AJAS / SPAS

ISSN 2316-0861

African Journal of Applied Statistics
Vol. 4 (1), 2017, pages 273–287.
DOI: http://dx.doi.org/10.16929/ajas/2017.273.215

A comparative study on the regularized versions of discriminant analysis: An application to gene expression data

Olusola Samuel Makinde[*]

Department of Statistics, Federal University of Technology, P.M.B 704, Akure, Nigeria

Received November 25, 2017; Accepted December 30, 2017; Published Online February 8, 2018

Abstract. Discriminant analysis has been used in many application for classification and dimension reduction when the ratio of sample size to dimension diverges. However, the applicability of this method is almost impossible whenever sample size is bigger than dimension of the data. Efforts have been made to circumvent this problem by either regularise or penalise sample covariance matrices of the competing classes of observations. However, presence of redundant features in the data raises misclassification rates of discriminant rule. In this paper, we explore shrunken centroid regularised discriminant analysis for gene selection and regularised discriminant analysis as classification method based on various versions of regularised covariance matrices of competing classes of gene expression levels. The performance of the regularised linear and quadratic discriminant analysis in comparison with some other classification methods is illustrated using some gene expression data sets as well as simulated data.

Key words: Regularised covariance matrices; discriminant analysis; high dimensional data; shrunken centroid; gene expression data.
AMS 2010 Mathematics Subject Classification : 62H30, 62H10, 60E05.

Presented by Professor Gane Samb LO
University Gaston Berger, Saint-Louis (Sénégal)
Member of the Editors Board (Chief Editor).

osmakinde@futa.edu.ng
[*]Corresponding Author: Olusola Samuel Makinde
Email Corresponding Author : osmakinde@futa.edu.ng

Résumé. L'analyse discriminante a été utilisée dans beaucoup d'application pour la classification and das la réduction de dimension lorsque le rapport taille de l'échantillon/Dimension diverge. Toutefois, l'applicabilité de cette méthode est problématique si la taille de l'échantillon est plus grande que la dimension de données. Des efforts ont été faits pour régler cette difficulté soit en régularisant soit en pénalisant la matrice empirique des variances-covariances des classes d'observations en compétition. Cependant, la présence de caractéristiques redonnantes conduit à accroitre le taux de mal classement dans la discrmination. Dans ce papier, we explorons la méthode dite shrunken centroid regularized discriminant Analysis pour l'expression des gènes et celle de la méthode de l'analyse discriminate régularisée come outil de classement, relative à plusieurs versions de régularisatin des matrices de covariances des classes en compétitions relatives aux niveau d'espression des gènes. La performance de la régularization linéaire et quadratique de l'analyse discriminante en comparaison avec certaines autres méthodes de classification est illustrée par une étude de cas avec des jeux de données rélles et une étude de simulation.

1. Introduction

Classification methods like discriminant analysis has been considered in many important applications. For example, in classifying textile fabric data(Kiruthika and Chandrasekaran, 2012), iris data(Johnson et al., 2007), etc. Linear and quadratic discriminant analysis are known for their optimal performance in classification when competing populations are normally distributed. In practice, one may have to work with samples from competing populations. These methods become practically difficult to implement when dimension of the data is greater than sample sizes. This is because covariance matrix $\mathbf{\Sigma}$ of the data becomes singular, making it practically impossible to compute distance function $(\mathbf{x}-\boldsymbol{\mu})^\top\mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})$ in executing discriminant analysis, where $\boldsymbol{\mu}$ is the mean of the data.

In handling singularity of estimate of covariance matrix, approaches in literature to circumvent these problem can be categorised into three. One may choose to penalise the covariance matrix $\mathbf{\Sigma}$(Hastie et al., 1995; Witten and Tibshirani, 2011). Second, Dudoit et al. (2002) suggested that correlation among variables should be ignored and demonstrated a better performance of the approach using simulation and real data. This approach, referred to as Diagonal linear discriminant analysis (DLDA), involves replacing $\mathbf{\Sigma}$ in the discriminant function by $\mathbf{D}$, the diagonal of pooled covariance matrix of the competing classes. Then the discriminant rule based on the transformed discriminant function is carried out on the test data. Bickel and Levina (2004) argued that DLDA is not much lower in performance compared to Bayes rule in terms of proportion of correct classification. Fan et al. (2012) argued that it may perform very poor when using all the features in data because of accumulation of noise in estimating population centroids in high dimensional feature space and showed that optimal risk using DLDA increases as correlation among features increases.

Another approach is to regularise $\mathbf{\Sigma}$. Use of regularized $\mathbf{\Sigma}$ has gained significant attention. Friedman (1989) considered this regularisation for classification in low dimension and refer to it as regularized discriminant analysis. Guo et al. (2007) presented some regularized versions of covariance matrix and a detailed discussion on regularized discriminant analysis with application to gene expression data. Baldovin et al. (1997) presented a comparative study among some classifiers: regularized discriminant analysis(RDA), linear and quadratic discriminant analysis, nearest mean classifier, nearest weighted mean classifier and partial least square classifiers using industrial pollution datasets. In another development, Fan et al. (2012) proposed regularised optimal affine discriminant. Calis and Erol (2012) modified discriminant analysis and proposed a per-field classification method based on Gaussian mixture discriminant analysis for classifying remotely sensed multispectral image data.

Execution of usual discriminant analysis for classifying gene expression data is not possible because gene expression data is characterised by huge number of genes (as features) with very small gene profiles (as sample points). One may reduce the dimension of the data and carry out the usual discriminant analysis. Data reduction techniques available in literature include sparse principal component analysis (Zou et al., 2002) and multivariate adaptive stochastic search method (Tian et al., 2010). However, data reduction methods will lead to loss of important information. Reduced data may also suffer from poor interpretability.

Another intuitive feature of gene expression data is the presence of redundant genes. These are noisy genes and are not contributing to classification in terms of accuracy of the method employed. In dealing with this problem, Tibshirani et al. (2002) suggested shrinking each class mean vector towards overall mean vector and proposed a classification method based on distance to each shrunken mean. In this way, only contributing genes are extracted and used for classification. Guo et al. (2007) proposed shrunken centroid regularized discriminant analysis (SCRDA), a gene selection technique and classification method which combine shrunken mean vector and regularized covariance matrix. However, the performance of SCRDA depends on the choice of parameters employed in its execution.

In this paper, SCRDA with parameters (α, Δ) is employed to extract genes that best contribute to classification. That is, SCRDA is employed in this study as tool for selecting informative gene subset and not as a classifier. In extracting the best subset of genes that contribute to classification by minimising the cross validation error of training samples, the number of genes in the subset may be greater than sample size. This implies that the number of contributing genes $p_* > n$, the sample size. In this case, we suggest performing some versions of regularized linear discriminant analysis on the reduced data. We also apply the notion of regularizing covariance matrices in non linear classification case. In this case, regularized covariance matrices of competing classes is suggested to replace $\hat{\mathbf{\Sigma}}_j$ in quadratic discriminant function.

2. Methodology

Suppose $\mathbf{X}$ is a random vector in $\mathbb{R}^p$. Classification rule based on linear discriminant analysis can be expressed, for two competing classes π_1 and π_2, as

$$\text{assign } \mathbf{x} \text{ to } \pi_1 \text{ if } (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^\top \mathbf{\Sigma}^{-1} (\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2 - 2\mathbf{x}) > \log\left(\frac{p_2}{p_1}\right), \tag{1}$$

where p_1 and p_2 are prior probabilities, $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ are mean vectors of π_1 and π_2 respectively, and $\mathbf{\Sigma}$ is pooled covariance matrix. The RHS of (1) is zero if $p_1 = p_2$. The empirical classification rule uses estimates of $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$ and $\boldsymbol{\Sigma}$. The classification rule in (1) can be extended for $J > 2$ classes. Similarly, the quadratic discriminant rule for discriminating between observations in $\pi_1, \pi_2, \ldots, \pi_J$ can be expressed as

$$\text{assign } \mathbf{x} \text{ to } \pi_j \text{ if } (\mathbf{x} - \boldsymbol{\mu}_j)^\top \boldsymbol{\Sigma}_j^{-1} (\mathbf{x} - \boldsymbol{\mu}_j) + \log_e |\boldsymbol{\Sigma}_j| - 2 \log p_j \text{ is minimum,} \tag{2}$$

where $\boldsymbol{\Sigma}_j$, $\boldsymbol{\mu}_j$ and p_j are covariance matrices, mean vector and prior probability of π_j respectively. However, the estimate of $\boldsymbol{\Sigma}_j$, $j = 1, 2, \ldots, J$, denoted by $\hat{\boldsymbol{\Sigma}}_j$, becomes ill-conditioned or singular when the dimension is greater than sample size $(p > n)$.

A number of regularized versions of $\hat{\boldsymbol{\Sigma}}_j$ have been suggested, starting from the study of Friedman (1989) who suggested the use of $\tilde{\boldsymbol{\Sigma}}_j^{(0)}$ in place of the usual $\hat{\boldsymbol{\Sigma}}_j$, where

$$\tilde{\boldsymbol{\Sigma}}_j^{(0)} = (1 - \alpha)\hat{\boldsymbol{\Sigma}}_j + \frac{\alpha}{p} trace \left(\hat{\boldsymbol{\Sigma}}_j \right) \mathbf{I},$$

$\alpha \in (0, 1)$, $\hat{\boldsymbol{\Sigma}}_j$ is the estimate of covariance matrix of jth class in low dimension and $\mathbf{I}$ is a $p \times p$ identity matrix. However, this can be extended to very high dimension $(p >> n)$.

In another development, Guo et al. (2007) suggested the four regularized versions of $\hat{\boldsymbol{\Sigma}}_j$. These are

$$\begin{aligned}
\tilde{\boldsymbol{\Sigma}}_j^{(1)} &= \alpha \hat{\boldsymbol{\Sigma}}_j + (1 - \alpha)\mathbf{I} \text{ for } \alpha \in (0, 1); \\
\tilde{\boldsymbol{\Sigma}}_j^{(2)} &= \lambda \hat{\boldsymbol{\Sigma}}_j + \mathbf{I} \text{ for } \lambda > 0; \\
\tilde{\boldsymbol{\Sigma}}_j^{(3)} &= \hat{\boldsymbol{\Sigma}}_j + \lambda \mathbf{I} \text{ for} \lambda > 0; \\
\tilde{\boldsymbol{\Sigma}}_j^{(4)} &= \hat{\mathbf{D}}_j^{1/2} \tilde{\mathbf{R}}_j \hat{\mathbf{D}}_j^{1/2},
\end{aligned}$$

where $\tilde{\mathbf{R}}_j = \alpha \hat{\mathbf{R}}_j + (1 - \alpha)\mathbf{I}$, $\hat{\mathbf{R}}_j = \hat{\mathbf{D}}_j^{-1/2} \hat{\boldsymbol{\Sigma}}_j \hat{\mathbf{D}}_j^{-1/2}$ and $\hat{\mathbf{D}}_j = diag \left(\hat{\boldsymbol{\Sigma}}_j \right)$. It is observed that $\tilde{\boldsymbol{\Sigma}}_j^{(1)}$ and $\tilde{\boldsymbol{\Sigma}}_j^{(2)}$ tend to look like $\mathbf{I}$ as α and λ are very close to 0. It is also observed that $\tilde{\boldsymbol{\Sigma}}^{(2)} = \tilde{\boldsymbol{\Sigma}}_j^{(3)}$ when $\lambda = 1$.

Wu et al. (1996) considered another regularized covariance matrix $(\tilde{\boldsymbol{\Sigma}}_j^{(5)})$ for discriminant analysis, where

$$\tilde{\boldsymbol{\Sigma}}_j^{(5)} = (1 - \lambda)\hat{\boldsymbol{\Sigma}}_j + \lambda \hat{\boldsymbol{\Sigma}}_{pooled}$$

and $\hat{\boldsymbol{\Sigma}}_{pooled}$ is the covariance matrix of pooled sample. Baldovin et al. (1997) proposed

$$\tilde{\boldsymbol{\Sigma}}_j^{(6)} = \frac{(1 - \lambda)\hat{\boldsymbol{\Sigma}}_j + \lambda \hat{\boldsymbol{\Sigma}}_{pooled}}{(1 - \lambda)n_j + \lambda n}$$

where n_j is the sample size of jth group and $n = \sum_j n_j$. It is observed that the performance of $\tilde{\boldsymbol{\Sigma}}_j^{(6)}$ in discriminant analysis is equivalent to that of $\tilde{\boldsymbol{\Sigma}}_j^{(5)}$ when $n_j / \sum_{j=1}^{J} n_j \to 1/J$.

We refer to the classification rule in (1) using regularized pooled covariance matrix in place of usual covariance matrix as regularized linear discriminant analysis(denoted by

RLDA). Similarly, we refer to the classification rule in (2) using regularized covariance matrices of competing classes in place of usual covariance matrices as regularized quadratic discriminant analysis(denoted by RQDA). Wu et al. (1996) illustrated the use of LDA, QDA and RLDA based on $\tilde{\mathbf{\Sigma}}_j^{(0)}$ for spectrum NIR data. Due to ill-condition of the datasets, feature selection was employed. In their deduction, it was shown that LDA and QDA perform well and should be recommended for practical use. As noted in Wu et al. (1996), RLDA based on $\tilde{\mathbf{\Sigma}}_j^{(0)}$ performs well for spectrum NIR data but the optimisation is time consuming.

The use of $\tilde{\mathbf{\Sigma}}_j^{(1)}$ and $\tilde{\mathbf{\Sigma}}_j^{(4)}$ for some values of α was explored in Guo et al. (2007) for SCRDA. SCRDA combines shrunken centroid approach to remove noisy features from the data and perform classification on the reduced data. However, performance of $\tilde{\mathbf{\Sigma}}_j^{(k)}$, $k = 0, 1, 2, 3, 4$ in regularized linear and quadratic discriminant analysis needs to be explored.

SCRDA is similar to nearest centroid classifier (NCC) of Hastie et al. (2001), except for use of shrunken centroid in place of the usual mean vector and individual test vector by its transformed version. In SCRDA, regularised pooled covariance matrix $\tilde{\mathbf{\Sigma}}$ is employed in shrinking class mean vector $\bar{\mathbf{x}}$ towards overall mean vector. That is,

$$\bar{\mathbf{x}}^{'} = sign(\bar{\mathbf{x}}^{*})(|\bar{\mathbf{x}}^{*}| - \Delta)_{+},$$

where

$$\bar{\mathbf{x}}^{*} = \tilde{\mathbf{\Sigma}}^{-1}\bar{\mathbf{x}}.$$

Tibshirani et al. (2002) proposed nearest shrunken centroid (NSC) classifier. In NSC, $\bar{\mathbf{x}}^{'} = sign(\bar{\mathbf{x}})(|\bar{\mathbf{x}}| - \Delta)_{+}$. The corresponding classification rules for SCRDA and NSC are to assign $\mathbf{x}$ to class j if

$$j = \arg\min_{k}(\mathbf{x}^{*} - \bar{\mathbf{x}}_k^{'})^{\top}\hat{D}^{-1}(\mathbf{x}^{*} - \bar{\mathbf{x}}_k^{'}) - \log p_k$$

and

$$j = \arg\min_{k}(\mathbf{x} - \bar{\mathbf{x}}_k^{'})^{\top}\hat{D}^{-1}(\mathbf{x} - \bar{\mathbf{x}}_k^{'}) - \log p_k$$

respectively, where $\mathbf{x}^{*} = \tilde{\mathbf{\Sigma}}^{-1}\mathbf{x}$,$\hat{D} = \{s_1^2, s_2^2, \ldots, s_p^2\}$, $\bar{\mathbf{x}}_k^{'}$ is the shrunken centroid of class k, s_p^2 is the pooled variance of gene p.
Dudoit et al. (2002) suggested diagonal linear discriminant analysis (DLDA) which assumes no correlation among features. DLDA involves replacing pooled covariance matrix $\hat{\mathbf{\Sigma}}$ by its diagonal matrix $\hat{D} = \text{diag}\left(\hat{\mathbf{\Sigma}}\right)$. Pang et al. (2009) suggested a modified diagonal linear and quadratic discriminant analysis. Ackermann and Strimmer (2009) argued that correlation among features, especially micro-array data and clinical outcomes, is an essential characteristic and is not always negligible. To illustrate this, we applied SCRDA with parameters $\alpha = 0.1$ and $\Delta = 2.5$ to Lymphoma data. 20 genes were returned. We computed correlation matrix for the 20 returned genes. Also, we applied SCRDA with parameters $\alpha = 0.2$ and $\Delta = 0.3$ to colon cancer data. 28 genes were returned. We also computed the correlation matrix for the 28 genes.

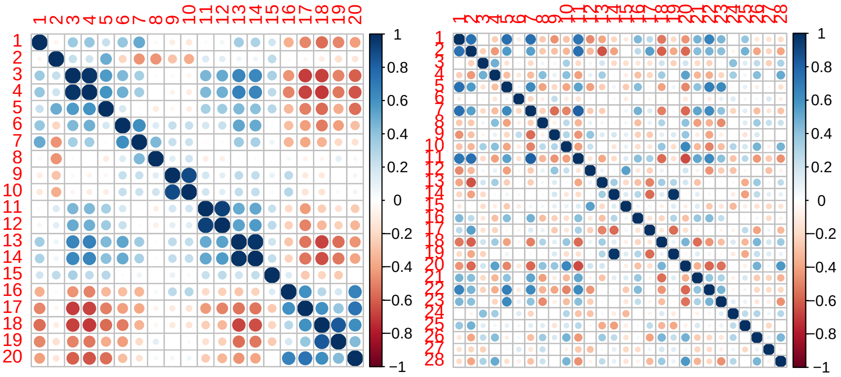

Fig. 1. Correlation plots of genes in lymphoma (left) and colon cancer (right) datasets.

Figure 1 presents plot of correlation matrices for lymphoma and colon cancer genes. It can be seen from the figure that correlation among genes in micro-array experiment is not always negligible and any classification rule based on $\hat{D}$, and not on covariance matrix $\hat{\mathbf{\Sigma}}$, may have high misclassification rate.

We denote the classification rule in (1) based on $\tilde{\mathbf{\Sigma}}^{(k)}$ by RLDA-k and classification rule in (2) based on $\tilde{\mathbf{\Sigma}}_j^{(k)}$ by RQDA-k, where $k = 0, 1, 2, 3, 4$, $j = 1, 2, \ldots, J$. It is observed that RLDA-2 and RLDA-3 are equivalent when $\lambda = 1$ because $\tilde{\mathbf{\Sigma}}_j^{(2)} = \tilde{\mathbf{\Sigma}}_j^{(3)}$ when $\lambda = 1$.

3. Data analysis and result

3.1. Application to gene expression data

Classification of gene expression data was considered in Yeang et al. (2001), Guo et al. (2007), among others. Here we apply RLDA and RQDAs to gene expression data, which are colon cancer data, leukaemia data and lymphoma data.

Regularised linear and quadratic discriminant analyses as well as diagonal linear and quadratic discriminant analyses can not be applied directly to gene expression data be-

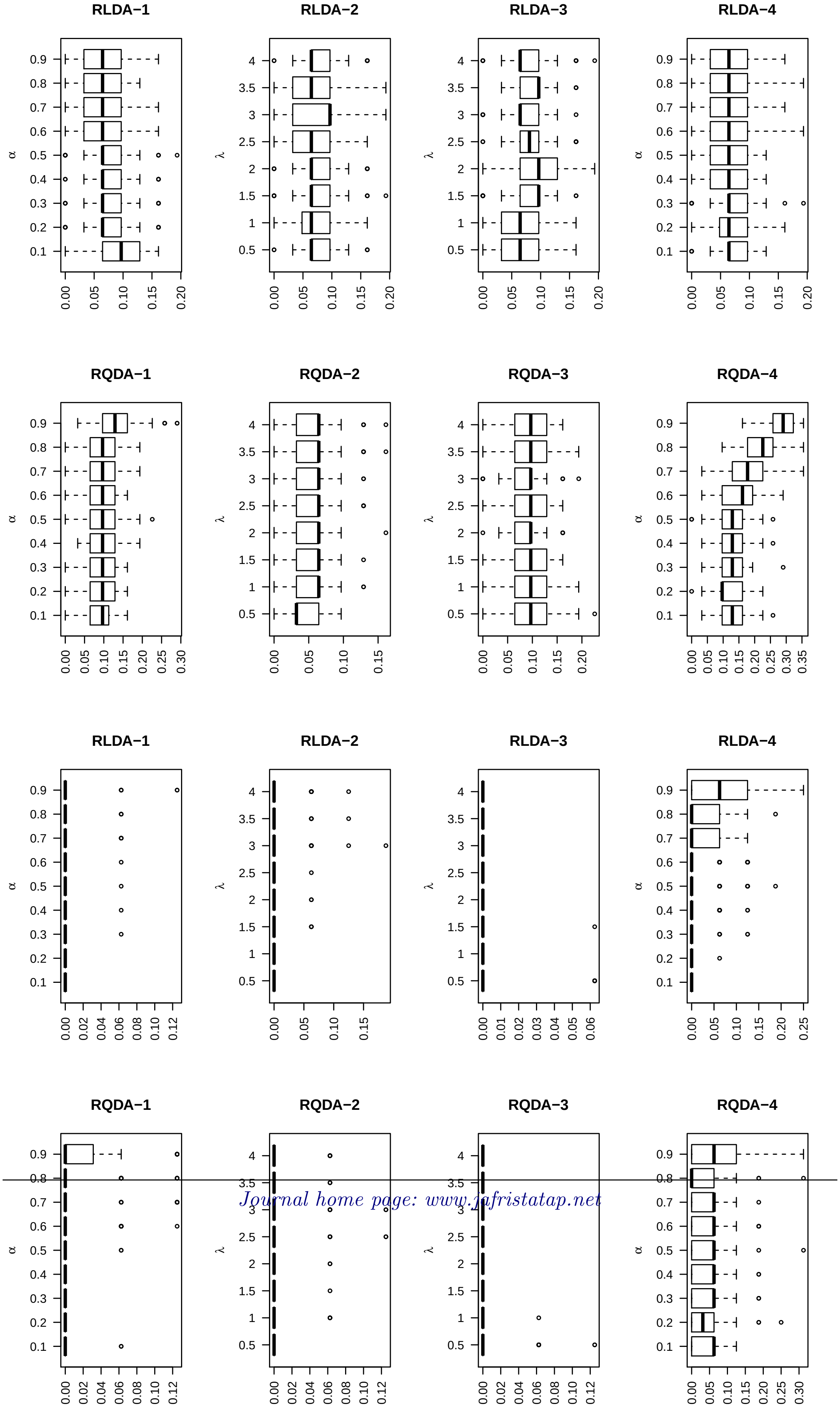

Fig. 2. Boxplots of misclassification rates of regularized linear and quadratic discriminant

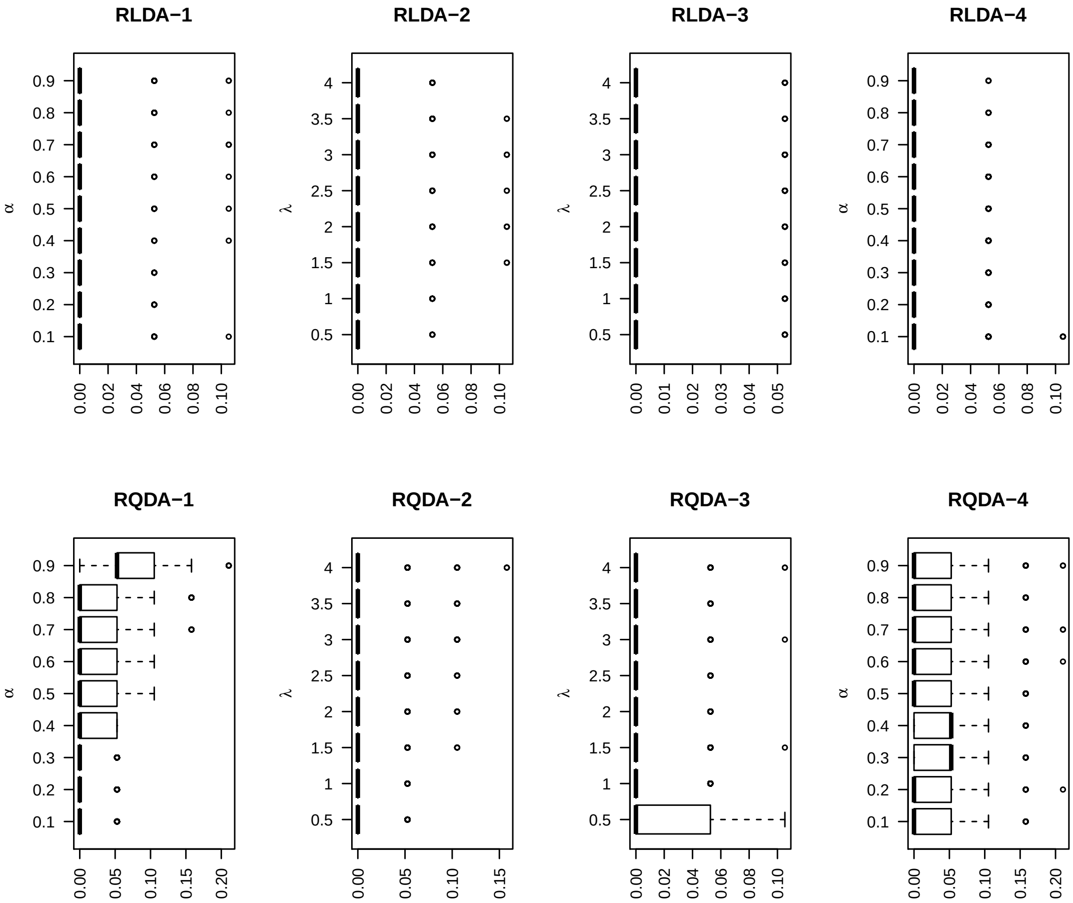

Fig. 3. Boxplots of misclassification rates of regularized linear and quadratic discriminant analysis obtained from Lymphoma data for various values of α and λ.

cause some genes are zero-valued. Possibility of using SCRDA with parameters α and Δ for extracting subsets of genes that contribute best to classification was raised in Guo et al. (2007). It is important to mention that Guo et al. (2007) applied SCRDA for classification and chose the parameters of the model by minimising cross validation error of the training samples. Here, SCRDA is only employed to select gene expression features that contribute most to classification, RLDAs and RQDAs are then performed on the reduced data.

Leukemia data consists of two classes of sizes 27 and 11 with 3051 genes. SCRDA was employed to select informative gene subset. The parameters α and Δ are taken to be 0.1 and 0.9 respectively. SCRDA returns 28 genes on which RLDAs and RQDAs are applied. Random training samples of sizes 15 and 7 and random test samples of sizes 12 and 4 are

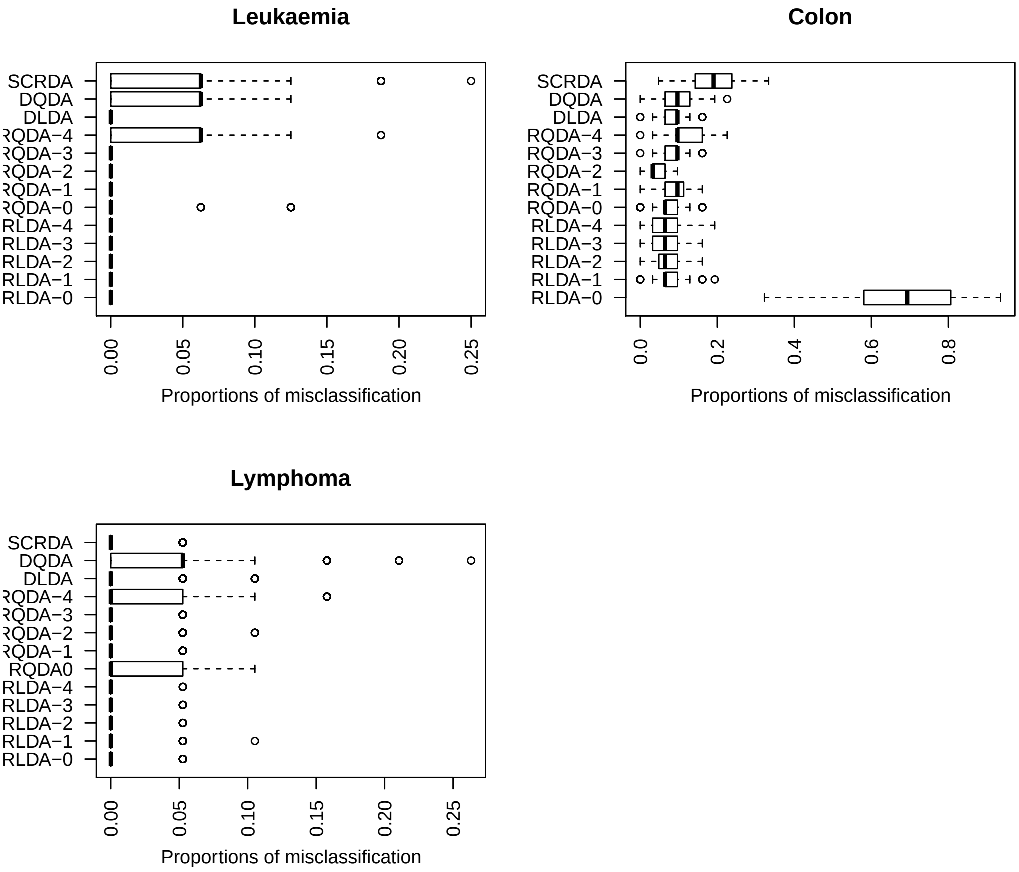

Fig. 4. Box plots for the misclassification rates of classifiers obtained from leukaemia data, colon cancer data and lymphoma data.

selected from the two classes respectively. The experiment is repeated 999 times and the misclassification error rates are computed and presented using boxplot. Figure 2 (row 1 and 2) presents boxplots of misclassification error rates of some regularized linear discriminant analyses (RLDA-1, RLDA-2, RLDA-3 and RLDA-4) and regularized quadratic discriminant analyses (RQDA-1, RQDA-2, RQDA-3 and RQDA-4) for some values of regularization parameters (α and λ) for leukaemia. It is observed that RLDA-0, RLDA-1, RLDA-2 and RLDA-3 achieve 100% proportion of correct classification when $\alpha \in [0.1, 0.2]$, $\alpha \in [0.1, 0.2]$, $\lambda \in [0.5, 1.0]$ and $\lambda \in [1, 4]$ respectively. RLDA-4 achieves 100% proportion of correct classification when $\alpha = 0.1$. This indicates that leukaemia data is linearly separable. Similarly, RQDA-1, RQDA-2 and RQDA-3 perform well for some values of α and λ. Misclassification error rates of RQDA-4 is a bit high.

Colon cancer microarray data consists of two classes of sizes 22 and 40 with 2000 genes. SCRDA(α, Δ) was employed to select informative gene subset. The parameters α and Δ are taken to be 0.2 and 0.3 respectively. 28 genes were returned. Random training samples of sizes 11 and 20 and random test samples of sizes 11 and 20 are selected from classes 1 and 2 respectively. The experiment is repeated 999 times and the misclassification error rates are computed and presented using boxplot. Figure 2 (row 3 and 4) presents boxplots of misclassification error rates of some regularized linear discriminant analyses (RLDA-1, RLDA-2, RLDA-3 and RLDA-4) and regularized quadratic discriminant analyses (RQDA-1, RQDA-2, RQDA-3 and RQDA-4) for various values of regularization parameters for colon. It is observed that error rates of RLDA-1 are similar for all values of α except $\alpha = 0.1$ where the mean error rates is a bit high. It is evident that RQDA-1, RQDA-2 and RQDA-4 perform much better for lower values of α and λ.

Lymphoma data consists of three classes of sizes 42, 9 and 11 with 4026 genes. SCRDA(0.1, 2.5) was employed to select 20 informative genes. Random training samples of sizes 30, 6 and 7; and random test samples of sizes 12, 3 and 4 are selected from classes 1, 2 and 3 respectively. This experiment is also repeated 999 times and the misclassification error rates are computed and presented using boxplot. Figure 3 presents boxplots of proportions of misclassification of RLDAs and RQDAs for leukaemia for various values of regularization parameters.

Figure 4 presents comparison of RLDA and RQDA with DLDA, DQDA and SCRDA for Leukaemia data, colon cancer data and lymphoma data. For Leukemia data, we set α to be 0.7, 0.2, 0.3, 0.4, 0.1 and 0.1 for RQDA-0, RQDA-1, RQDA-4, RLDA-0, RLDA-1 and RLDA-4 respectively. λ is chosen to be 0.5, 1.5, 0.5 and 1.0 for RQDA-2, RQDA-3, RLDA-2 and RLDA-3 respectively. DLDA also achieves 100% proportion of correct classification. SCRDA achieves mean proportion of misclassification of 0.0075. All the competing classification methods perform equivalently as shown in the figure. For colon data, we set α to be 0.2, 0.1, 0.2, 0.8, 0.5 and 0.8 for RQDA-0, RQDA-1, RQDA-4, RLDA-0, RLDA-1 and RLDA-4 respectively. λ is chosen to be 1.0, 2.0, 1.0 and 1.0 for RQDA-2, RQDA-3, RLDA-2 and RLDA-3 respectively. RLDAs and RQDAs perform competitively with DLDA and SCRDA except RLDA-0. RLDA-0 performs worst. For lymphoma data, α to be 0.6, 0.1, 0.8, 0.4, 0.6 and 0.9 for RQDA-0, RQDA-1, RQDA-4, RLDA-0, RLDA-1 and RLDA-4 respectively. λ is chosen to be 1.5, 2.5, 0.5 and 3.5 for RQDA-2, RQDA-3, RLDA-2 and RLDA-3 respectively. All the classifiers perform well except RQDA-0.

Table 1 presents the average and standard error of misclassification error rates of RLDAs, RQDAs DLDA, DQDA and SCRDA for 1000 repetitions. It is observed that use of $\tilde{\mathbf{\Sigma}}^{(0)}$, $\tilde{\mathbf{\Sigma}}^{(1)}$, $\tilde{\mathbf{\Sigma}}^{(2)}$ and $\tilde{\mathbf{\Sigma}}^{(3)}$ in classification rules defined in (1) and (2) for some values of tuning parameters yields perfect classification of Leukemia data. Similarly, DLDA and DQDA-4 achieve perfect classification while their quadratic forms do not. For colon data, all the classifiers perform equivalently except RQDA-4 and SCRDA. For lymphoma data, we observe that regularized linear discriminant analyses (RLDA-0, RLDA-1, RLDA-2, RLDA-3, RLDA-4) outperform their quadratic counterpart. RQDA-0 performs worst.

Table 1. Average error rates, with standard error of the proportions of misclassification in parenthesis, of classifiers for Leukemia data, Colon cancer data and Lymphoma data.

Classifiers	Leukemia	Colon	Lymphoma
RQDA-0	0.014(0.037)	0.075(0.037)	0.252(0.036)
RQDA-1	0.000(0.000)	0.088(0.034)	0.003(0.012)
RQDA-2	0.000(0.000)	0.044(0.030)	0.010(0.021)
RQDA-3	0.000(0.000)	0.085(0.036)	0.007(0.018)
RQDA-4	0.043(0.047)	0.116(0.048)	0.019(0.025)
RLDA-0	0.000(0.000)	0.682(0.144)	0.000(0.006)
RLDA-1	0.000(0.000)	0.081(0.039)	0.001(0.006)
RLDA-2	0.000(0.000)	0.075(0.041)	0.000(0.005)
RLDA-3	0.000(0.000)	0.070(0.038)	0.000(0.004)
RLDA-4	0.000(0.000)	0.064(0.045)	0.000(0.003)
DLDA	0.000(0.000)	0.082(0.035)	0.014(0.026)
DQDA	0.043(0.040)	0.098(0.049)	0.039(0.046)
SCRDA	0.059(0.058)	0.179(0.062)	0.009(0.020)

3.2. Simulation study

Simulation studies are presented to illustrate the performance of various regularized versions of discriminant analysis in high dimension. This will in practice, provide an answer to the question that which regularized covariance matrix be employed for discriminant analysis in high dimension. Suppose C_1 and C_2 are two competing classes of observations in high dimension. Each class has training sample and test sample of size 50. Each experiment consists of assigning an observation in the test set to each of the two competing classes based on measurements on p features of each class.

[Simulation 1] Suppose ith observation is in kth class, then $\mathbf{Y}_i \sim N(\boldsymbol{\mu}_k, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu}_k = (\mu_{k1}, \mu_{k2}, \ldots, \mu_{kp})^\top$ with $\mu_{1j} = 0$ for $1 \leq j \leq p$, $\mu_{2j} = 0.7$ if $1 \leq j \leq 100$ and $\mu_{2j} = 0$ otherwise and $k = 1, 2$. The covariance structure $\boldsymbol{\Sigma}$ consists of 5×5 blocks, each block of dimension 100×100 with (j, j') element $0.6^{|j-j'|}$.

[Simulation 2] Suppose each experiment consists of measurements on independent features such that $X_{kj} \sim N(\mu_{kj}, 1)$ where $\mu_{1j} = 0$ for $1 \leq j \leq p$, $\mu_{2j} = 0.7$ if $1 \leq j \leq 100$ and $\mu_{2j} = 0$ otherwise.

Simulation 1 consists of dependent features with block. Simulation 2 consists of independent and identically normally distributed features. Figure 5 present the comparison of RLDA-1, RLDA-2, RLDA-3 and RLDA-4 for various values of α and λ in simulation 1. In Figure 5, proportions of misclassification in RLDA-2 are similar for some values of λ. It is observed that higher values of λ yield lower misclassification error rates. Misclassification error rates of RLDA-1 and RLDA-4 decrease as values of α decrease.

For simulation 2, mean error rates of RLDA-1 and RLDA-4 increase monotonically with increase in $\alpha \in (0, 1)$ while mean error rates of RLDA-3 decreases monotonically with increase in the value of λ. Performance of RLDA-1, RLDA-2, RLDA-3 and RLDA-4 are similar in terms of averages of proportions of misclassification. That is, mean error rate of each of RLDA-1, RLDA-2, RLDA-3 and RLDA-4 approaches zero for all values of α and λ.

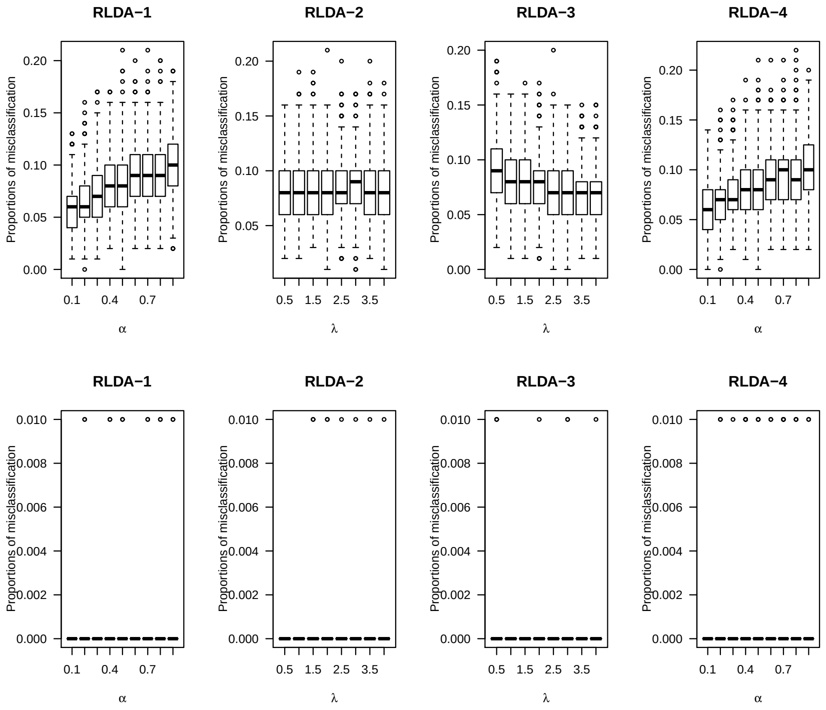

Fig. 5. Box plots for the misclassification rates of various versions of regularised linear discriminant analysis obtained from simulated data example 1(first row) and 2(second row) for various values of α and λ.

Comparing various versions of RLDA with NSC, DLDA and SCRDA, we observe that the performances of RLDA-1, RLDA-2, RLDA-3 and RLDA-4 are competitive with NSC and DLDA except for RLDA-0 in simulation 1 as shown in Figure 6. Except DLDA and RDLA-0 in simulation 2, all the classifiers achieve perfect classification. However, all the classifiers perform well.

4. Conclusion

In low dimension, testing significant difference among mean vectors of competing classes of observations if the classes have a common covariance matrix before carrying out classification exercise can provide a clue whether linear discriminant rule will make sense or not. In high dimension, where dimension of each observation is bigger than sample size, there is no simple way of conducting such test due to singularity of common covariance matrix. We have considered notions of various regularized covariance matrices in regularized discriminant analysis. The performance of various versions of regularized linear and quadratic discriminant analyses is illustrated using simulated data as

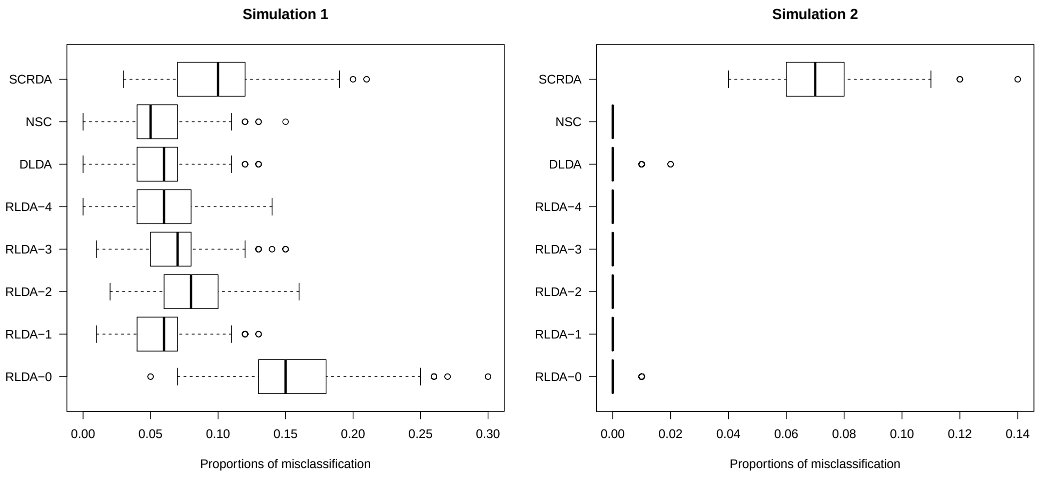

Fig. 6. Box plots for the misclassification rates of classifiers obtained from simulated data examples.

well as real data. The real data are gene expression data that arose from micro-array studies.

From real data and simulated data examples, regularized linear and quadratic discriminant analyses compete favourably with SCRDA and diagonal discriminant analysis. To choose values of turning parameter(λ or γ) for $\tilde{\mathbf{\Sigma}}^{(k)}$, $k = 0, 1, \ldots, 4$, that maximise the proportion of correct classification, we suggest use of leave-one-out cross validation of error rates, as also suggested by Baldovin et al. (1997).

Acknowledgments The author wishes to address his special thanks to the presenter of the paper, for the valuable suggestions, comments and recommendations on the manuscript, which have been used to improve on the version.

References

Ackermann M. and K. Strimmer (2009). A general modular framework for gene set enrichment analysis. *BMC Bioinformatics* 10:47

Baldovin A, Wen W, Massart DL, Turello A (1997). Regularised discriminant analysis - Modelling for the binary discrimination between pollution types. *Chemometrics and Intelligence Systems*, 38(1), 25–37.

Bickel, P. and E. Levina (2004). Some theory for Fisher's linear discriminant function, naive Bayes, and some alternatives when there are many more variables than observations. *Bernoulli*, 10(6), 989–1010.

Calis N. and H. Erol (2012). A new per-field classification method using mixture discriminant analysis, *J. Appl. Stat.*, 39, 2129–2140.

Dudoit S., J. Fridlyand and T.P. Speed (2002). Comparison of discrimination methods for the classification of tumors using gene expression data. *J Am Stat Assoc*, 97, 77–87.

Fan J, Y. Fan, Y. Wu (2011). High-dimensional classification. *In Cai TT, Shen X (eds.): High-dimensional Data Analysis, World Scientific, New Jersey.* 3–37.

Fan J. and Y. Fan (2008). High-dimensional classification using features annealed independence rules. *Ann. Statist.*, 36, 2605–2637

Fan J., Y. Feng and X. Tong (2012), A road to classification in high dimensional space: the regularized optimal affine discriminant. *Journal of the Royal Statistical Society: Series B*, 74(4), 745–771.

Fisher R.A. (1936). The use of multiple measurements in taxonomic problems. *Ann. Eugenics*, 7, 179–188.

Friedman J. (1989). Regularized discriminant analysis. *J Am Stat Assoc*, 84, 165–175.

Gorecki T. (2015). Sequential combining in discriminant analysis. *J. Appl. Stat.*, 42, pp. 398–408.

Guo Y., T. Hastie and R. Tibshirani (2007). Regularized linear discriminant analysis and its application in micro-arrays, *Biostatistics*, 8, 86–100.

Hastie T., A. Buja and R. (1995) Tibshirani, Penalized Discriminant Analysis. *The Annals of Statistics*, 23, 73–102.

Hastie T., R. Tibshirani and J.H. Friedman (2001): The Elements of Statistical Learning: Data Mining, Inference and Prediction, 2nd ed., Springer, New York.

Johnson, R. A. and D. W. Wichern (2007): Applied multivariate statistical analysis. Sixth edition, Pearson Prentice Hall inc. New Jersey.

Kiruthika C. and Chandrasekaran R.(2012). Classification of textile fabrics using statistical multivariate techniques. *J. Appl. Stat.*, Vol. 39, No. 5, 1129–1138.

LeCun Y., L. Bottou, Y. Bengio, and P. Haffner (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86, 2278–2324.

Mahalanobis P.C. (1936). On the generalized distance in statistics. *Proceedings of the National Institute of Science of India*, 2, 49–55.

Pang H., T. Tong, and H. Zhao. (2009). Shrinkage-based diagonal discriminant analysis and its applications in high-dimensional data. *Biometrics.* 65(4), 1021–1029.

Sohn I., J. Shim, C. Hwang, S. Kimd, and J.W. Lee (2014). Transcription factor-binding site identification and gene classification via fusion of the supervised-weighted discrete kernel clustering and support vector machine. *J. Appl. Stat.*, 41, 573–581.

Tian T.S., G.M. James and R.R. Wilcox (2010). A multivariate adaptive stochastic search method for dimensionality reduction in classification. *Annals of Applied Statistics*, 4, 340–365.

Tibshirani R., T. Hastie, B. Narasimhan and G. Chu (2002). Diagnosis of multiple cancer type by shrunken centroid, *Proceedings of the National Academy of Sciences*, 99(10), pp. 6567–6572.

Witten, D. M. and R. Tibshirani (2011). Penalized classification using Fisher's linear discriminant, *Journal of the Royal Statistical Society: Series B*, 73(5), 753–772.

Wu W., Y. Mallet, B. Walczak, W. Penninckx, D.L. Massart, S. Heuerding and F. Erni (1996). Comparison of regularized discriminant analysis, linear discriminant analysis and quadratic discriminant analysis, applied to NIR data. *Analytica Chimica Acta*, 329, 257–265

Yeang C.H., S. Ramaswamy, P. Tamayo, S. Mukherjee, R.M. Rifkin, M. Angelo, M. Reich, E. Lander, J. Mesirov and T. Golub (2001). Molecular classification of multiple turmor types. *Bioinformatics*, 17, S316–S322.

Zou H., T. Hastie and R. Tibshirani (2006): Sparse Principal Component Analysis. *Journal Of Computational And Graphical Statistics*, 15(2), 265–286.