Journal of Multivariate Analysis 133 (2015) 356–376

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

A robust predictive approach for canonical correlation analysis

Jorge G. Adrover [a,*], Stella M. Donato [b]

[a] FAMAF, Universidad Nacional de Córdoba, CIEM and CONICET, Argentina
[b] Instituto de Cálculo, Universidad de Buenos Aires and CONICET, Argentina

ARTICLE INFO

Article history:
Received 10 June 2013
Available online 5 November 2014

AMS 1991 subject classifications:
primary 62F35
secondary 62H12.0

Keywords:
Canonical correlation analysis
S-estimation
M-scales
Mean relative prediction error

ABSTRACT

Canonical correlation analysis (CCA) is a dimension-reduction technique in which two random vectors from high dimensional spaces are reduced to a new pair of low dimensional vectors after applying linear transformations to each of them, retaining as much information as possible. The components of the transformed vectors are called canonical variables. One seeks linear combinations of the original vectors maximizing the correlation subject to the constraint that they are to be uncorrelated with the previous canonical variables within each vector. By these means one actually gets two transformed random vectors of lower dimension whose expected square distance has been minimized subject to have uncorrelated components of unit variance within each vector. Since the closeness between the two transformed vectors is evaluated through a highly sensitive measure to outlying observations as the mean square loss, the linear transformations we are seeking are also affected. In this paper we use a robust univariate dispersion measure (like an M-scale) based on the distance of the transformed vectors to derive robust S-estimators for canonical vectors and correlations. An iterative algorithm is performed by exploiting the existence of efficient algorithms for S-estimation in the context of Principal Component Analysis. Some convergence properties are analyzed for the iterative algorithm. A simulation study is conducted to compare the new procedure with some other robust competitors available in the literature, showing a remarkable performance. We also prove that the proposal is Fisher consistent.

1. Introduction

Principal component analysis (PCA) and canonical correlation analysis (CCA) are two dimension-reduction techniques of widespread use in statistics. Though the principal component analysis relates to an internal analysis, i.e. within-group spectral decomposition for the study of dispersion, and the canonical correlations to an external analysis, i.e. between-group interrelations or correlations, conceptually they are interrelated. We will further explore this relationship. For a random vector $\mathbf{x}$ in the Euclidean space of dimension q, with positive definite dispersion matrix Σ, PCA looks for the spectral decomposition of Σ, the eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_q$ associated with the corresponding eigenvalues in decreasing order $\delta_1 \geq \delta_2 \geq \cdots \geq \delta_q > 0$, that is,

$$\Sigma = \sum_{i=1}^{q} \delta_i \mathbf{v}_i \mathbf{v}_i^t. \tag{1}$$

* Corresponding author.
E-mail addresses: adrover@famaf.unc.edu.ar (J.G. Adrover), stelladonato@yahoo.com.ar (S.M. Donato).

http://dx.doi.org/10.1016/j.jmva.2014.09.007
0047-259X/

The variables $\mathbf{v}_1^t(\mathbf{x}-E\mathbf{x}),\ldots,\mathbf{v}_q^t(\mathbf{x}-E\mathbf{x})$ are usually referred as principal components. The spectral decomposition gives the orthonormal directions of maximum dispersion for $\mathbf{x}$, where the eigenvalues and eigenvectors can be defined through an optimization scheme,

$$\begin{aligned}\delta_1 &= \max_{\mathbf{a}\in\mathbb{R}^q,\ \|\mathbf{a}\|=1} \mathrm{Var}(\mathbf{a}^t(\mathbf{x}-E\mathbf{x})), \qquad \mathbf{v}_1 = \arg\max_{\mathbf{a}\in\mathbb{R}^q,\ \|\mathbf{a}\|=1} \mathrm{Var}(\mathbf{a}^t(\mathbf{x}-E\mathbf{x}))\\ \delta_j &= \max_{\|\mathbf{a}\|=1,\ \mathrm{Cov}(\mathbf{a}^t(\mathbf{x}-E\mathbf{x}),\mathbf{v}_k^t(\mathbf{x}-E\mathbf{x}))=0,\ k=1,\ldots,j-1} \mathrm{Var}(\mathbf{a}^t(\mathbf{x}-E\mathbf{x})), \quad j>1\\ \mathbf{v}_j &= \arg\max_{\|\mathbf{a}\|=1,\ \mathrm{Cov}(\mathbf{a}^t(\mathbf{x}-E\mathbf{x}),\mathbf{v}_k^t(\mathbf{x}-E\mathbf{x}))=0,\ k=1,\ldots,j-1} \mathrm{Var}(\mathbf{a}^t(\mathbf{x}-E\mathbf{x})),\end{aligned} \tag{2}$$

where Var and Cov stand for the variance and the covariance operators for random variables. On the other hand, the principal components are the best linear predictors for $\mathbf{z}=\mathbf{x}-E\mathbf{x}$ when looking for linear combinations $\sum_{k=1}^p(\mathbf{a}_k^t\mathbf{z})\mathbf{a}_k$ based on an orthonormal set $\{\mathbf{a}_1,\ldots,\mathbf{a}_p,\mathbf{a}_{p+1},\ldots,\mathbf{a}_q\}$, $p<q$. More precisely, principal components solve the optimization problem

$$\begin{aligned}(\boldsymbol{\mu}_{\mathbf{x}},V_p) &= \arg\min_{\boldsymbol{\mu}\in\mathbb{R}^p,V} E\,\|(\mathbf{x}-\boldsymbol{\mu})-P_V(\mathbf{x}-\boldsymbol{\mu})\|^2\\ &= \arg\min_{\boldsymbol{\mu}\in\mathbb{R}^p,V} E\,\left\|P_{V^\perp}(\mathbf{x}-\boldsymbol{\mu})\right\|^2,\end{aligned} \tag{3}$$

where P_V stands for the orthogonal projection on a subspace V of dimension $p<q$, $V=\langle\mathbf{a}_1,\ldots,\mathbf{a}_p\rangle$ means that V is generated by the orthonormal set $\{\mathbf{a}_1,\ldots,\mathbf{a}_p\}$ and $V^\perp=\langle\mathbf{a}_{p+1},\ldots,\mathbf{a}_q\rangle$ denotes the orthogonal complement of V. Then, the solutions $(\boldsymbol{\mu}_{\mathbf{x}},V_p)$ for (3) are given by

$$\boldsymbol{\mu}_{\mathbf{x}}=E\mathbf{x}, \qquad V_p=\langle\mathbf{v}_1,\ldots,\mathbf{v}_p\rangle \quad\text{and}\quad P_{V_p}(\mathbf{z})=\sum_{k=1}^p(\mathbf{v}_k^t\mathbf{z})\mathbf{v}_k.$$

CCA was proposed by Hotelling [10] to determine the relationship between two sets of variables obtained by transforming the vectors $\mathbf{x}$ and $\mathbf{y}$ into two vectors $\mathbf{z}$ and $\mathbf{w}$ in lower dimensions whose association has been greatly strengthened (see Das and Sen [5] for a very thorough account on CCA and their wide variety of applications). In recent years, CCA has also gained popularity as a method for the analysis of genomic data, since CCA has the potential to be a powerful tool for identifying relationships between genotype and gene expression. It has also been used in geostatistical applications (see Furrer and Genton [8]). CCA is closely related to multivariate regression when the vectors $\mathbf{x}$ and $\mathbf{y}$ are not treated symmetrically (see Yohai and García Ben [20]). Given the two random vectors $\mathbf{x}$ and $\mathbf{y}$ of dimensions p and q respectively, with dispersion matrix given by

$$\Sigma=\begin{pmatrix}E(\mathbf{x}-E\mathbf{x})(\mathbf{x}-E\mathbf{x})^t & E(\mathbf{x}-E\mathbf{x})(\mathbf{y}-E\mathbf{y})^t\\ E(\mathbf{y}-E\mathbf{y})(\mathbf{x}-E\mathbf{x})^t & E(\mathbf{y}-E\mathbf{y})(\mathbf{y}-E\mathbf{y})^t\end{pmatrix}=\begin{pmatrix}\Sigma_{\mathbf{xx}} & \Sigma_{\mathbf{xy}}\\ \Sigma_{\mathbf{yx}} & \Sigma_{\mathbf{yy}}\end{pmatrix}, \tag{4}$$

$$\det(\Sigma_{\mathbf{xx}})>0<\det(\Sigma_{\mathbf{yy}}), \qquad 0<r=\mathrm{rank}(\Sigma_{\mathbf{xy}})\le\min(p,q)=s. \tag{5}$$

CCA seeks linear combinations of the variables in $\mathbf{x}$ and the variables in $\mathbf{y}$ that are maximally correlated with each other, that is, the first canonical vectors $\boldsymbol{\alpha}_1$ and $\boldsymbol{\beta}_1$ are defined (except for the signs) as

$$(\boldsymbol{\alpha}_1,\boldsymbol{\beta}_1)=\arg\max_{(\mathbf{a},\mathbf{b})\in(\mathbb{R}^p-\{\mathbf{0}\})\times(\mathbb{R}^q-\{\mathbf{0}\})}\mathrm{Corr}\left(\mathbf{a}^t\mathbf{x},\mathbf{b}^t\mathbf{y}\right). \tag{6}$$

Since the correlation measure is scale invariant, we can define the first canonical vectors $\boldsymbol{\alpha}_1$, $\boldsymbol{\beta}_1$ as solutions to the optimization problem,

$$(\boldsymbol{\alpha}_1,\boldsymbol{\beta}_1)=\arg\max_{(\mathbf{a},\mathbf{b})\in\mathcal{A}_1}\mathrm{Corr}\left(\mathbf{a}^t\mathbf{x},\mathbf{b}^t\mathbf{y}\right), \tag{7}$$

with

$$\mathcal{A}_1=\left\{(\mathbf{a},\mathbf{b})\in\mathbb{R}^p\times\mathbb{R}^q:\mathrm{Var}\left(\mathbf{a}^t\mathbf{x}\right)=\mathbf{a}^t\Sigma_{\mathbf{xx}}\mathbf{a}=1,\mathrm{Var}\left(\mathbf{b}^t\mathbf{y}\right)=\mathbf{b}^t\Sigma_{\mathbf{yy}}\mathbf{b}=1\right\}. \tag{8}$$

The variables $\boldsymbol{\alpha}_1^t(\mathbf{x}-E\mathbf{x})$ and $\boldsymbol{\beta}_1^t(\mathbf{y}-E\mathbf{y})$ are called the first canonical variables and its positive correlation $\rho_1=\mathrm{Corr}(\boldsymbol{\alpha}_1^t\mathbf{x},\boldsymbol{\beta}_1^t\mathbf{y})$ is called the first canonical correlation. Canonical vectors and variables of higher order are defined recursively. Given $k>1$, let us take the first $k-1$ canonical variables $\boldsymbol{\alpha}_1^t(\mathbf{x}-E\mathbf{x}),\ldots,\boldsymbol{\alpha}_{k-1}^t(\mathbf{x}-E\mathbf{x})$ and $\boldsymbol{\beta}_1^t(\mathbf{y}-E\mathbf{y}),\ldots,\boldsymbol{\beta}_{k-1}^t(\mathbf{y}-E\mathbf{y})$ based on canonical vectors $\{\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_{k-1}\}\subset\mathbb{R}^p$ and $\{\boldsymbol{\beta}_1,\ldots,\boldsymbol{\beta}_{k-1}\}\subset\mathbb{R}^q$. Then, the kth canonical variables $\boldsymbol{\alpha}_k^t(\mathbf{x}-E\mathbf{x})$ and $\boldsymbol{\beta}_k^t(\mathbf{y}-E\mathbf{y})$ can be obtained by seeking the vectors $\boldsymbol{\alpha}_k\in\mathbb{R}^p$ and $\boldsymbol{\beta}_k\in\mathbb{R}^q$ so that the linear combinations $\boldsymbol{\alpha}_k^t\mathbf{x}$ and $\boldsymbol{\beta}_k^t\mathbf{y}$ with unit variance, uncorrelated to $\boldsymbol{\alpha}_1^t\mathbf{x},\ldots,\boldsymbol{\alpha}_{k-1}^t\mathbf{x}$ and $\boldsymbol{\beta}_1^t\mathbf{y},\ldots,\boldsymbol{\beta}_{k-1}^t\mathbf{y}$, maximize the correlation coefficient between them. More precisely, we look for vectors defined as

$$(\boldsymbol{\alpha}_k,\boldsymbol{\beta}_k)=\arg\max_{(\mathbf{a},\mathbf{b})\in\mathcal{A}_k}\mathrm{Corr}\left(\mathbf{a}^t\mathbf{x},\mathbf{b}^t\mathbf{y}\right), \tag{9}$$