

6.435 Final Project

Spring 2022

Important Dates:

- Pre-proposal title and abstract due: Monday, March 7, 11:59pm (Boston time zone)
- Proposal pdf due: Monday, March 14, 11:59pm (Boston time zone)
- Project progress report pdf due: Wednesday, April 6, 11:59pm (Boston time zone)
- Project final report pdf due: Friday, April 29, 11:59pm (Boston time zone)
- Project presentations: final class meetings

Note: All written components should be emailed to the course staff email (see [6435_overview.pdf](#)) with a clear subject line: “6.435 Project Pre-Proposal”, “6.435 Project Proposal”, “6.435 Project Progress Report”, and “6.435 Project Final Report”.

Process

The final project in 6.435 will consist of the following components: a pre-proposal, a proposal, a research project, a progress report, a project report and a brief final presentation. We will allow project teams with one, two, or three people.

The pre-proposal (one per team) will be a written plain-text document or pdf document consisting of a title and abstract. It should be about one paragraph in length per team member, and it should detail the high-level idea for the project.

The proposal (one per team) will be a written pdf document. The length should be at most one page per team member. The document should outline the work to be done. You should answer the following questions:

- What do you want to do? What question are you answering?
- How does your question and your approach relate to this class?
- What data will you use? Give a specific description of any data sets you will use and confirm that you already have any data **in hand** at this point.
- What is some relevant related work? (Brief summary.) You should clearly have read (in-depth) at least one research paper outside of the assigned class readings, and it should be referenced in your proposal.

- What is your project plan? You should include at least 4 steps per team member and should indicate rough internal deadlines for each of those steps. Each step should be substantive and specific to your project; “reading”, “writing”, “preparing my presentation” do not count as steps. And “data collection” should not be a step since you should already have your data by the proposal time. If there is more than one team member, each step should be labeled with the corresponding team member.
- What are the risks? What might turn out to be more difficult than you anticipated? And how might you mitigate these risks?

If you are going to do an empirical study, be sure that you think about how you will evaluate your method—ideally in comparison to other methods.

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.

The progress report (one per team) will be a written pdf document. The length should be at most 2 pages per team member. The report should cover what has been accomplished so far and outline the work that remains to be done. You should answer the following questions:

- What have you done so far? Report any preliminary findings (positive or negative).
- What do you plan to do with the time remaining? Include at least 2 steps per team member and indicate rough internal deadlines for each step. Each step should be substantive and specific to your project, as above. If there is more than one team member, each step should be labeled with the corresponding team member.
- Describe the general layout of at least one plot that will appear in your final report. What are the horizontal and vertical axes? Why is your proposed plot useful? What will the reader learn from it? You do not need to have actually created the plot at this point. (Note: make sure you get pre-approval from the course staff if you anticipate having a project without plots in the final report.)

The project report (one per team) will be a written pdf document of at most 4 pages per team member. It should include any graphs and tables that 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 expository style of a technical research 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. Be precise enough to allow someone to replicate your experiments (or verify your proofs).

Projects

You have most of the semester to work on this project. 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 the analysis of real data. Compare their performance and discuss 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.

You don't necessarily have to implement all (or even any) of the algorithms you use. Many authors post their code publicly, and there are a number of toolkits available with algorithms already implemented (e.g., Stan, Edward, PyMC, BUGS, R packages on CRAN, etc). However, if you don't do any implementation yourself, something deeper is expected in the way of analysis, problem formulation, or modeling.

If you decide to implement an algorithm, keep in mind that there may be numerical problems; for instance, problems may be ill-conditioned, or products of probabilities may go to zero (necessitating the use of logs for intermediate values).

There are repositories of data available; links will appear on a project info post on Piazza.

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. If you're going to pick your own paper, please run it by the staff before you hand in your project proposal.

The project info post on Piazza will have a list of papers you might try to replicate. It's ultimately an arbitrary selection and in no way exhaustive. Many of the papers are journal articles, and may be too big to replicate all of; if so, pick a subpart. If one of these topics interests you, check the references in the paper; maybe you'll find a precursor paper that is a better starting point.

Note that replicating someone else's results is notoriously difficult. There are often a lot of things left unsaid in technical papers, many of which have a real effect on the outcome. Part of the value of trying to do this replication 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, this type of project can be great. It would be a good idea to discuss your plans in office hours before the proposal in this event.

Collaboration

Make completely clear in every submission which software you wrote and which software you used but did not write.

You may work on your project in groups of up to three. If you have a *very* strong proposal *and* a good reason, larger groups may be allowed. 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 the update must be specific and accurate.
- Be sure that all participants understand all of the work.

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, as well as to acknowledge other individuals (students, professors, etc.) who helped you substantially.

Relationship to other classes You may be able to use a single project for 6.435 and another class that you are taking concurrently, or a graduate or undergraduate research project. If it is one project for two classes you must: (1) produce a project that is twice as large in depth and content as would have been required for either class individually; (2) obtain permission from the instructor of the other class; and (3) make clear to us which other class this project is being used for. If it is for your graduate or undergraduate research, make the context clear, and delineate the part of the overall project that is to be considered your project for this class.

Grading

The grading will be broken down as follows:

- pre-proposal (5%)
- proposal (10%)
- progress report (20%)
- research project: technical content (30%)
- project report: presentation, writing, clarity (25%)
- presentation (10%)

2% will be deducted from the grade of the relevant component (e.g., final report) for each hour late.

More Writing Details and Advice

- References do not count toward any page limits; be sure to reference appropriately.
- Keep your writing below the page limits. Think about how to convey rich ideas succinctly. Editing down to the page limit will take extra time; plan ahead.
- You may use appendices if needed, but assume the staff will not read anything after the page limit given in the project description to come up with your grade. Note that this is standard advice to reviewers at typical machine learning conferences.
- Write clearly! You only hurt yourself by not communicating your ideas. Aim at your fellow students in this class who know nothing about your project. Equivalently, aim for an ICML reviewer with some vague knowledge of the area but not necessarily the specific background for your paper.

- You are welcome, but not required, to use the ICML style file linked from here: <https://icml.cc/Conferences/2022/StyleAuthorInstructions>. Note that the two-column format (as in this style file) is a way to make great use of limited page space. Please do *not* use weird fonts or margins in an effort to gain more space; if you make font or margin choices that are difficult to read (in your main text or any tables or figures, etc.), these choices may dramatically decrease your presentation score.