



Challenges with Big Data: Computational and Statistical Issues

Martin J. Wainwright
Departments of Statistics and EECS
UC Berkeley - wainwrig@berkeley.edu

The scale and richness of massive data sets provides challenges that are both statistical and computational in nature. For instance, massive data sets are often too large to be stored on a single processor. This fact leads the use of distributed methods for statistical inference, as well as various questions with their use. What are optimal methods for distributed inference?

What is the statistical price of decentralization? Similarly, we have often a range of methods that can be ordered in terms of their computational complexity, ranging from simple methods (e.g., thresholding) through to exponential-time search (e.g., all-subset selection). How to characterize trade-off between the computational cost and statistical accuracy? In this talk, we discuss various approaches to these and related questions, thereby providing an overview of this rapidly evolving area of research.