THE UNIVERSITY OF CHICAGO

Department of Statistics

STATISTICS COLLOQUIUM

RAHUL MAZUMBER

Operations Research and Statistics Group
MIT Sloan School of Management

Sparse Regression at Scale: Convex and Mixed Integer Programming Perspectives

MONDAY, October 12, 2020 at 4:00 PM
Via Zoom (session information will be e-mailed to subscribers)
Virtual Reception After Colloquium

ABSTRACT

Many fundamental high-dimensional statistics (aka sparse learning) estimators can be naturally expressed as discrete optimization problems, posing computational challenges. Recently, mixed integer programming methods have been shown to be promising candidates in formulating and solving (to optimality) small/moderate instances of the best-subset selection problem for instances much larger than what was considered possible in the statistics community. This sheds interesting insights into the statistical properties of best-subsets. Furthermore, compared to specialized L1-based methods, current high-performance integer programming methods rely on commercial solvers that are harder to deploy at scale in an open-source (non-academic) environment. To this end, I will discuss (a) methods to compute approximate solutions for best-subset type problems at scales comparable to current L1-based methods; and (b) tailored branch and bound methods to solve these problems to optimality. These algorithms employ techniques from first order methods in convex optimization and integer programming.

For further information and inquiries about building access for persons with disabilities, please contact Jonathan Rodriguez at 773.702.8333 or send him an email at jgrodriguez@galton.uchicago.edu. If you wish to subscribe to our email list, please visit the following website: https://lists.uchicago.edu/web/subscribe/statseminars.