Multivariate data analysis based on two location vectors and two scatter matrices

Hannu Oja

Tampere School of Public Health

FIN-33014 University of Tampere, Finland.

Abstract

The regular sample mean vector and covariance matrix are popular tools to describe location and scatter of a multivariate data cloud. Weighted mean vectors and covariance matrices, M and S location and scatter estimates, for example, provide often less efficient but more robust estimates of the population mean vector and covariance matrix in the multivariate normal and elliptic case. In families of non-elliptical (e.g. skew) distributions, different location and scatter estimates estimate different population quantities. In this talk, two location vectors and two scatter matrices (suitably chosen) are used together to analyze non-elliptical data. We show how these can be used for a description of multivariate data cloud (location, scatter, skewness and kurtosis). Two scatter matrices together give an invariant coordinate system (ICS) and then the transformation-retransformation (TR) technique can be used for further nonparametric analyses. Two scatter matrices with the so called independence property can be used to find independent components in the independent component analysis (ICA). Invariant coordinate system helps in hunting for clusters and outliers, and it can be used in dimension reduction. Several examples are given.