

# Significance of the control parameters of a nonparametric linear mapping procedure

Mayer E. Aladjem

*Department of Electrical and Computer Engineering, Ben-Gurion University of the Negev, P.O. Box 653, 84105 Beer-Sheva, Israel*

Received 3 March 1992

## *Abstract*

Aladjem, M.E., Significance of the control parameters of a nonparametric linear mapping procedure, Pattern Recognition Letters 14 (1993) 637-645.

We describe computer-based experiments evaluating the significance of the control parameters of a mapping procedure which is an extension of most of the known linear mapping techniques. The guidelines for control parameter variations are found. By this means the number of trials needed in the search for high quality mapping is reduced to a great extent.

*Keywords.* Mapping of multidimensional data, nonparametric scatter matrices, nearest neighbour technique, measures of association.

## 1. Introduction

In a companion paper (Aladjem (1993)) a method for evaluating the significance of control parameters of mapping procedures was proposed. It is based on a statistical analysis of the mapping results for various data sets. The significance of the control parameters is estimated in terms of a measure of association between values of the control parameters and the data sets. This provides a strategy for objective evaluation, as opposed to the estimation by experience that is typically used.

In this paper we describe an application of this method to a nonparametric linear mapping pro-

cedure previously proposed by the author (see Aladjem (1991a,b)). The results obtained were consistent for different data sets. This allows us to rate the significance of the parameters, to select the appropriate values and to find the guidelines for parameter variations. The required variations are reduced to a great extent which leads to a fast search for a high quality mapping.

## 2. Nonparametric linear mapping procedure

In this section we present a short description of the nonparametric linear mapping method proposed by the author (Aladjem (1991a,b)). This method is oriented to the discrimination of two classes  $\omega_1$  and  $\omega_2$ . It is applied to data after previous normalization

$$x = A^T \hat{x}, \quad (1)$$

where  $A^T \Sigma A = I$  and  $\Sigma$  is the covariance matrix estimated over the original data set  $\hat{x}_j, j=1, 2, \dots, N$ .

*Correspondence to:* M.E. Aladjem, Department of Electrical and Computer Engineering, Ben-Gurion University of the Negev, P.O. Box 653, 84105 Beer-Sheva, Israel.

This work was supported in part by the Israeli Ministry of Science and Technology under contract number 3528, and in part by the Paul Ivanier Center for Robotics and Production Management, Ben Gurion University of the Negev, Israel.

Here,  $I$  denotes the unit matrix. Throughout the rest of this section, all the expressions relate to normalized data.

The mapping is executed by matrix multiplication

$$y = Hx, \quad x \in \mathbb{R}^n, y \in \mathbb{R}^2, \quad n > 2, \quad (2)$$

where

$$H = [r_1 \quad r_2]^T. \quad (3)$$

Here,  $r_1$  and  $r_2$  are the direction vectors obtained by maximizing the class-distance

$$J(r) = (1 - \beta)r^T S_{bk} r + \beta |r^T S_{wk}^{(-)} r|, \quad (4)$$

where  $\beta$  ( $0 \leq \beta \leq 1$ ) is a user-supplied scalar,  $S_{bk}$  is a between-class scatter matrix, and  $S_{wk}^{(-)}$  is the difference between the within-class scatter matrices  $S_{wk1}$  and  $S_{wk2}$  ( $S_{wk}^{(-)} = S_{wk1} - S_{wk2}$  or  $S_{wk2} - S_{wk1}$ ) estimated over the data sets in the normalized space  $x_j \in \omega_1$  for  $j = 1, 2, \dots, N_1$  and  $x_j \in \omega_2$  for  $j = N_1 + 1, N_1 + 2, \dots, N = N_1 + N_2$  with  $r$  being a directional vector. The motivation for the definition of  $J(r)$  is to maximize the distance between the mapped classes ( $\max_r \{(1 - \beta)r^T S_{bk} r\}$ ), while causing one class to be highly scattered and the other to be concentrated ( $\max_r \{\beta |r^T S_{wk}^{(-)} r|\}$ ).

The matrices  $S_{bk}$ ,  $S_{wk1}$  and  $S_{wk2}$ , called nonparametric scatter matrices, are constructed using a nearest neighbor (NN) technique proposed by Fukunaga and Mantock (1983). Their expressions are

$$S_{bk} = \frac{P(\omega_1)}{N_1} \sum_{j=1}^{N_1} u(x_j) [x_j - m_{k2}(x_j)] [x_j - m_{k2}(x_j)]^T + \frac{P(\omega_2)}{N_2} \sum_{j=N_1+1}^{N_1+N_2} u(x_j) [x_j - m_{k1}(x_j)] [x_j - m_{k1}(x_j)]^T, \quad (5)$$

$$S_{wk1} = \frac{1}{N_1} \sum_{j=1}^{N_1} u(x_j) [x_j - m_{k1}(x_j)] [x_j - m_{k1}(x_j)]^T \quad (6)$$

and

$$S_{wk2} = \frac{1}{N_2} \sum_{j=N_1+1}^{N_1+N_2} u(x_j) [x_j - m_{k2}(x_j)] [x_j - m_{k2}(x_j)]^T. \quad (7)$$

Here,  $m_{ki}(x_j)$  is the  $k$ -NN's mean vector from class  $\omega_i$  to sample  $x_j$ ,  $u(x_j)$  is the weighting func-

tion which specifies to what extent the vector  $x_j$  belongs to the Bayes classification boundary, and  $P(\omega_1)$  and  $P(\omega_2)$  are the prior probabilities of the classes  $\omega_1$  and  $\omega_2$ . The expression of the weighting function is

$$u(x_j) = \frac{\min\{d^\alpha(x_j, n_{k1}(x_j)), d^\alpha(x_j, n_{k2}(x_j))\}}{d^\alpha(x_j, n_{k1}(x_j)) + d^\alpha(x_j, n_{k2}(x_j))}, \quad (8)$$

where  $d(x_j, n_{ki}(x_j))$  is the distance from  $x_j$  to its  $k$ -NN's vector  $n_{ki}(x_j)$  from class  $\omega_i$ , and  $\alpha$  is a user-supplied parameter ( $\alpha > 0$ ), which controls the decreasing rate of the weighting function. The value of  $u(x)$  is close to 0.5 for the samples along the classification boundary between the classes  $\omega_1$  and  $\omega_2$ . For samples far from the boundary, the value of  $u(x)$  is close to zero.

It should be pointed out that matrices  $S_{bk}$  (5),  $S_{wk1}$  (6) and  $S_{wk2}$  (7) are extensions of the classical scatter matrices. Substituting the weighting function  $u(x)$  and the local means  $m_{ki}(x)$  by the constant ( $u(x) = 1$ ) and the global means  $m_i$ , respectively, in equations (5), (6) and (7) the nonparametric scatter matrices reduce to classical ones (Fukunaga and Mantock (1983)). The purpose of applying nonparametric scatter matrices is to reflect the local data structure in the case of a significantly non-Gaussian distribution of the samples.

The measure  $J(r)$  is an extension of most of the known scatter measures of class-separation including the classical Fisher measure, Malina's measure (Malina (1981)), the measures for combined mapping (Kissiov and Aladjem (1984), Longstaff (1987)) and finally the nonparametric discriminant measures proposed by Fukunaga and Mantock (1983) and by Huang et al. (1984).

Four control parameters are used in this mapping method, namely  $\beta$  (see equation (4)),  $\alpha$  (see equation (8)) and  $k_b, k_w$ , which are the number  $k$  of the nearest neighbours used for the computation of the between-class scatter matrix  $S_{bk}$  (5) and the number  $k$  of the within-class scatter matrices  $S_{wk1}$  and  $S_{wk2}$  (equations (6) and (7)), respectively. In order to give explicit expression of the mapping criteria (4) we rewrite it in the form

$$J(r, \beta, \alpha, k_b, k_w) = (1 - \beta)r^T S_{bk} r + \beta |r^T S_{wk}^{(-)} r|. \quad (9)$$

From experience the following ranges of the control parameters are defined:

$$0 \leq \beta \leq 1, \quad 1 \leq \alpha \leq 5, \\ 2 \leq k_b \leq 10 \quad \text{and} \quad 2 \leq k_w \leq 10. \quad (10)$$

By means of the mapping the projection subspace is spanned on the directional vectors

$$r_1 = \arg \left\{ \max_r J(r, \beta_1, \alpha_1, k_{b1}, k_{w1}) \right\} \quad (11)$$

and

$$r_2 = \arg \left\{ \max_r J(r, \beta_2, \alpha_2, k_{b2}, k_{w2}) \right\}. \quad (12)$$

They are obtained sequentially by maximizing  $J(r, \beta, \alpha, k_b, k_w)$  (9) through the values  $\beta_1, \alpha_1, k_{b1}, k_{w1}$  and  $\beta_2, \alpha_2, k_{b2}, k_{w2}$  of the control parameters, that lead to a low misclassification error of the sample projections.

In order to increase the significance of the second direction  $r_2$ , it is created on the base of samples which are not well separated along the first direction. For this purpose two different sample distances  $d(\cdot, \cdot)$  are used in the computation of  $m_{ki}(x)$  and  $u(x)$  (8). The criterion  $J(r, \beta_1, \alpha_1, k_{b1}, k_{w1})$  is obtained using the Euclidean distance in  $\mathbb{R}^n$  ( $d(x_i, x_j) = \|x_i - x_j\|$ ). The criterion  $J(r, \beta_2, \alpha_2, k_{b2}, k_{w2})$  is computed using the distance  $d(x_i, x_j) = \|r_1^T x_i - r_1^T x_j\|$ , which is the Euclidean distance in the subspace  $\mathbb{R}^1$  spanned on the first direction  $r_1$ . The application of this distance leads to NN-selection and assignment of large weights  $u(x)$  for the samples near the classification boundary along the first direction  $r_1$ . Consequently, the second direction discriminates between samples from different classes, which are mixed or close to each other along  $r_1$ .

The detailed description of the mapping algorithm has been published previously by the author (Aladjem (1991b)).

### 3. Data sets

Three artificially constructed and two real data sets were used for evaluating the significance of the control parameters of the mapping method described above.

#### 3.1. Artificially constructed data sets

Two data classes were constructed using time sampling waveforms with random parameters. The classes were obtained by corresponding sampling of two signals,  $x_1(t)$  and  $x_2(t)$  of the form

$$x_1(t) = a \exp\{- (t - b)^2 / 2c^2\} \quad (13)$$

and

$$x_2(t) = a \exp\{- |t - b|^\delta / 2c^2\}, \quad (14)$$

with uniform distribution of the random parameters  $a, b$  and  $c$ , and namely

$$0.7 \leq a \leq 1.3, \\ 0.3 \leq b \leq 0.7 \quad \text{and} \\ 0.2 \leq c \leq 0.4. \quad (15)$$

The parameter  $\delta$  ( $1 \leq \delta \leq 2$ ) controls the overlapping of the classes. If  $\delta = 2$ , full covering exists. In the case of  $\delta = 1$  the classes are entirely separated. The samples of each class were represented in an eight-dimensional space by sampling the time signals (13) and (14) at eight uniformly spaced times. The total time interval used for both classes was  $0.0 \leq t \leq 1.05$ . The time signals (13) and (14) have a slightly different form in comparison with those used by Fukunaga and Mantock (1983). The difference consists of introducing the parameter  $\delta$ . Using different values of  $\delta$ , the following data sets were constructed:

- (1) *GI.0*: This data was generated with  $\delta = 1$ . It contains the entirely separated classes.
- (2) *GI.5*: This data set was generated with  $\delta = 1.5$ . It contains the weakly overlapping classes.
- (3) *GI.9*: This data set was generated with  $\delta = 1.9$ . It contains the widely overlapping classes. For each class 50 samples were generated.

#### 3.2. Real data sets

The real data sets concern the medical diagnosis of neurological and cardiological diseases, namely the cerebrovascular accident (*CVA*) and organic heart defects (*HEART*).

- (1) *CVA*: This data set contains pathologically verified *CVA*-cases including a first class of 91 cases with haemorrhages and a second

class of 122 cases with infarction due to ischaemia. Twenty variables were considered as sufficient for the objective description of each disease.

(2) *HEART*: This data set concerns a design problem of a binary tree classifier for cardiological diseases. In the experiment the data available for one of the nonterminal nodes of the designed tree were used. The first class includes 52 cases with healthy control subjects and 34 cases with pulmonary stenosis (a total of 86 cases). The second class consists of 49 cases with mitral stenosis and 40 cases with interatrial and interventricular defects (a total of 89 cases). Sixteen binary variables considered as most important for differentiation of the diseases were chosen.

#### 4. Significance of control parameters

The significance of the control parameters  $\beta$ ,  $\alpha$ ,  $k_b$  and  $k_w$  was analyzed by the technique proposed in a companion paper (Aladjem (1993)). For this purpose the projections into one-dimensional space spanned on  $r_1$  and into two-dimensional space spanned on  $r_1$  and  $r_2$  were run. The experiment was performed by full combination of the following values of the control parameters:

- for directional vector  $r_1$

$$\begin{aligned} \beta &= 0, 0.33, 0.67, 1, \\ \alpha &= 1, 2, 3, 3.7, 5, \\ k_b &= 3, 5, 7, \\ k_w &= 3, 5, 7, 8, 9 \end{aligned} \quad (16)$$

- for directional vector  $r_2$

$$\begin{aligned} \beta &= 0, 0.33, 0.67, 1, \\ \alpha &= 1, 2, 3, 3.7, 5, \\ k_b &= 3, 5, 7, \\ k_w &= 3, 5, 7. \end{aligned} \quad (17)$$

Thus, in all, 240 trials of one-dimensional projections and 144 trials of two-dimensional projections were carried out for each data set.

A priori class information about the data sets was used to estimate the classification error of the projected samples. The leave-one-out-method for

error estimation, based on the 2-NN error counting rule proposed by Fukunaga and Flick (1985) was adopted.

The extent to which data sets affect the control parameters was tested on the basis of 10, 20 and 40 results at low error rates selected from trials available for each data set.

The frequency tables of the findings were computed by cross-classification of the discrete values of control parameters  $\beta$ ,  $\alpha$ ,  $k_b$ , and  $k_w$  ((16) and (17)) and data sets *HEART*, *CVA*, *G1.0*, *G1.5* and *G1.9*. The measures of association described in the companion paper (Aladjem (1993)) were computed, namely:

- Chi square test for independence,
- Cramer's  $V$ ,
- Goodman and Kruskal's  $\lambda$ ,
- Goodman and Kruskal's  $\tau$ .

Two variants of the analysis were carried out. In the first variant all data sets were used. In the second variant the analysis was done separately for artificial (*G1.0*, *G1.5* and *G1.9*) and for real data sets (*HEART* and *CVA*).

In Table 1 the  $P$ -values of the chi square test for independence of the obtained cross-classification tables are summarized.  $P$ -values are the tail probabilities under the hypothesis of independence. The significant dependence between control parameter and data sets is marked by '\*'. The result is assumed as significant if the  $P$ -value is less than a level of significance 0.05. The smallest expected value for each several cross-tabulations is shown in parenthesis under the  $P$ -value. Values less than 5 indicate that the decision for dependence or independence may not be correct. The symbol 'c' indicates that the result is obtained after combining the values of the control parameters.

In Tables 2 and 3 the values of the measures of association  $V$ ,  $\lambda$  and  $\tau$  are shown. Table 2 summarizes the results for directional vector  $r_1$  and Table 3 for directional vector  $r_2$ . The rating of the control parameters is marked with symbols '\*' and '+'. The symbol '\*' indicates the largest value of the measure of association in a particular experiment defined by the data set and number of findings used. The symbol '+' shows the value which follows the biggest one.

All of the computations for cross-classification

Table 1  
*P*-values of the chi square statistic

Number of the cross-tabulated findings per data set	Directional vector	Control parameter	Real data sets	Artificial data sets	All data sets
40	$r_1$	$\beta$	* 0.0017 (11.33)	* 0.0070 c (11.00)	* 0.0000 c (11.14)
		$\alpha$	* 0.0044 (6.00)	0.2949 (7.50)	* 0.0070 (8.14)
		$k_b$	* 0.0000 (7.00)	* 0.0000 (7.75)	* 0.0000 (7.43)
		$k_w$	0.6762 (9.33)	0.7926 (12.00)	0.8209 (9.50)
	$r_2$	$\beta$	* 0.0061 (7.33)	* 0.000 (8.50)	* 0.0000 (9.14)
		$\alpha$	* 0.0290 (8.33)	0.1200 (6.75)	* 0.0189 (8.14)
		$k_b$	* 0.0000 (10.67)	* 0.0000 (11.00)	* 0.0000 (10.86)
		$k_w$	0.3226 (10.00)	0.8938 (12.50)	0.7900 (11.57)
20	$r_1$	$\beta$	* 0.0011 (5.67)	* 0.0232 c (5.25)	* 0.0011 c (5.48)
		$\alpha$	* 0.0084 (4.00)	0.1813 (3.75)	* 0.0063 (4.00)
		$k_b$	* 0.0002 (4.00)	* 0.0000 (3.25)	* 0.0000 (3.57)
		$k_w$	0.8795 (5.00)	0.9448 (6.25)	0.5197 (5.65)
	$r_2$	$\beta$	0.0592 (3.33)	* 0.0000 (3.25)	* 0.0000 (4.29)
		$\alpha$	* 0.0086 (3.67)	0.1578 (2.75)	* 0.0128 (3.86)
		$k_b$	* 0.0000 (3.33)	* 0.0000 (4.75)	* 0.0000 (4.14)
		$k_w$	0.5809 (5.67)	0.7548 (4.75)	0.8128 (5.33)
10	$r_1$	$\beta$	* 0.0002 (2.67)	* 0.0451 c (2.75)	* 0.0007 c (3.14)
		$\alpha$	0.1205 c (2.67)	0.4559 c (2.00)	* 0.0259 (2.00)
		$k_b$	* 0.0040 (1.67)	* 0.0054 (2.00)	* 0.0006 (1.43)
		$k_w$	0.3551 (2.33)	0.8731 (2.75)	0.5202 (2.50)
	$r_2$	$\beta$	* 0.0035 (2.00)	* 0.0001 (1.75)	* 0.0000 (2.00)
		$\alpha$	0.0559 (2.33)	* 0.0147 (2.00)	* 0.012 (2.14)
		$k_b$	* 0.0038 c (3.33)	* 0.0000 (2.75)	* 0.0000 (2.00)

Table 1 (contd.)

*P*-values of the chi square statistic

Number of the cross-tabulated findings per data set	Directional vector	Control parameter	Real data sets	Artificial data sets	All data sets
10	$r_2$	$k_w$	0.3574 (1.67)	0.2780 (1.50)	0.2884 (1.57)

\* The *P*-value indicates significant dependence between control parameter and data sets; the level of significance is 0.05.

c Indicates results after collapsing two discrete values of control parameters in order to gain bigger minimum expected value.

() The values in parenthesis are minimum expected values.

Table 2

Values of measures of association for directional vector  $r_1$ 

Number of the cross-tabulated findings per data set	Data sets	Control parameters	Values of measures of association		
			$V$	$\lambda$	$\tau$
40	Real	$\beta$	0.268	+0.167	+0.076
		$\alpha$	+0.280	0.085	0.048
		$k_b$	* 0.439	* 0.388	* 0.211
	Artificial	$\beta$	+0.217	+0.080	+0.057
		$\alpha$	0.149	0.062	0.020
		$k_b$	* 0.353	* 0.165	* 0.145
	All data	$\beta$	+0.233	+0.151	+0.058
		$\alpha$	0.207	0.095	0.041
		$k_b$	* 0.404	* 0.267	* 0.188
20	Real	$\beta$	+0.330	+0.176	+0.171
		$\alpha$	0.299	0.143	0.099
		$k_b$	* 0.427	* 0.267	* 0.194
	Artificial	$\beta$	+0.330	+0.260	+0.122
		$\alpha$	0.229	0.121	0.026
		$k_b$	* 0.451	* 0.364	* 0.247
	All data	$\beta$	+0.366	+0.226	+0.146
		$\alpha$	0.294	0.129	0.085
		$k_b$	* 0.444	* 0.324	* 0.228
10	Real	$\beta$	* 0.603	* 0.368	* 0.323
		$\alpha$	0.344	+0.222	0.128
		$k_b$	+0.506	0.167	+0.223
	Artificial	$\beta$	* 0.401	+0.320	+0.176
		$\alpha$	0.181	0.080	0.055
		$k_b$	+0.368	* 0.381	* 0.265
	All data	$\beta$	+0.494	* 0.356	+0.245
		$\alpha$	0.430	0.229	0.184
		$k_b$	* 0.496	+0.303	* 0.262

\* The largest value of the measure of association in a particular experiment defined by the number of findings and the data set.

+ The value which follows the biggest one.

analysis were performed using the 4F procedure of the BMDP-computer programs (Dixon (1987)).

In our previous study (Aladjem (1991a)) the minimal error rates of the projected samples for each data set had been reported. The values of the control parameters corresponding to them were given as well. The best mappings were shown and a comparative study with some known from the literature of mapping procedures was done. We will use these results in the conclusions made in the following section.

### 5. Discussion and concluding remarks

The main results of this application can be summarized as follow:

(1)  $P$ -values of the chi square statistic (Table 1) define  $k_w$  as nonsignificant control parameter for all variants of the study (various data sets: real, artificial and real + artificial, and various number of cross-tabulated findings: 40, 20 and 10). The parameters  $k_b$  and  $\beta$  are assigned to be significant for all variants excluding only one case—20 cross-

Table 3  
Values of measures of association for  $r_2$

Number of the cross-tabulated findings per data set	Data sets	Control parameters	Values of measures of association		
			$V$	$\lambda$	$\tau$
40	Real	$\beta$	+0.274	0.071	+0.045
		$\alpha$	0.242	+0.108	0.040
		$k_b$	* 0.393	* 0.203	* 0.173
	Artificial	$\beta$	* 0.404	* 0.210	* 0.175
		$\alpha$	0.171	0.045	0.026
		$k_b$	+0.361	+0.170	+0.132
	All data	$\beta$	+0.354	+0.202	+0.130
		$\alpha$	0.197	0.104	0.039
		$k_b$	* 0.391	* 0.202	* 0.157
20	Real	$\beta$	0.318	0.140	0.057
		$\alpha$	+0.379	+0.167	+0.096
		$k_b$	* 0.488	* 0.250	* 0.301
	Artificial	$\beta$	+0.515	* 0.370	* 0.299
		$\alpha$	0.234	0.074	0.046
		$k_b$	* 0.536	+0.367	+0.293
	All data	$\beta$	+0.450	+0.304	+0.206
		$\alpha$	0.284	0.115	0.078
		$k_b$	* 0.547	* 0.329	* 0.316
10	Real	$\beta$	+0.569	+0.350	+0.205
		$\alpha$	0.453	0.190	0.134
		$k_b$	* 0.700	* 0.600	* 0.490
	Artificial	$\beta$	+0.524	+0.407	+0.319
		$\alpha$	0.414	0.250	0.180
		$k_b$	* 0.625	* 0.478	* 0.388
	All data	$\beta$	+0.505	+0.396	+0.276
		$\alpha$	0.402	0.240	0.162
		$k_b$	* 0.614	* 0.424	* 0.393

• The largest value of the measure of association in a particular experiment defined by the number of findings and the data set.  
+ The value which follows the biggest one.

Table 4  
Guidelines for control parameter variations

Directional vector	Control parameter	Range of variations	Number of variations	An example of parameter variations
$r_1$	$\beta$	$0 \leq \beta \leq 0.7$	8	0, 0.1, 0.2, ..., 0.7
	$\alpha$	$2 \leq \alpha \leq 4$	1	3
	$k_b$	$3 \leq k_b \leq 7$	3	3, 5, 7
	$k_w$	$5 \leq k_w \leq 8$	1	6
$r_2$	$\beta$	$0.3 \leq \beta \leq 1$	8	0.3, 0.4, 0.5, ..., 1
	$\alpha$	$3 \leq \alpha \leq 5$	1	4
	$k_b$	$3 \leq k_b \leq 5$	3	3, 4, 5
	$k_w$	$3 \leq k_w \leq 5$	1	4

tabulated findings for directional vector  $r_2$  using only real data sets. The control parameter  $\alpha$  is significant in most experiments with minimum expected values greater than 4.

Therefore the decision for the significance of the control parameters is consistent in the various combinations of the data sets. This is the reason for generalizing that the control parameter  $k_w$  is nonsignificant for the mapping quality.

(2) The results in Tables 2 and 3 show that the rating of the significance of  $\beta$ ,  $\alpha$  and  $k_b$  can be assumed consistent in the various data sets (real, artificial and all data together). The control parameter  $k_b$  corresponds to the biggest values of  $V$ ,  $\lambda$  and  $\tau$ . The parameters  $\beta$  and  $\alpha$  correspond to the similar values of  $V$ ,  $\lambda$  and  $\tau$ .  $\beta$  can be seen to slightly prevail over  $\alpha$ . Therefore the most important control parameter is  $k_b$  followed by  $\beta$  and  $\alpha$ .

(3) Based on the analysis of the values of the control parameters for the mappings with high quality the ranges of the control parameters were found. They are summarized in Table 4.

(4) The required number of the control parameter variations are shown in Table 4. They are assigned to be proportional to the significance of the control parameters. The nonsignificant parameter  $k_w$  and low significance parameter  $\alpha$  take constant values in the combinations. The most significant parameter  $k_b$  is varied in three levels. The parameter  $\beta$  is permitted a great number of variations. This choice is connected with the computational complexity. The variations of  $k_b$ ,  $k_w$  and  $\alpha$  need recomputation of the scatter matrices

(5), (6) and (7) which significantly influences the computation time of the mapping procedure. The variations of  $\beta$  do not require complicated computations (see Aladjem (1991b)).

The guidelines summarised in Table 4 were used in a binary tree classifier design concerning cardiological diseases (see Aladjem (1994)). Following them, high quality mapping at each node of the tree was found by a small number of trials. The values of the control parameters were found to be around the guidelines defined. This confirms the effectiveness of the technique for evaluating the significance of the control parameters of mapping procedures. It can be viewed as a powerful tool for reducing the number of the trials in mapping experiments.

## References

- Aladjem, M.E. (1991a). Parametric and nonparametric linear mappings of multidimensional data. *Pattern Recognition* 24, 543-553.
- Aladjem, M.E. (1991b). PNM: A program for parametric and nonparametric mapping of multidimensional data. *Computers in Biology and Medicine* 21, 321-343.
- Aladjem, M.E. (1993). A statistical technique for evaluating the significance of control parameters of mapping procedures. *Pattern Recognition Lett.* 14, 631-636.
- Aladjem, M.E. (1994). Multiclass discriminant mappings. *Signal Process.* 35 (1), in press.
- Dixon, W.J. (Ed.) (1987). *BMDP Statistical Software 1987*. University of California Press, Berkeley, CA.
- Fukunaga, K. and J.M. Mantock (1983). Nonparametric discrimination analysis. *IEEE Trans. Pattern Anal. Machine Intell.* 5, 671-678.

- Fukunaga, K. and T.E. Flick (1985). The 2-NN rule for more accurate NN risk estimation. *IEEE Trans. Pattern Anal. Machine Intell.* 7, 107-111.
- Huan Zhen-hua, Li Ming-hong and N. Laveen (1984). A non-parametric feature extraction algorithm. *Proc. Internat. Conf. Syst. Man Cybern.* 1, 591-595.
- Kissiov, V.T. and M.E. Aladjem (1984). Experimental study of combined linear mapping method. *Proc. Seventh Internat. Conf. on Pattern Recognition II*, 696-698.
- Longstaff, I.D. (1987). On extensions to Fisher's linear discriminant function. *IEEE Trans. Pattern Anal. Machine Intell.* 9, 321-325.
- Malina, W. (1981). On an extended Fisher criterion for feature selection. *IEEE Trans. Pattern Anal. Machine Intell.* 3, 611-614.