Advanced Methods in Probabilistic Modeling

David M. Blei
Princeton University

September 13, 2013

We will study how to use probability models to analyze data, focusing both on mathematical details of the models and the technology that implements the corresponding algorithms. We will study advanced methods, such as large scale inference, model diagnostics and selection, and Bayesian nonparametrics. Our goals are to understand the cutting edge of modern probabilistic modeling, to begin research that makes contributions to this field, and develop good practices for specifying and applying probabilistic models to analyze real-world data.

The centerpiece of the course will be the student project. Over the course of the semester, students will develop an applied case study, ideally one that is connected to their graduate research. Each project must involve using probabilistic models to analyze real-world data.

Prerequisites

I assume you are familiar with the basic material from COS513 (Foundations of Proababilistic Modeling). For example, you should be comfortable with

- probabilistic graphical models
- basic statistics
- mixture modeling
- linear regression
- hidden Markov models
- exponential families
- the expectation-maximization algorithm

We will study again some of the advanced material that was touched on in COS513, such as variational inference and Bayesian nonparametrics.

I assume you are comfortable writing software to analyze data and learning about new tools for that purpose. For example, you should be familiar with a statistical programming language such as R and a scripting language such as Python.

Administrative Details

The instructor is David Blei (`blei@cs.princeton.edu`).

The course meets Monday from 1:30PM - 4:20PM in Room 302 of the CS building.

Office hours are Mondays from 10:30AM - 12:30PM in Room 419 of the CS building.

The course website is `www.cs.princeton.edu/courses/archive/fall13/cos597A/`. We will use Piazza to distribute readings and have online discussion.

Requirements and Grading

There are four requirements for the course.

1. **Participate in class.** This is a seminar based on discussion. Each student must participate substantially to class.

2. **Weekly paper.** Each student will submit a two part weekly paper. First, write on what you thought about the week's reading. Second, write about your progress on the class project. This might include summaries of additional reading or interesting intermediate results.

   These papers should be no longer than two pages. They can be as short as needed.

3. **Project.** There will be short progress reports about the final project due throughout the semester. A final report about your project is due by Dean's Date. I grade final reports on both content and writing quality. Two good books about writing are Strunk and White (1979) and Williams (1981).

4. **Demonstration.** Each student is required to give a 15 minute demonstration of a tool or technique. As much as possible, he or she should walk us through code or run an interpreter. I do not expect the student to be an expert in the tool; the idea is to have some experience with it and then lead a discussion.

   Some example demonstrations include "Processing text data with `nltk`", "Keeping track of research with iPython notebooks", "Exploring results with `plyr`", "Interactive data visualization with `D3`", "Probabilistic programming with Stan", and "Shell scripts I cannot live without".

Your course grade will be mainly based on the final report, but I will also consider the other requirements. The weekly papers will not be individually graded. Please prepare all written work using LaTex. I will provide LaTex templates before the first assignment.

There are no auditors. Those that cannot enroll (for example, postdocs or visiting researchers) must still complete all of the work.

Schedule

There is an assigned reading each week, sometimes a choice of readings and sometimes additional optional readings. Students are also expected to read outside of the syllabus in the service of their final projects.

Below is a tentative schedule of course topics and readings. These may change depending on student interests and the overall trajectory of the course.

1. Introduction and overview of the course

2. Applied probabilistic modeling
   (Blei, 2013)

3. Model specification
   (Lehmann, 1990; Varian, 1997)

4. Variational inference and stochastic optimization
   (Wainwright and Jordan, 2008; Hoffman et al., 2013)

5. Hierarchical models, shrinkage, and empirical Bayes
   (Gelman and Hill, 2007; Efron, 2010)

6. Mixed-membership models
   (Pritchard et al., 2000; Blei, 2012; Rusch et al., 2013)

7. Model fitness: Posterior predictive checks and predictive likelihood
   (Gelman et al., 1996; Box, 1980; Rubin, 1984; Geisser, 1975)

8. Data visualization, the grammar of graphics (and `ggplot2`)
   (Wilkinson, 2009)

9. Bayesian nonparametrics: Clustering models
   (Gershman and Blei, 2012; Teh and Jordan, 2008)

10. Bayesian nonparametrics: Latent feature models
    (Griffiths and Ghahramani, 2011; Broderick et al., 2013)

11. Variational inference with nonconjugate models
    (Braun and McAuliffe, 2010; Wang and Blei, 2013)

12. Bayesian statistics and the philosophy of science
    (Gelman and Shalizi, 2012)

In addition to the assigned papers, consider reading Gelman et al. (1995), Bishop (2006), and Murphy (2013). These are excellent sources on applied probabilistic modeling.

References

Bishop, C. (2006). *Pattern Recognition and Machine Learning*. Springer New York.

Blei, D. (2012). Probabilistic topic models. *Communications of the ACM*, 55(4):77–84.

Box, G. (1980). Sampling and Bayes' inference in scientific modeling and robustness. *Journal of the Royal Statistical Society, Series A*, 143(4):383–430.

Braun, M. and McAuliffe, J. (2010). Variational inference for large-scale models of discrete choice. *Journal of the American Statistical Association*.

Broderick, T., Jordan, M. I., and Pitman, J. (2013). Cluster and feature modeling from combinatorial stochastic processes. *Statistical Science*, 28(3):289–312.

Efron, B. (2010). *Large-scale inference: empirical Bayes methods for estimation, testing, and prediction*, volume 1. Cambridge University Press.

Geisser, S. (1975). The predictive sample reuse method with applications. *Journal of the American Statistical Association*, 70:320–328.

Gelman, A., Carlin, J., Stern, H., and Rubin, D. (1995). *Bayesian Data Analysis*. Chapman & Hall, London.

Gelman, A. and Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press.

Gelman, A., Meng, X., and Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica*, 6:733–807.

Gelman, A. and Shalizi, C. (2012). Philosophy and the practice of Bayesian statistics. *British Journal of Mathematical and Statistical Psychology*.

Gershman, S. and Blei, D. (2012). A tutorial on Bayesian nonparametric models. *Journal of Mathematical Psychology*, 56:1–12.

Griffiths, T. and Ghahramani, Z. (2011). The indian buffet process: An introduction and review. *Journal of Machine Learning Research*, 12:1185–1224.

Hoffman, M., Blei, D., Wang, C., and Paisley, J. (2013). Stochastic variational inference. *Journal of Machine Learning Research*, 14(1303–1347).

Lehmann, E. (1990). Model specification: The views of Fisher and Neyman, and later developments. *Statistical Science*, 5(2):160–168.

Murphy, K. (2013). *Machine Learning: A Probabilistic Approach*. MIT Press.

Pritchard, J., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155:945–959.

Rubin, D. (1984). Bayesianly justifiable and relevant frequency calculations for the applied statistician. *The Annals of Statistics*, 12(4):1151–1172.

Rusch, T., Hofmarcher, P., Hatzinger, R., and Hornik, K. (2013). Model trees with topic model preprocessing: An approach for data journalism illustrated with the wikileaks afghanistan war logs. *The Annals of Applied Statistics*, 7(2):613–639.

Strunk, W. and White, E. (1979). *Elements of Style*. Longman Press.

Teh, Y. and Jordan, M. (2008). Hierarchical Bayesian nonparametric models with applications.

Varian, H. R. (1997). How to build an economic model in your spare time. *The American Economist*, pages 3–10.

Wainwright, M. and Jordan, M. (2008). Graphical models, exponential families, and variational inference. *Foundations and Trends in Machine Learning*, 1(1–2):1–305.

Wang, C. and Blei, D. (2013). Variational inference in nonconjugate models. *Journal of Machine Learning Research*, 14:1005–1031.

Wilkinson, L. (2009). *The Grammar of Graphics*. Springer.

Williams, J. (1981). *Style: Towards Clarity and Grace*. University of Chicago Press.