Final Project Guidelines

David M. Blei
Columbia University

August 11, 2025

The final project is an important part of this course. There is no better way to learn about probabilistic modeling than by applying it to a research question of your own.

We expect you to choose your topic. The only criterion is that it deeply involves the concepts, or extensions of them, that we discuss in class.

From past offerings of the class, most projects involve applying probabilistic models to a real-world problem. You choose a dataset, an interesting question about it (e.g., from your research), and address it with probabilistic modeling.

There are other types of projects too. Some projects are algorithmic, such as those which develop and study a new inference algorithm for a class of probabilistic models. Others are theoretical, such as those that investigate mathematical properties of a class of probabilistic models. Some projects combine algorithmic goals, theoretical goals, and applied goals.

The project involves three assignments. For each, please use the LaTex template. Page limits are without figures; include as many pages of figures as needed.

Project Proposal

The project proposal is an abstract that imagines the completed project. We understand that your project will evolve and change over the semester, but writing an abstract early is a good way to plan and think about what it would mean to successfully complete it.

We encourage you to refer to computer science conferences (such as *Neural Information Processing Systems* and *International Conference of Machine Learning*) or journals (such as *The Annals of Applied Statistics*, *Journal of the American Statistical Association*, *Journal of Machine Learning Research*) to get a sense of how to write an abstract.

Consider the following questions.

- What is the problem that I am solving?
- Why is this problem worth solving?
- What is my strategy for solving it?
- What did I find?

A good abstract answers all of these questions, usually in this order, and at a high level.

(We do not grade the project proposal.)

Project Milestone

The project milestone describes the problem you are addressing in detail and discusses some preliminary results. Include what you have completed and what you plan to finish by the end of the semester.

The project milestone is 1–3 pages long. (We grade the milestone.)

Project Report

The project report is due at the end of the semester.

- The report should include an abstract.
- The report can be up to 5 pages. You can also include appendices (of any length) to which you can banish the details. Note: We may not carefully read the appendix when grading your work.
- We encourage you to put source code online and open source, but please do not hand in hard copies of source code. In the report, you can discuss the implementation at a high level. (For example, you can discuss if you used existing open-source implementations or implemented the algorithm yourself.)

Project Evaluation

We evaluate each project on the following dimensions. Very few projects excel in all of them. But the best projects will have high marks along many.

- Ambition : How ambitious would this project seem by describing the main idea?
- Significance:
  - (High) The project would be seriously considered at a major conference.
  - (Medium) The project would be seriously considered at a good workshop.
  - (Low) The project would not be considered at a conference or workshop.
- Technical depth:
  - (High) The project requires significant preprocessing and coding from scratch, or deriving novel mathematics from scratch.
  - (Medium) The project relies on existing programming tools, like Stan or Pytorch, or existing mathematical results.
  - (Low) The project relies on simple libraries, like regression or others in scikitlearn.
- Originality:

  - (High) The project represents a brand new problem space, where there have been no more than a few papers trying to achieve the same goals.
  - (Medium) The problem/method is not totally novel, but has added wrinkles and complications that make the method interesting.
  - (Low) The project does not shed new light on a problem. It reimplements an existing method.
- Results:
  - (High) The project involves many quantitative experiments, tuned hyperparameters, evaluated results, and provides insightful qualitative analysis.
  - (Medium) The project involves some, but not many, experiments. It does not sufficiently evaluate results or provide much qualitative analysis.
  - (Low) There are few results or analysis.
- Writing quality and clarity:
  - (High) The report is clear. It is easy to understood all background, methods, and experiments.
  - (Medium) The report is generally clear, but leaves the reader with some important questions..
  - (Low) The project is unclear and difficult to understand. There are many typos.
- Relevance to class
  - (Yes) There is a clear connection to probabilistic modeling.
  - (No) The project does not take a probabilistic modeling perspective.