The field of Statistics has historically been linked with Probability Theory. However, some of the central problems of classification, regression and estimation can naturally be written as optimization problems. While continuous optimization approaches has had a significant impact in Statistics, mixed integer optimization (MIO) has played a very limited role, primarily based on the belief that MIO models are computationally intractable.

The period 1991–2015 has witnessed a) algorithmic advances in mixed integer optimization (MIO), which coupled with hardware improvements have resulted in an astonishing 2 trillion factor speedup in solving MIO problems, b) significant advances in our ability to model and solve very high dimensional robust and convex optimization models.

In this talk, we demonstrate that modern convex, robust and especially mixed integer optimization methods, when applied to a variety of classical Machine Learning (ML)/Statistics (S) problems can lead to certifiable optimal solutions for large scale instances that have often significantly improved out of sample accuracy compared to heuristic methods used in ML/S.

Specifically, we report results on

• The classical variable selection problem in regression currently solved by Lasso heuristically.
• We show that robustness and not sparsity is the major reason of the success of Lasso in contrast to widely held beliefs in ML/S.
• A systematic approach to design linear and logistic regression models based on MIO.
• Optimal trees for classification solved by CART heuristically.
• Robust classification including robust Logistic regression, robust optimal trees and robust support vector machines.
• Sparse matrix estimation problems: Principal Component Analysis, Factor Analysis and Covariance matrix estimation.

In all cases we demonstrate that optimal solutions to large scale instances (a) can be found in seconds, (b) can be certified to be optimal in minutes and (c) outperform classical approaches. Most importantly, this body of work suggests that linking ML/S to modern optimization leads to significant advances.

**Friday, March 24 at 3:00pm**
Rackham Building, 4th floor, Amphitheater