Invariant Coordinate Selection and Fisher discriminant subspace beyond the case of two groups

Aurore Archimbaud[1], **Colombe Becquart**[2], **Anne Ruiz-Gazen**[2], **Luka Prilc**[2], **Klaus Nordhausen**[3]

[1] *TBS Business School, Toulouse, France*
[2] *Toulouse School of Economics, Toulouse, France*
[3] *University of Jyvaskyla, Finland*

Abstract

Invariant Coordinate Selection (ICS) is a powerful unsupervised multivariate method designed to identify the structure of multivariate datasets on a subspace. It relies on the joint diagonalization of two scatter matrices and is particularly relevant as a dimension reduction tool prior to clustering or outlier detection. It goes beyond the well-known Principal Components Analysis (PCA) method by not relying on maximizing the inertia but on optimizing a generalized kurtosis and is not only invariant by an orthogonal transformation of the data but by any affine transformation. Unlike PCA, ICS has a theoretical foundation that explains why the identified subspace should contain relevant information and under what conditions it might fail to reveal such subspace. More precisely, some theoretical results proved that under some elliptical mixture models, the subspace spanned by the first and/or last components carries the information regarding the multivariate structure and recovers the Fisher discriminant subspace, whatever the choice of scatter matrices. These general results have traditionally been examined in detail primarily for specific scatter combinations within a two-cluster framework. In this study, we expand these investigations to include more clusters and scatter combinations. Based on these expanded theoretical insights and supported by numerical studies, we conclude that ICS is indeed suitable for dimension reduction with some guidelines on which cases it might fail.