CSCI 2420: Probabilistic Graphical Models

Brown University, Fall 2016

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. We will discuss a range of efficient algorithms for approximate inference, including optimization-based variational methods, and simulation-based Monte Carlo methods. Several approaches to learning from data will be covered, including conditional models for discriminative learning, and Bayesian methods for controlling model complexity.

Many readings will be drawn from David Barber's *Bayesian Reasoning and Machine Learning*; an electronic version is freely available. Advanced topics will be supported by tutorial and survey articles, and illustrated with state-of-the-art research results and applications. Overall grades will be assigned based on homework assignments combining statistical analysis and implementation of learning algorithms, as well as a final research project involving probabilistic graphical models.

Prerequisites: An introductory course in statistical machine learning, such as Brown's *CSCI 1420: Machine Learning* or *APMA 1690: Computational Probability and Statistics*, or permission of instructor. Programming experience required for homeworks and projects.

Administrative Information

Lectures: Tuesdays and Thursdays from 2:30-3:50pm, CIT room 368.

Lecture attendance and participation does not directly impact course grades, but an in-depth understanding of lecture concepts will be needed for homework assignments and projects.

Instructor: Prof. Erik Sudderth (sudderth@cs.brown.edu; 401-863-7660; CIT room 555)

Graduate Teaching Assistant: Zhile Ren (ren@cs.brown.edu; CIT room 545)

Office Hours: See the course website for a detailed schedule.

Course Credit and Learning Goals

The study of probabilistic graphical models integrates statistical theory, efficient optimization algorithms, and applications based on real-world data. Because we want you to have an intuitive high-level perspective on the key conceptual tools linking theory and applications, lecture attendance is essential. Weekly readings will give you a deeper understanding of these concepts, as well as provide hands-on experience with the machine learning research literature. Because we want you to gain expertise in deriving learning algorithms and using them to analyze real data, there will be five homework assignments integrating programming with mathematics. The final course project will allow you to apply what you've learned to a more complex data analysis problem, and prepare you to independently develop novel probabilistic graphical models in academia or industry.

Successful completion of CSCI 1420 provides 4 semester credit hours. Over 14 weeks, students will spend 3 hours per week in lectures (42 hours total). Background readings for lectures, homeworks, and projects are estimated to require 4 hours per week (56 hours total). During 10 weeks of the course, homework assignments are estimated to require 6 hours per week (60 hours total). Final projects are estimated to require roughly 50 hours to complete, with most of this effort coming in the final month of the semester.

Grading: Homework Assignments

Homework assignments will count towards 60% of overall grades. There will be five equally weighted assignments, each of which will be available for two weeks before its due date. Homeworks will involve a combination of mathematical derivations, algorithm design, programming, and real data analysis. Further details, and homework due dates, are available on the course webpage.

Late Submission Policy: Homework assignments are due by 11:59pm on Wednesday evenings. Your answers may be submitted up to one week late; after this point, solutions will be distributed and handins will no longer be accepted. You may submit one late assignment without penalty. For each subsequent late assignment, 20 points (out of a maximum of 100) will be deducted from the overall score. *Exceptions to this policy are only given in very unusual circumstances, and must be requested in advance by e-mail to the instructor.*

Grading: Final Projects

The final project will count towards 40% of overall grades. Of these points, 5% will be based on a 2-3 page project proposal, due midway through the semester; 10% will be based on a short oral presentation, given on the last class; and 25% will be based on a technical report describing the results. This technical report should be between 6-10 pages long, in the style of top machine learning conferences. Although the results need not be sufficiently novel for publication, the presentation and experimental protocols should be of high quality. Projects which apply graphical models to the student's own research interests are particularly encouraged. See the course webpage for due dates and additional information.

Late Submission Policy: Final project proposals and reports will not be accepted after the due dates announced at the start of the semester. *Exceptions to this policy are only given in very unusual circumstances, and must be requested in advance by e-mail to the instructor.*

Syllabus: Summary of Course Topics

Please see the course webpage for a more detailed weekly schedule and reading list.

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 algorithms

Learning from data exponential families and sufficient statistics; parameter estimation in directed and undirected graphical models; conjugate priors

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

Sequential Monte Carlo importance sampling; particle filters; non-sequential models; nonparametric belief propagation, particle belief propagation

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

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

Loopy belief propagation variational interpretations, Bethe entropies; historical perspectives and implementation; reparameterization and convergence; expectation propagation

MAP Estimation max-product algorithm; loopy max-product BP; variational interpretations, linear programming relaxations

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

Discriminative Learning conditional random fields; max-margin Markov networks and structural support vector machines (SVMs)

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