Bocconi
Università Commerciale Luigi Bocconi
Department of Decision Sciences

Statistics Seminar

On the Computational and Statistical Interface and "Big Data"

Michael I. Jordan
University of California, Berkeley

Thursday, 16th January 2014
12:30pm Room 3-E4-SR03 Via Rontgen 1 Milano

Abstract

The rapid growth in the size and scope of datasets in science and technology has created a need for novel foundational perspectives on data analysis that blend the statistical and computational sciences. That classical perspectives from these fields are not adequate to address emerging problems in "Big Data" is apparent from their sharply divergent nature at an elementary level---in computer science, the growth of the number of data points is a source of "complexity" that must be tamed via algorithms or hardware, whereas in statistics, the growth of the number of data points is a source of "simplicity" in that inferences are generally stronger and asymptotic results can be invoked. We wish to blend these perspectives. Indeed, if data are a data analyst's principal resource, why should more data be burdensome in some sense? Shouldn't it be possible to exploit the increasing inferential strength of data at scale to keep computational complexity at bay? I present three research vignettes that pursue this theme, the first involving the deployment of resampling methods such as the bootstrap on parallel and distributed computing platforms, the second involving large-scale matrix completion, and the third introducing a methodology of "algorithmic weakening," whereby hierarchies of convex relaxations are used to control statistical risk as data accrue.
[Joint work with Venkat Chandrasekaran, Ariel Kleiner, Lester Mackey, Purna Sarkar, and Ameet Talwalkar].

Department of Decision Sciences

Via Röntgen 1 - 20136Milano

Tel. 02 5836.5632
Fax 02 5836.5630