Robust and sparse multiclass classification by the optimal scoring approach

I. Hoffmann[1*], P. Filzmoser[1], C. Croux[2]

[1] *Institute of Statistics and Mathematical Methods in Economics, Vienna University of Technology, Wiedner Hauptstrasse 8-10, 1040 Vienna, Austria; irene.hoffmann@tuwien.ac.at, p.filzmoser@tuwien.ac.at*
[2] *Faculty of Business and Economics, KU Leuven, Leuven, Belgium; christophe.croux@econ.kuleuven.be*
[*] *Presenting author*

Keywords. *Multiclass discriminant analysis; Outlier detection; Variable selection;*

1 Introduction

In regression analysis a great variety of so-called sparse methods have been developed, which perform simultaneous model estimation and variable selection due to restrictions on the coefficient estimate. Thereby estimation precision can be increased and the selected models are easier to interpret. This is especially useful for data sets with a large number of predictor variables.

1.1 Sparse LTS regression

Sparse least trimmed squares (LTS) introduced by **?** is a robust regression estimator with a lasso penalty (L_1 norm penalty on the coefficient estimate), which has good robustness properties and a fast algorithm.

1.2 Optimal Scoring

Several formulations for linear discriminant analysis (LDA) exist leading to the same discriminant rules. Optimal scoring being one of them recasts the classification problem into a regression framework and iteratively models the class-membership as continuous variable. Let Q be smaller than the number of groups G, commonly

$Q = G - 1$. Then solve for $q = 1, ..., Q$

$$\min_{\boldsymbol{\beta}_q, \boldsymbol{\theta}_q} \{\|\boldsymbol{Y}\boldsymbol{\theta}_q - \boldsymbol{X}\boldsymbol{\beta}_q\|^2\} \quad \text{s.t.} \quad \frac{1}{n}\boldsymbol{\theta}_q^T \boldsymbol{Y}^T \boldsymbol{Y} \boldsymbol{\theta}_q = 1, \quad \boldsymbol{\theta}_q^T \boldsymbol{Y}^T \boldsymbol{Y} \boldsymbol{\theta}_l = 0 \quad \forall l < q. \tag{1}$$

where $\boldsymbol{X}$ is the data matrix with n observations and p variables and $\boldsymbol{Y}$ an $n \times G$ matrix of dummy variables coding the class membership of the observations. Adding an L_1 penalty for $\boldsymbol{\beta}_k$ to the minimization problem leads to sparse discriminant analysis as proposed by **?**.

2 Methodology

We propose to recast the classification problem into a robust regression framework by optimal scoring in order to obtain a robust discriminant model. The minimization problem for $\boldsymbol{\beta}_q$ and $\boldsymbol{\theta}_q$ is solved iteratively. For fixed $\boldsymbol{\theta}_q$ (a vector of random values in the first iteration) $\boldsymbol{\beta}_q$ is estimated robustly by a sparse fast LTS algorithm (using starting observations representing all classes). Then $\boldsymbol{\theta}_q$ is calculated using $\boldsymbol{X}\boldsymbol{\beta}_q$ as response but excluding the observations detected as outliers by sparse LTS in the former step. Robust LDA is then applied to $(\boldsymbol{X}\boldsymbol{\beta}_1, ..., \boldsymbol{X}\boldsymbol{\beta}_Q)$. For the selection of the sparsity parameter cross validation is performed and a robust misclassification rate (excluding potential outliers) is used to decide for the best model.

3 Evaluation

The proposed algorithm is evaluated on simulated data with few relevant variables and a large number of noise variables. Outliers are included to demonstrate the stability of the model estimation. False negative and false positive rates for the selected variables and the detected outliers are reported. The prediction performance is compared to classical sparse LDA and to robust LDA evaluated only on the relevant variables (oracle estimator).

References

Alfons, A., Croux, C. & Gelper, S. (2013). Sparse least trimmed squares regression for analysing high-dimensional large data sets. *The Annals of Applied Statistics*, **7(1)**, 226–248.

Clemmensen, L., Hastie, T., Witten, D. & Ersboll, B. (2011). Sparse discriminant analysis. *Technometrics*, **53**, 406–413.

Acknowledgement: This work is supported by the Austrian Science Fund (FWF), project P 26871-N20.