Learning and Inference in Probabilistic Graphical Models

CSCI 2950-P: Special Topics in Machine Learning, Spring 2010

Probabilistic graphical models provide a flexible framework for modeling large, complex, heterogeneous collections of random variables. Graphs are used to decompose multivariate, joint distributions into a set of local interactions among small subsets of variables. These local relationships produce conditional independencies which lead to efficient learning and inference algorithms. Moreover, their modular structure provides an intuitive language for expressing domain-specific knowledge, and facilitates the transfer of modeling advances to new applications.

After a brief introduction to their representational power, this course will provide a comprehensive survey of state-of-the-art methods for statistical learning and inference in graphical models. Our primary focus will be variational methods, which adapt tools from optimization theory to develop efficient, possibly approximate, inference algorithms. We will also discuss a complementary family of Monte Carlo methods, based on stochastic simulation. Advanced topics may include graph structure learning and Bayesian nonparametric models, depending on time and student interest.

Course readings will be drawn from the text *Probabilistic Graphical Models: Principles and Techniques*, by Koller and Friedman; tutorial and survey papers; and research papers describing state-of-the-art applications. Overall grades will be assigned based on classroom participation, as well as a final research project involving probabilistic graphical models.

Prerequisites: Completion of an introductory course in statistics or machine learning, such as Brown's *CSCI 1950-F: Introduction to Machine Learning*. Sufficient comfort with calculus, linear algebra, and probability to read mathematically sophisticated research papers. For most course projects, programming abilities will also be required.

Administrative Information

Lectures: Mondays and Wednesdays from 10:30-11:50am, CIT room 506, 115 Waterman St.

Instructor:
    Erik Sudderth (sudderth@cs.brown.edu; 401-863-7660)
    Office Hours: Tuesdays 11:00am-12:00pm, Wednesdays 3:00pm-4:00pm, CIT room 509.

Grading: Class Participation

Because this is a seminar course on advanced topics, class attendance and discussion is critical, and will count towards 30% of overall grades. In addition to expecting regular class attendance and involvement in course discussions, participation will be evaluated as follows:

1. Each week, Prof. Sudderth will lecture for roughly half of the class meeting time. The other half will be devoted to student presentations of related reading materials. Over the course of the term, students should expect to give an overview presentation, and lead discussion, about two research papers (depending on course enrollment and paper length).

2. For every course reading, all students are expected to submit brief comments about its strengths, its weaknesses, and the questions it raises. Detailed instructions for the electronic

submission process will be provided later. *Comments are due by 8:00am on the day that paper is discussed.* Late comments will not be given credit, but students can skip comments for three readings over the course of the semester without penalty.

Grading: Final Projects

The final project will count towards 70% of overall grades. Of these points, 10% will be based on a 1-2 page project proposal, due in March; 10% will be based on a short oral presentation given in May; and 50% will be based on a technical report describing the results. Specific due dates will be announced later. This technical report should be between 8-12 pages long, in the style of top machine learning conferences. Although the results need not be sufficiently novel for publication at such conferences, the presentation and experimental protocols should be. Projects which apply graphical models to the student's own research interests are particularly encouraged.

Tentative Syllabus

Graphical models directed, undirected, and factor graph representations; factorization, Markov properties; common temporal, spatial, hierarchical, and relational models

Exact inference variable elimination; message passing algorithms; junction tree algorithm

Gaussian graphical models Bayesian networks, Markov random fields; inference, Kalman filters; linear dynamical systems, factor analysis, probabilistic PCA

Exponential families sufficient statistics, ML estimation; parameter estimation in directed and undirected graphical models; conjugate priors, MAP estimation

Variational methods inference as optimization; entropy, information; mean field methods; parameter estimation, variational EM, variational Bayes

Loopy belief propagation variational interpretations, Bethe entropy approximation; historical perspectives, implementation details, applications

Advanced variational methods clustering, generalized BP, Kikuchi entropy approximations; convex relaxations, fractional BP, undirected parameter estimation

MAP Estimation max-product algorithm; loopy max-product; variational interpretations, linear programming relaxations; graph cut algorithms

Sequential Monte Carlo importance sampling; particle filters; non-sequential models

Markov Chain Monte Carlo Metropolis-Hastings algorithm, detailed balance; Gibbs samplers; Rao-Blackwellization, blocking

Continuous Message Passing nonparametric belief propagation; expectation propagation

Graph Structure Learning structure scoring and search; pseudo-likelihood, local optimization

Bayesian Nonparametrics Dirichlet processes, stick-breaking, Chinese restaurant process; finite approximations; variational and Monte Carlo inference