ISTD 50.570 (Machine Learning): Final Project

Important Notes

- Proposal due: due March 3, 5PM (via Email),
- Proposal discussion: March 4 (sign up for time, via this link: https://goo.gl/ddsZ3i)
- Project report due: April 22, 5PM (via Email)

Process

Similar to the final project done by MIT graduate students, the final project in ISTD 50.570 will consist of the following components: proposal and proposal discussion, research project and project report.

The proposal (one per team) will be a written document, 1–2 pages long, outlining the work to be done. It should include a plan with at least 4 steps, and indicate your internal deadlines for each of those steps. If there are multiple participants, the division of responsibility should be made clear. In addition, please include your assessment of what the "risks" to the project are: that is, what things do you think might turn out to be more difficult than planned, and what thoughts do you have about how to mitigate the risks?

If you are going to do an empirical study, be sure that you think about what method to use as a "baseline". It might be running a simple off-the-shelf algorithm or comparing to what happens if you predict the most common class.

Remember that almost anything will turn out to be harder and more time-consuming than you expect. Try to arrange your project so that there are intermediate milestones that can serve as alternative finishing points, in case you don't get to the end. It will be much better to turn in a polished version of a small-scale project than to find yourself at the end of the term with a three-quarters implemented system of great depth and scope.

A proposal interview (one per team) of about 10 minutes will be scheduled for all teams. Note that the proposals are due the night before the interviews so that the instructor will have a chance to review them.

The project report (one per team) will be a written document of no more than $5n$ pages (where n is the number of members in your team), including whatever graphs and tables are necessary to make your point. 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.

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 and to be precise enough to allow someone to replicate your experiments (or verify your proofs).

Projects

You have 1 month to do this project, and we expect it to take about 8 person-hours per week; so that's 40 person-hours for a single-person project and 80 person-hours for a two-person project, etc. You'll have to make a plan and stick to it, to avoid getting behind and doing a bad rush job at the end.

Here are some ideas for types of projects. Replicating an experiment is probably the best option if you don't have a concrete idea of something different to do.

Apply a technique Take one or more of the methods that we have talked about in class, or that we are about to cover, and apply them to a problem. Compare their performance and elucidate why they perform differently, if they do. Do they do a good job on the problem? This is most interesting if you can apply it to some other research question or problem you know about. A big issue here is being sure that you can get the data you need. You don't necessarily have to implement all (or even any) of the algorithms you use. 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. There are repositories of data available. **Running existing implementations of algorithms on standard data sets is a bare minimum project which cannot earn more than a C.**

Replicate an experiment Often, the best way to understand something is to replicate an experiment reported in the literature. There are huge numbers of papers in these journals and conferences, some of which are good, some bad, some hard, some easy.

Note that replicating someone else's results is notoriously difficult. There are often a lot of things left unsaid in technical papers, which have a real effect on the outcome. Part of the value of trying to do this is learning how to be clearer and more complete in future papers you may write.

Something else If you are more theoretically inclined, and have an idea for a theoretical direction to pursue, or want to do something else different, we're certainly open to the idea. The proposal / discussion process will be very important in this event.

Collaboration

You may do your project in groups of up to 3 (if you have a very strong proposal and a good reason, we would consider groups of larger size). If you work in a group, you must: Make clear before you start what the division of labor will be. Make clear in the written report what the division of labor actually was (it's fine if it deviates from the proposal, but it must be specific and accurate). Be sure that all participants understand all of the work (it will not be a sufficient answer, to a question in the potential after-submission interview, that you don't know something because that's the part your partner worked on). Projects done by n people will be expected to have n times as much technical depth and content as those done by a single person. For joint projects, the written work may be done jointly. Be sure to cite all papers and web sites consulted during the course of your project and acknowledge anyone that provided help.

Grading

The grading will be broken down as follows: research project (70%) project report (30%).