



Climate Program Remote Sensing Workshop February 12-14, 2018

SPEAKER TITLES/ABSTRACTS

Venkat Chandrasekaran

Caltech

“Computational and Statistical Trade-offs in Data Analysis”

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 computer science and statistics. That classical perspectives from these fields are not adequate to address emerging challenges with massive datasets 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. In classical statistics, one usually considers the increase in inferential accuracy as the number of data points grows (with little formal consideration of computational complexity), while in classical numerical computation, one typically analyzes the improvement in accuracy as more computational resources such as space or time are employed (with the size of a dataset not formally viewed as a resource). In this talk we describe some of our research efforts towards addressing the question of trading off the amount of data and the amount of computation required to achieve a desired inferential accuracy.

This is joint work with Michael Jordan, Yong Sheng Soh, and Quentin Berthet