STCS 6701: Probabilistic Models and Machine Learning

David M. Blei
Columbia University

August 26, 2022

Description. *Probabilistic Models and Machine Learning* is a PhD-level course about how to design and use probability models. We study their mathematical properties, algorithms for computing with them, and applications to real problems. We study both the foundations and modern methods in this field. Our goals are to understand probabilistic modeling, to begin research that makes contributions to this field, and to develop good practices for building and applying probabilistic models. (Note: The course is officially listed as *Foundations of Graphical Models*.)

Prerequisites. The prerequisites are: knowledge of basic probability and statistics, calculus, and some optimization; comfort writing software to analyze data; familiarity with a good programming language for statistics and machine learning, such as R or Python.

Requirements and Grades. The requirements are weekly reader reports, several homework assignments, and a final project.

- Each week, you write what you thought about the reading. (Later in the semester, you also write about your progress on the class project.) These papers can be up to one page; they can be as short as one paragraph. (Short is good.) They are required, but not graded. They must be handed in during class on Thursday. Late papers are not accepted.

- The homework assigments involve problems, programming, and data analysis. There will be three assignments throughout the semester. Each student has five late days to use for late homework.

- The main requirement is the class project. Most projects involve using and developing probabilistic models to analyze real-world data. Some projects will develop novel theoretical research in probabilistic models. Ideally, your project connects to your doctoral research.

  A project report is due at the end of the semester. There will also be some intermediate assignments to check your progress. We grade your project on both content and writing quality.

Please prepare all written work using the LaTeX templates we provide.

Your grade is decomposed as follows:

- Final Project: 60%
- Homeworks: 30%
- Reader reports: 10%

Independent of this breakdown, you must hand in all homework, at least 10 reader reports, and submit a final project. Without satisfying these requirements, you may not receive a passing grade.

Syllabus

Below are the subjects we cover and in what order. (It may change.)

The Basics of Probabilistic Machine Learning

1. Introduction and the ingredients of probabilistic models
2. Basic distributions, conjugate priors, and the exchangeable data model
3. Conditional models: Linear and logistic regression, stochastic MAP estimation

Motifs and Algorithms in Probabilistic Models

4. Bayesian mixtures and the Gibbs sampler
5. Mixed-membership, topic models, and variational inference
6. Matrix factorization, recommendation systems, and efficient MAP

Generalizations and Advanced Ideas

7. Exponential families, conjugacy, and generalized linear models
8. Deep learning: A probabilistic perspective
9. Black box variational inference
10. Graphical models: Semantics, independence, and inference

Further Topics

11. Model criticism and model diagnostics
12. An introduction to Bayesian nonparametrics
13. Summary (and wiggle room)