

Vichi, Maurizio; Saporta, Gilbert

Clustering and disjoint principal component analysis. (English) Zbl 1453.62230
Comput. Stat. Data Anal. 53, No. 8, 3194-3208 (2009).

Summary: A constrained principal component analysis, which aims at a simultaneous clustering of objects and a partitioning of variables, is proposed. The new methodology allows us to identify components with maximum variance, each one a linear combination of a subset of variables. All the subsets form a partition of variables. Simultaneously, a partition of objects is also computed maximizing the between cluster variance. The methodology is formulated in a semi-parametric least-squares framework as a quadratic mixed continuous and integer problem. An alternating least-squares algorithm is proposed to solve the clustering and disjoint PCA. Two applications are given to show the features of the methodology.

MSC:

62-08 Computational methods for problems pertaining to statistics
62H30 Classification and discrimination; cluster analysis (statistical aspects)
62H25 Factor analysis and principal components; correspondence analysis

Cited in **20** Documents

Full Text: [DOI](#)

References:

- [1] Cattell, R.B., The scree test for the number of factors, *Multivariate behavioral research*, 1, 245-276, (1966)
- [2] DeSarbo, W.S.; Jedidi, K.; Cool, K.; Schendel, D., Simultaneous multidimensional unfolding and cluster analysis: an investigation of strategic groups, *Marketing letters*, 2, 129-146, (1990)
- [3] De Soete, G.; Carroll, J.D., K-means clustering in a low-dimensional Euclidean space, (), 212-219
- [4] De Soete, G.; Heiser, W.J., A latent class unfolding model for analyzing single stimulus preference ratings, *Psychometrika*, 58, 545-565, (1993) · [Zbl 0826.62098](#)
- [5] Gabriel, K.R., The biplot graphic display of matrices with application to principal component analysis, *Biometrika*, 58, 453-467, (1971) · [Zbl 0228.62034](#)
- [6] Heiser, W.J., Clustering in low-dimensional space, (), 162-173
- [7] Heiser, W.J.; Groenen, P.J.F., Cluster differences scaling with a within-clusters loss component and a fuzzy successive approximation strategy to avoid local minima, *Psychometrika*, 62, 63-83, (1997) · [Zbl 0889.92037](#)
- [8] Kaiser, H.F., The varimax criterion for analytic rotation in factor analysis, *Psychometrika*, 23, 187-200, (1958) · [Zbl 0095.33603](#)
- [9] Milligan, G.W.; Cooper, M., An estimation of procedures for determining the number of clusters in a data set, *Psychometrika*, 50, 159-179, (1985)
- [10] Vichi, M.; Kiers, H.A.L., Factorial k -means analysis for two way data (2001), *Computational statistics and data analysis*, 37, 49-64, (2001) · [Zbl 1051.62056](#)
- [11] Vichi, M., Double k -means clustering for simultaneous classification of objects and variables, (), 43-52
- [12] Vichi, M., Discrete and continuous models for two way data (2002), (), 139-147
- [13] Vichi, M.; Rocci, R.; Kiers, H.A.L., Simultaneous component and clustering models for three-way data: within and between approaches, *Journal of classification*, 24, 1, 71-98, (2007) · [Zbl 1144.62045](#)
- [14] Vigneau, E.; Qannari, E.M., Clustering of variables around latent component — application to sensory analysis, *Communications in statistics, simulation and computation*, 32, 4, 1131-1150, (2004) · [Zbl 1100.62582](#)
- [15] Zou, H.; Hastie, T.; Tibshirani, R., Sparse principal component analysis, *Journal of computational and graphical statistics*, 15, 2, 262-286, (2006)

This reference list is based on information provided by the publisher or from digital mathematics libraries. Its items are heuristically matched to zbMATH identifiers and may contain data conversion errors. It attempts to reflect the references listed in the original paper as accurately as possible without claiming the completeness or perfect precision of the matching.