Model-based clustering of high dimensional data via Random Projections

A seminar by Laura Anderlucci

Department of Statistical Sciences "Paolo Fortunati"
University of Bologna

Thursday 18 Apr 2024 | 2.30 p.m.
Room Benvenuti
Department of Statistical Sciences

Random projections (RPs) have shown to provide promising results in the context of high-dimensional supervised classification. When information on the class membership is not available, the classification issue can be addressed by exploiting the general idea of RP ensemble. Specifically, we proposed to aggregate (via consensus) the clustering results obtained on a selected set of low-dimensional independent random projections of the original data; such method proved to accurately recover the `true' underlying group structure.
In this work, we address the problem of clustering high dimensional data by resorting to a RP-based ensemble of low-rank estimates for the group-specific covariance matrix.
The performances of the proposal are assessed in terms of both clustering accuracy and estimate precision through numerical studies.

Joint work with Angela Montanari.

Università degli Studi di Padova

Dipartimento di Scienze Statistiche