COS513: FOUNDATIONS OF PROBABILISTIC MODELS

DAVID M. BLEI

Probabilistic modeling is a mainstay of modern artificial intelligence research, providing essential tools for analyzing the vast amount of data that have become available in science, scholarship, and everyday life. This course will cover the mathematical and algorithmic foundations of this field, as well as methods underlying the current state of the art.

Over the last century, problems that have been partially solved with probabilistic models include:

(1) Automatically grouping genes into clusters
(2) Identifying email that is likely to be spam
(3) Transcribing speech from the recorded signal
(4) Identifying recurring patterns in gene sequences
(5) Predicting books or movies that a user will like based on his or her previous purchases
(6) Tracking an object's position by radar
(7) Determining the structure of the evolutionary tree of a set of species
(8) Diagnosing a disease from its symptoms
(9) Decoding the original message from a noisy transmission
(10) Understanding the phase transitions in a physical system of electrons

Each of these applications of probabilistic modeling has involved the determination of a statistical model, a method for fitting that model to observed data, and a method for using the fitted model to solve the task at hand. As one might expect from the diversity of applications listed above, each model has been developed and studied within a different intellectual community.

Over the past two decades, scholars working in the field of machine learning have sought to unify such data analysis activities. Their focus has been on developing tools for devising, analyzing, and implementing probabilistic models in generality. These efforts have lead to the body of work on *probabilistic graphical models*, a marriage of graph theory and probability theory. Graphical models provide both a language for expressing assumptions about data, and a suite of efficient algorithms for reasoning and computing with those assumptions.

As a consequence, graphical models research has forged connections between signal processing, coding theory, computational biology, natural language processing, computer vision, statistics, and many other fields. Knowledge of graphical models is essential to academics working in artificial intelligence and machine learning, and is of increased importance to those in the other scientific and engineering fields to which these methods have been applied.

Week	Topic	Reading
01A	Graphical models	Jordan, 2004
01B	Conditional independence	ITGM Ch 2
02A	Elimination	ITGM Ch 3
02B	Propagation	ITGM Ch 4
03A	Junction tree I	ITGM Ch 17
03B	Junction tree II	ITGM Ch 17
04A	Junction tree III	ITGM Ch 17
04B	Statistical concepts	ITGM Ch 5
05A	Linear regression	ITGM Ch 6
05B	Regularized linear regression	
06A	GLMs	ITGM Ch 8
06B	GLMs	ITGM Ch 10
07A	Conjugacy and mixture models	ITGM Ch 11
07B	Mixture models and expectation maximization	ITGM Ch 11
08A	Factor analysis	ITGM Ch 13-14
08B	MCMC sampling I	Neal, 1993
09A	MCMC sampling II	Neal, 1993
09B	Guest lecture from Frank Wood (UCL)	
10A	Variational methods I	Blei, 2004 (Ch 2)
10B	Exponential families revisited	GMEFVI Ch 3
11A	Sum product and BP	GMEFVI Ch 4
11B	Variational methods II	GMEFVI Ch 5
12A	Language modeling	Manning and Schutze, Ch 6
12B	Nonparametric Bayes I	Sudderth Thesis (95–118)
		Neal, 2000 (optional)
12C	Nonparametric Bayes II	Teh et al., 2007

References

Blei, D. (2004). *Probabilistic Models of Text and Images*. PhD thesis, U.C. Berkeley, Division of Computer Science.

Blei, D. and Lafferty, J. (2009). Topic models. In Srivastava, A. and Sahami, M., editors, *Text Mining: Theory and Applications*. Taylor and Francis.

Jordan, M. (2004). Graphical models. *Statistical Science*, 19(1):140–155.

Jordan, M. (2009). An introduction to probabilistic graphical models.

Jordan, M., Ghahramani, Z., Jaakkola, T., and Saul, L. (1999). Introduction to variational methods for graphical models. *Machine Learning*, 37:183–233.

Manning, C. and Schutze, H. (1999). *Foundations of Statistical Natural Language Processing*. The MIT Press, Cambridge, MA.

Neal, R. (2000). Markov chain sampling methods for Dirichlet process mixture models. *Journal of Computational and Graphical Statistics*, 9(2):249–265.

Sudderth, E. (2006). *Graphical Models for Visual Object Recognition and Tracking*. PhD thesis.

Teh, Y., Jordan, M., Beal, M., and Blei, D. (2007). Hierarchical Dirichlet processes. *Journal of the American Statistical Association*, 101(476):1566–1581.

Wainwright, M. and Jordan, M. (2003). Graphical models, exponential families, and variational inference. Technical Report 649, U.C. Berkeley, Dept. of Statistics.