Robust Selection of Variables in the Linear Discriminant Analysis

V. Todorov[1]

[1] Austro Control, Vienna, Austria

Keywords: Discriminant Analysis, variable selection, Wilks' lambda, MCD

1 Abstract

A commonly used procedure for reduction of the number of variables in the linear discriminant analysis is the stepwise method for variable selection. Although often criticized, when used carefully, this method can be a useful prelude to a further analysis. The contribution of a variable to the discriminatory power of the model is usually measured by the maximum likelihood ratio criterion, reffered to as Wilks' lambda.

It is well known that the Wilks' lambda statistic is extremely sensitive to the influence of outliers. Krusinska et al. (1989) proposed a robust version of this statistic using the M-estimates of the covariance matrix. Todorov et al. (1990) computed the Wilks' lambda relative to the Minimum Volume Ellipsoid (MVE) estimates and showed on examples that it can resist the influence of outliers, but the computational burden did not allow to carry out a more thorough study of its performance.

In this work will be constructed a robust version of the Wilks' lambda statistic based on the Minimum Covariance Discriminant (MCD) estimator and its reweighed version which has a higher efficiency. Taking advantage of the availability of a fast algorithm for computing the MCD (Rousseeuw et al. (1999)) a simulation study will be done to evaluate the performance of this statistic.

References

E. Krusinska and J. Liebhart (1989). Some Further Remarks on the Robust Selection of Variables in Discriminant Analysis. *Biom. J.*, 31, 2, 227–233.

P. Rousseeuw, P.J. and K. Van Driessen (1999). A Fast Algorithm for the Minimum Covariance Determinant Estimator. *Technometrics*, 41, 212–223.

V.K. Todorov, N.M. Neykov and P.N. Neytchev (1990). Robust Selection of Variables in the Discriminant Analysis Based on MVE and MCD Estimators. in *Proceedings in Computational statistics, COMPSTAT'90* Physica Verlag, Heidelberg.