Speaker: Rahul Mazumder, Columbia University

Title: Exact Subset Selection in Regression via a Modern Optimization Lens

Abstract:

We consider the usual linear regression set up where the task is to find the best least squares fit to data subject to a cardinality constraint in the regression coefficient vector. This problem, popularly dubbed as the best-subset selection problem is known to be NP-hard and has been widely dismissed as computationally intractable. In this talk, I will propose a novel framework via which the best-subset selection problem can indeed be provably solved to global (or near global) accuracy in problems of practical interest. At the core of our proposal is a computationally tractable framework that brings to bear the power of modern discrete optimization methods: (a) Mixed Integer Optimization (MIO) techniques and (b) Discrete first-order methods, which may be viewed as a “discrete” extension of first order continuous optimization methods. Superiority of best-subset selection procedures over popularly used computationally friendlier alternatives will be discussed.

If time permits, I will describe how our framework can be adapted to solve another classic but computationally difficult procedure in robust statistics, namely, the Least Median of Squares Regression (Rousseeuw ’84).

Joint work with Dimitris Bertsimas (MIT)