A Robust Approach to Regularized Discriminant Analysis

Moritz Gschwandtner
Vienna University of Technology, Austria

Peter Filzmoser
Vienna University of Technology, Austria

Cristophe Croux
University Center of Statistics, K. U. Leuven, Belgium

Gentiane Haesbroeck
University of Liege, Belgium

Abstract: Machine learning is a wide open field in today's statistics. In the classification setting, the goal is to find a decision or classification rule that assigns an observation to one - and only one - of k groups. If the decision rule is based on a training set that contains labelled observations, the learning process is called 'supervised', otherwise 'unsupervised'. For supervised learning problems, the performance of classification rules is most often measured by estimating the misclassification rate, which can be achieved by techniques like cross validation or bootstrap. It is desirable to minimize not only the training error, but also the misclassification rate of independent test data.

Linear discriminant analysis assumes that the groups are normally distributed with common covariance matrix Σ, i.e.

$$\Sigma_1 = \Sigma_2 = \ldots = \Sigma_k = \Sigma$$

If Σ and $\Theta := \Sigma^{-1}$ are unknown, they have to be estimated from the data. Especially in the high dimensional case, where $n < p$, the classical estimations may become unstable; they can be improved by modern techniques like regularization, though. Furthermore, classical methods often suffer from the presence of outliers in the data. Therefore, in many cases 'robustification' is a must have in the advanced statistical analysis process.

Sparse inverse covariance matrices are estimated following a method proposed by Friedman et al. (2007). Possible outliers are dealt with by using a subset of the data in a MCD-like manner as suggested by Croux et al. (2010). A combination of both regularization and robustification of Θ can be achieved by centering the data and maximazing a penalized log-likelihood function

$$\mathcal{L}(H, (\mu, \boldsymbol{\Theta})) = \log\det(\boldsymbol{\Theta}) - \frac{1}{h}\sum_{i \in H}(\mathbf{x}_i - \mu)^\top \boldsymbol{\Theta}(\mathbf{x}_i - \mu) - \lambda ||\boldsymbol{\Theta}||_1$$

where H denotes the index subset of the data ($|H| = h < n$), μ is the location parameter, and $\lambda > 0$ is a penalty parameter, controlling the sparseness of the resulting estimate $\hat{\Theta}$.

Focusing on 'high dimension small sample size problems', the method is presented, comparisons to standard methods are given, and results of extensive simulation studies are discussed.

References

Croux, C., Gelper, S., & Haesbroeck, G. (2010). Robust scatter regularization. In G. Saporta & Y. Lechevallier (Eds.), *Compstat 2010, book of abstracts* (p. 138). Paris: Conservatoire National des Arts et Métiers (CNAM) and the French National Institute for Research in Computer Science and Control (INRIA).

Filzmoser, P., Maronna, R., & Werner, M. (2008). Outlier identification in high dimensions. *Computational Statistics and Data Analysis*, *52*, 1694-1711.

Friedman, J. H. (1989). Regularized discriminant analysis. *Journal of the American Statistical Association*, *84*, 165-175.

Friedman, J. H., Hastie, T., & Tibshirani, R. (2007). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, *9*, 432-441.