Implementing a new method for Discriminant Analysis when Group Covariance Matrices are nearly singular

Winfried Theis[1], Christian Röver[1], and Britta Pouwels[2]

[1] Sonderforschungsbereich 475, Fachbereich Statistik, Universität Dortmund, D-44221 Dortmund, Germany
[2] Universitäts- und Landesbibliothek Münster, D-48143 Münster, Germany

Abstract. We consider a unified description of classification rules for nearly singular covariance matrices. When the covariance matrices of the groups or the pooled covariance matrix become nearly singular, bayesian classification rules become seriously unstable. Several procedures have been proposed to tackle this problem, e.g. SIMCA, and Regularized Discriminant Analysis. Næs and Indahl (1998) discovered common properties for all of these procedures and proposed a unified classifier that incorporates the functionality of them all. Since the unified approach needs many parameters, they also proposed an alternative classifier with fewer parameters. We implemented both classifiers and compared them in a simulation study to the procedures RDA, LDA, and QDA. To enhance the comparability of our results we based the simulation study on the study of Friedman (1989).

In the implementation, we used a combination of the Nelder-Mead Simplex-algorithm and Simulated Annealing (Bohachevsky et al. (1986)) to optimize the classification error directly.

Keywords

DISCRIMINATION, MULTICOLLINEAR DATA, CLASSIFICATION ERROR, OPTIMIZATION

1 Introduction

The classical approaches to discriminant analysis are based on estimates of the inverse covariance matrices for each group (Quadratic Discriminant Analysis, QDA) or on a common inverse covariance matrix (Linear Discriminant Analysis, LDA). The estimates of the covariance matrix can become badly conditioned for several reasons. First of all the explanatory variables may be multicollinear. Another situation encountered is an insufficient number of observations per group, which leads to a similar problem. Since these problems are encountered in many studies, a wide variety of methods has been developed to tackle this problem. There are mainly two typical approaches: On the one hand dimension reduction techniques like Principal Component

Analysis (PCA), Partial Least Squares (PLS) or Minimal Error Rate Classifiers (MEC) (Röhl et al. (2002)) are applied. On the other hand there are methods which preserve the dimensionality but manipulate the covariance matrix estimates, like MCA, SIMCA, DASCO or Regularized Discriminant Analysis (RDA) (please refer to Næs and Indahl (1998) for these methods). Næs and Indahl (1998) recognized that the latter methods could be described in a unified way.

In this paper we consider first the unified description of Næs and Indahl in section 2. Next we explain our approach for an implementation of the method (section 3). We tested our algorithm in an extensive simulation study and compared it to RDA, and the classical LDA and QDA (sections 4 and 5). Finally we conclude with some comments and suggestions for further research.

2 Unified Description

Næs and Indahl recognized three "conceptual dimensions" in which the different methods for manipulating the covariance matrix can be organized. We reformulated these "dimensions" in the following way:

1. Manipulation of singular eigenvalues λ_k , $k = 1, \ldots, d$.
2. Shrinking of the group-specific covariance matrices S_j , $j = 1, \ldots, G$, towards the pooled covariance matrix S_p.
3. Shrinking of the covariance matrices S_j , $j = 1, \ldots, G$, towards the identitiy matrix I_d.

The first dimension yields the opportunity to replace some eigenvalues by values which result in more stable inverses or reduce the typical bias in eigenvalue-estimates. The aforementioned bias corresponds to the fact that small eigenvalues are estimated too small and large eigenvalues too large.

The second dimension focuses on stabilizing the estimate by moving towards an estimator which comprises more knowledge from the data, and is therefore more stable.

Finally the third dimension tries to solve the multicollinearity problem by reducing the influence of covariances between the explanatory variables.

Formula 1 gives the complete form of the covariance estimator for the jth group, which includes all regularisation possibilities.

$$\hat{\Sigma}_U^j = c_j \left[\delta (U_1^j U_2^j) \begin{pmatrix} m\Lambda_1^j + \alpha_j I_K & 0 \\ 0 & (\alpha_j + \beta_j) I_{d-K} \end{pmatrix} (U_1^j U_2^j)^T + (1 - \delta) S_p \right] \tag{1}$$

where $(U_1^j U_2^j)$ is a partition of the $d \times d$ eigenvector matrix of the j-th group covariance matrix, with U_1^j corresponding to the large eigenvalues in the K-dimensional diagonal matrix Λ_1^j and U_2^j corresponding to the $d - K$ small eigenvalues. The other parameters correspond to the different regularisations in the following way:

- $K \in \{0, \ldots, d\}$ is the dimension of the space of high variability.
- $\delta \in [0,1]$ models the shrinking towards the pooled covariance matrix.
- $\alpha_j \in [0, \infty)$ models the shrinking towards the identity matrix.
- m manipulates the large eigenvalues similarly for all groups.
- $\beta_j \in [0, \infty)$ is the replacement for the small eigenvalues.
- $c_j \in [0, \infty)$ is an additional scaling factor.

Thus all mentioned directions of regularisation can be achieved with this estimator. A great problem are the additional $3 + 2G$ parameters which have to be derived from the data to use this estimator. When a covariance estimate for LDA is to be regularized, the pooled covariance matrix is used instead of the group covariance matrices. Thus the group indices j and parameter δ are omitted in formula 1, and the eigenvalue-decomposition is applied directly to the pooled covariance matrix.

To avoid the large number of parameters of the unified approach, Næs and Indahl proposed an alternative estimator by omitting the parameters c_j, α_j and by not using a group-specific choice of β:

$$\hat{\Sigma}_A^j = \delta(U_1^j U_2^j) \begin{pmatrix} m\Lambda_1^j & 0 \\ 0 & \beta I_{d-K} \end{pmatrix} (U_1^j U_2^j)^T + (1-\delta)S_p \tag{2}$$

This parameter reduction leads mainly to the loss of the possibility of regularization in the third of our dimensions, the shrinking towards the identity matrix. Furthermore the group-specific manipulation of the small eigenvalues is lost.

In our simulations (see section 4) we compare these two approaches which will be called **U**nified **R**egularized **C**lassification method (URC), or **A**lternative **R**egularized **C**lassification method (ARC), resp., to standard LDA and QDA, and as a further competitor we chose Regularized Discriminant Analysis (Friedman (1989)), since RDA especially shrinks towards the identity matrix and so it is of great interest to compare the performance of the alternative estimator to RDA. RDA regularizes in dimensions 2 and 3 and actually has two paramters δ, γ steering the shrinkage. The first step is the shrinkage towards the pooled covariance matrix:

$$\hat{\Sigma}_j(\delta) = \delta S_j + (1-\delta)S_p \, ,$$

and the second step is towards a weighted identity matrix:

$$\hat{\Sigma}_j(\delta, \gamma) = (1-\gamma)\hat{\Sigma}_j(\delta) + \frac{\gamma}{d} tr(\hat{\Sigma}_j(\delta)) I_d, \; j = 1, \ldots, G$$

The unified estimate equals the RDA estimate, if

$$K = d, \; c_j = 1 - \gamma, \; m = 1, \; \alpha_j = \frac{\gamma tr(\delta S_j + (1-\delta)S_p)}{d(1-\gamma)\delta}.$$

Before the simulation study will be described, we point out some important issues about the implementation.