The University of Chicago
Department of Statistics

PhD Thesis Presentation

RINA FOYGEL BARBER

Department of Statistics
The University of Chicago

Bayesian Model Selection and the Extended BIC for High-dimensional Regression and Graphical Models

THURSDAY, May 3, 2012, at 1:30 PM

110 Eckhart Hall, 5734 S. University Avenue

ABSTRACT

We are interested in the model selection problem in two high-dimensional settings: sparse regression and sparse graphical models. In a sparse regression, we would like to find a small number of predictive variables that allow us to accurately estimate the response, when the total number of possible predictors is large—for example, we may want to filter out spam emails based on their content, where each word appearing in an email is a potential predictor. In a graphical setting, we are given a large number of variables with a sparse dependency structure, meaning that most variables are only directly dependent on a small number of other variables—we can consider the example of modeling precipitation at a cluster of nearby weather stations. In each setting, our goal is to find an efficient procedure that reliably discovers the sparse dependency pattern in the data. We show that the extended Bayesian Information Criterion (BIC) is a reliable model selection tool in each setting, and that it outperforms other model selection procedures for the spam email and weather data applications. Furthermore, in the high-dimensional sparse regression setting, we prove that the extended BIC is an accurate estimate of the Bayesian posterior probability distribution over the set of sparse models.

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