

HOW TO REDUCE DIMENSIONALITY OF DATA: ROBUSTNESS POINT OF VIEW

Dita Rensová
Melbourne, Australia

Abstract

Data analysis in management applications often requires to handle data with a large number of variables. Therefore, dimensionality reduction represents a common and important step in the analysis of multivariate data by methods of both statistics and data mining. This paper gives an overview of robust dimensionality procedures, which are resistant against the presence of outlying measurements. A simulation study represents the main contribution of the paper. It compares various standard and robust dimensionality procedures in combination with standard and robust methods of classification analysis. While standard methods turn out not to perform too badly on data which are only slightly contaminated by outliers, we give practical recommendations concerning the choice of a suitable robust dimensionality reduction method for highly contaminated data. Namely the highly robust principal component analysis based on the projection pursuit approach turns out to yield the most satisfactory results over four different simulation studies. At the same time, we give recommendations on the choice of a suitable robust classification method.

Keywords: data analysis, dimensionality reduction, robust statistics, principal component analysis, robust classification analysis

1. PRINCIPLES OF DIMENSIONALITY REDUCTION

Multivariate data with a large number of variables are commonly encountered in management or econometrics (Kalina,

2014b), where a trend has been observed towards analyzing more massive data. They may be called big data or high-dimensional data. The concept of big data is preferred in data mining context for data in a database with a very large number of observations n .

On the other hand, the concept of high-dimensional data is preferred in multivariate statistics and is used for data with a large number of variables p . Thus, the two terms partially coincide. Unfortunately, various popular methods of data mining and multivariate statistics are unsuitable for data with a large number of variables and suffer from the so-called curse of dimensionality (Martinez et al., 2011).

Dimensionality reduction is commonly used as a preliminary or assistive step within the information extraction from big or high-dimensional data (Ma & Zhu, 2013). Its methods may bring several benefits. First of all, they simplify a consequent data analysis. Some of the methods allow a clearer interpretation or enable to divide variables to clusters. An important property of some approaches is their ability to reduce or remove correlation among variables, which may again much simplify a consequent data analysis and reduce uncertainty in estimation of parameters or increase the power of hypothesis tests.

Dimensionality reduction methods can be divided to two major groups (Hastie et al., 2001):

- Variable selection
- Feature extraction

Variable selection methods include wrappers, filters, embedded methods (Guyon & Elisseeff, 2003), the minimum redundancy maximum relevance (MRMR) approach or algorithms based on information theory. In linear regression, statistical hypothesis tests or sliced inverse regression may be used as a tool for reducing the number of independent variables. The major drawback of variable selection methods is lack of stability (Breiman, 1996). Some variable selection procedures (e.g. common procedures in

linear regression) suffer from the correlations among variables.

Feature extraction methods search for combinations of variables. Most important examples include principal component analysis (PCA), factor analysis, correspondence analysis, multidimensional scaling, independent component analysis or partial least squares regression. In contrary to variable selection, feature extraction methods decorrelate the observed variables, i.e. remove their correlation structure. On the other hand, their resulting combinations of variables do not always allow a clear interpretation.

Besides, there is a variety of specific statistical or data mining methods which perform a dimensionality reduction as its by-product, e.g. lasso and least angle regression. Some of the methods would not even be considered as dimensionality reduction tools themselves, e.g. cluster analysis and linear discriminant analysis (Martinez et al., 2011).

There has been used a vast number of various dimensionality reduction methods in numerous management applications. Often, the aim is not dimensionality reduction itself, but finding latent factors or components and their interpretation. This is true in data mining performed within a decision making process with the aim to find relevant information from a large database, e.g. in operations management. In strategic management, there have been attempts to incorporate expert knowledge into the dimensionality reduction process. The same has been performed in marketing e.g. in the analysis of customers. In econometrics, dimensionality reduction is commonly performed to reduce the set of regressors in linear regression with the aim to prevent multicollinearity, to simplify solving sets of linear equations, or in economic time series

analysis (see e.g. Greene, 2012).

This paper has the following structure. Section 2 overviews some robust versions of principal component analysis (PCA), which are resistant to the presence of outlying measurements (outliers) in the data. Some of their properties and computational aspects are discussed. Four numerical simulation studies are described and discussed in Section 3 comparing various robust approaches to PCA and a consequent robust classification analysis. Finally, Section 4 concludes the paper.

2. ROBUST DIMENSIONALITY REDUCTION

Sensitivity of the standard PCA to the presence of outlying measurements in the data has been repeatedly reported as a serious problem e.g. by Croux and Haesbroeck (2000). The aim of robust statistics is to develop and investigate alternative statistical procedures which are resistant to the influence of noise and to the presence of outliers (Jurečková & Picek, 2006). This section overviews important methods of robust dimensionality reduction, primarily based on a modification (robustification) of PCA.

Robust dimensionality reduction procedures remain rather rare in econometrics (Kalina, 2012a) or optimization (Xanthopoulos et al., 2013). On the other hand, they have spread to applied tasks in chemometrics, image analysis, or bioinformatics. This is true primarily for robust versions of PCA, because the classical PCA is very intuitive as well as happens to be the most commonly used feature extraction method at all. General principles of robust dimensionality reduction have been

formulated by Hubert et al. (2008) and Filzmoser and Todorov (2011), without a systematic comparison of the performance of different methods on data.

Croux and Haesbroeck (2000) applied robust estimation of the covariance or correlation matrix to obtain a robust version of PCA and replaced eigenvalues and eigenvectors by their counterparts computed from M-estimates and S-estimates of the covariance matrix Σ . Their methods possess only a local robustness in terms of the influence function, which quantifies the influence of an individual observation on the resulting method.

Other methods are based on the projection pursuit technique, which can be described as a general method of Rousseeuw and Leroy (1987) for finding the most informative directions or components for multivariate data. Candidate directions for these principal components are selected by a grid algorithm optimizing an objective function in an effective way in a small dimension, adding subsequently more and more dimensions. One example of such robust PCA is the ROBPCA of Hubert et al. (2005), which includes assessing the outlyingness of each data point. The method is robust also in a global sense in terms of a high breakdown point, which is a statistical measure of sensitivity against severe outliers in the data.

PCA-PROJ of Croux and Ruiz-Gazen (2005) represents another robust version of PCA based on projection pursuit. It is a computationally efficient method not requiring the whole empirical covariance matrix S , but computing only its first components. Its improved version PCA-GRID (Croux et al., 2007) uses principal directions in the data in a more suitable way for $n < p$. Spherical principal component

analysis (Locantore et al., 1999) is based on projecting data onto a sphere with a unit radius and performing classical PCA on such transformed data.

Other methods can be obtained directly as the singular value decomposition of a certain robust estimate of the covariance matrix. Such construction is straightforward allowing to construct a robust PCA with the robustness properties inherited from the robust estimator of Σ . Various robust versions of PCA will be used in simulations in Section 3. They are denoted and defined in the following way.

- PCA-OGK: PCA based on the OGK-estimator of Gnadadesikan and Kettenring (1972).
- PCA-SD: PCA based on the SD-estimator of Stahel (1981) and Donoho (1982).
- PCA-MVE: PCA based on the minimum volume ellipsoid estimator (Rousseeuw, 1985).
- PCA-MCD: PCA based on the minimum covariance determinant estimator (Rousseeuw, 1985).
- PCA-S1: PCA based on S-estimator with Tukey's biweight function (Davies, 1987).
- PCA-S2: PCA based on S-estimator with translated biweight function (Davies, 1987).
- PCA-LWS (Kalina, 2012b): PCA based on the least weighted squares regression (Višek, 2011), assigning weights to individual observations with the aim to down-weight less reliable data likely to be outliers.

Besides, it has been claimed that robustness of multivariate statistical methods for high-dimensional data can be ensured by a suitable regularization (Tibshirani, 1996). Various statistical methods suitable for $n < p$

work with a regularized covariance matrix (Pourahmadi, 2013) in the form

$$S^* = (1 - \lambda)S + \lambda T, \lambda \in (0,1), \quad (1)$$

where T is a regular symmetric positive definite matrix. However, if PCA is computed from S^* , it can be easily verified to be exactly equal to the classical (and non-robust) PCA under the assumption that T is a unit matrix or a diagonal matrix with a common variance. However, regularization does not guarantee robustness, at least not in a general situation and regularized procedures (overviewed e.g. by Kalina (2014a)) cannot be considered robust, while robustness properties of regularized methods have not been systematically investigated.

From the computational point of view, robust PCA is based on singular value decomposition, which can be computed in a numerically stable way even for $n < p$ (Barlow et al., 2005). Thus, suitable matrix manipulations do not require tailor-made adaptations of PCA for high-dimensional data (Rencher, 2002). Nevertheless, already the classical PCA appears in some software packages implemented in an unsuitable way. In R software package, specialized packages as HDMD and FactoMineR can be recommended for computing PCA for $n < p$, compared to other common implementations, which fail for $n < p$ (McFerrin, 2013). Robust versions of PCA should be implemented carefully fulfilling requirements of numerical linear algebra on numerical stability and suitability of a particular implementation for high-dimensional data should be explicitly documented by the author or software provider.

Examples of robust methods for dimensionality reduction not based on PCA

include robust versions of methods mentioned in Section 1, e.g. robust factor analysis or robust partial least squares (Liebmann et al., 2009). Because performance of available robust versions of PCA has not been systematically compared, we will now do so in a simulation study.

3. SIMULATION STUDY

We performed a simulation study with the aim to compare the performance of various robust versions of PCA on data with $p=6$. Each of the generated data sets comes from $K=3$ groups and after reducing the dimensionality by a robust PCA, a classification method is used to classify the observations into the groups. Here, linear discriminant analysis (LDA), which suffers from the presence of outliers in the data, is compared with its several robust counterparts. Then, classification performance of various approaches is compared.

The loss of information due to a dimensionality reduction is evaluated by means of the performance of a consequently computed classification analysis. We use the fact that dimensionality reduction can describe differences among groups of the data, reveal the dimensionality of the separation among groups, and express the contribution of individual variables to this separation (Hastie et al., 2001).

We will need the following notation.

$$\mu_1 = (0,0,0,0,0,0)^T, \mu_2 = (1,1,1,0,0,0)^T, \mu_3 = (0,0,0,1,1,1)^T,$$

$$v_1 = (0,5,5,0,0,0)^T, v_2 = (1,1,1,0,5,5)^T, v_3 = (5,5,0,1,1,1)^T,$$

$$\Sigma_2 = \text{diag} \{0.5, \dots, 0.5\},$$

$$\Sigma = \begin{pmatrix} P & 0 \\ 0 & Q \end{pmatrix},$$

$$P = \begin{pmatrix} 1 & 0.95 & 0.3 \\ 0.95 & 1 & 0.1 \\ 0.3 & 0.1 & 1 \end{pmatrix},$$

$$Q = \begin{pmatrix} 1 & -0.499 & -0.499 \\ -0.499 & 1 & -0.499 \\ -0.499 & -0.499 & 1 \end{pmatrix} \quad (2)$$

In four different simulation studies (A,B,C,D), training data are generated from the distributions given below. Simulation A considers Gaussian (normally distributed) data and the others contaminated data by outliers. Let us use the notation $N(\mu, \Sigma)$ for the Gaussian distribution. In each group, the total number of n_k observations is generated from the k -th group for $k=1,2,3$.

A. $N(500, \Sigma_2), n_1 = n_2 = n_3 = 500$

B. $(1 - \varepsilon) N(\mu_k, \Sigma) + \varepsilon N(\mu_k, 25 \Sigma_2),$

where $n_1 = n_2 = n_3 = 500, \varepsilon = 1/5$

C. $(1 - \varepsilon) N(\mu_k, \Sigma) + \varepsilon N(v_2, \Sigma_2),$

where $n_1 = 750, n_2 = 500, n_3 = 250, \varepsilon = 1/5$

D. $(1 - \varepsilon_k) N(\mu_k, \Sigma) + \varepsilon_k N(v_2, \Sigma_2), \quad (3)$

where $n_1 = n_2 = n_3 = 500, \varepsilon_1 = 1/5, \varepsilon_2 = 1/4, \varepsilon_3 = 1/3.$

The classification rule learned on the training data is consequently used to assign each observation from a validation set to one of the three groups. Validation data in the k -th group are randomly generated from the Gaussian distribution $N(\mu_k, \Sigma)$. This is performed 250 times for each of the four simulation studies. Tables 1 and 2 evaluate the classification performance by means of the average classification error.

We use several robust versions of LDA to learn the classification rules in a robust way. All of them have are based on the robust multivariate estimators, which were already used in Section 3 to define robust versions of PCA. Thus, e.g. LDA-MVE denotes a robust

version of LDA, which replaces the mean and covariance matrix by their robust counterparts computed by means of the MVE estimator. These methods can be interpreted as methods based on a deformed (i.e. robustified) version of the Mahalanobis distance (Hubert et al., 2008; Filzmoser & Todorov, 2011). Robust classification procedures are computed used the package `rrcov` in R software (Todorov & Filzmoser, 2009) using default values of parameters.

Parts of the results of the simulation are shown in Table 1 and Table 2. For simulation A, we give results only for LDA and LDA-MCD, because other robust LDA versions do not greatly differ from LDA-MCD. If LDA is

Table 1. Average classification error for selected methods of dimensionality reduction and classification analysis in simulation studies A, B and C

Method	A		B		C	
	LDA	LDA-MCD	LDA	LDA-MCD	LDA	LDA-MCD
PCA	0.491	0.527	0.520	0.529	0.826	0.428
PCA-MVE	0.515	0.522	0.516	0.523	0.604	0.533
PCA-MCD	0.518	0.523	0.518	0.523	0.604	0.533
PCA-S1	0.517	0.522	0.493	0.528	0.606	0.533
PCA-S2	0.517	0.524	0.519	0.523	0.606	0.534
PCA-SD	0.518	0.524	0.519	0.524	0.606	0.528
PCA-OGK	0.486	0.528	0.493	0.527	0.608	0.522
PCA-PROJ	0.480	0.527	0.494	0.528	0.560	0.496
PCA-GRID	0.481	0.525	0.493	0.528	0.560	0.496
ROBPCA	0.489	0.528	0.493	0.528	0.606	0.522
SPCA	0.482	0.528	0.499	0.527	0.638	0.538

Table 2. Average classification error for selected methods of dimensionality reduction and classification analysis in simulation study D

Method	LDA	LDA-MVE	LDA-MCD	LDA-S1	LDA-S2	LDA-SD	LDA-OGK
PCA	0.747	0.528	0.455	0.528	0.527	0.518	0.630
PCA-MVE	0.663	0.564	0.542	0.555	0.550	0.565	0.566
PCA-MCD	0.665	0.572	0.538	0.564	0.552	0.565	0.576
PCA-S1	0.683	0.565	0.529	0.558	0.552	0.571	0.573
PCA-S2	0.668	0.571	0.539	0.556	0.552	0.565	0.568
PCA-SD	0.672	0.586	0.535	0.573	0.557	0.573	0.584
PCA-OGK	0.691	0.586	0.512	0.590	0.549	0.577	0.602
PCA-PROJ	0.722	0.545	0.469	0.552	0.548	0.519	0.710
PCA-GRID	0.717	0.543	0.465	0.556	0.548	0.527	0.686
ROBPCA	0.681	0.570	0.526	0.557	0.549	0.566	0.574
SPCA	0.724	0.532	0.465	0.534	0.519	0.510	0.635

used for classification, robust versions of PCA not based on a direct estimation of the covariance matrix yield the best results. If LDA-MCD is used, all versions of PCA yield very similar results. Thus, the choice of the classification rule is more important than the choice of the dimensionality reduction procedure.

In simulations B and C, PCA is outperformed by each of the robust PCA versions if LDA is used as a classification method. If LDA-MCD is used instead, various PCA versions do not greatly differ in simulation B and robust PCA is outperformed by its classical counterpart in simulation C. We can say that contamination B does not complicate the classification analysis substantially. Simulation C is the only one with unequal sample sizes and robust procedures seem to suffer from it more than their classical counterparts. Nevertheless, it is rather a property of classification methods that they suffer from an unbalanced design.

Finally, the results are less obvious to interpret in simulation D. The worst results are obtained with a non-robust approach. PCA-PROJ combined with LDA-OGK is almost as bad. On the other hand, PCA-PROJ together with LDA-MCD give the very best classification result. Thus, PCA-PROJ has a high potential, but should not be combined with unsuitable classification procedures.

The simulation results indicate that LDA-MCD gives generally the most suitable classification procedure. PCA-PROJ or PCA-GRID should be combined with LDA-MCD to reduce the dimensionality reliably in a highly robust way. Besides, the simulations show that the combination of PCA with robust LDA leads surprisingly weak results. Indeed, a robust approach may

lead to worse results compared to classical methods, even for data contaminated by outliers. Therefore, we strongly recommend to combine a robust classification method with a robust version of PCA.

4. CONCLUSIONS

Data analysis in the situation with a large number of variables becomes an important task in a variety of applications in management or econometrics (Kalina, 2013). However, numerous data analysis in the context of both data mining and statistics are unsuitable for the task of information extraction from such data because of the curse of dimensionality. Although dimensionality reduction represents a popular solution, standard dimensionality reduction procedures are very sensitive to the presence of outlying measurements in the data, which makes robustness with respect to outliers an important requirement.

In this paper, we give an overview of available robust dimensionality reduction methods. They are known to be resistant against the presence of outliers in the data. They are usually computationally intensive, but their implementations are nowadays available. We believe that the most reliable implementation nowadays appears in R software. Some of the methods are suitable for data with a very large p , e.g. for $p > 1000$ or $p > 10000$, but there is no guarantee that a particular implementation in a commercial software is valid also for such a large p .

The main contribution of this paper is a simulation study comparing various robust classification methods, which are computed after a prior dimensionality reduction. The loss of information caused by particular dimensionality reduction procedures

complicates a consequent information extraction from the data and we compare it by a classification accuracy measured in a consequently perform classification analysis. To the best of our knowledge, a systematic overview of the performance of robust dimensionality reduction procedures followed by robust classification methods has not been performed. Based on the results of our study, we conclude that there is no uniformly optimal method for all possible structures of multivariate data and standard methods do not perform too badly on data which are only slightly contaminated.

On rather severely contaminated data, the simulations show that robust methods outperform the classical PCA. If we do not have the information about the presence of outliers in the data, it is recommended to use the robust method. Dimensionality reduction does lead to a worse classification result. As a good solution, it turns out to combine the robust PCA based on projection pursuit with a robust classification analysis based on the MCD estimator. Thus, highly robust dimensionality procedures PCA-PROJ and PCA-GRID turn out to be the winners of the simulation studies. The highly robust classification method LDA-MCD turns out to be the winner among all classification analysis procedures.

Dimensionality reduction is not the only solution allowing to analyze multivariate data with a large number of variables. Regularized methods represent an alternative for $n < p$ not requiring a dimensionality reduction. They have obtained recent attention (Kalina, 2014a) and it has been sometimes claimed that some of the regularized methods have been empirically observed to possess reasonable robustness properties (Tibshirani, 1996). It is true that regularization may cause a certain local

robustness against small departures in the observed data, but cannot ensure robustness against severe outliers. Robustness properties of regularized methods remain to be a topic of our future systematic research.

Of course, methods overviewed in this paper have their limitations. Their hidden assumption is that the data come from a Gaussian distribution contaminated by some other distribution, commonly a different Gaussian distribution with a much larger variance. Besides, in specific applications, robust PCA may be outperformed by robust tailor-made methods for the given context. For example in linear regression it may be recommendable to use a robust version of partial least squares, because robust PCA performed on the independent variables does not and cannot take into account the response variable.

Acknowledgements

The work of J. Kalina was supported by the grant GA13-17187S of the Czech Science Foundation. The work was financially supported by the Neuron Fund for Support of Science.

References

- Barlow, J.L., Bosner, N., & Drmač, Z. (2005). A new stable bidiagonal reduction algorithm. *Linear Algebra and its Applications*, 397, 35-84.
- Breiman, L. (1996). Heuristics of instability and stabilization in model selection. *Annals of Statistics*, 24, 2350-83.
- Croux, C., Filzmoser, P., & Oliveira, M.R. (2007). Algorithms for projection-pursuit robust principal component analysis.

Chemometrics and Intelligent Laboratory Systems, 87 (2), 218-225.

Croux, C., & Haesbroeck, G. (2000). Principal component analysis based on robust estimators of the covariance or correlation matrix: influence functions and efficiencies. *Biometrika*, 87, 603-618.

Croux, C., & Ruiz-Gazen, A. (2005). High breakdown estimators for principal components: The projection pursuit approach revisited. *Journal of Multivariate Analysis*, 95 (1), 206-226.

Davies, P.L. (1987). Asymptotic behaviour of S-estimates of multivariate location parameters and dispersion matrices. *Annals of Statistics*, 15 (3), 1269-1292.

Donoho, D.L. (1982). Breakdown properties of multivariate location estimators. Ph.D. qualifying paper. Boston,

MA, USA: Harvard University.

Filzmoser, P., & Todorov, V. (2011). Review of robust multivariate statistical methods in high dimension. *Analytica Chimica Acta*, 705, 2-14.

Gnadadesikan, R., & Kettenring J.R. (1972). Robust estimates, residuals, and outlier detection with multiresponse data. *Biometrics*, 28, 81-124.

Greene, W.H. (2012). *Econometric analysis*. Seventh edition. Upper Saddle River, NJ, USA: Pearson.

Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of machine learning research*, 3, 1157-1182.

Hastie, T., Tibshirani, R., & Friedman, J. (2001). *The elements of statistical learning*. New York, NY, USA: Springer.

Hubert, M., Rousseeuw, P.J., & Vanden Branden, K. (2005). ROBPCA: A new approach to robust principal component analysis. *Technometrics*, 47, 64-79.

Hubert, M., Rousseeuw, P.J., & Van Aelst, S. (2008). High-breakdown robust multivariate methods. *Statistical Science*, 23, 92-119.

Jurečková, J., & Picek, J. (2006). *Robust statistical methods with R*. Chapman & Hall/CRC, Boca Raton.

Kalina, J. (2014a). Classification methods for high-dimensional genetic data. *Biocybernetics and Biomedical Engineering*, 34 (1), 10-18.

Kalina, J. (2014b). Machine learning and robustness aspects. *Serbian Journal of Management*, 9 (1), 131-144.

Kalina, J. (2013). Highly robust methods in data mining. *Serbian Journal of Management*, 8 (1), 9-24.

Kalina, J. (2012a). On multivariate methods in robust econometrics. *Prague Economic Papers*, 21 (1), 69-82.

Kalina, J. (2012b). Implicitly weighted methods in robust image analysis. *Journal of Mathematical Imaging and Vision*, 44 (3), 449-462.

Lempert, R.J., Groves, D.G., Popper, S.W., & Bankes, S.C. (2006). A general, analytic method for generating robust strategies and narrative scenarios. *Management Science*, 52 (4), 514-528.

Liebmann, B., Filzmoser, P., & Varmuza, K. (2009). Robust and classical PLS regression compared. Preprint. Available: <http://www.statistik.tuwien.ac.at/forschung/CS/CS-2009-8complete.pdf> (July 25, 2014).

Locantore, N., Marron, J.S., Simpson, D.G., Tripoli, N., Zhang, J.T., & Cohen, K.L. (1999). Robust principal component analysis for functional data. *Test*, 8 (1), 1-73.

Ma, Y., & Zhu, L. (2013). A review on

dimension reduction. *International Statistical Review*, 81 (1), 134-150.

Martinez, W.L., Martinez, A.R., & Solka, J.L. (2011). *Exploratory data analysis with MATLAB*. Second edition. London, UK: Chapman & Hall/CRC.

McFerrin, L. (2013). Package HDMD. Available: <http://cran.r-project.org/web/packages/HDMD/HDMD.pdf> (June 14, 2013).

Pourahmadi, M. (2013). *High-dimensional covariance estimation*. Hoboken, NJ, USA: Wiley.

Rencher, A.C. (2002). *Methods of multivariate analysis*. Second edn. New York, NY, USA: Wiley.

Rousseeuw, P.J. (1985). Multivariate estimation with high breakdown point. Pp. 283-297 in Grossmann W., Pflug G., Vincze I., & Wertz W. (Eds.), *Mathematical Statistics and Applications*, Vol. B. Dordrecht, NL: Reidel Publishing Company.

Rousseeuw, P.J., & Leroy, A.M. (1987). *Robust regression and outlier detection*. New York, NY, USA: Wiley.

Stahel, W.A. (1981). Breakdown of covariance estimators. Research report 31, Fachgruppe für Statistik. Zurich, CH: ETH.

Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society B*, 58 (1), 267-288.

Todorov, V., & Filzmoser, P. (2009). An object oriented framework for robust multivariate analysis. *Journal of Statistical Software*, 32 (3), 1-47.

Višek, J.Á. (2011). Consistency of the least weighted squares under heteroscedasticity. *Kybernetika*, 47 (2), 179-206.

Xanthopoulos, P., Pardalos, P.M., & Trafalis, T.B. (2013). *Robust data mining*. New York, NY, USA: Springer.