The University of Chicago
Department of Statistics

Seminar Series

BENJAMIN MARLIN

Pacific Institute for the Mathematical Sciences

Machine Learning, Collaborative Filtering and the Missing at Random Assumption

FRIDAY, January 21, 2011, at 2:30 PM

Ryerson 251, 1100 East 58th Street

Refreshments will be served following the talk at 3:30 in Ryerson 255.

ABSTRACT

Model estimation, inference and prediction in the presence of incomplete data are pervasive problems in machine learning and statistical data analysis. In this talk, I will focus on collaborative filtering and recommender systems, which were recently popularized through the million dollar Netflix Prize. In recommender systems like Netflix, users provide ratings for items like movies and the goal is to produce personalized recommendations for each user. Typical online collections contain thousands to millions of items while typical users rate a small fraction of items resulting in massively incomplete data sets. Building on the theory of missing data due to Little and Rubin, I will present a line of research investigating the implicit mechanisms that people employ when selecting the items they rate in recommender systems. I will describe how the properties of these mechanisms can invalidate the common practice of ignoring missing data when estimating and evaluating collaborative filtering models. I will present a probabilistic framework for rating prediction that can mitigate these issues by explicitly modeling both the ratings and the underlying missing data mechanism, resulting in substantially increased prediction accuracy. This work received the best technical paper prize at the ACM Conference on Recommender Systems in 2009 and was recently selected for the best papers track at the 2011 International Joint Conference on Artificial Intelligence. I will conclude with a very brief overview of the other areas that I actively work in including Gaussian graphical models, unsupervised feature induction and non-likelihood-based model estimation.

This is a joint talk sponsored by the Departments of Statistics and Computer Science.

Host: Pedro Felzenszwalb

For further information and about building access for persons with disabilities, please contact Laura Rigazzi at 773.702.8333 or send email (lrigazzi@galton.uchicago.edu). If you wish to subscribe to our email list, please visit the following web site: https://lists.uchicago.edu/web/info/statseminars.