Sparse and robust modeling for high-dimensional data

I. Hoffmann[1], P. Filzmoser[1], C. Croux[2], S. Serneels[3], F.S. Kurnaz[4], K. Varmuza[1]

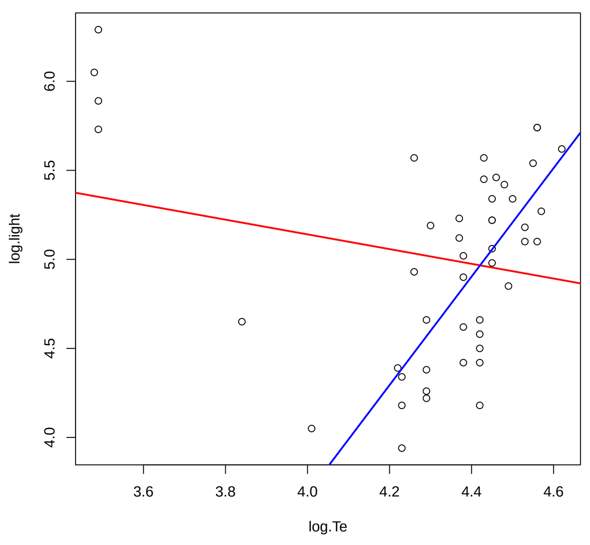
[1]TU Wien, [2]KU Leuven, [3]BASF Corporation, [4]Yildiz Technical University

irene.hoffmann@tuwien.ac.at

Data problems and approaches

- **Outliers** in the data may distort models heavily.

Robust statistics: model estimation based on the majority of the data.



- Reduction of the influence of single observations on the model estimation.
- Identification of outliers, i.e. observations which are different than the majority.

Ordinary least squares (–): minimize sum of squared residuals
Least trimmed squares [7] (–): minimize sum of smallest 75% of squared residuals

- **Uninformative variables** do not contribute to the explanation of the response, but increase model uncertainty.

Sparse modeling: estimation with intrinsic variable selection.

Example: Lasso regression [8]

$$\min_{\beta} \frac{1}{n}\|\boldsymbol{X}\beta - \boldsymbol{y}\|^2 + \lambda\|\beta\|_1 \qquad (1)$$

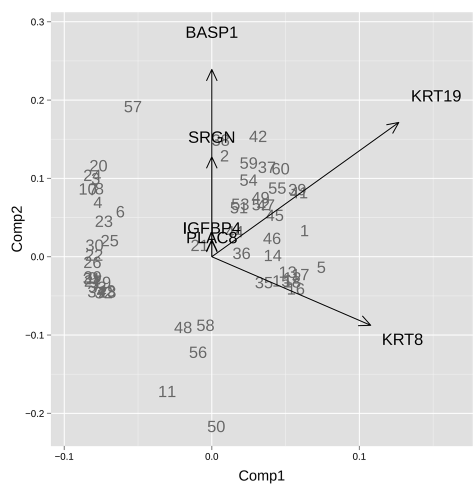
- Applicable also for data sets $\boldsymbol{X}$ with less observations n than variables p.
- Favors zeros in coefficient vector β. Reduction of noise $\rightarrow$ increasing model precision.
- Identification of relevant variables, easier to interpret.

- **Correlated predictor variables** lead to ill conditioned covariance matrix of $\boldsymbol{X}$ and unstable estimates of β.
  - In partial least squares regression (**PLS**) [9] uncorrelated latent variables are constructed from linear combinations of the original variables, such that the squared covariance to the response $\boldsymbol{y}$ is maximized. Then a linear regression model is estimated on the latent variables.
  - In **elastic net** [10] regression an L_2 penalty is added to (1). Correlated variables tend to obtain similar coefficient estimates.

Sparse partial robust M regression [1]

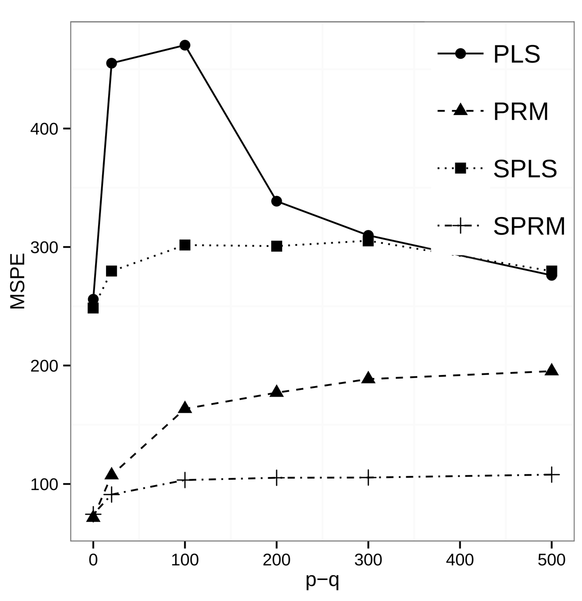
Combines ideas from

- **Robust PLS** [2]
  Iteratively down-weight the influence of outliers on the construction of latent variables as well as the regression model.
- **Sparse PLS** [3]
  Latent variables are constructed only with a subset of the original variables.

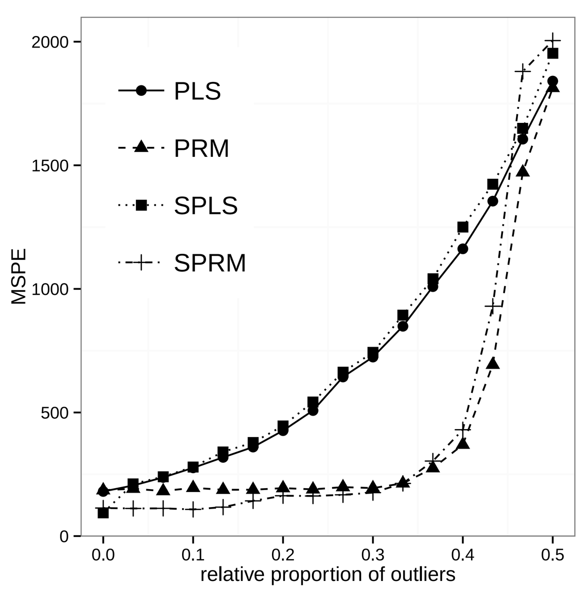


Biplot of a sparse and robust PLS model with p = 5571 variables.

Simulation results: comparison with PLS, sparse PLS (SPLS) and robust PLS (PRM) in terms of mean squared prediction error (MSPE)



Fixed proportion of 10% outliers, increasing number of uninformative variables.

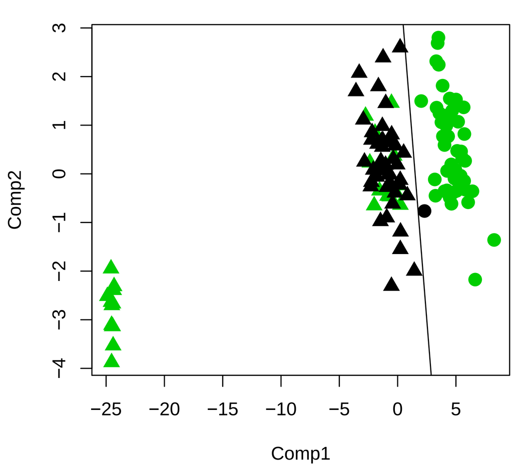


Fixed number of 500 uninformative variables, increasing number of outliers.

Sparse and robust PLS for binary classification [4]

- Initial down-weighting of potential outliers.
- **Step 1:** Iteratively re-weighting to estimate robust sparse PLS directions.
- **Step 2:** Robust LDA in the sparse score space.

- Down-weighting of outliers in the predictor space and for observations with potentially wrong class labels.
- Outlier detection group wise.
- Use outlier weights for robust LDA in the score space.



Robust and sparse multi-group classification [5]

- **The optimal scoring approach**
  - Iteratively transform categorical class membership into continuous values: the optimal scores.
  - Optimal scores are used as response in a regression model.

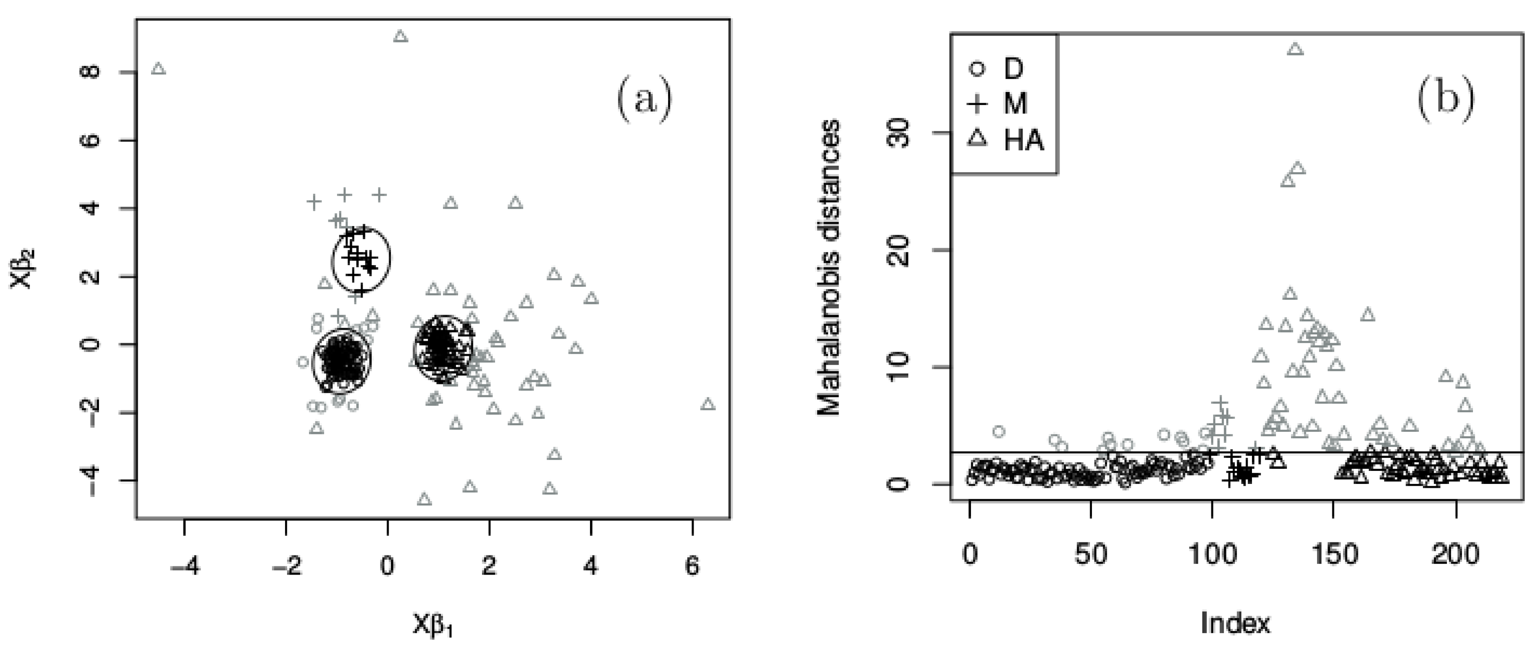
  Solve for $h = 1, \dots, H$

$$\min_{\beta_h, \theta_h} \frac{1}{n}\|\boldsymbol{Y}\theta_h - \boldsymbol{X}\beta_h\|^2 \quad \text{s.t.} \quad \theta_h^T \boldsymbol{D}\theta_h = 1, \quad \boldsymbol{Q}_h^T \boldsymbol{D}\theta_h = \boldsymbol{0},$$

  where $\boldsymbol{Q}_h = [\boldsymbol{Q}_{h-1}, \hat{\theta}_{h-1}]$ is a $K \times h$ matrix, $\boldsymbol{D} = \frac{1}{n}\boldsymbol{Y}^T\boldsymbol{Y}$ is a $K \times K$ diagonal matrix of class proportions and $\boldsymbol{Y}$ the dummy matrix of class memberships.

- Developments in robust sparse regression can be transferred to multi-group classification problems.
- We combine Lasso regression with the iterative reweighting algorithm to down-weight the influence of outliers.

Robust and sparse multi-group classification via the optimal scoring approach



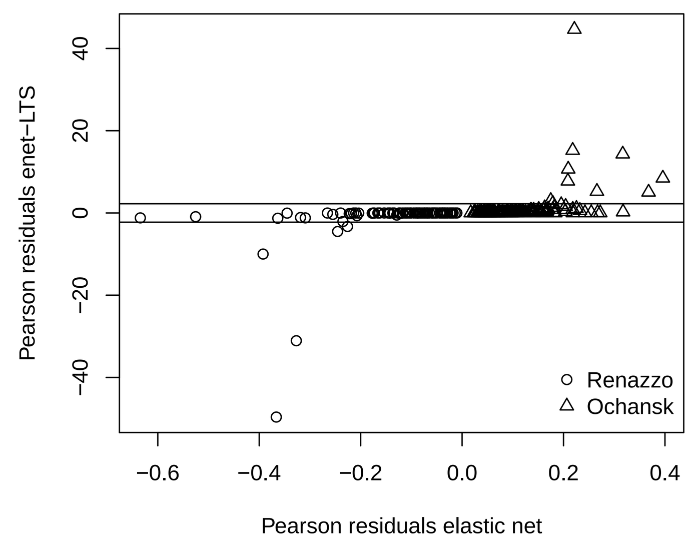
(a) Visualization of the robustly transformed subspace; (b) Mahalanobis distances to the group centers in the subspace.

Robust linear and logistic models with elastic net [6]

- First robust and sparse methods with elastic net penalty.
  - Induces sparsity on the coefficients by L_1 penalty.
  - Favors similar coefficient values for correlated variables due to L_2 penalty.
- Based on C-step algorithm from fast least trimmed squares regression [7].

- Mass spectra of 2 groups: meteorites Renazzo and Ochansk.
- with n_R = 110 and n_O = 160 observations, respectively.
- Number of variables: p = 1540
- Evaluation by trimmed mean negative loglikelihood (MNLL).

model typ	# variables	trim. MNLL
elastic net	136	0.00866
enet-LTS	397	0.00014



References

[1] Hoffmann, I., Serneels, S., Filzmoser, P., Croux, C. (2015). *Sparse partial robust M regression.* Chemometrics and Intelligent Laboratory Systems, 149, 50-59.

[2] Serneels, S., Croux, C., Filzmoser, P., Van Espen, P. J. (2005). *Partial robust M-regression.* Chemometrics and Intelligent Laboratory Systems, 79(1), 55-64.

[3] Chun, H., Keleş, S. (2010). *Sparse partial least squares regression for simultaneous dimension reduction and variable selection.* Journal of the Royal Statistical Society: Series B (Statistical Methodology), 72(1), 3-25.

[4] Hoffmann, I., Filzmoser, P., Serneels, S., Varmuza, K. (2016). *Sparse and robust PLS for binary classification.* Journal of Chemometrics, 30(4), 153-162.

[5] Hoffmann, I., Filzmoser, P., Croux, C. *Robust and sparse multi-group classification by the optimal scoring approach.* Submitted for publication.

[6] Kurnaz, F. S. , Hoffmann, I., Filzmoser, P. *Robust and sparse estimation methods for high-dimensional linear and logistic regression.* Submitted for publication.

[7] Rousseeuw, P. J., Van Driessen, K. (2006). *Computing LTS regression for large data sets.* Data mining and knowledge discovery, 12(1), 29-45.

[8] Hastie, T., Tibshirani, R., Wainwright, M. (2015). *Statistical learning with sparsity: the lasso and generalizations.* CRC Press.

[9] Wold, S., Sjöström, M., Eriksson, L. (2001). *PLS-regression: a basic tool of chemometrics.* Chemometrics and Intelligent Laboratory Systems, 58(2), 109-130.

[10] Zou, H., Hastie, T. (2005). *Regularization and variable selection via the elastic net.* Journal of the Royal Statistical Society: Series B (Statistical Methodology), 67(2), 301-320.

ACKNOWLEDGMENTS

This research is supported by the *Austrian Science Fund (FWF)*, project number P 26871-N20, by grant TUBITAK 2214/A from the Scientific and Technological Research Council of Turkey (TUBITAK) and by BASF Corporation.