Discriminant Procedures Based on Efficient Robust Discriminant Coordinates

K. Crimin[1], J.W. McKean [2] and S.J. Sheather[3]

[1] Pfizer
[2] Western Michigan University
[3] Texas A&M University

Keywords: Affine-equivariant estimators, Least squares, Linear discriminant rule, Nearest center rule, Rank-based analysis, Wilcoxon analysis, Visualization.

Abstract

For multivariate data collected over groups, discriminant analysis is a two-stage procedure: separation and allocation. For the traditional least squares procedure, separation of training data into groups is accomplished by the maximization of the Lawley-Hotelling test for differences between group means. This produces a set of discriminant coordinates which are used to visualize the data. Using the nearest center rule, the discriminant representation can be used for allocation of data of unknown group membership. In this talk, we propose a discriminant analysis based on efficient robust discriminant coordinates. These coordinates are obtained by the maximization of a Lawley-Hotelling test based on robust estimates. The design matrix used in the fitting is the usual one-way incidence matrix of zeros and ones; hence, our procedure uses highly efficient robust estimators to do the fitting. This produces efficient robust discriminant coordinates for visualizing the data. Further, the allocation is based on the robust discriminant representation of the data using the nearest robust center rule. While our procedure is general, we discuss three specific procedures including an affine-equivariant procedure. The robustness of our procedures is verifed in several examples. We present the results of a Monte Carlo study on probabilities of misclassifications of the procedures over a variety of error distributions. The robust discriminant analyses perform practically as well as the traditional procedure for good data and they are much more efficient than the traditional procedure in the presence of outliers and heavy tailed error distributions.