Title: Sparse model-based clustering of three-way data via lasso-type penalties

Dr. Andrea Cappozzo

Abstract:

Mixtures of matrix Gaussian distributions provide a probabilistic framework for clustering continuous matrix-variate data, which are increasingly common in various fields. Despite their widespread use and successful applications, these models suffer from over-parameterization, making them not suitable for even moderately sized matrix-variate data. To address this issue, we introduce a sparse model-based clustering approach for three-way data. Our approach assumes that the matrix mixture parameters are sparse and have different degrees of sparsity across clusters, enabling the induction of parsimony in a flexible manner. Estimation relies on the maximization of a penalized likelihood, with specifically tailored group and graphical lasso penalties. These penalties facilitate the selection of the most informative features for clustering three-way data where variables are recorded over multiple occasions, as well as allowing the identification of cluster-specific association structures. We validate the effectiveness of the proposed methodology through an application to time-dependent crime patterns across multiple U.S. cities.