The University of Chicago
Department of Statistics

PhD Thesis Proposal Presentation

RINA FOYGEL
Department of Statistics
The University of Chicago

Comparing Applied Settings and Theoretical Results in Sparse and Low-Rank Modeling

WEDNESDAY, May 11, 2011, at 10:30 AM
110 Eckhart Hall, 5734 S. University Avenue

ABSTRACT

Applications of high-dimensional regression and graphical modeling, such as as genetic prediction of disease risk and inference of gene regulatory networks, have led to extensive research on L1-regularization techniques for model selection and model fitting in sparse high-dimensional settings. Similarly, the Netflix challenge has stimulated much theoretical and computational work on regularized low-rank matrix completion, where a set of observed entries within a matrix are used to predict the values of the remaining entries, most commonly using the matrix trace-norm for regularization.

While these methods have been used extensively in a wide range of applications, the supporting theoretical results have for the most part made strict assumptions about the nature of the data, which are often violated in the same applied settings where these methods have been so successful. The strict theoretical assumptions generally require (1) true sparsity in the regression or the graphical model (or true low-rank structure in the matrix setting), (2) approximate orthogonality among the covariates (or "incoherence" among the singular vectors, in the matrix setting), and (3) independent and subgaussian noise (and, in the matrix setting, an additional assumption on the distribution of sampled entries within the matrix).

For each setting, I will discuss progress on open problems that concern the necessity of these assumptions as well as guarantees of good performance based on weaker (and more realistic) assumptions.

Motivated in part by the problem of tuning regularization parameters in L1-based techniques, I will also outline possible extensions of existing results about the theoretical properties of the BIC and of the Bayesian marginal model likelihood for both (generalized) linear regressions and graphical modeling, in the "small n, large p" asymptotic setting.

Information about building access for persons with disabilities may be obtained in advance by calling Sandra Romero at 773.702-0541 or by email (sandra@galton.uchicago.edu).