Project Proposal Guidelines

Due Date: Nov 12, 2022 at 11:59pm EST

Instructions: Putting your thoughts on paper via a project proposal allows you to evaluate the scope, feasibility, strengths, and weaknesses of your project idea (and re-evaluate it if necessary) before getting too deep into the implementation phase. It also gives me a chance to give you early feedback along roughly the same lines. Being able to write short and insightful project proposals is also a valuable skill to develop regardless of your future career trajectory, whether it be one in academia, industry, or government; entire projects are often funded on the basis of short pitches, abstracts, or grants that read very much like a project proposal.

The proposal should be written independently, and submitted on Gradescope. A minimal template for the proposal is provided on Piazza. Please remember that you cannot use late days for this submission, and late submissions will be assigned a score of 0% (except under extenuating circumstances.)

The assignment is graded on a pass/fail basis. You will receive full points if your proposal meets the standards set out in the requirements below, and 0 otherwise.

Requirements

Your proposal should be 1-1.5 pages long (no more than 2 pages), and cover the following key points:

1. **Background**: provide some brief background about the problem you are interested in tackling; why is it important, and to whom?
2. **Hypotheses and goals**: frame one or more hypotheses of how you think machine learning can help improve the current state-of-the-art and assist humans in decision-making for your problem of interest. While framing your hypotheses, ask yourself what the limitations of existing methods are, and what gap you are trying to address. For example, in the protein binding case study, we talked briefly about the enormity of the search space, and how machine learning might help augment the current understanding of oncologists. Note: I don't expect you to do an extensive literature review at the same depth of a full publication, but investing some effort here helps motivate your problem better, and can also ultimately help you build better models by leveraging domain knowledge.
3. **Sources of data**: What kind of data sources have you looked into, and list one or more datasets (with references) that you feel might be good candidates for your analysis.
4. **Sources of complication**: What kind of complications do you foresee in building ML models for your problem? This could include messiness of the data (e.g., missing data, selection bias), small sample sizes making it difficult to fit more complex models, class imbalance, ethical considerations etc.
5. **Rough plan of analysis**: Are there specific methods that you are interested in applying? What are the main steps you intend to pursue to overcome some of the challenges you have written about? This

does not have to be fully sketched out yet; we will also talk more about helpful software tools you might consider using for the analysis in your final project.