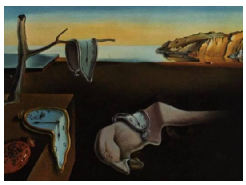

Journal of Statistical Software
August 2015, Volume 66, Issue 10. http://www.jstatsoft.org/

dawai: An R Package for Discriminant Analysis with Additional Information

David Conde
Universidad de Valladolid

Miguel A. Fernández
Universidad de Valladolid

Bonifacio Salvador
Universidad de Valladolid

Cristina Rueda
Universidad de Valladolid

Abstract

The incorporation of additional information into discriminant rules is receiving increasing attention as the rules including this information perform better than the usual rules. In this paper we introduce an R package called **dawai**, which provides the functions that allow to define the rules that take into account this additional information expressed in terms of restrictions on the means, to classify the samples and to evaluate the accuracy of the results. Moreover, in this paper we extend the results and definitions given in previous papers (Fernández, Rueda, and Salvador 2006, Conde, Fernández, Rueda, and Salvador 2012, Conde, Salvador, Rueda, and Fernández 2013) to the case of unequal covariances among the populations, and consequently define the corresponding restricted quadratic discriminant rules. We also define estimators of the accuracy of the rules for the general more than two populations case. The wide range of applications of these procedures is illustrated with two data sets from two different fields, i.e., biology and pattern recognition.

Keywords: classification rules, order-restricted inference, true error rate, R package **dawai**, R.

1. Introduction

The incorporation of additional information, often available in applications, into multivariate statistical procedures through order restrictions is receiving increasing attention during the last years as it allows to improve the performance of the procedures. Good examples of this trend are the papers by Rueda, Fernández, and Peddada (2009), Fernández, Rueda, and Peddada (2012) and Barragán, Fernández, Rueda, and Peddada (2013), where this information

is used to improve statistical procedures for circular data applied to cell biology, El Barmi, Johnson, and Mukerjee (2010), where the information is used for estimating cumulative incidence functions in survival analysis, Ghosh, Banerjee, and Biswas (2008), where it is used to make inferences on tumor size distributions, or Davidov and Peddada (2013), where a test for multivariate stochastic order is applied to dose-response studies.

In this work, we deal with the incorporation of additional information into discriminant rules. Discriminant analysis is a well-known technique, first established by Fisher (1936), used in many scientific fields to define rules that allow to classify samples into a small number of populations based on a sample of observations where the population membership is known, usually called training set. To the best of our knowledge, the first paper considering additional information under the usual equal covariances assumption, which leads to linear discriminant rules, is Long and Gupta (1998). However, that paper provided limited results for the case of two populations with simple order restrictions and identity covariance matrices only. In a series of papers, the rules appearing in that initial paper have been improved, first to deal with more general types of information expressed in terms of cones of restrictions and general covariance matrices in Fernández *et al.* (2006) and later to the case of more than two populations in Conde *et al.* (2012). The robustness of the rules has also been studied in Salvador, Fernández, Martín, and Rueda (2008) and good estimators of the performance of the rules (which is an essential issue in discriminant analysis) have been provided in Conde *et al.* (2013). From now on, we will refer to these rules as restricted rules as the additional information is incorporated through restrictions on the population means.

The purpose of the present paper is twofold. The first is to introduce the **dawai** package, programmed in R (R Core Team 2015), which is available from the Comprehensive R Archive Network (CRAN) at `http://CRAN.R-project.org/package=dawai`. This package provides all the functions needed to take advantage of the rules that incorporate additional information. The functions in the package allow to define the restricted rules, to classify the samples and to evaluate the accuracy of the results. The second contribution of this paper is the extension of the ideas given in previous papers from the case of equal covariances in the different populations to the case of unequal covariances among the populations and consequently the definition of the corresponding restricted quadratic discriminant rules, and also the definition of estimators of the accuracy of the rules for the general case where more than two populations appear in the problem.

In Section 2 we describe the statistical problem and the methodology of Fernández *et al.* (2006), Conde *et al.* (2012) and Conde *et al.* (2013), which we extend to the above mentioned situations. In Section 3 we introduce the **dawai** package and explain some details about the main user-lever functions it includes. The wide range of applications of the **dawai** package is illustrated in Section 4 using two data sets coming from two different fields, i.e., biology and pattern recognition. Some concluding remarks are provided in Section 5.

2. Discriminant analysis with additional information

We consider a finite number $k \geq 2$ of distinct populations of items $\Pi_1, \ldots, \Pi_k$. Each item is assumed to belong to one and only one of the populations. Let Z be a categorical variable identifying the population and let $\boldsymbol{X} = (X_1, \ldots, X_p)^\top$ be the p-dimensional vector of predictors. Denote also as $P_{\boldsymbol{X}Z}$ the joint distribution of $(\boldsymbol{X}, Z)$, and as P_j the distribution of $\boldsymbol{X}$

in population Π_j with density function f_j, $j = 1, \ldots, k$. The classical discrimination problem deals with the classification of an observation $\boldsymbol{U} = (U_1, \ldots, U_p)^\top$, whose origin is unknown, into one of those populations. If we consider a 0–1 loss function and a priori probability π_j for the population Π_j, $j = 1, \ldots, k$, it is well known that the optimal classification rule, also called Bayes rule, is given by:

$$\text{Classify } \boldsymbol{U} \text{ in } \Pi_j \text{ iff } \pi_j f_j(\boldsymbol{U}) \geq \pi_l f_l(\boldsymbol{U}),\ l = 1, 2, \ldots, k.$$

In applications, the density functions $f_j, j = 1, \ldots, k$, are unknown although there is sample information available. This sample information is contained in the so-called training sample given by a set of items for which both the predictor values and the correct population they belong to are registered. We represent the training sample as $M_n = \{(\boldsymbol{Y}_i, Z_i), i = 1, \ldots, n\}$, where n is the item sample size, $\boldsymbol{Y}_i$ is the value that vector $\boldsymbol{X}$ takes at the ith item in the sample and Z_i is the population the ith item belongs to. Then, a classification rule is an application $R_n : \{\mathbb{R}^p \times \{1, \ldots, k\}\}^n \times \mathbb{R}^p \to \{1, \ldots, k\}$, that assigns a new observation $\boldsymbol{U} \in \mathbb{R}^p$ for which the population is unknown to one of the k populations, $R_n(M_n, \boldsymbol{U}) \in \{1, \ldots, k\}$.

From now on, we assume that $\pi_j = \frac{1}{k}$, $j = 1, \ldots, k$ (the case of unequal a priori probabilities is a trivial extension). If we further assume multivariate normality, i.e., $P_j \sim N_p(\boldsymbol{\mu}_j, \boldsymbol{\Sigma})$, $j = 1, \ldots, k$, where $\boldsymbol{\mu}_j = (\mu_{j1}, \ldots, \mu_{jp})^\top$ is the mean of vector $\boldsymbol{X}$ in population Π_j, $j = 1, \ldots, k$, and $\boldsymbol{\Sigma}$ is a common covariance matrix, then the optimal classification rule (the one with lowest expected loss) may be written as:

$$\text{Classify } \boldsymbol{U} \text{ in } \Pi_j \text{ iff } (\boldsymbol{U} - \boldsymbol{\mu}_j)^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{U} - \boldsymbol{\mu}_j) \leq (\boldsymbol{U} - \boldsymbol{\mu}_l)^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{U} - \boldsymbol{\mu}_l),\ l = 1, \ldots, k.$$

Unfortunately, this rule cannot be used in practice as the mean vectors $\boldsymbol{\mu}_j$, $j = 1, \ldots, k$, and the common covariance matrix $\boldsymbol{\Sigma}$ are unknown. However, as we have a training sample these parameters may be estimated using respectively the sample vectors means $\overline{\boldsymbol{Y}}_j$ and the pooled sample covariance matrix $\boldsymbol{S}$,

$$\overline{\boldsymbol{Y}}_j = (\overline{Y}_{j1}, \ldots, \overline{Y}_{jp})^\top = \frac{1}{n_j} \sum_{l=1}^{n} \boldsymbol{Y}_l I_{(Z_l = j)} \quad \text{and}$$

$$\boldsymbol{S} = \frac{1}{n-k} \sum_{j=1}^{k} \sum_{l=1}^{n} \left(\boldsymbol{Y}_l - \overline{\boldsymbol{Y}}_j\right) \left(\boldsymbol{Y}_l - \overline{\boldsymbol{Y}}_j\right)^\top I_{(Z_l = j)},$$

where $n_j = \sum_{l=1}^{n} I_{(Z_l = j)}$ is the sample size of population Π_j, $j = 1, \ldots, k$, and $n = \sum_{j=1}^{k} n_j$. As this estimated rule, obtained by plugging the estimators into the initial rule, is linear in the predictors, it is usually known as linear discriminant rule or Fisher's rule:

$$\text{Classify } \boldsymbol{U} \text{ in } \Pi_j \text{ iff}$$
$$(\boldsymbol{U} - \overline{\boldsymbol{Y}}_j)^\top \boldsymbol{S}^{-1} (\boldsymbol{U} - \overline{\boldsymbol{Y}}_j) \leq (\boldsymbol{U} - \overline{\boldsymbol{Y}}_l)^\top \boldsymbol{S}^{-1} (\boldsymbol{U} - \overline{\boldsymbol{Y}}_l), l = 1, \ldots, k. \quad (1)$$

If the covariance matrices are not assumed to be equal, i.e., $P_j \sim N_p(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)$, $j = 1, \ldots, k$, the optimal rule can be written as:

$$\text{Classify } \boldsymbol{U} \text{ in } \Pi_j \text{ iff } -\frac{1}{2} \log(|\boldsymbol{\Sigma}_j|) - \frac{1}{2} \left\{ (\boldsymbol{U} - \boldsymbol{\mu}_j)^\top \boldsymbol{\Sigma}_j^{-1} (\boldsymbol{U} - \boldsymbol{\mu}_j) \right\} \geq$$
$$-\frac{1}{2} \log(|\boldsymbol{\Sigma}_l|) - \frac{1}{2} \left\{ (\boldsymbol{U} - \boldsymbol{\mu}_l)^\top \boldsymbol{\Sigma}_l^{-1} (\boldsymbol{U} - \boldsymbol{\mu}_l) \right\},\ l = 1, \ldots, k. \quad (2)$$

Again, if we replace in this rule the unknown parameters $\boldsymbol{\mu}_j$ and $\boldsymbol{\Sigma}_j$ by their corresponding estimators $\overline{\boldsymbol{Y}}_j$ and $\boldsymbol{S}_j = \frac{1}{n_j-1}\sum_{l=1}^{n}\left(\boldsymbol{Y}_l - \overline{\boldsymbol{Y}}_j\right)\left(\boldsymbol{Y}_l - \overline{\boldsymbol{Y}}_j\right)^\top I_{(Z_l=j)}$, $j = 1, \ldots, k$, we obtain a rule that depends on the predictor in a quadratic way and it is therefore known as the quadratic discriminant rule:

$$\begin{aligned}\text{Classify } \boldsymbol{U} \text{ in } \Pi_j \text{ iff } & -\frac{1}{2}\log(|\boldsymbol{S}_j|) - \frac{1}{2}\left\{(\boldsymbol{U} - \overline{\boldsymbol{Y}}_j)^\top \boldsymbol{S}_j^{-1}(\boldsymbol{U} - \overline{\boldsymbol{Y}}_j),\right\} \geq \\ & -\frac{1}{2}\log(|\boldsymbol{S}_l|) - \frac{1}{2}\left\{(\boldsymbol{U} - \overline{\boldsymbol{Y}}_l)^\top \boldsymbol{S}_l^{-1}(\boldsymbol{U} - \overline{\boldsymbol{Y}}_l),\right\},\ l = 1, \ldots, k. \quad (3)\end{aligned}$$

2.1. Restricted discriminant rules

In the introduction we referred to applications where it is usual that some additional information is available. In many of these cases the information can be written as inequality restrictions among the population means. In the literature these restrictions are usually represented by a polyhedric cone (cf. Robertson, Wright, and Dykstra 1988 or Silvapulle and Sen 2005). In our case, our pk-dimensional population means will belong to a cone C in $\mathbb{R}^{pk}$,

$$(\boldsymbol{\mu}_1^\top, \ldots, \boldsymbol{\mu}_k^\top)^\top \in C = \left\{\boldsymbol{x} \in \mathbb{R}^{pk} : \boldsymbol{a}_j^\top \boldsymbol{x} \geq 0,\ j = 1, \ldots, q\right\}, \quad (4)$$

where the q vectors $\boldsymbol{a}_j \in \mathbb{R}^{pk}$, $j = 1, \ldots, q$, are determined by the restrictions imposed on the means.

Polyhedral cones are widely used in the restricted inference literature, because they cover the most interesting cases from a practical standpoint. Among these cones, those representing order relations among the means are especially interesting. For example, it is not unusual to know that the observations from one of the populations, for example, Π_1 (which may be the control population in a medical study), take, in mean, lower values than those coming from any of the other populations for a subset $L \subseteq \{1, \ldots, p\}$ of predictor variables. In the usual restricted statistical terminology, we can say that there is a "tree order" among the population means of the variables in L. In this case, we can write

$$(\boldsymbol{\mu}_1^\top, \ldots, \boldsymbol{\mu}_k^\top)^\top \in C_{TO} = \left\{\boldsymbol{x} \in \mathbb{R}^{pk} : x_l \leq x_{l+rp},\ r = 1, \ldots, k-1,\ l \in L\right\}. \quad (5)$$

Another usual situation appears when it is known that there is an increase in the means of a subset L of predictors (for example, due to increasing severity level in an illness study). This is known as a "simple order" among the population means of the variables in L, and may be represented in $\mathbb{R}^{pk}$ using the cone

$$(\boldsymbol{\mu}_1^\top, \ldots, \boldsymbol{\mu}_k^\top)^\top \in C_{SO} = \left\{\boldsymbol{x} \in \mathbb{R}^{pk} : x_l \leq x_{l+p} \leq \cdots \leq x_{l+(k-1)p},\ l \in L\right\}. \quad (6)$$

Restricted linear discriminant rules

As mentioned above, in this case we assume $\boldsymbol{\Sigma}_j = \boldsymbol{\Sigma}$, $j = 1, \ldots, k$. Fernández *et al.* (2006) deal with this situation when the number of populations $k = 2$. They propose a family of classification rules whose expected loss (total probability of misclassification) is lower than that of the linear discriminant rule (1). These rules are based on the use of additional

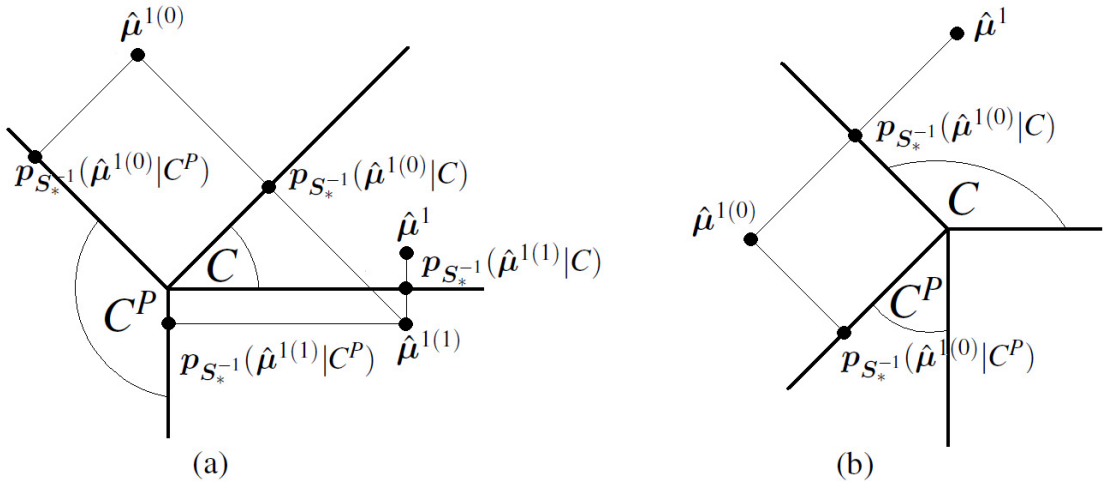

Figure 1: Examples of the iterative procedure for mean vector estimation for an acute (a) and a non-acute cone (b).

information to obtain alternative estimators of the vector means. The generalization to the $k > 2$ populations case appears in Conde *et al.* (2012). These alternative estimators are defined via an iterative procedure whose convergence is shown in Fernández *et al.* (2006) and that is described here for completeness.

Consider the pk square matrix $\boldsymbol{S}_*^{-1} = \left[\text{diag}\left(\frac{\boldsymbol{S}}{n_1}, \frac{\boldsymbol{S}}{n_2}, \ldots, \frac{\boldsymbol{S}}{n_k}\right)\right]^{-1} = \bigoplus_{j=1}^{k} \left(\frac{\boldsymbol{S}}{n_j}\right)^{-1}$.

Definition 1 (Conde *et al.* 2012)

For $\gamma \in [0,1]$, let $\widehat{\boldsymbol{\mu}}^{\gamma}$ be the limit value, when $m \to \infty$, of the following iterative procedure:

$$\widehat{\boldsymbol{\mu}}^{\gamma(m)} = \boldsymbol{p}_{\boldsymbol{S}_*^{-1}}\left(\widehat{\boldsymbol{\mu}}^{\gamma(m-1)}|C\right) - \gamma \boldsymbol{p}_{\boldsymbol{S}_*^{-1}}\left(\widehat{\boldsymbol{\mu}}^{\gamma(m-1)}|C^P\right), \; m = 1, 2, \ldots,$$

where $\widehat{\boldsymbol{\mu}}^{\gamma(0)} = \left(\overline{\boldsymbol{Y}}_1^{\top}, \ldots, \overline{\boldsymbol{Y}}_k^{\top}\right)^{\top} \in \mathbb{R}^{pk}$, $\boldsymbol{p}_{\boldsymbol{S}_*^{-1}}(\boldsymbol{Y}|C)$, is the projection of $\boldsymbol{Y} \in \mathbb{R}^{pk}$ onto the cone C using the metric given by the matrix $\boldsymbol{S}_*^{-1}$, and $C^P = \{\boldsymbol{y} \in \mathbb{R}^{pk} : \boldsymbol{y}^{\top}\boldsymbol{S}_*^{-1}\boldsymbol{x} \leq 0, \; \boldsymbol{x} \in C\}$ is the polar cone of C.

The computation of the projection of a vector onto a polyhedral cone can be carried out using the `lsConstrain.fit` method contained in the R package **ibdreg** (Sinnwell and Schaid 2013). Figure 1 shows, in $\mathbb{R}^2$, the cones C and C^P and the estimators defined when $\gamma = 1$ and $\boldsymbol{S}_* = I$, indicating the need for an iterative procedure when C is an acute cone.

These estimators $\widehat{\boldsymbol{\mu}}^{\gamma} = \left(\widehat{\boldsymbol{\mu}}_1^{\gamma\top}, \ldots, \widehat{\boldsymbol{\mu}}_k^{\gamma\top}\right)^{\top}$ are plugged into the original rule to obtain the restricted lineal discriminant rules $R_l(\gamma)$:

$$\text{Classify } \boldsymbol{U} \text{ in } \Pi_j \text{ iff } (\boldsymbol{U} - \widehat{\boldsymbol{\mu}}_j^{\gamma})^{\top}\boldsymbol{S}^{-1}(\boldsymbol{U} - \widehat{\boldsymbol{\mu}}_j^{\gamma}) \leq (\boldsymbol{U} - \widehat{\boldsymbol{\mu}}_l^{\gamma})^{\top}\boldsymbol{S}^{-1}(\boldsymbol{U} - \widehat{\boldsymbol{\mu}}_l^{\gamma}), \quad l = 1, \ldots, k,$$

for $\gamma \in [0,1]$.

For more details on these restricted linear rules and their properties the reader is referred to Fernández *et al.* (2006) and Conde *et al.* (2012).

Restricted quadratic discriminant rules

The problem of incorporating additional information into classification rules when the covariance matrices cannot be assumed to be equal has not, to the best of our knowledge, been considered in the literature up to this date. This problem is important as equality among covariance matrices cannot be assumed in many applications. In this subsection, we extend the ideas appearing in the definition of restricted linear discriminant rules to the unequal covariances case, thus defining the corresponding restricted quadratic discriminant rules. The main novelty is the definition of the appropriate projection matrix for the computation of the restricted estimators. We have checked that the matrix that correctly extends the restricted linear discriminant rules is $\boldsymbol{S}_{**}^{-1} = \bigoplus_{j=1}^{k} \left(\frac{\boldsymbol{S}_j}{n_j}\right)^{-1}$. Consequently, in this case for each $\gamma \in [0,1]$ the estimator $\widehat{\boldsymbol{\mu}}^{\gamma} = \left(\widehat{\boldsymbol{\mu}}_1^{\gamma\top}, \ldots, \widehat{\boldsymbol{\mu}}_k^{\gamma\top}\right)^{\top}$ of $\boldsymbol{\mu} = (\boldsymbol{\mu}_1^{\top}, \ldots, \boldsymbol{\mu}_k^{\top})^{\top} \in C$ is obtained using the iterative procedure described in Definition 1, replacing the matrix $\boldsymbol{S}_{*}^{-1}$ by $\boldsymbol{S}_{**}^{-1}$. Again, the estimators of the means $\widehat{\boldsymbol{\mu}}_j^{\gamma}$ and the covariance matrices $\boldsymbol{S}_j$ are plugged into the original rule (3) to obtain the restricted quadratic discriminant rules $R_q(\gamma)$:

$$\text{Classify } \boldsymbol{U} \text{ in } \Pi_j \text{ iff } -\frac{1}{2}\log|S_j| - \frac{1}{2}\left\{(\boldsymbol{U}-\widehat{\boldsymbol{\mu}}_j^{\gamma})^{\top}\boldsymbol{S}_j^{-1}(\boldsymbol{U}-\widehat{\boldsymbol{\mu}}_j^{\gamma}),\right\} \geq$$
$$-\frac{1}{2}\log|S_l| - \frac{1}{2}\left\{(\boldsymbol{U}-\widehat{\boldsymbol{\mu}}_l^{\gamma})^{\top}\boldsymbol{S}_l^{-1}(\boldsymbol{U}-\widehat{\boldsymbol{\mu}}_l^{\gamma}),\right\},\ l=1,\ldots,k,$$

for $\gamma \in [0,1]$.

The computational complexity of these rules is clearly not significantly higher than that of the restricted linear discriminant rules, and the simulations studies we have performed show, as we will see in the applications appearing in Section 4, that this rule improves the corresponding unrestricted quadratic discriminant rule.

The `rlda` function and `predict` method for '`rlda`' objects in R package **dawai** allow to define restricted linear discriminant rules and to classify samples, respectively. The `rqda` function and `predict` method for '`rqda`' objects are the corresponding versions for performing restricted quadratic discrimination.

2.2. True error rate estimation

From an applied point of view, the evaluation of the classification rule for a given training sample is even more important than the expected loss of the rule. The true error rate, E_n, of the rule R_n is the probability of misclassification of the rule given the training sample, i.e., $E_n = P_{\boldsymbol{X}Z}(R_n(M_n, \boldsymbol{U}) \neq Z \mid M_n)$. It is well known that the best way of estimating the true classification error of a classification rule is the use of an independent and large enough sample, usually called test sample. However, in practice it is common that the sample size is not large enough to split the sample into a training and a test sample as that would decrease the efficiency of the rule. For this reason, the estimation of E_n for the usual rules such as for example Fisher's linear rule (1), the quadratic discriminant rule (3), the nearest neighbors rules (Cover and Hart 1967) or random forest rules (Breiman 2001), is a widely studied topic in the literature. Parametric and non-parametric estimators of E_n have been proposed and non-parametric estimators based on resampling have shown a good performance for the above mentioned rules. Schiavo and Hand (2000) summarize the work made on this topic until that

date. More recent references are, for instance, Steele and Patterson (2000), Wehberg and Schumacher (2004), Fu, Carroll, and Wang (2005), Molinaro, Simon, and Pfeiffer (2005), Kim and Cha (2006), Kim (2009) or Borra and Di Ciaccio (2010).

Conde *et al.* (2013) propose four new estimators of E_n specific to the restricted linear discriminant rule $R_l(\gamma)$ for $k = 2$ populations. Two of them, $BT2$ and $BT3$, are generated from the leave-one-out bootstrap ($LOOBT$, see Efron 1983). The other two, $BT2CV$ and $BT3CV$, are cross-validation after bootstrap (BCV, see Fu *et al.* 2005) versions of $BT2$ and $BT3$ respectively. In the following, we describe the generalization of these estimators to $k > 2$ populations and to restricted quadratic discrimination cases. This is the second theoretical novelty appearing in this paper. As happened when the restricted discriminant rule was extended to the more than 2 populations case in Conde *et al.* (2012), the extension is not immediate as the appropriate parameter spaces and projection matrices have to be considered. Following the arguments already appearing in the original paper, Efron (1979), or in Boos (2003), the underlying idea in the definition of the new estimators of the true error rate is that the "bootstrap world" should mirror the "real world". We present two proposals: the first one is to modify the restrictions cone, the second one is to adapt the training sample.

The BT2 and BT2CV estimators

Assume that the additional information is written as in (4). Let us denote as $\overline{C}$ the following random cone generated by the sample mean vectors $\overline{\boldsymbol{Y}} = \left(\overline{\boldsymbol{Y}}_1^\top, \ldots, \overline{\boldsymbol{Y}}_k^\top\right)^\top$:

$$\overline{C} = \left\{ \boldsymbol{x} \in \mathbb{R}^{pk} : \begin{array}{lll} \boldsymbol{a}_j^\top \boldsymbol{x} \geq 0 & \text{if} & \boldsymbol{a}_j^\top \overline{\boldsymbol{Y}} \geq 0 \\ \boldsymbol{a}_j^\top \boldsymbol{x} \leq 0 & \text{if} & \boldsymbol{a}_j^\top \overline{\boldsymbol{Y}} < 0 \end{array}, \ j = 1, \ldots, q \right\},$$

i.e., the cone determined by the restrictions verified by the sample means.

The true error rate estimator $BT2$ of the restricted linear or quadratic classification rules $(R_l(\gamma),\ R_q(\gamma))$ is computed as follows. A bootstrap training sample $M_n^* = \{(\boldsymbol{Y}_i^*, Z_i^*),\ i = 1, \ldots, n\}$ is a size n randomly obtained (with replacement) sample from the original training sample (i.e., $\mathsf{P}((\boldsymbol{Y}_i^*, Z_i^*) = (\boldsymbol{Y}_s, Z_s)) = \frac{1}{n}$ with $s, i \in \{1, \ldots, n\}$). B such bootstrap samples $M_n^{*b} = \{(\boldsymbol{Y}_i^{*b}, Z_i^{*b}), i = 1, \ldots, n\}$, $b = 1, \ldots, B$, are obtained from M_n. For each bootstrap training sample we define the bootstrap version of the estimator of $\boldsymbol{\mu} = (\boldsymbol{\mu}_1^\top, \ldots, \boldsymbol{\mu}_k^\top)^\top$ that we denote as $\boldsymbol{\mu}_\gamma^{*b}$ (with $\gamma \in [0, 1]$), as the limit when $m \to \infty$ of the following iterative procedure similar to the one considered in Definition 1. Let $\widehat{\boldsymbol{\mu}}_\gamma^{(0)b} = \overline{\boldsymbol{Y}}$ and

$$\widehat{\boldsymbol{\mu}}_\gamma^{(m)b} = \boldsymbol{p}_{\boldsymbol{A}}\left(\widehat{\boldsymbol{\mu}}_\gamma^{(m-1)b} | \overline{C}\right) - \gamma \boldsymbol{p}_{\boldsymbol{A}}\left(\widehat{\boldsymbol{\mu}}_\gamma^{(m-1)b} | \overline{C}^P\right),\ m = 1, 2, \ldots,$$

where matrix $\boldsymbol{A}$ is equal to $\boldsymbol{S}_*^{-1}$ for the restricted linear discriminant rule and equal to $\boldsymbol{S}_{**}^{-1}$ for the restricted quadratic discriminant rule.

Now, we denote as $R_l^{*b}(\gamma)$ and $R_q^{*b}(\gamma)$ the bootstrap versions of the classification rules $R_l(\gamma)$ and $R_q(\gamma)$, respectively. For each of the B bootstrap rules we classify the observations in the original training sample that do not belong to the corresponding bootstrap sample M_n^{*b}. The true error rate estimator $BT2$ is the proportion of observations wrongly classified.

The $BT2CV$ estimator is the BCV (Fu *et al.* 2005) version of $BT2$. For each of the B bootstrap training samples, let CV_b be the true error rate estimator obtained using the cross-validation method on sample M_n^{*b}. Then $BT2CV = \frac{1}{B}\sum_{b=1}^{B} CV_b$.

The BT3 and BT3CV estimators

The true error rate estimator denoted as $BT3$ is based on adapting the original training sample, instead of modifying the cone C like in $BT2$, as follows. Assume that the original training sample $M_n = \{(\boldsymbol{Y}_i, Z_i), i = 1, \ldots, n\}$ does not verify the restrictions, i.e., $\overline{\boldsymbol{Y}} \notin C$. For any $\gamma \in [0, 1]$, let $\widehat{\boldsymbol{\mu}}_j^{\gamma}$ be the restricted estimator of $\boldsymbol{\mu}_j$ obtained in Definition 1. Now, we transform the original training sample in such a way that the new sample means belong to C. The transformed training sample is $\{(\boldsymbol{W}_i, Z_i), i = 1, \ldots, n\}$, where

$$\boldsymbol{W}_i = \boldsymbol{Y}_i - \overline{\boldsymbol{Y}}_j + \widehat{\boldsymbol{\mu}}_j^{\gamma} \quad \text{if } Z_i = j,$$

for $i = 1, \ldots, n$ and $j = 1, \ldots, k$.

In this way $\overline{\boldsymbol{W}} = \left(\overline{\boldsymbol{W}}_1^{\top}, \ldots, \overline{\boldsymbol{W}}_k^{\top}\right)^{\top}$, $\overline{\boldsymbol{W}}_j = \frac{1}{n_j} \sum_{l=1}^{n} \boldsymbol{W}_l I_{(Z_l = j)}$, $j = 1, \ldots, k$. Now, the estimator denoted as $BT3$ is computed in a similar way to that of $BT2$ but replacing the original training sample by the transformed one and C by $\overline{C}$.

The $BT3CV$ estimator is the cross-validation after bootstrap version of $BT3$.

These four estimators of the true error rates of the restricted linear and quadratic discriminant rules can be obtained with the `err.est` methods for '`rlda`' and '`rqda`' objects in the R package **dawai**.

3. Package dawai

We start this section giving some background on R packages for performing discriminant analysis. We then explain some details about the functions of this package.

3.1. Related packages

As discriminant analysis is a well-known and widely used technique there are many packages in R for performing discriminant analysis. The basic procedures are in the following package:

- **MASS** (Ripley 2015): Support functions and datasets for Venables and Ripley's MASS.

Some more recent packages including new features and discrimination in specific conditions are:

- **mda** (Hastie, Tibshirani, Leisch, Hornik, and Ripley 2015): Mixture and flexible discriminant analysis.
- **rrlda** (Gschwandtner, Filzmoser, Croux, and Haesbroeck 2012): Robust regularized linear discriminant analysis.
- **sparsediscrim** (Ramey 2014): Sparse discriminant analysis.

Since none of the existing packages for discriminant analysis are applicable to perform discriminant analysis under restrictions, we introduce in this article the package "*d*iscriminant *a*nalysis *w*ith *a*dditional *i*nformation", with the acronym **dawai**.

Our package depends on package **boot** (Canty and Ripley 2015) for bootstrapping, package **ibdreg** (Sinnwell and Schaid 2013) for computing the projection of a vector onto a polyhedral

cone with `lsConstrain.fit`, and package **mvtnorm** (Genz, Bretz, Miwa, Mi, Leisch, Scheipl, and Hothorn 2015) for computing multivariate normal densities. These packages should be installed before loading **dawai**.

3.2. Functions provided in dawai

The R package **dawai** consists of a total of six functions, three for each of the two restricted discrimination analysis situations: equal or unequal covariances in the populations. The three functions for each case are: one to define the rules that take into account the additional information expressed in terms of restrictions on the population means and to classify the samples in the training set; a second one which predicts the populations of new samples using the previously defined rule; and, finally, a third one which can evaluate the accuracy of the rules associated to the training set.

The R help files provide the definitive reference. Here we explain some details about the main user-level functions and their specific arguments.

The first function for linear discrimination is the the `rlda` function that can be used to build restricted linear classification rules with additional information expressed as inequality restrictions on the populations means, using the methodology developed in Fernández *et al.* (2006) and Conde *et al.* (2012). It creates an object of class '`rlda`':

```
rlda(x, grouping, subset = NULL, resmatrix = NULL, restext = NULL,
  gamma = c(0, 1), prior = NULL, ...)
```

or

```
rlda(formula, data, subset = NULL, resmatrix = NULL, restext = NULL,
  gamma = c(0, 1), prior = NULL, ...)
```

Argument `resmatrix` collects the additional information on the means vectors, as follows: $\texttt{resmatrix} \cdot (\boldsymbol{\mu}_1^\top, \ldots, \boldsymbol{\mu}_k^\top)^\top \leq \mathbf{0}$. Each of the rows of `resmatrix` corresponds to each of the q restrictions included in the model (4). Obviously, the number of columns of `resmatrix` is the total number of parameters kp, with the first p columns corresponding to the mean values of the predictors for the first population and so on.

The purpose of `restext` is to facilitate the specification of the two most usual cones of restrictions, the tree order (5) and the simple order (6) cones. The first element of `restext` must be either "s" (simple order) or "t" (tree order), the second element must be either "<" (increasing componentwise order) or ">" (decreasing componentwise order), and the rest of the elements must be numbers from the set $\{1, \ldots, p\}$, separated by commas, specifying among which variables the restrictions hold.

Argument `gamma` is the vector of values (in the unit interval) used to determine the restricted rules $R_l(\gamma)$.

The second function is the `predict` method for '`rlda`' objects and classifies multivariate observations contained in a data frame `newdata` using the restricted linear classification rules defined in an `object` of class '`rlda`':

```
predict(object, newdata, prior = object$prior, gamma = object$gamma,
  grouping = NULL, ...)
```

Finally, the third function is the `err.est` method for '`rlda`' objects and estimates the true error rate of the restricted linear classification rules defined in an object `x` of class '`rlda`', using the methodology developed in Conde *et al.* (2013) and in Section 2.2:

```
err.est(x, nboot = 50, gamma = x$gamma, prior = x$prior, ...)
```

Argument `nboot` is the number of bootstrap samples used.

The `rqda` function and the `predict` and `err.est` methods for '`rqda`' objects are the corresponding versions of the `rlda` and the `predict` and `err.est` methods for '`rlda`' objects for performing restricted quadratic discrimination. The `rqda` function builds restricted quadratic classification rules using the methodology developed in Section 2.1. The `predict` method for '`rqda`' objects classifies multivariate observations with restricted quadratic classification rules. Finally, the `err.est` method for '`rqda`' objects estimates the true error rate of restricted quadratic classification rules using the methodology developed in Section 2.2.

Examples to illustrate these functions are provided in Section 4.

4. Applications

There is a wide range of applications of the **dawai** package, which we illustrate in this section using two data sets coming from two different fields, i.e., biology and pattern recognition.

4.1. Biological application

In patient care, as for example in cancer treatment, an important component is the correct classification of the patient into one of the disease stages. The disease stages correspond to increasingly advanced levels of the disease, so it is reasonable to expect the mean values of some variables to increase or decrease with the severity of the illness. This is the case for primary biliary cirrhosis (PBC), an autoimmune liver disease causing liver inflammation and a gradual destruction of the intrahepatic bile ducts found within the liver. PBC is a progressive disease, with four sucessive stages as time passes (Scheuer 1967).

The data set we will use now, called `pbc`, is in the R package **survival** (Therneau 2015), taken from Therneau and Grambsch (2000). This data set is from the Mayo Clinic trial in PBC of the liver conducted between 1974 and 1984, and it has 418 cases and 20 variables.

We will use this data set to exemplify the restricted linear discriminant rules. We consider three variables as predictors ($p = 3$), `bili`, `albumin` and `platelet` (the amounts of serum bilirubin (mg/dl) and serum albumin (g/dl) and platelet count, respectively), and three populations ($k = 3$), joining the original stages 1 and 2 into one so that the classes have enough elements to split the sample into training and test data sets of reasonable sizes, as seen below.

```
R> data("pbc", package = "survival")
R> data <- pbc[, c("bili", "albumin", "platelet", "stage")]
```

Let us take complete observations only.

```
R> data <- na.omit(data)
```

We transform logarithmically the values of the explicative variables so that the variables are approximately normally distributed.

```
R> data <- data.frame(stage = factor(data$stage),
+    logBili = log(data[["bili"]]), logAlbumin = log(data[["albumin"]]),
+    logPlatelet = log(data[["platelet"]]))
R> levels(data$stage)

[1] "1" "2" "3" "4"

R> table(data$stage)

  1   2   3   4
 20  86 153 142
```

These are the number of elements in each of the four classes. Notice that there is a low number in the first class and that 401 cases out of the 418 initial ones have no missing values in the three predictor variables. As mentioned before, we join stages "1" and "2" and relabel them.

```
R> levels(data$stage) <- c(1, 1, 2, 3)
R> table(data$stage)

  1   2   3
106 153 142
```

We will consider the restrictions between population means given by the whole data set: $\mu_{1,1} \leq \mu_{2,1} \leq \mu_{3,1}$, $\mu_{1,2} \geq \mu_{2,2} \geq \mu_{3,2}$ and $\mu_{1,3} \geq \mu_{2,3} \geq \mu_{3,3}$, i.e., the amount of serum bilirubin increases and the amount of serum albumin and platelet count decreases with PBC stage. Notice that as two orderings are decreasing and one increasing, we cannot use the `restext` argument. These restrictions need to be expressed as $\texttt{resmatrix} \cdot (\boldsymbol{\mu}_1^\top, \boldsymbol{\mu}_2^\top, \boldsymbol{\mu}_3^\top)^\top \leq \mathbf{0}$, where $\boldsymbol{\mu}_i^\top = (\mu_{i,1}, \mu_{i,2}, \mu_{i,3})$ is the population Π_i mean vector, $i = 1, 2, 3$. We have $q = 6$ restrictions, $p = 3$ predictors and $k = 3$ populations so that `resmatrix` is a 6×9 matrix. Then, we define the following restrictions matrix (`resmatrix`):

```
R> A <- matrix(0, ncol = 9, nrow = 6)
R> A[matrix(c(1, 1, 4, 4, 2, 5, 3, 6, 5, 8, 6, 9), ncol = 2,
+    byrow = TRUE)] <- 1
R> A[matrix(c(1, 4, 4, 7, 2, 2, 3, 3, 5, 5, 6, 6), ncol = 2,
+    byrow = TRUE)] <- -1
R> A

     [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9]
[1,]    1    0    0   -1    0    0    0    0    0
[2,]    0   -1    0    0    1    0    0    0    0
[3,]    0    0   -1    0    0    1    0    0    0
[4,]    0    0    0    1    0    0   -1    0    0
[5,]    0    0    0    0   -1    0    0    1    0
[6,]    0    0    0    0    0   -1    0    0    1
```

We split the data set into a randomly selected training set and a test set, fixing a seed in order to get the same results as the reader.

```
R> set.seed(-5436)
R> values <- runif(nrow(data))
R> trainsubset <- (values < 0.25)
R> testsubset <- (values >= 0.25)
```

Now we can build the restricted linear discriminant rules. Let us consider equal a priori probabilities.

```
R> library("dawai")
R> obj <- rlda(stage ~ logBili + logAlbumin + logPlatelet, data,
+    subset = trainsubset, gamma = c(0, 0.75, 1), resmatrix = A,
+    prior = c(1/3, 1/3, 1/3))
R> obj

Restrictions:
   mu1,1 - mu2,1 <= 0
 - mu1,2 + mu2,2 <= 0
 - mu1,3 + mu2,3 <= 0
   mu2,1 - mu3,1 <= 0
 - mu2,2 + mu3,2 <= 0
 - mu2,3 + mu3,3 <= 0

Prior probabilities of classes:
   class1    class2    class3
0.3333333 0.3333333 0.3333333

Apparent error rate (%):
   gamma=0 gamma=0.75    gamma=1
  45.09804   47.05882   47.05882
```

The apparent error rates suggest that the classes are not completely separated in the training sample. This is a situation, usual in practice, where the restricted rules are expected to perform better than the unrestricted ones, as we will see.

Now we consider the test set, containing the observations in `data` not present in the training set, and classify them. As we know the classes that the observations in the test set belong to, we can estimate the true error rate.

```
R> pred <- predict(obj, newdata = data[testsubset, ],
+    grouping = data[testsubset, "stage"])
R> pred$error.rate

                      gamma=0 gamma=0.75  gamma=1
True error rate (%): 49.83278   48.82943 47.82609
```

The fact that the apparent error rate increases and the true error rate decreases with γ is a typical effect for these restricted rules, see Fernández *et al.* (2006), Conde *et al.* (2012) and Conde *et al.* (2013).

Finally, we estimate the true error rate from the training sample with `nboot` = 50 (the default value).

```
R> err.est(obj)

Restrictions:
   mu1,1 - mu2,1 <= 0
 - mu1,2 + mu2,2 <= 0
 - mu1,3 + mu2,3 <= 0
   mu2,1 - mu3,1 <= 0
 - mu2,2 + mu3,2 <= 0
 - mu2,3 + mu3,3 <= 0

Prior probabilities of classes:
   class1    class2    class3
0.3333333 0.3333333 0.3333333

True error rate estimation (%):
       gamma=0 gamma=0.75  gamma=1
BT2   45.12459   45.12459 44.85374
BT3   49.45829   49.62080 49.18743
BT2CV 42.96078   42.72549 42.66667
BT3CV 46.58824   48.00000 47.68627
```

We can see that in this case $BT3$ and $BT3CV$ provide reasonable estimates of the true error rate.

We can also compare these results with the error rates for some standard unrestricted classifiers such as LDA (**MASS** package) and random forest (**randomForest** package). For $\gamma = 1$, the test error rates for the restricted linear rules are 7.74% lower than for LDA (51.84) and 15.38% lower than for random forest (56.52).

4.2. Pattern recognition application

As an example of pattern recognition, we will use the data set contained in **dawai** package called `Vehicle2`.

```
R> data("Vehicle2", package = "dawai")
```

This data set is a subset from the `Vehicle` data set, available in the R package **mlbench** (Leisch and Dimitriadou 2010) and taken from the UCI Machine Learning Repository (Bache and Lichman 2013), originally gathered in Siebert (1987). The purpose of the data set is to study how to distinguish 3D objects from a 2D image, i.e., how to classify a given silhouette as viewed from a camera from different angles and elevations into one of four types of vehicles, using a set of features extracted from the silhouette. The vehicles used were a double-decker

bus, a Cheverolet van, a Saab 9000 and an Opel Manta 400, with the expectation that the bus, the van and either one of the cars would be readily distinguishable, but it would be more difficult to distinguish between the cars.

`Vehicle2` is a data frame with 846 observations on 5 variables, four numerical and one nominal defining the class of the objects, i.e., the vehicle. The variables are `Skew.maxis` (skewness about minor axis), `Kurt.Maxis` (kurtosis about major axis), `Holl.Ra` (quotient hollows area/bounding polygon area), `Sc.Var.maxis` (quotient 2nd order moment about minor axis/area) and `Class`.

We will use this data set to exemplify the restricted quadratic discriminant rules. We consider the four variables as predictors ($p = 4$) and the four available populations ($k = 4$).

```
R> data <- Vehicle2[, 1:4]
R> grouping <- Vehicle2$Class
R> levels(grouping)

[1] "bus"  "opel" "saab" "van"

R> levels(grouping) <- c(4, 2, 1, 3)
```

We have "ordered" the populations in terms of the vehicle size. It could be reasonable to think that the means of the first three variables decrease with the vehicle size (in fact, this ordering is verified by the whole data set), so let us suppose the following restrictions on the means: $\mu_{11} \geq \mu_{21} \geq \mu_{31} \geq \mu_{41}$, $\mu_{12} \geq \mu_{22} \geq \mu_{32} \geq \mu_{42}$, $\mu_{13} \geq \mu_{23} \geq \mu_{33} \geq \mu_{43}$. As this is a classical simple order on these predictors and all orderings are decreasing, we easily specify these restrictions by `restext = "s>1,2,3"`.

We split the data set into a randomly selected training set and test set, fixing a seed in order to get the same results as the reader.

```
R> set.seed(-9152)
R> values <- runif(nrow(data))
R> trainsubset <- (values < 0.25)
```

Now we can build the restricted quadratic discriminant rules:

```
R> obj <- rqda(data, grouping, subset = trainsubset, restext = "s>1,2,3")
R> obj

Restrictions:
 - mu1,1 + mu2,1 <= 0
 - mu1,2 + mu2,2 <= 0
 - mu1,3 + mu2,3 <= 0
 - mu2,1 + mu3,1 <= 0
 - mu2,2 + mu3,2 <= 0
 - mu2,3 + mu3,3 <= 0
 - mu3,1 + mu4,1 <= 0
 - mu3,2 + mu4,2 <= 0
```

```
 - mu3,3 + mu4,3 <= 0

Prior probabilities of classes:
   class1    class2    class3    class4
0.2431193 0.2752294 0.2385321 0.2431193

Apparent error rate (%):
 gamma=0  gamma=1
28.89908 29.81651
```

Note that, as we have neither specified `gamma` nor `prior`, restricted rules are by default obtained for $\gamma = 0, 1$, and the class proportions of the training set are used as the prior probabilities of class membership.

Now we consider the test set, containing the observations in `data` not present in the training set, and classify them. As we know the classes of the observations in the test set, we can estimate the true error rate.

```
R> testsubset <- (values >= 0.25)
R> pred <- predict(obj, newdata = data[testsubset, ],
+    grouping = grouping[testsubset])
R> pred$error.rate

                      gamma=0  gamma=1
True error rate (%): 32.80255 31.21019
```

Finally, we estimate the true error rate from the training sample.

```
R> err.est(obj)

Restrictions:
 - mu1,1 + mu2,1 <= 0
 - mu1,2 + mu2,2 <= 0
 - mu1,3 + mu2,3 <= 0
 - mu2,1 + mu3,1 <= 0
 - mu2,2 + mu3,2 <= 0
 - mu2,3 + mu3,3 <= 0
 - mu3,1 + mu4,1 <= 0
 - mu3,2 + mu4,2 <= 0
 - mu3,3 + mu4,3 <= 0

Prior probabilities of classes:
   class1    class2    class3    class4
0.2431193 0.2752294 0.2385321 0.2431193

True error rate estimation (%):
       gamma=0  gamma=1
BT2   36.20905 35.75894
```

```
BT3   36.30908 36.68417
BT2CV 32.25688 32.21101
BT3CV 32.29358 32.28440
```

In this case, $BT2CV$ and again $BT3CV$ provide good estimates of the true error rate.

Again, if we compare these results with the error rates for some standard unrestricted classifiers such as QDA (**MASS** package) and random forest, for $\gamma = 1$ the test error rates for the restricted quadratic rules are 7.98% lower than for QDA (33.92) and 14.78% lower than for random forest (36.62).

In both examples we can see that there is a significant improvement with respect to usual methods in statistical practice that do not take into account the additional information given by the restrictions.

5. Conclusions

In this paper the R package **dawai** has been presented. The package provides the functions needed to define linear or quadratic classification rules under order restrictions, to classify the samples and to evaluate the accuracy of the rules.

We have also extended in this paper the definitions given in previous papers (Fernández *et al.* 2006, Conde *et al.* 2012, Conde *et al.* 2013) from the case of equal covariances in the different populations to the case of unequal covariances among the populations and consequently defined the corresponding restricted quadratic discriminant rules. Another novelty is the definition of estimators of the accuracy of the rules for the general more than two populations case, for restricted linear and quadratic discriminant rules, thus completing the procedures presented in those three previous papers.

Though we have illustrated the proposed methodology using examples from biology and pattern recognition, the software can obviously be applied to a wide range of contexts such as medical image analysis, drug discovery and development, optical character and handwriting recognition, document classification, credit scoring,.... We expect the software described to be useful for researchers working in any of those fields.

Acknowledgments

This research was partially supported by the Spanish MCI grant MTM2012-37129. We thank the anonymous reviewers, the associate editor and the editor for several useful comments which improved the presentation of this manuscript.

References

Bache K, Lichman M (2013). "UCI Machine Learning Repository." URL http://archive.ics.uci.edu/ml/.

Barragán S, Fernández MA, Rueda C, Peddada SD (2013). "**isocir**: An R Package for Constrained Inference using Isotonic Regression for Circular Data, with an Application to Cell

Biology." *Journal of Statistical Software*, **54**(4), 1–17. URL http://www.jstatsoft.org/v54/i04/.

Boos DD (2003). "Introduction to the Bootstrap World." *Statistical Science*, **18**(2), 168–174.

Borra S, Di Ciaccio A (2010). "Measuring the Prediction Error. A Comparison of Cross-Validation, Bootstrap and Covariance Penalty Methods." *Computational Statistics and Data Analysis*, **54**(12), 2976–2989.

Breiman L (2001). "Random Forests." *Machine Learning*, **45**(1), 5–32.

Canty A, Ripley B (2015). ***boot**: Bootstrap R (S-PLUS) Functions*. R package version 1.3-17, URL http://CRAN.R-project.org/package=boot.

Conde D, Fernández MA, Rueda C, Salvador B (2012). "Classification of Samples into Two or More Ordered Populations with Application to a Cancer Trial." *Statistics in Medicine*, **31**(28), 3773–3786.

Conde D, Salvador B, Rueda C, Fernández MA (2013). "Performance and Estimation of the True Error Rate of Classification Rules built with Additional Information. An Application to a Cancer Trial." *Statistical Applications in Genetics and Molecular Biology*, **12**(5), 583–602.

Cover T, Hart P (1967). "Nearest Neighbor Pattern Classification." *IEEE Transactions on Information Theory*, **13**(1), 21–27.

Davidov O, Peddada SD (2013). "Testing for the Multivariate Stochastic Order among Ordered Experimental Groups with Application to Dose-Response Studies." *Biometrics*, **101**(9), 1903–1909.

Efron B (1979). "Bootstrap Methods: Another Look at the Jackknife." *The Annals of Statistics*, **7**, 1–26.

Efron B (1983). "Estimating the Error Rate of a Prediction Rule: Improvement on Cross-Validation." *Journal of the American Statistical Association*, **78**, 316–331.

El Barmi H, Johnson M, Mukerjee H (2010). "Estimating Cumulative Incidence Functions When the Life Distributions are Constrained." *Journal of Multivariate Analysis*, **101**(9), 1903–1909.

Fernández MA, Rueda C, Peddada SD (2012). "Identification of a Core Set of Signature Cell Cycle Genes Whose Relative Order of Time to Peak Expression is Conserved Across Species." *Nucleic Acids Research*, **40**(7), 2823–2832.

Fernández MA, Rueda C, Salvador B (2006). "Incorporating Additional Information to Normal Linear Discriminant Rules." *Journal of the American Statistical Association*, **101**(474), 569–577.

Fisher RA (1936). "The Use of Multiple Experiments in Taxonomic Problems." *Annals of Eugenics*, **7**, 179–188.

Fu WJ, Carroll RJ, Wang S (2005). "Estimating Misclassification Error with Small Samples via Bootstrap Cross-Validation." *Bioinformatics*, **21**, 1979–1986.

Genz A, Bretz F, Miwa T, Mi X, Leisch F, Scheipl F, Hothorn T (2015). ***mvtnorm**: Multivariate Normal and t Distributions*. R package version 1.0-3, URL http://CRAN.R-project.org/package=mvtnorm.

Ghosh D, Banerjee M, Biswas P (2008). "Inference for Constrained Estimation of Tumor Size Distributions." *Biometrics*, **64**(4), 1009–1017.

Gschwandtner M, Filzmoser P, Croux C, Haesbroeck G (2012). ***rrlda**: Robust Regularized Linear Discriminant Analysis*. R package version 1.1, URL http://CRAN.R-project.org/package=rrlda.

Hastie T, Tibshirani R, Leisch F, Hornik K, Ripley BD (2015). ***mda**: Mixture and Flexible Discriminant Analysis*. R package version 0.4-7, URL http://CRAN.R-project.org/package=mda.

Kim J (2009). "Estimating Classification Error Rate: Repeated Cross-Validation, Repeated Hold-Out and Bootstrap." *Computational Statistics and Data Analysis*, **53**(11), 3735–3745.

Kim J, Cha E (2006). "Estimating Prediction Errors in Binary Classification Problem: Cross-Validation versus Bootstrap." *Korean Communications in Statistics*, **13**, 151–165.

Leisch F, Dimitriadou E (2010). ***mlbench**: Machine Learning Benchmark Problems*. R package version 2.1-1, URL http://CRAN.R-project.org/package=mlbench.

Long T, Gupta RD (1998). "Alternative Linear Classification Rules under Order Restrictions." *Communications in Statistics – Theory and Methods*, **27**(3), 559–575.

Molinaro AM, Simon R, Pfeiffer RM (2005). "Prediction Error Estimation: A Comparison of Resampling Methods." *Bioinformatics*, **21**(15), 3301–3307.

Ramey JA (2014). ***sparsediscrim**: Sparse Discriminant Analysis*. R package version 0.2, URL http://CRAN.R-project.org/package=sparsediscrim.

R Core Team (2015). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Ripley B (2015). ***MASS**: Support Functions and Datasets for Venables and Ripley's MASS*. R package version 7.3-43, URL http://CRAN.R-project.org/package=MASS.

Robertson T, Wright FT, Dykstra RL (1988). *Order Restricted Statitical Inference*. John Wiley & Sons.

Rueda C, Fernández MA, Peddada S (2009). "Estimation of Parameters Subject to Order Restrictions on a Circle with Application to Estimation of Phase Angles of Cell-Cycle Genes." *Journal of the American Statistical Association*, **104**(485), 338–347.

Salvador B, Fernández MA, Martín I, Rueda C (2008). "Robustness of Classification Rules that Incorporate Additional Information." *Computational Statistics & Data Analysis*, **52**(5), 2489–2495.

Scheuer P (1967). "Primary Biliary Cirrhosis." *Proceedings of the Royal Society of Medicine*, **60**(12), 1257–1260.

Schiavo RA, Hand DJ (2000). "Ten More Years of Error Rate Research." *International Statistical Review*, **68**(3), 295–310.

Siebert JP (1987). "Vehicle Recognition Using Rule Based Methods." *Technical Report TIRM-87-018*, Turing Institute, Glasgow, Scotland.

Silvapulle MJ, Sen PK (2005). *Constrained Statistical Inference*. John Wiley & Sons.

Sinnwell JP, Schaid DJ (2013). ***ibdreg**: Regression Methods for IBD Linkage with Covariates*. R package version 0.2.5, URL http://CRAN.R-project.org/package=ibdreg.

Steele BM, Patterson DA (2000). "Ideal Bootstrap Estimation of Prediction Error for K-Nearest Neighbor Classifiers: Applications for Classification and Error Assessment." *Statistics and Computing*, **10**(4), 349–355.

Therneau T, Grambsch P (2000). *Modeling Survival Data: Extending the Cox Model*. Springer-Verlag.

Therneau TM (2015). ***survival**: A Package for Survival Analysis in S*. R package version 2.38-3, URL http://CRAN.R-project.org/package=survival.

Wehberg S, Schumacher M (2004). "A Comparison of Nonparametric Error Rate Estimation Methods in Classification Problems." *Biometrical Journal*, **46**(1), 35–47.

Affiliation:

David Conde, Miguel A. Fernández, Bonifacio Salvador, Cristina Rueda
Departamento de Estadística e Investigación Operativa
Instituto de Matemáticas (IMUVA)
Universidad de Valladolid
Valladolid, Spain
E-mail: dconde@eio.uva.es, miguelaf@eio.uva.es, bosal@eio.uva.es, crueda@eio.uva.es
URL: http://www.eio.uva.es/~david/, http://www.eio.uva.es/~miguel/, http://www.eio.uva.es/~boni/, http://www.eio.uva.es/~cristina/

Journal of Statistical Software http://www.jstatsoft.org/
published by the American Statistical Association http://www.amstat.org/
Volume 66, Issue 10 *Submitted:* 2014-02-18
August 2015 *Accepted:* 2014-11-21