how can we efficiently use use non-linearities within graphical models). In this case, the contribution would be the new technique and an evaluation of its performance.
- **Generalization to new setting**: you pick a specific machine learning technique, and try to extend it to a new setting (for example, making a graphical model version of random forests). In this case, the contribution could be the new technique and an evaluation of its performance, or could be a comparison of different ways to address the problem.
- **Perspective paper**: you pick a specific topic in ML, read a large number of 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.
- **Coding project**: you pick a specific method or set of methods (like independent component analysis), 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.
- A **reproducibility report** of a recent paper, as in the ML Reproducibility Challenge: we missed the deadline for the official challenge, but you can draw inspiration from the challenge and past submissions there. This is pretty similar to the "coding project," but with slightly different aims.
- **Theory**: you pick a theoretical topic (like the variance of cross-validation or the convergence of stochastic gradient in non-smooth and non-convex setting), 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.

Any one of these project "types" is enough to get a reasonable project grade. Some are naturally more ambitious than others, though; grading will take into account the ambition of your scope as well as how well you execute it.