J. S. Marron

Joint and Individual Variation Explained

A major challenge in the age of Big Data is the integration of disparate data types into a data analysis. That is tackled here in the context of data blocks measured on a common set of experimental subjects. This data structure motivates the simultaneous exploration of the joint and individual variation within each data block. This is done here in a way that scales well to large data sets, using principal angle analysis, careful formulation of the underlying linear algebra, and differing outputs depending on the analytical goals. Ideas are illustrated using mortality, cancer and neuroimaging data sets.