

MANAGEMENT SCIENCES SEMINAR SERIES

Optimization for Big Data Analysis: Complexity and Scalability

Qihang Lin
Carnegie Mellon University

February 22, 2013
2:30-3:20 p.m.

W181 Pappajohn Business Building

Abstract

The development of modern technology has enabled collecting data of unprecedented size and dimensionality. To learn from these large-scale and high-dimensional data, different machine learning models have been developed in the past decade, most of which are intrinsically optimization problems. To solve these machine learning problems, traditional optimization techniques often suffer from unaffordable computation time or excessive memory requirements.

To facilitate big data analysis, we need novel optimization methods for proper efficiency and scalability. In this talk, I present two research threads that tackle this challenge under different settings: (1) A gradient homotopy method for L1-regularized least-squares problems, which achieves a geometric convergence rate. (2) A stochastic first-order method for structured regressions, which can learn from streaming data with a low memory requirement.

I will also talk about applications of the proposed methods in conversion rate predictions for marketing data analysis. If time permits, I will also highlight my research on crowdsourcing technique for big data classification.