University of Chicago and Toyota Technological Institute at Chicago
Machine Learning Seminar Series

PRESENTS:

Madeleine Udell
Assistant Professor Operations Research & Information Engineering
Cornell

Title: Big data is low rank

Abstract: Matrices of low rank are pervasive in big data, appearing in recommender systems, movie preferences, topic models, medical records, and genomics. While there is a vast literature on how to exploit low rank structure in these datasets, there is less attention on explaining why low rank structure appears in the first place. In this talk, we explain the abundance of low rank matrices in big data by proving that certain latent variable models associated to piecewise analytic functions are of log-rank. Any large matrix from such a latent variable model can be approximated, up to a small error, by a low rank matrix. Armed with this theorem, we show how to use a low rank modeling framework to exploit low rank structure even for datasets that are not numeric, with applications in the social sciences, medicine, and automated machine learning.

Bio: Madeleine Udell is Assistant Professor of Operations Research and Information Engineering and Richard and Sybil Smith Sesquicentennial Fellow at Cornell University. She studies optimization and machine learning for large scale data analysis and control, with applications in marketing, demographic modeling, medical informatics, engineering system design, and automated machine learning. Her research in optimization centers on detecting and exploiting novel structures in optimization problems, with a particular focus on convex and low rank problems. Her research in machine learning centers on methods for imputing missing data in large tabular data sets. Her work on generalized low rank models (GLRMs) extends principal components analysis (PCA) to embed tabular data sets with heterogeneous (numerical, Boolean, categorical, and ordinal) types into a low dimensional space, providing a coherent framework for compressing, denoising, and imputing missing entries.
Host: Rebecca Willett

Friday, October 18, 10:30 – 11:30 am
Crerar Library, Room 390