

Final Project and Proposal

Due: Nov 6/Dec 4-11; 16

The goal of the final project is for you to explore some topic or topics in more depth than we'll have time for in this class. While I don't expect you to produce publishable quality work in half a semester, I *do* expect you to demonstrate some initiative and creativity. Specifically, I want you to clearly formulate a question, develop a plan for answering or addressing that question, carry out your plan, evaluate your results, be able to describe the relationship of your work to previous efforts in this or closely related areas, and be able to present the results of your study to me and your classmates in both written and oral forms. Basically, you can think of this project as a microcosm of a research or development project that you might do in academia or for a company.

First and foremost, I encourage you to talk to me about your project! Especially before the proposal, but also as your project progresses, if you have any questions about what you're doing, things you should consider, scope, etc., please ask me. Remember that a half-hour of your time talking to me early, could save you a week of going down the wrong direction or prevent you from spending a month working on an insurmountable (or maybe inadequate) project!

1 Topics and Types of Projects

The topic of your project is largely up to you — I want you to choose a question or an area that intrigues you and that you'd like to explore in some depth. That said, I *do* require you to spend at least some effort on a topic that we haven't covered in class. Possibilities include (but are not limited to) “core” ML topics that we don't have time to cover in class (such as relational learning, grammar induction, cognitive modeling, metalearning, etc.) as well as “application” areas (such as bioinformatics, real-time game agents, dialog systems, stock prediction, etc.) You are free to do a hybrid project that combines two or more areas (including some from the topics we *have* discussed in class), so long as some aspect of your project reaches outside the scope of what we've covered in class.

There are also many different *types* of projects that are available to you. The following are some suggestions, though others are possible:

Development/implementation The purpose of this project is to learn about a handful of learning algorithms in depth by implementing them and testing them on some benchmark data. The number of things you'll need to implement depends on their complexity — a single, very complex or difficult to understand algorithm could be sufficient, though more often 3 or 4 algorithms would be required. The benchmark data should be chosen to demonstrate that the algorithm performs as advertised and to highlight performance characteristics.

Empirical The goal of this project is to understand the performance of a number of different competing algorithms by testing them on a common benchmark suite of data sets or problems. The implementations of the algorithms can be off-the-shelf, but the data/example problems should be carefully chosen to highlight differences between the algorithms and demonstrate each algorithm's strengths and weaknesses. The data may be a mixture of carefully developed synthetic data or problems with established benchmark data/problems.

Data or problem driven In this project, the goal is to address some specific, hard, data set or problem. The data/problem may be novel or may be known in the literature, so long as it's a current and difficult problem, with no established solution (e.g., some bioinformatic data sets are currently widely studied, but no consensus yet exists on how best to analyze them). In this project, you are free to use off-the-shelf algorithms, so long as you try at least one thing that hasn't yet been applied to this specific problem.

Reading and synthesis Your project need not include a development or data component. A very interesting project can be built by examining two or more different topics (at least one of which falls outside of the scope of the class lectures and standard textbooks) and describing their interactions. E.g., you might examine the intersection between support vector machines and relational learning or between learning and vision. For almost any pair of topics you can find, people have probably examined something in the intersection. This project involves doing fairly substantial reading in the two fields and their intersection (possibly 20–30 references), reporting on how they interact, and, hopefully, proposing novel directions for future research in the intersection of the fields.

Mathematical If you're mathematically inclined, there are a number of fascinating open mathematical questions in ML. If you're interested in something along these lines, please see me for details.

2 The Proposal

The first concrete element of the project that is due is the proposal. The purposes of the proposal are basically to ensure that you've given some thought to the project, that you have some clear idea of what you intend to do, that you've done enough background reading to understand some of the context of the problem you're interested in, and for you to state your goals clearly so that I can provide feedback. Essentially, I want to see that you've done enough thinking and background reading to ensure that you're on the right track and that you've got a question in mind that's the right size — neither too big nor too small. What you are to hand in is a short (perhaps 3–5 pages) description of what you intend to do and how you'll evaluate your results when you're done.

Like any proposal, you're trying to convince some audience (me, in this case), that (a) what you intend to do is interesting or useful, (b) it is within your capabilities, and (c) your methods or results will be useful, innovative, interesting, or otherwise notable. (Note that those criteria are the same whether you're making a thesis proposal or a proposal for a multi-million dollar development contract.) In order to do that, there are some specific things you need to accomplish in your document:

Define the problem You need to clearly lay out what question you're trying to answer, what you intend to develop, what areas you'll be reporting in, etc.

Establish context and importance The next thing the reader needs to know is why we should care about what you intend to do. What will it accomplish? What will you and we learn from it? Why is that important? How will it benefit the world? To establish this, you usually have to demonstrate that someone cares about doing what you intend to do. Sometimes it's

self-evident (“my study will contribute to the quest for a cure for AIDS”), but often you have to make more of a case for it. Since an in-class project is of limited scope, it may be enough for you to show that a number of other people have studied the problem already or that some other group considers this to be an interesting problem.

Examine related work You should spend at least a little time discussing what people have already tried in this area. The idea here is both to show that you’ve looked enough at the problem to know what you’re talking about, and to establish that your approach isn’t reinventing the wheel. I.e., show not only what other people *have* done, but also that what you’re doing is *different* in some way.

Describe your methods Lay out what you plan to do and show how it will answer or address the question you’re posing. While you clearly won’t know *everything* about how your project will go at this point, you should be able to say some high-level things like “I will implement algorithm X, using the following programming language and supporting tools/libraries”, “I will test on the following data sets”, “I will do the following experiments and measure the following quantities”, “I will investigate the following mathematical properties”, etc. You should also describe how each of the parts of your methods contributes to addressing your initial question.

Define the criteria for success and the evaluation methods How will you know if your project has succeeded? What specific, measurable goals are you trying to achieve? What constitutes an answer to your question or a solution to your problem? And, since this is just a class project and you don’t necessarily expect it to succeed, how should we evaluate partial progress? If you don’t fully answer the problem, what will you be able to demonstrate as progress? **Note!** Since I will grade you, in part, on how well you achieve your *evaluation* criteria, it behooves you to think carefully about this and describe them clearly.

There is no hard and fast requirement on the length of your proposal. Your proposal should be long enough to address all of those points, but no longer! To do a truly thorough job on all of these can require a substantial document — thesis proposals run to dozens of pages, while development proposals for hundred-million dollar contracts can be hundreds of pages. For this class, I expect that you should be able to answer each of those points in about 3–5 pages, but you’re welcome to take more or less space as necessary. To address the importance and related work questions will probably require you to cite 4–10 papers or related works, depending on what you’re discussing.

3 Grading, Deliverables, and Due Dates

I will grade you both on the content of the project and on delivering/reporting it. The content part of the grade is an evaluation of how well you define and lay out the problem (largely in the proposal), how well you address it, and how well you meet the evaluation criteria you establish in the proposal. The delivery part of the grade depends on how clearly you present your material in written and oral forms. It will cover things like organization of your material, how well you argue your points, correct use of spelling and grammar, proper use of citations, oral presentation organization and style, etc.

The items you'll have to turn in are:

Proposal due Nov 6. (Described in this handout.) *Approximately 20-30% of the project grade.*

Oral report due in class, sometime in the last 2 weeks of classes. This will be an approximately 15 minute presentation of what you did for your project and what your results were. I will make additional details available on this later. *Approximately 30% of the project grade.*

Final report due Dec 16 (the day of the final exam). This will be a written report that documents in detail your progress on your final project and describes your results. Again, details will be available as the time grows closer. *Approximately 40-50% of the project grade, including the next item.*

Code, data, etc. due Dec 16 (with the final report). You should turn in a copy of any code you developed, any novel data sets you employed (i.e., data that is not already widely available), any results of experiments that aren't included in the report, etc.