Stat Papers (2009) 50:817–835
DOI 10.1007/s00362-009-0259-z

# Discriminant analysis of multivariate repeated measures data with a Kronecker product structured covariance matrices

**Mirosław Krzyśko · Michał Skorzybut**

Received: 26 September 2008 / Accepted: 19 May 2009 / Published online: 5 August 2009


**Abstract** This paper proposes new classifiers under the assumption of multivariate normality for multivariate repeated measures data with Kronecker product covariance structures. These classifiers are especially effective when the number of observations is not large enough to estimate the covariance matrices, and thus the traditional classifiers fail. Computational scheme for maximum likelihood estimates of required class parameters are also given. The quality of these new classifiers are examined on some real data.

**Keywords** Classifiers · Repeated measures data · Kronecker product covariance structure · Maximum likelihood estimates

**Mathematics Subject Classification (2000)** Primary 62H30 · Secondary 62H12

## 1 Introduction

*Discrimination* or *classification* is about predicting the unknown class to which an observation is to be allocated. An observation is a collection of numerical measurements represented by a $d$-dimensional vector $\boldsymbol{x}$. The unknown nature of the observation is called a *class*. It is denoted by $y$ and takes values in the set $\mathcal{Y} \in \{1, 2, \ldots, K\}$.

Dedicated with best wishes to Professor Tadeusz Caliński on the occasion of his 80th birthday.

M. Krzyśko (✉) · M. Skorzybut
Faculty of Mathematics and Computer Science, Adam Mickiewicz University,
ul. Umultowska 87, 61-614 Poznan, Poland
e-mail: mkrzysko@amu.edu.pl

M. Skorzybut
e-mail: skorzmic@amu.edu.pl

The mapping

$$g : \mathbb{R}^d \to \mathcal{Y}$$

is called a *classifier* or a *classification rule*.

The optimal Bayes classifier is

$$g(\boldsymbol{x}) = \arg \max_{1 \le i \le K} \pi_i f_i(\boldsymbol{x}),$$

where $\pi_i = P(y = i)$ is the prior probability that $\boldsymbol{x}$ is a member of a class $i$, $\pi_1 + \pi_2 + \cdots + \pi_K = 1$, and $f_i(\boldsymbol{x}) = f(\boldsymbol{x}|y = i)$ is the probability density function associated with the random vector $\boldsymbol{x}$ for a class $i$, $i = 1, 2, \ldots, K$. In our model, we have access to a training data observed in the past. The training data set is given by

$$D_n = \{(\boldsymbol{x}_i, y_i)\}, \quad i = 1, 2, \ldots, n.$$

That is, a set of $n$ observations is available for which the true categorization is known. If $\pi_i$ and $f_i(\boldsymbol{x})$ are unknown then a classifier is constructed on the basis of the training data set $D_n$ and is denoted by $g_n$. The process of constructing $g_n$ is called *learning*.

In the usual classification problems a training data are taken at a given time point. Moreover, in many practical situations in medicine, agriculture, psychology and education a training data are taken repeatedly over time. Such learning data are often referred to in the statistical and behavioral science literature as the multivariate repeated measures data or doubly multivariate data.

Suppose there are $p$ response variables and on each of them, observations are taken over $T$ time points. We denote information on a typical subject by $\boldsymbol{x}$, a $(pT \times 1)$-dimensional column vector obtained by stacking all $p$ responses at the first time point, then stacking all $p$ responses at the second time point below it and so on.

Next, assume that $\boldsymbol{x} \sim N_{pT}(\boldsymbol{\nu}, \boldsymbol{\Omega})$ with $pT \times pT$ positive definite covariance matrix $\boldsymbol{\Omega}$. When $\boldsymbol{\nu}$ and $\boldsymbol{\Omega}$ are unknown and completely unspecified, a total of $pT + pT(pT + 1)/2$ unknown parameters must be estimated. This number increases very rapidly with $p$ and $T$. For example, when $p = 4$ and $T = 5$, the number of unknown parameters equals two hundred thirty. Estimation of so many parameters will require a very large sample, which may not always be feasible.

Hence, we assume $\boldsymbol{\Omega}$ to be of the form (Roy and Khattree 2005b; Roy and Khattree 2008):

$$\boldsymbol{\Omega} = \boldsymbol{V} \otimes \boldsymbol{\Sigma},$$

where $\boldsymbol{V}$ is a $T \times T$ positive definite covariance matrix and $\boldsymbol{\Sigma}$ is $p \times p$ positive definite covariance matrix. The matrix $\boldsymbol{V}$ represents the covariance between repeated measures on a given subject and for a given response variable. Likewise, $\boldsymbol{\Sigma}$ represents the covariance between all response variables on a given subject and for a given time point. The above covariance structure is subject to an implicit assumption that for all the response variables, the correlation structure between repeated measures remains the same and that for covariance between all the response variables does not depend on time and remains constant for all time points. No structure whatsover on $\boldsymbol{\Sigma}$ is assumed except that it is positive definite.

Classification rules for univariate repeated measures data were given by Roy and Khattree (2005a). Classification rules in case of multivariate repeated measures data under the assumption of multivariate normality for classes and with compound symmetric correlation structure on the matrix $\boldsymbol{V}$ were given by Roy and Khattree (2005b). Next, Roy and Khattree (2008) gave the solution of this problem for the case when the correlation matrix $\boldsymbol{V}$ has the first order autoregressive (AR(1)) structure.

From Roy and Khattree (2008) study it is clear that taking the correct correlation structure on the repeated measurements in the classification rule is very important to reduce the misclassification error rates. Also, testing the equality of the Kronecker product structure variance-covariance matrices is equally important. Hypotheses testings for Kronecker product structures $\boldsymbol{V} \otimes \boldsymbol{\Sigma}$ with $\boldsymbol{V}$ as CS structure are discussed in Roy and Khattree (2003) for both one population and two population cases. The paper also discusses the hypotheses testings for Kronecker product structures $\boldsymbol{V} \otimes \boldsymbol{\Sigma}$ for one population case with both $\boldsymbol{V}$ and $\boldsymbol{\Sigma}$ as unstructured. Similar hypotheses testings for Kronecker product structures $\boldsymbol{V} \otimes \boldsymbol{\Sigma}$ with $\boldsymbol{V}$ as AR(1) structure are discussed in Roy (2006) for both one population and two population cases.

In this paper, we propose the new classification rules applicable in the case of multivariate repeated measures data under the following assumptions: (1) multivariate normality for classes, (2) Kronecker product structure of the covariance matrix $\boldsymbol{\Omega}$, (3) no structures whatsover are imposed on $\boldsymbol{V}$ and $\boldsymbol{\Sigma}$ except that they are positive definite.

These new general classifiers are presented in Sect. 2. For completness, in Sect. 3, we describe the compound symmetric case for matrix $\boldsymbol{V}$, and in Sect. 4, we describe the AR(1) case for matrix $\boldsymbol{V}$. Section 5 examines the quality of the various classifiers on some real data.

## 2 General classifiers

Suppose that no structures whatsover are assumed on $\boldsymbol{V}$ and $\boldsymbol{\Sigma}$ except that they are positive definite. In this case the classifier has the form

$$g(\boldsymbol{x}) = \arg \max_{1 \le i \le K} \ln(\pi_i f_i(\boldsymbol{x})),$$

where

$$f_i(\boldsymbol{x}) = (2\pi)^{-\frac{pT}{2}} |\boldsymbol{V}_i|^{-\frac{p}{2}} |\boldsymbol{\Sigma}_i|^{-\frac{T}{2}} \exp\left[-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\nu}_i)' \left(\boldsymbol{V}_i^{-1} \otimes \boldsymbol{\Sigma}_i^{-1}\right) (\boldsymbol{x} - \boldsymbol{\nu}_i)\right],$$

and $\pi_i$ is the prior probability that an observation $\boldsymbol{x}$ is from class $i$.

The parameters $\boldsymbol{\nu}_i$, $\boldsymbol{V}_i$ and $\boldsymbol{\Sigma}_i$ are unknown and should be estimated relying on $K$ training samples of sizes $n_1, n_2, \ldots, n_K$ from the respective classes.

Let $\boldsymbol{x}_{ijk}$ $(k = 1, 2, \ldots, T;\ j = 1, 2, \ldots, n_i;\ i = 1, 2, \ldots, K)$ be a $p \times 1$ column vector of measurements on the $j$th individual in the $i$th class at the $k$th time point and

$$\boldsymbol{x}_{ij} = \left(\boldsymbol{x}'_{ij1}, \boldsymbol{x}'_{ij2}, \ldots, \boldsymbol{x}'_{ijT}\right)'.$$

Then $\boldsymbol{x}_{ij}$ is a $pT \times 1$ random observational vector corresponding to the $j$th individual in the $i$th class.

We consider a model described as follows:

$$\text{all observations } \boldsymbol{x}_{ij} \text{ are independent and } \boldsymbol{x}_{ij} \sim N_{pT}(\boldsymbol{\nu}_i, \boldsymbol{V}_i \otimes \boldsymbol{\Sigma}_i), \tag{1}$$

where $\boldsymbol{V}_i$ is a $T \times T$ positive definite matrix and $\boldsymbol{\Sigma}_i$ is a $p \times p$ positive definite matrix, $j = 1, 2, \ldots, n_i, n_i > \max(p, T), i = 1, 2, \ldots, K$.

*Case 1* Matrix $\boldsymbol{V}_i = (v_{rs}^{(i)})$ has all the diagonal elements equal to one. In this case the maximum likelihood estimate of $\boldsymbol{\nu}_i$ is given by

$$\hat{\boldsymbol{\nu}}_i = \bar{\boldsymbol{x}}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} \boldsymbol{x}_{ij}, \quad i = 1, \ldots, K, \tag{2}$$

and we can estimate $\boldsymbol{\Sigma}_i$ by

$$\tilde{\boldsymbol{\Sigma}}_i = \boldsymbol{S}_i = \frac{1}{n_i T}\sum_{j=1}^{n_i}\sum_{k=1}^{T}(\boldsymbol{x}_{ijk} - \bar{\boldsymbol{x}}_{ik})(\boldsymbol{x}_{ijk} - \bar{\boldsymbol{x}}_{ik})', \tag{3}$$

where

$$\bar{\boldsymbol{x}}_{ik} = \frac{1}{n_i}\sum_{j=1}^{n_i} \boldsymbol{x}_{ijk}, \quad i = 1, \ldots, K, \quad k = 1, \ldots, T. \tag{4}$$

Similarly, we can estimate $v_{rs}^{(i)}$, $r \neq s$, by

$$\tilde{v}_{rs}^{(i)} = \frac{1}{n_i p}\sum_{j=1}^{n_i} \operatorname{tr}(\boldsymbol{S}_i^{-1}(\boldsymbol{x}_{ijr} - \bar{\boldsymbol{x}}_{ir})(\boldsymbol{x}_{ijs} - \bar{\boldsymbol{x}}_{is})'), \tag{5}$$

$r, s = 1, \ldots, T, r \neq s, i = 1, \ldots, K$.

**Theorem 1** (Srivastava et al. 2008) *If $\boldsymbol{V}_i = (v_{rs}^{(i)})$ has all the diagonal elements equal to one then $\tilde{v}_{rs}^{(i)}$ defined in (5) is a consistent estimate of $v_{rs}^{(i)}$, $r \neq s$, $r, s = 1, \ldots, T$, and $(n_i/(n_i - 1))\boldsymbol{S}_i$ is an unbiased and consistent estimate of $\boldsymbol{\Sigma}_i$, where $\boldsymbol{S}_i$ is given by* (3).

The form of the obtained classifier is presented in the following theorem.

**Theorem 2** *The classifier based on K training samples of sizes $n_1, n_2, \ldots, n_K$ from the respective classes has the form*

$$\hat{g}_1(\boldsymbol{x}) = \arg \max_{1 \leq i \leq K} \hat{\delta}_{i1}(\boldsymbol{x}), \tag{6}$$

*where*

$$\hat{\delta}_{i1}(\boldsymbol{x}) = -\frac{p}{2} \ln |\tilde{\boldsymbol{V}}_i| - \frac{T}{2} \ln |\tilde{\boldsymbol{\Sigma}}_i| - \frac{1}{2}(\boldsymbol{x} - \bar{\boldsymbol{x}}_i)'(\tilde{\boldsymbol{V}}_i^{-1} \otimes \tilde{\boldsymbol{\Sigma}}_i^{-1})(\boldsymbol{x} - \bar{\boldsymbol{x}}_i) + \ln \hat{\pi}_i, \tag{7}$$

*is the quadratic classification function, $\bar{\boldsymbol{x}}_i$ is given by (2), $\hat{\pi}_i = n_i / \sum_{j=1}^{K} n_j$, and where $\tilde{\boldsymbol{\Sigma}}_i$ is given by (3) and the nondiagonal elements of matrix $\tilde{\boldsymbol{V}}_i$ are given by (5).*

*Case 2* For $\boldsymbol{V}_i = (v_{rs}^{(i)})$, we only assume that $v_{TT}^{(i)} = 1$.

Let

$$\boldsymbol{X}_{ij} = (\boldsymbol{x}_{ij1}, \boldsymbol{x}_{ij2}, \ldots, \boldsymbol{x}_{ijT}), \tag{8}$$

$$\bar{\boldsymbol{X}}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \boldsymbol{X}_{ij}, \tag{9}$$

$$\boldsymbol{X}_{ijc} = \boldsymbol{X}_{ij} - \bar{\boldsymbol{X}}_i, \tag{10}$$

and

$$\boldsymbol{X}_{ijc} = (\boldsymbol{X}_{ijc1} : \boldsymbol{X}_{ijcT}) : (p \times (T-1) : p \times 1), \tag{11}$$

$j = 1, \ldots, n_i, i = 1, \ldots, K.$

In this case the maximum likelihood estimation equations are of the form (Srivastava et al. 2008):

$$\hat{\boldsymbol{\nu}}_i = \bar{\boldsymbol{x}}_i = \text{vec}(\bar{\boldsymbol{X}}_i), \tag{12}$$

$$\begin{aligned} \hat{\boldsymbol{V}}_i &= \frac{1}{n_i p} \begin{bmatrix} \sum_{j=1}^{n_i} \boldsymbol{X}'_{ijc1} \hat{\boldsymbol{\Sigma}}_i^{-1} \boldsymbol{X}_{ijc1} & \sum_{j=1}^{n_i} \boldsymbol{X}'_{ijc1} \hat{\boldsymbol{\Sigma}}_i^{-1} \boldsymbol{X}_{ijcT} \\ \sum_{j=1}^{n_i} \boldsymbol{X}'_{ijcT} \hat{\boldsymbol{\Sigma}}_i^{-1} \boldsymbol{X}_{ijc1} & \sum_{j=1}^{n_i} \boldsymbol{X}'_{ijcT} \hat{\boldsymbol{\Sigma}}_i^{-1} \boldsymbol{X}_{ijcT} \end{bmatrix} \\ &= \frac{1}{n_i p} \sum_{j=1}^{n_i} \boldsymbol{X}'_{ijc} \hat{\boldsymbol{\Sigma}}_i^{-1} \boldsymbol{X}_{ijc}, \end{aligned} \tag{13}$$

where

$$\sum_{j=1}^{n_i} \boldsymbol{X}'_{ijcT} \hat{\boldsymbol{\Sigma}}_i^{-1} \boldsymbol{X}_{ijcT} = n_i p, \tag{14}$$

and

$$\hat{\boldsymbol{\Sigma}}_i = \frac{1}{n_i T} \sum_{j=1}^{n_i} \boldsymbol{X}_{ijc} \hat{\boldsymbol{V}}_i^{-1} \boldsymbol{X}'_{ijc}, \tag{15}$$

respectively, $i = 1, 2, \ldots, K$.

In this case no explicit maximum likelihood estimates of $\boldsymbol{V}_i$ and $\boldsymbol{\Sigma}_i$ are available. The MLEs of $\boldsymbol{V}_i$ and $\boldsymbol{\Sigma}_i$ are obtained by solving simultaneously and iteratively the Eqs. (13) and (15) subject to the condition (14). This is the so called "flip-flop" algorithm.

The results given above are summarized in the following theorem:

**Theorem 3** (Srivastava et al. 2008) *In the model* (1) *with* $v_{TT}^{(i)} = 1$, *if* $n_i > \max(p, T)$ *then the maximum likelihood estimation equations given by* (13) *and* (15) *subject to the condition* (14) *will always converge to the unique maximum.*

The following iterative steps are suggested to get the maximum likelihood estimates of $\boldsymbol{V}_i$ and $\boldsymbol{\Sigma}_i$, $i = 1, 2, \ldots, K$.

**Algorithm 1**

*Step 1* Get the initial covariance matrix $\boldsymbol{\Sigma}_i$ of the form (3), $i = 1, 2, \ldots, K$.

*Step 2* On the basis the initial covariance matrix $\boldsymbol{S}_i$ compute the matrix $\hat{\boldsymbol{V}}_i$ given by (13) and replace all the elements $\hat{v}_{rs}^{(i)}$ by $\hat{v}_{rs}^{(i)} / \hat{v}_{TT}^{(i)}$.

*Step 3* Compute the matrix $\hat{\boldsymbol{\Sigma}}_i$ from the Eq. (15) using $\hat{\boldsymbol{V}}_i$ obtained in step 2.

*Step 4* Repeat steps 2 and 3 until convergence is attained.

We have selected the following stopping rule. Compute two matrices: (a) a matrix of difference between two successive solutions of (13), and (b) a matrix of difference between two successive solutions of (15). Continue the iterations until the maxima of the absolute values of the elements of the matrices in (a) and (b) are smaller than the pre-specified quantities.

The form of the obtained classifier is presented in the following theorem.

**Theorem 4** *The classifier based on* $K$ *training samples of sizes* $n_1, n_2, \ldots, n_K$ *from the respective classes has the form*

$$\hat{g}_2(\boldsymbol{x}) = \arg \max_{1 \le i \le K} \hat{\delta}_{i2}(\boldsymbol{x}), \tag{16}$$

*where*

$$\hat{\delta}_{i2}(\boldsymbol{x}) = -\frac{p}{2} \log |\hat{\boldsymbol{V}}_i| - \frac{T}{2} \log |\hat{\boldsymbol{\Sigma}}_i| - \frac{1}{2} (\boldsymbol{x} - \bar{\boldsymbol{x}}_i)' \left( \hat{\boldsymbol{V}}_i^{-1} \otimes \hat{\boldsymbol{\Sigma}}_i^{-1} \right) (\boldsymbol{x} - \bar{\boldsymbol{x}}_i) + \log \hat{\pi}_i \tag{17}$$

*is the quadratic classification function,* $\bar{\boldsymbol{x}}_i$ *is given by* (2), $\hat{\pi}_i = n_i / \sum_{j=1}^{K} n_j$, *and where* $\hat{\boldsymbol{V}}_i$ *and* $\hat{\boldsymbol{\Sigma}}_i$ *are obtained by solving simultaneously and iteratively the Eqs.* (13) *and* (15) *subject to the condition* (14).

## 3 Classifier for compound symmetry case (Roy and Khattree 2005b)

In repeated measurement designs, one often assume that the correlation matrix $\boldsymbol{V}_i$ has compound symmetry structure. The compound symmetry correlation structure assumes equal correlation among all the repeated measurements:

$$\boldsymbol{V}_i = (1-\varrho_i)\boldsymbol{I}_T + \varrho_i \mathbf{1}_T \mathbf{1}_T', \quad i = 1, 2, \ldots, K. \tag{18}$$

The determinant of $\boldsymbol{V}_i$ is given by

$$|\boldsymbol{V}_i| = (1-\varrho_i)^{T-1}[1 + (T-1)\varrho_i],$$

and the inverse of $\boldsymbol{V}_i$ is given by

$$\boldsymbol{V}_i^{-1} = \frac{1}{1-\varrho_i}\left(\boldsymbol{I}_T - \frac{\varrho_i}{1+(T-1)\varrho_i}\mathbf{1}_T\mathbf{1}_T'\right).$$

Since $\boldsymbol{V}_i$ must be positive definite, we also require that

$$-\frac{1}{T-1} < \varrho_i < 1, \quad i = 1, 2, \ldots, K.$$

Let

$$\boldsymbol{A}_i = \left(a_{rs}^i\right) = \sum_{j=1}^{n_i}(\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_i)(\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_i)' = \begin{pmatrix} \boldsymbol{A}_{11}^i & \boldsymbol{A}_{12}^i & \ldots & \boldsymbol{A}_{1T}^i \\ \boldsymbol{A}_{21}^i & \boldsymbol{A}_{22}^i & \ldots & \boldsymbol{A}_{2T}^i \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{A}_{T1}^i & \boldsymbol{A}_{T2}^i & \ldots & \boldsymbol{A}_{TT}^i \end{pmatrix} \tag{19}$$

is the block matrix containing $T^2$ blocks, where $\boldsymbol{A}_{jk}^i = (\boldsymbol{A}_{kj}^i)'$.

Let

$$\boldsymbol{A}_i^* = \left(a_{rs}^{i*}\right) = \begin{pmatrix} \boldsymbol{A}_{11}^{i*} & \boldsymbol{A}_{12}^{i*} & \ldots & \boldsymbol{A}_{1p}^{i*} \\ \boldsymbol{A}_{21}^{i*} & \boldsymbol{A}_{22}^{i*} & \ldots & \boldsymbol{A}_{2p}^{i*} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{A}_{p1}^{i*} & \boldsymbol{A}_{p2}^{i*} & \ldots & \boldsymbol{A}_{pp}^{i*} \end{pmatrix}, \tag{20}$$

where $\boldsymbol{A}_{jk}^{i*} = (\boldsymbol{A}_{kj}^{i*})'$.

The blocks

$$\boldsymbol{A}_{jk}^{i*} = \left(a_{rs}^{i*jk}\right)$$

are constructed with the elements of the matrix $\boldsymbol{A}_i = \left(a^i_{rs}\right)$:

$$a^{i*jk}_{rs} = a^i_{j+(r-1)p,k+(s-1)p},$$

$r, s = 1, 2, \ldots, T;\ j, k = 1, 2, \ldots, p;\ i = 1, 2, \ldots, K$.
Let

$$\boldsymbol{B}_i = \left(\mathrm{tr}\boldsymbol{A}^i_{jk}\boldsymbol{\Sigma}^{-1}_i\right) \tag{21}$$

and

$$\boldsymbol{C}_i = \left(\mathrm{tr}\boldsymbol{A}^{i*}_{jk}\boldsymbol{V}^{-1}_i\right), \tag{22}$$

$i = 1, 2, \ldots, K$.

The maximum likelihood estimates $\hat{\boldsymbol{\Sigma}}_i$ and $\hat{\varrho}_i$ are obtained by simultaneously and iteratively solving the following Eqs. (23) and (24):

$$\boldsymbol{\Sigma}_i = \frac{1}{Tn_i}\boldsymbol{C}_i \tag{23}$$

$$\frac{T(T-1)^2}{2}pn_i\varrho_i^3 - \left[\frac{T(T-1)(T-2)}{2}pn_i - \frac{(T-1)(T-2)}{2}c_{i0} + (T-1)c_{i3}\right]\varrho_i^2 + \left[(T-1)c_{i0} - \frac{T(T-1)}{2}pn_i\right]\varrho_i - c_{i3} = 0, \tag{24}$$

where

$$c_{i0} = \sum_{k=1}^{T} \mathrm{tr}\left(\boldsymbol{A}^i_{kk}\boldsymbol{\Sigma}^{-1}_i\right), \tag{25}$$

$$c_{i1} = \sum_{\substack{k,l=1\\k<l}}^{T} \mathrm{tr}\left(\boldsymbol{A}^i_{kl}\boldsymbol{\Sigma}^{-1}_i\right), \tag{26}$$

$i = 1, 2, \ldots, K$.

The following iterative steps are suggested to get the maximum likelihood estimators of $\varrho_i$ and $\boldsymbol{\Sigma}_i$, $i = 1, 2, \ldots, K$.

**Algorithm 2**

*Step 1* Compute $\boldsymbol{A}_i$ and $\boldsymbol{A}^*_i$ from the Eqs. (19) and (20), respectively and compute

$$\boldsymbol{H}_i = \frac{1}{p}\sum_{j=1}^{p}\boldsymbol{H}_{ij}, \tag{27}$$

where $\boldsymbol{H}_{ij}$ is the correlation matrix based on the training sample of size $n_i$ from the $i$th class between $T$ time points for the $j$th variable, $i = 1, 2, \ldots, K$, $j = 1, 2, \ldots, p$.

*Step 2* Obtain an initial estimate of $\varrho_i$ as

$$\varrho_i^{(0)} = (\mathbf{1}_T' \boldsymbol{H}_i \mathbf{1}_T - \mathrm{tr}\boldsymbol{H}_i)/(T(T-1)).$$

Thus, $\boldsymbol{V}_i^{(0)}$, an initial estimate of $\boldsymbol{V}_i$, is obtained by replacing $\varrho_i$ by $\varrho_i^{(0)}$ in Eq. (18).

*Step 3* Compute $\boldsymbol{C}_i$ from (22) and next compute $\boldsymbol{\Sigma}_i$ from (23).

*Step 4* Compute $c_{i0}$ and $c_{i1}$ from (25) and (26), respectively, using $\boldsymbol{\Sigma}_i$ obtained in step 3.

*Step 5* Compute the value of $\varrho_i$ by solving the cubic equation (24).

*Step 6* Compute the revised estimate of $\boldsymbol{V}_i$ from $\varrho_i$.

*Step 7* Compute the revised estimate of $\boldsymbol{\Sigma}_i$ from (23) using $\boldsymbol{V}_i$ obtained in step 6.

*Step 8* Repeat steps 4 through 7 until convergence is attained.

## 4 Classifier for AR(1) case (Roy and Khattree 2008)

Suppose the repeated measures are modeled using the first order autoregressive (AR(1)) covariance structure. In this case we write the correlation matrix $\boldsymbol{V}_i$ as

$$\boldsymbol{V}_i = \left(\varrho_i^{|r-s|}\right)_{r,s=1}^{T}, \quad i = 1, 2, \ldots, K. \tag{28}$$

We have

$$|\boldsymbol{V}_i| = \left(1 - \varrho_i^2\right)^{T-1},$$

and

$$\boldsymbol{V}_i^{-1} = \left(1 - \varrho_i^2\right)^{-1}\left(\boldsymbol{I}_T + \varrho_i^2 \boldsymbol{K}_1 - \varrho_i \boldsymbol{K}_2\right),$$

where $\boldsymbol{K}_1 = \mathrm{diag}\left(0, \mathbf{1}_{T-2}', 0\right)$ and $\boldsymbol{K}_2$ is a tridiagonal $T \times T$ matrix with 0 on the diagonal, 1 on the first superdiagonal and on the first subdiagonal.

As $\boldsymbol{V}_i$ is positive definite, we should have

$$-1 < \varrho_i < 1, \quad i = 1, 2, \ldots, K.$$

The maximum likelihood estimates $\hat{\boldsymbol{\Sigma}}_i$ and $\hat{\varrho}_i$ are obtained by simultaneously and iteratively solving the Eqs. (28) and (29):

$$n_i(T-1)p\varrho_i^3 - c_{i2}\varrho_i^2 - [n_i(T-1)p - c_{i0} - c_{i1}]\varrho_i - c_{i2} = 0, \;\; i = 1, 2, \ldots, K, \tag{29}$$

where

$$c_{i0} = \mathrm{tr}\Big[\Big(\boldsymbol{I}_T \otimes \boldsymbol{\Sigma}_i^{-1}\Big)\boldsymbol{A}_i\Big] = \sum_{k=1}^{T} \mathrm{tr}\Big(\boldsymbol{A}_{kk}^i \boldsymbol{\Sigma}_i^{-1}\Big), \tag{30}$$

$$c_{i2} = \mathrm{tr}\Big[\Big(\boldsymbol{K}_1 \otimes \boldsymbol{\Sigma}_i^{-1}\Big)\boldsymbol{A}_i\Big] = \sum_{k=2}^{T-1} \mathrm{tr}\Big(\boldsymbol{A}_{kk}^i \boldsymbol{\Sigma}_i^{-1}\Big), \tag{31}$$

$$c_{i3} = \mathrm{tr}\Big[\Big(\boldsymbol{K}_2 \otimes \boldsymbol{\Sigma}_i^{-1}\Big)\boldsymbol{A}_i\Big] = \sum_{k=1}^{T-1} \mathrm{tr}\Big(\boldsymbol{A}_{k,k+1}^i \boldsymbol{\Sigma}_i^{-1}\Big) \tag{32}$$

with $\boldsymbol{K}_1$, $\boldsymbol{K}_2$ as defined earlier.
The following iterative steps are suggested to get the maximum likelihood estimators of $\varrho_i$ and $\boldsymbol{\Sigma}_i$, $i = 1, 2, \ldots, K$.
**Algorithm 3**

*Step 1* Compute $\boldsymbol{A}_i$ and $\boldsymbol{A}_i^*$ from the Eqs. (19) and (20), respectively, and compute $\boldsymbol{H}_i$ from (27).
*Step 2* Get the average of the first superdiagonal elements of $\boldsymbol{H}_i$, say $\varrho_{1*}^{(i)}$. Then get the average of the second superdiagonal elements of $\boldsymbol{H}_i$, say $\varrho_{2*}^{(i)}$, and so on. The initial estimate of $\varrho_i$ is obtained as

$$\varrho_i^{(0)} = \left[\frac{1}{T-1}\sum_{m=1}^{T-1} \mathrm{sgn}\left(\left(\varrho_{m*}^{(i)}\right)^{(T-1)/m}\right)\left|\left(\varrho_{m*}^{(i)}\right)^{(T-1)/m}\right|\right]^{1/(T-1)},$$

and thus $\boldsymbol{V}_i^{(0)}$, an initial estimate of $\boldsymbol{V}_i$, is obtained by replacing $\varrho_i$ by $\varrho_i^{(0)}$ in Eq. (28).
*Step 3* Compute the matrix $\boldsymbol{C}_i$ from (22) and next compute the matrix $\boldsymbol{\Sigma}_i$ from Eq. (23).
*Step 4* Compute $c_{i0}$, $c_{i2}$ and $c_{i3}$ from (30)–(32), respectively, using $\boldsymbol{\Sigma}_i$ obtained in step 3.
*Step 5* Compute the value of $\varrho_i$ by solving the cubic equation (29).
*Step 6* Compute the revised estimate of $\boldsymbol{V}_i$ from $\varrho_i$.
*Step 7* Compute the revised estimate of $\boldsymbol{\Sigma}_i$ from (23) using $\boldsymbol{V}_i$ obtained in step 6.
*Step 8* Repeat steps 4 through 7 until convergence is attained.

## 5 Examples

To illustrate our classifiers and their effectiveness, we take three real data sets into consideration.

**Table 1** False flax (Camelina sativa L.) measurements

| Line 1 | | | | Line 2 | | | |
|---|---|---|---|---|---|---|---|
| 1st time point | | 2nd time point | | 1st time point | | 2nd time point | |
| Height | # leaves | Height | # leaves | Height | # leaves | Height | # leaves |
| 98 | 59 | 117 | 226 | 74 | 28 | 98 | 56 |
| 85 | 135 | 97 | 344 | 68 | 27 | 95 | 43 |
| 92 | 118 | 112 | 277 | 70 | 28 | 98 | 51 |
| 95 | 51 | 118 | 145 | 74 | 29 | 105 | 56 |
| 90 | 34 | 110 | 86 | 72 | 30 | 105 | 56 |
| 94 | 36 | 118 | 76 | 64 | 29 | 95 | 58 |
| 94 | 34 | 116 | 164 | 65 | 27 | 94 | 48 |
| 95 | 33 | 116 | 82 | 66 | 28 | 93 | 58 |
| 101 | 46 | 121 | 132 | 68 | 31 | 95 | 51 |
| 99 | 38 | 116 | 191 | 75 | 27 | 101 | 58 |

## 5.1 Example 1

Table 1 contains false flax (Camelina sativa L.) measurements. There were two classes (lines) of plants. Each plant was measured at two time points for each of two features: height of plant and number of leaves. The data set was made available by professor Tadeusz Łuczkiewicz from Poznań University of Life Sciences (Department of Genetics and Plant Breeding).

In this example $K = 2$, $p = 2$, $T = 2$, $n_1 = n_2 = 10$.

Multivariate normality with $\boldsymbol{\Omega}_i = \boldsymbol{V}_i \otimes \boldsymbol{\Sigma}_i$, $i = 1, 2$ is assumed.

The maximum likelihood estimates of $\boldsymbol{\nu}_1$ and $\boldsymbol{\nu}_2$ in the two classes are:

$$\hat{\boldsymbol{\nu}}_1 = (94.3\ 58.4\ 114.1\ 172.3)',$$
$$\hat{\boldsymbol{\nu}}_2 = (69.6\ 28.4\ \ 97.9\ \ 53.5)',$$

respectively.

*Case 1* Matrix $\boldsymbol{V}_i$ has all the diagonal elements equal to one.
The estimates of $\boldsymbol{V}_1$, $\boldsymbol{\Sigma}_1$, $\boldsymbol{V}_2$, $\boldsymbol{\Sigma}_2$ in the two classes are:

$$\tilde{\boldsymbol{V}}_1 = \begin{pmatrix} 1.0000 & 0.7493 \\ 0.7493 & 1.0000 \end{pmatrix}, \quad \tilde{\boldsymbol{\Sigma}}_1 = \begin{pmatrix} 30.1500 & -224.4300 \\ -224.4300 & 4168.6000 \end{pmatrix},$$

$$\tilde{\boldsymbol{V}}_2 = \begin{pmatrix} 1.0000 & 0.4256 \\ 0.4256 & 1.0000 \end{pmatrix}, \quad \tilde{\boldsymbol{\Sigma}}_2 = \begin{pmatrix} 15.9650 & 3.6550 \\ 3.6550 & 12.4450 \end{pmatrix},$$

respectively.

Table 2 gives the confusion matrix. We see that all plants were classified correctly.

**Table 2** Confusion matrix for the case in which matrix $\boldsymbol{V}_i$ has all the diagonal elements equal to one (Example 1)

| | Assigned | | Total | % Errors |
|---|---|---|---|---|
| | 1 | 2 | | |
| True = 1 | 10 | 0 | 10 | 0.00 |
| True = 2 | 0 | 10 | 10 | 0.00 |
| Total | 10 | 10 | 20 | 0.00 |

**Table 3** Confusion matrix, where for $\boldsymbol{V}_i$ we only assume that $v_{TT}^{(i)} = 1$ (Example 1)

| | Assigned | | Total | % Errors |
|---|---|---|---|---|
| | 1 | 2 | | |
| True = 1 | 10 | 0 | 10 | 0.00 |
| True = 2 | 0 | 10 | 10 | 0.00 |
| Total | 10 | 10 | 20 | 0.00 |

*Case 2* For $\boldsymbol{V}_i$, we only assume that $v_{TT}^{(i)} = 1$.
The maximum likelihood estimates of $\boldsymbol{V}_1$, $\boldsymbol{\Sigma}_1$, $\boldsymbol{V}_2$, $\boldsymbol{\Sigma}_2$ in the two classes are:

$$\hat{\boldsymbol{V}}_1 = \begin{pmatrix} 0.3067 & 0.4776 \\ 0.4776 & 1.0000 \end{pmatrix}, \quad \hat{\boldsymbol{\Sigma}}_1 = \begin{pmatrix} 50.6250 & -322.9900 \\ -322.9900 & 5621.2000 \end{pmatrix},$$

$$\hat{\boldsymbol{V}}_2 = \begin{pmatrix} 0.3155 & 0.2756 \\ 0.2756 & 1.0000 \end{pmatrix}, \quad \hat{\boldsymbol{\Sigma}}_2 = \begin{pmatrix} 27.0900 & 0.7652 \\ 0.7652 & 16.8920 \end{pmatrix},$$

respectively.

Table 3 gives the confusion matrix. We see that all plants were classified correctly.

*Case 3* The correlation matrix $\boldsymbol{V}_i$ has the compound symmetric structure.
The maximum likelihood estimates of $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ in the two classes are:

$$\hat{\boldsymbol{\Sigma}}_1 = \begin{pmatrix} 27.2 & -257.3 \\ -257.3 & 5361.3 \end{pmatrix}, \quad \hat{\boldsymbol{\Sigma}}_2 = \begin{pmatrix} 12.8552 & 2.5411 \\ 2.5411 & 15.8353 \end{pmatrix}.$$

The maximum likelihood estimates of $\varrho_1$ and $\varrho_2$ are 0.7797 and 0.5160, respectively.

Table 4 gives the confusion matrix. In this model all plants were also classified correctly.

*Case 4* As we are taking into consideration only two time points, structures of the first order autoregressive and the compound symmetry are equivalent (the Eqs. (24) and (29) are equivalent).

**Table 4** Confusion matrix for the compound symmetry correlation structure case (Example 1)

| | Assigned | | Total | % Errors |
|---|---|---|---|---|
| | 1 | 2 | | |
| True = 1 | 10 | 0 | 10 | 0.00 |
| True = 2 | 0 | 10 | 10 | 0.00 |
| Total | 10 | 10 | 20 | 0.00 |

## 5.2 Example 2

SAS Institute Inc (1990, example 9, p. 988) provides data for two responses, Y1 and Y2, measured three times for each subject (at pre, post, and follow-up). Each subject reviewed one of three treatments: A, B, or the control C. In our case $K = 3$, $p = 2$, $T = 3$, $n_1 = n_2 = n_3 = 6$.

Multivariate normality with $\boldsymbol{\Omega}_i = \boldsymbol{V}_i \otimes \boldsymbol{\Sigma}_i, i = 1, 2, 3$ is assumed.

The maximum likelihood estimates of $\boldsymbol{\nu}_1$, $\boldsymbol{\nu}_2$ and $\boldsymbol{\nu}_3$ in the three classes are:

$$\hat{\boldsymbol{\nu}}_1 = (4.5000\ 6.6667\ 11.0000\ 4.8333\ 11.3330\ \ 6.0000)',$$

$$\hat{\boldsymbol{\nu}}_2 = (5.0000\ 6.5000\ 12.0000\ 2.6667\ 12.8330\ 19.8330)',$$

$$\hat{\boldsymbol{\nu}}_3 = (6.3333\ 4.6667\ \ 8.8333\ 4.5000\ 11.8330\ 12.0000)',$$

respectively.

*Case 1* Matrix $\boldsymbol{V}_i$ has all the diagonal elements equal to one.
The estimates of $\boldsymbol{V}_1$, $\boldsymbol{\Sigma}_1$, $\boldsymbol{V}_2$, $\boldsymbol{\Sigma}_2$, $\boldsymbol{V}_3$ and $\boldsymbol{\Sigma}_3$, in the three classes are:

$$\tilde{\boldsymbol{V}}_1 = \begin{pmatrix} 1.0000 & 0.0464 & -0.1416 \\ 0.0464 & 1.0000 & -0.6802 \\ -0.1416 & -0.6802 & 1.0000 \end{pmatrix}, \quad \tilde{\boldsymbol{\Sigma}}_1 = \begin{pmatrix} 9.9352 & 2.2778 \\ 2.2778 & 11.2310 \end{pmatrix},$$

$$\tilde{\boldsymbol{V}}_2 = \begin{pmatrix} 1.0000 & -0.3340 & 0.3410 \\ -0.3340 & 1.0000 & 0.0664 \\ 0.3410 & 0.0664 & 1.0000 \end{pmatrix}, \quad \tilde{\boldsymbol{\Sigma}}_2 = \begin{pmatrix} 7.1574 & 2.9907 \\ 2.9907 & 23.9810 \end{pmatrix},$$

$$\tilde{\boldsymbol{V}}_3 = \begin{pmatrix} 1.0000 & 0.6540 & -0.0470 \\ 0.6540 & 1.0000 & 0.2878 \\ -0.0470 & 0.2878 & 1.0000 \end{pmatrix}, \quad \tilde{\boldsymbol{\Sigma}}_3 = \begin{pmatrix} 9.5000 & -5.8796 \\ -5.8796 & 12.6020 \end{pmatrix},$$

respectively.

Table 5 gives the confusion matrix. In this case 11.11% of subjects was classified incorrectly.

**Table 5** Confusion matrix for the case in which matrix $\boldsymbol{V}_i$ has all the diagonal elements equal to one (Example 2)

| | Assigned | | | Total | % Errors |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | | |
| True = 1 | 6 | 0 | 0 | 6 | 0.00 |
| True = 2 | 1 | 4 | 1 | 6 | 33.33 |
| True = 3 | 0 | 0 | 6 | 6 | 0.00 |
| Total | 7 | 4 | 7 | 18 | 11.11 |

**Table 6** Confusion matrix, where for $\boldsymbol{V}_i$ we only assume that $v_{TT}^{(i)} = 1$ (Example 2)

| | Assigned | | | Total | % Errors |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | | |
| True = 1 | 6 | 0 | 0 | 6 | 0.00 |
| True = 2 | 1 | 5 | 0 | 6 | 16.67 |
| True = 3 | 0 | 0 | 6 | 6 | 0.00 |
| Total | 7 | 5 | 6 | 18 | 5.56 |

*Case 2* For $\boldsymbol{V}_i$, we only assume that $v_{TT}^{(i)} = 1$.
The maximum likelihood estimates of $\boldsymbol{V}_1, \boldsymbol{\Sigma}_1, \boldsymbol{V}_2, \boldsymbol{\Sigma}_2, \boldsymbol{V}_3$ and $\boldsymbol{\Sigma}_3$, in the three classes are:

$$\hat{\boldsymbol{V}}_1 = \begin{pmatrix} 1.2890 & 0.0034 & -0.1846 \\ 0.0034 & 1.5329 & -0.9594 \\ -0.1846 & -0.9594 & 1.0000 \end{pmatrix}, \quad \hat{\boldsymbol{\Sigma}}_1 = \begin{pmatrix} 6.2317 & 1.8963 \\ 1.8963 & 11.8660 \end{pmatrix},$$

$$\hat{\boldsymbol{V}}_2 = \begin{pmatrix} 0.4574 & -0.3297 & 0.2750 \\ -0.3297 & 0.6366 & -0.1510 \\ 0.2750 & -0.1510 & 1.0000 \end{pmatrix}, \quad \hat{\boldsymbol{\Sigma}}_2 = \begin{pmatrix} 12.7220 & 13.5300 \\ 13.5300 & 45.2480 \end{pmatrix},$$

$$\hat{\boldsymbol{V}}_3 = \begin{pmatrix} 1.0775 & 0.5566 & -0.1195 \\ 0.5566 & 0.5561 & 0.2261 \\ -0.1195 & 0.2261 & 1.0000 \end{pmatrix}, \quad \hat{\boldsymbol{\Sigma}}_3 = \begin{pmatrix} 11.0200 & -9.0898 \\ -9.0898 & 17.9780 \end{pmatrix},$$

respectively.

Table 6 gives the confusion matrix. In this case 5.56% of subjects was classified incorrectly.

**Table 7** Confusion matrix for the compound symmetry correlation structure case (Example 2)

| | Assigned | | | Total | % Errors |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | | |
| True = 1 | 6 | 0 | 0 | 6 | 0.00 |
| True = 2 | 0 | 4 | 2 | 6 | 33.33 |
| True = 3 | 0 | 1 | 5 | 6 | 16.67 |
| Total | 6 | 5 | 7 | 18 | 16.67 |

**Table 8** Confusion matrix for the autoregressive of order 1 correlation structure case (Example 2)

| | Assigned | | | Total | % Errors |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | | |
| True = 1 | 5 | 1 | 0 | 6 | 16.67 |
| True = 2 | 0 | 5 | 1 | 6 | 16.67 |
| True = 3 | 0 | 0 | 6 | 6 | 0.00 |
| Total | 5 | 6 | 7 | 18 | 11.11 |

*Case 3* The correlation matrix $\boldsymbol{V}_i$ has the compound symmetric structure. The maximum likelihood estimates of $\boldsymbol{\Sigma}_1$, $\boldsymbol{\Sigma}_2$ and $\boldsymbol{\Sigma}_3$ in the three classes are:

$$\hat{\boldsymbol{\Sigma}}_1 = \begin{pmatrix} 8.8557 & 3.2645 \\ 3.2645 & 13.6191 \end{pmatrix}, \quad \hat{\boldsymbol{\Sigma}}_2 = \begin{pmatrix} 7.2397 & 2.7346 \\ 2.7346 & 23.5208 \end{pmatrix},$$
$$\hat{\boldsymbol{\Sigma}}_3 = \begin{pmatrix} 10.7572 & -5.5036 \\ -5.5036 & 10.9386 \end{pmatrix}.$$

The maximum likelihood estimates of $\varrho_1, \varrho_2$ and $\varrho_3$ are $-0.2897$, $0.0341$ and $0.3409$, respectively.

Table 7 gives the confusion matrix. In this case 16.67% of subjects was classified incorrectly.

*Case 4* The correlation matrix $\boldsymbol{V}_i$ has an autoregressive structure of order 1.

The maximum likelihood estimates of $\boldsymbol{\Sigma}_1$, $\boldsymbol{\Sigma}_2$ and $\boldsymbol{\Sigma}_3$ in the three classes are:

$$\hat{\boldsymbol{\Sigma}}_1 = \begin{pmatrix} 8.9666 & 2.3714 \\ 2.3714 & 12.4522 \end{pmatrix}, \quad \hat{\boldsymbol{\Sigma}}_2 = \begin{pmatrix} 6.9794 & 6.2447 \\ 6.2447 & 30.7660 \end{pmatrix},$$
$$\hat{\boldsymbol{\Sigma}}_3 = \begin{pmatrix} 10.7609 & -6.0574 \\ -6.0574 & 11.8867 \end{pmatrix}.$$

The maximum likelihood estimates of $\varrho_1, \varrho_2$ and $\varrho_3$ are $-0.3073$, $-0.3124$ and $0.5504$, respectively.

Table 8 gives the confusion matrix. In this case 11.11% of subjects was classified incorrectly.

### 5.3 Example 3

A laboratory experiment was set up to investigate the effect of growth of inoculating paspalum grass with a fungal infection applied at four different temperatures (14, 18, 22, 26°C). For each pot of paspalum, measurements were made on three variables:

$x_1$ = the fresh weight of roots (gm),
$x_2$ = the maximum root length (mm),
$x_3$ = the fresh weight of tops (gm).

The inoculated group was compared with a control group and six three-dimensional observations were made on each treatment-temperature combination. These are given in Table 9.10 of Seber (1984) monograph.

The problem here is to classify an unknown subject into one of two groups: a control group and the inoculated group. In our case $K = 2$, $p = 3$, $T = 4$, $n_1 = n_2 = 6$.

Multivariate normality with $\boldsymbol{\Omega}_i = \boldsymbol{V}_i \otimes \boldsymbol{\Sigma}_i$, $i = 1, 2$ is assumed.

The maximum likelihood estimates of $\boldsymbol{\nu}_1$ and $\boldsymbol{\nu}_2$ in the two classes are:

$$\begin{aligned}\hat{\boldsymbol{\nu}}_1 &= (2.7000\ 23.0000\ \ 2.2667\ 15.3670\ 39.8330\ 20.6330\\ &\quad 11.6170\ 38.4170\ 25.5500\ \ 4.5000\ 27.2500\ 16.4170)',\\ \hat{\boldsymbol{\nu}}_2 &= (2.5333\ 22.5830\ \ 2.1000\ \ 8.4667\ 31.5830\ 13.4000\\ &\quad 9.4833\ 30.8330\ 23.4000\ \ 2.9500\ 23.9170\ 11.3830)',\end{aligned}$$

respectively.

*Case 1* Matrix $\boldsymbol{V}_i$ has all the diagonal elements equal to one.
The estimates of $\boldsymbol{V}_1$, $\boldsymbol{V}_2$, $\boldsymbol{\Sigma}_1$, $\boldsymbol{\Sigma}_2$ in the two classes are:

$$\tilde{\boldsymbol{V}}_1 = \begin{pmatrix} 1.0000 & -0.0538 & -0.0852 & 0.0195 \\ -0.0538 & 1.0000 & 0.3595 & 0.0898 \\ -0.0852 & 0.3595 & 1.0000 & -0.2511 \\ 0.0195 & 0.0898 & -0.2511 & 1.0000 \end{pmatrix},$$

$$\tilde{\boldsymbol{V}}_2 = \begin{pmatrix} 1.0000 & -0.0172 & -0.0452 & -0.0429 \\ -0.0172 & 1.0000 & -0.4011 & 0.0206 \\ -0.0452 & -0.4011 & 1.0000 & -0.1783 \\ -0.0429 & 0.0206 & -0.1783 & 1.0000 \end{pmatrix},$$

$$\tilde{\boldsymbol{\Sigma}}_1 = \begin{pmatrix} 6.3534 & 11.3530 & 8.8976 \\ 11.3530 & 45.4760 & 20.1780 \\ 8.8976 & 20.1780 & 18.4100 \end{pmatrix}, \quad \tilde{\boldsymbol{\Sigma}}_2 = \begin{pmatrix} 10.0150 & 6.0649 & 10.5890 \\ 6.0649 & 30.7690 & 8.0830 \\ 10.5890 & 8.0830 & 14.8870 \end{pmatrix},$$

respectively.

Table 9 gives the confusion matrix. We see that all subjects were classified correctly.

**Table 9** Confusion matrix for the case in which matrix $\boldsymbol{V}_i$ has all the diagonal elements equal to one (Example 3)

| | Assigned | | Total | % Errors |
|---|---|---|---|---|
| | 1 | 2 | | |
| True = 1 | 6 | 0 | 6 | 0.00 |
| True = 2 | 0 | 6 | 6 | 0.00 |
| Total | 6 | 6 | 12 | 0.00 |

**Table 10** Confusion matrix, where for $\boldsymbol{V}_i$ we only assume that $v_{TT}^{(i)} = 1$ (Example 3)

| | Assigned | | Total | % Errors |
|---|---|---|---|---|
| | 1 | 2 | | |
| True =1 | 6 | 0 | 6 | 0.00 |
| True= 2 | 0 | 6 | 6 | 0.00 |
| Total | 6 | 6 | 12 | 0.00 |

*Case 2* For $\boldsymbol{V}_i$, we only assume that $v_{TT}^{(i)} = 1$.
The maximum likelihood estimates of $\boldsymbol{V}_1$, $\boldsymbol{V}_2$, $\boldsymbol{\Sigma}_1$, $\boldsymbol{\Sigma}_2$ in the two classes are:

$$\tilde{\boldsymbol{V}}_1 = \begin{pmatrix} 0.1366 & -0.1466 & -0.1005 & 0.0197 \\ -0.1466 & 3.0201 & 0.9259 & 0.1211 \\ -0.1005 & 0.9259 & 2.6829 & -0.3556 \\ 0.0197 & 0.1211 & -0.3556 & 1.0000 \end{pmatrix},$$

$$\tilde{\boldsymbol{V}}_2 = \begin{pmatrix} 0.1433 & -0.0972 & -0.3183 & -0.1372 \\ -0.0972 & 14.0240 & -2.6968 & -0.0739 \\ -0.3183 & -2.6968 & 7.1557 & -0.4372 \\ -0.1372 & -0.0739 & -0.4372 & 1.0000 \end{pmatrix},$$

$$\tilde{\boldsymbol{\Sigma}}_1 = \begin{pmatrix} 3.1882 & 7.2789 & 4.6887 \\ 7.2789 & 33.1720 & 11.8650 \\ 4.6887 & 11.8650 & 10.5800 \end{pmatrix}, \quad \tilde{\boldsymbol{\Sigma}}_2 = \begin{pmatrix} 1.3018 & 1.5964 & 1.8687 \\ 1.5964 & 9.4255 & 3.5066 \\ 1.8687 & 3.5066 & 3.8256 \end{pmatrix},$$

respectively.

Table 10 gives the confusion matrix. We see that all subjects were classified correctly.

*Case 3* The correlation matrix $\boldsymbol{V}_i$ has the compound symmetric structure.

The maximum likelihood estimates of $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ in the two classes are:

$$\hat{\boldsymbol{\Sigma}}_1 = \begin{pmatrix} 6.3393 & 11.3247 & 8.8644 \\ 11.3247 & 45.3195 & 19.9976 \\ 8.8644 & 19.9976 & 18.3408 \end{pmatrix}, \quad \hat{\boldsymbol{\Sigma}}_2 = \begin{pmatrix} 9.7366 & 5.6052 & 9.7523 \\ 5.6052 & 29.6873 & 7.9287 \\ 9.7523 & 7.9287 & 13.8089 \end{pmatrix}.$$

**Table 11** Confusion matrix for the compound symmetry correlation structure case (Example 3)

| | Assigned | | Total | % Errors |
|---|---|---|---|---|
| | 1 | 2 | | |
| True = 1 | 6 | 0 | 6 | 0.00 |
| True = 2 | 1 | 5 | 6 | 16.67 |
| Total | 7 | 5 | 12 | 8.33 |

**Table 12** Confusion matrix for the autoregressive of order 1 correlation structure case (Example 3)

| | Assigned | | Total | % Errors |
|---|---|---|---|---|
| | 1 | 2 | | |
| True = 1 | 6 | 0 | 6 | 0.00 |
| True = 2 | 1 | 5 | 6 | 16.67 |
| Total | 7 | 5 | 12 | 8.33 |

The maximum likelihood estimates of $\varrho_1$ and $\varrho_2$ are 0.0148 and $-0.1293$, respectively.

Table 11 gives the confusion matrix. In this case 8.33% of subjects was classified incorrectly.

*Case 4* The correlation matrix $\boldsymbol{V}_i$ has an autoregressive structure of order 1. The maximum likelihood estimates of $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ in the two classes are:

$$\hat{\boldsymbol{\Sigma}}_1 = \begin{pmatrix} 6.3186 & 11.3008 & 8.8241 \\ 11.3008 & 45.1515 & 19.9586 \\ 8.8241 & 19.9586 & 18.3253 \end{pmatrix}, \quad \hat{\boldsymbol{\Sigma}}_2 = \begin{pmatrix} 9.5656 & 5.6329 & 9.5854 \\ 5.6329 & 29.4990 & 7.4972 \\ 9.5854 & 7.4972 & 13.5604 \end{pmatrix}.$$

The maximum likelihood estimates of $\varrho_1$ and $\varrho_2$ are 0.0169 and $-0.1509$, respectively.

Table 12 gives the confusion matrix. In this case also 8.33% of subjects was classified incorrectly.

**Acknowledgments** The authors wish to thank the two referees for their comments, which helped to improve the presentation of the paper.

## References

Roy A (2006) Testing of Kronecker product structured mean vectors and covariance matrices. J Stat Theory Appl 5(1):53–69

Roy A, Khattree R (2003) Tests for mean and covariance structures relevant in repeated measures based discriminant analysis. J Appl Stat Sci 12(2):91–104

Roy A, Khattree R (2005a) Discrimination and classification with repeated measures data under different covariance structures. Commun Stat Simul Comput 34:167–178

Roy A, Khattree R (2005b) On discrimination and classification with multivariate repeated measures data. J Stat Plan Inf 134:462–485

Roy A, Khattree R (2008) Classification rules for repeated measures data from biomedical research. In: Khattree R, Naik DN (eds) Computational methods in biomedical research. Chapman and Hall/CRC, London, pp 323–370

SAS Institute Inc (1990) SAS procedures guide, Version 6, 3rd edn. SAS Institute Inc, Cary, NC
Seber GAF (1984) Multivariate observations. Wiley, New York
Srivastava MS, von Rosen T, von Rosen D (2008) Models with a Kronecker product covariance structure: estimation and testing. Math Methods Stat 17(4):357–370