University of California, Irvine
Statistics Seminar

Data Integration Via Analysis of Subspaces (DIVAS)

J. S. Marron
Department of Statistics and O.R.
School of Data Science and Society
University of North Carolina

Noon, Friday, March 14, 2025
2011 Donald Bren Hall

A major challenge in the age of Big Data is the integration of disparate data types into a single data analysis. That is tackled here in the context of data blocks measured on a common set of experimental cases. Joint variation is defined in terms of modes of variation having identical scores across data blocks. That allows mathematically rigorous formulation of individual variation within each data block in terms of individual modes. These are mathematically defined through modes of variation with common scores. DIVAS improves earlier methods using a novel random direction approach to statistical inference, and by treating partially shared blocks. Usefulness is illustrated using mortality, cancer and neuroimaging data sets.