ROBUST PENALIZED MULTIVARIATE ANALYSIS FOR HIGH-DIMENSIONAL DATA

Pia Pfeiffer [1] and Peter Filzmoser[1]

[1] Institute of Statistics and Mathematical Methods in Economics, TU Wien, (e-mail: pia.pfeiffer@tuwien.ac.at, peter.filzmoser@tuwien.ac.at)

ABSTRACT: High-dimensional data sets, with fewer observations than variables, pose a challenge for statistical methods, particularly if outlying observations are present. Several proposals for robust and sparse estimation in the context of multivariate statistical methods are available, together with algorithms for the computation. We present a unified computational approach based on reformulating the problem as a constrained optimization problem, also incorporating sparsity constraints. Recent developments with adaptive gradient descent algorithms can efficiently solve such problems, and they are also scalable with data dimensionality. The procedures are illustrated in the example of canonical correlation analysis, where also higher-order directions can be directly computed, and the sparsity can be controlled easily. Extensions to other multivariate methods are possible.

KEYWORDS: high-dimensional data, robust multivariate analysis, sparse multivariate analysis.

1 Introduction

Classical methods for multivariate analyses, such as PCA (Principal Component Analysis), CCA (Canonical Correlation Analysis), and LDA (Linear Discriminant Analysis), are based on covariance estimation and aim to find projection directions in the data according to some criteria. This estimation procedure is not suitable for high-dimensional data sets, and therefore sparse methods have been proposed, e.g. by applying elastic net type penalties (Zou & Hastie, 2005) for the projection directions. Such methods are sensitive to outlying observations, and therefore methods combining sparsity with robust estimation have been proposed. In the context of CCA, for example, Wilms & Croux, 2015 suggest using alternating regressions with sparse and robust regression estimators. A disadvantage of this approach is that higher-order directions cannot be derived directly.

2 Methodology

In the example of CCA, we show how the objective can be reformulated as an optimization problem, directly stating the optimization conditions and offering a flexible choice of covariance estimator and penalty function. Let $\boldsymbol{x}$ and $\boldsymbol{y}$ denote a p- and q-dimensional random variable, respectively, and $\boldsymbol{\Sigma}_{xx}, \boldsymbol{\Sigma}_{yy}$ and $\boldsymbol{\Sigma}_{xy}$ the corresponding covariance matrices. The first canonical correlation coefficient ρ_1 and the first pair of canonical vectors $(\boldsymbol{a}_1, \boldsymbol{b}_1)$ are given as a solution of the optimization problem

$$\max_{\boldsymbol{a}\in\mathbb{R}^p, \boldsymbol{b}\in\mathbb{R}^q} \boldsymbol{a}'\boldsymbol{\Sigma}_{xy}\boldsymbol{b} \tag{1}$$

under the constraints

$$\boldsymbol{a}'\boldsymbol{\Sigma}_{xx}\boldsymbol{a} = 1 \quad and \quad \boldsymbol{b}'\boldsymbol{\Sigma}_{yy}\boldsymbol{b} = 1. \tag{2}$$

The k-th canonical correlation coefficient ρ_k and the respective pair of canonical vectors $(\boldsymbol{a}_k, \boldsymbol{b}_k)$ maximize (1) under the condition that they are uncorrelated with the previous $k-1$ directions, denoted as the constraints

$$\boldsymbol{a}'\boldsymbol{\Sigma}_{xx}\boldsymbol{a}_i = 0 \quad and \quad \boldsymbol{b}'\boldsymbol{\Sigma}_{yy}\boldsymbol{b}_i = 0, \text{ for } i = 1, \ldots, k-1. \tag{3}$$

Penalty terms are added as further constraints for a sparse setting,

$$P_{\alpha_1}(\boldsymbol{a}) \leq c_1 \quad and \quad P_{\alpha_2}(\boldsymbol{b}) \leq c_2 \tag{4}$$

where c_1 and c_2 denote positive constants, and the penalty terms (4) are given as elastic net (Zou & Hastie, 2005) penalties with mixing parameters α_1, α_2.

The augmented Lagrangian with $\boldsymbol{\lambda}$ denoting the Lagrange multipliers, and H summarizing the constraints, is then given as

$$\mathcal{L}_\rho(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{\lambda}) = -|\boldsymbol{a}'\boldsymbol{\Sigma}_{xy}\boldsymbol{b}| + \boldsymbol{\lambda}' \cdot H(\boldsymbol{a}, \boldsymbol{b}) + \frac{\rho}{2}\|H(\boldsymbol{a}, \boldsymbol{b})\|_2^2. \tag{5}$$

Then, a solution to (1)-(4) can be found by minimizing (5). For the optimization algorithm, the method of multipliers (see e.g. Bertsekas, 1982) is combined with an adaptive gradient descent algorithm as described by Reddi *et al.*, 2018 for an alternating update of $(\boldsymbol{a}, \boldsymbol{b})$ and $\boldsymbol{\lambda}$.

Our approach is not only flexible in the choice of covariance estimator and penalty type, but we can also directly state the necessary conditions for higher-order canonical correlations. The robustness of the resulting canonical correlations can be controlled by an appropriate choice of covariance estimators for $\mathbf{\Sigma}_{xx}, \mathbf{\Sigma}_{yy}$ and $\mathbf{\Sigma}_{xy}$. The penalty terms (4) induce sparsity in the resulting canonical directions. Conditions (3) ensure that higher-order directions are uncorrelated to lower-order canonical vectors. For the higher-order directions, again, a suitable level of sparsity can be chosen.

In a simulation study, we show the robustness and suitability of our approach for high-dimensional data in different simulation scenarios. Empirical applications from tribology underline the usefulness of this approach.

3 Outlook

The methodology can be adapted to other robust multivariate methods such as LDA or PCA for high-dimensional data. It is sufficient to formulate the optimization problem and the constraints in a joint Lagrangian problem. The advantage of using an adaptive gradient descent algorithm is its scalability to higher dimension, and it also leads to highly precise parameter estimates, especially for higher-order components.

References

BERTSEKAS, DIMITRI P. 1982. Constrained Optimization and Lagrange Multiplier Methods.

REDDI, SASHANK J., KALE, SATYEN, & KUMAR, SANJIV. 2018. On the Convergence of Adam and Beyond. *In: 6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings*. OpenReview.net.

WILMS, INES, & CROUX, CHRISTOPHE. 2015. Robust Sparse Canonical Correlation Analysis. *BMC Systems Biology*, **10**(01).

ZOU, HUI, & HASTIE, TREVOR. 2005. Regularization and Variable Selection via the Elastic Net. *Journal of the Royal Statistical Society: Series B*, **67(2)**, 301–320.