Error Rates for Multiple Group Robust Linear Discriminant Analysis

K. Joossens [1]

[1] Dept. of Applied Economics, K. U. Leuven, Naamsestraat 69, B-3000 Leuven, BELGIUM

Keywords: Error rate, Multiple Discriminant Analysis, Robustness.

Abstract

In multiple discriminant analysis one observes multiple groups of multivariate observations, forming together the *training sample*. For the data in the training sample, it is known to which group they belong. On the basis of the training sample, discriminant variables are constructed. Following Fishers' approach, these discriminant variables are obtained by projecting the original variables on the first few eigenvectors of a matrix constructed using the the *within groups* and *between groups* covariances. Robustifying the Fisher approach to multiple discriminant analysis is obvious by estimating the location and scatter in every group robustly and using these as plug-in estimates. In this work we focus on the associated error rates of this procedure.

By simulation experiments, we compare the performance in terms of error rate between the robust and standard procedure in a variety of sampling schemes, including t-distributions and contaminated normals. We focus on the three-group case for which we have an exact expression for the error rate. This allows us to check whether error rates estimates like (10-fold) cross-validation and the bootstrap yield good approximations to the true error rate. We also obtain some preliminary theoretical results on influence analysis on error rates for the multiple group case.

(This work is based on joint work with P. Filzmoser and C. Croux.)