ELSEVIER

Chemometrics and Intelligent Laboratory Systems 32 (1996) 151–164

Chemometrics and intelligent laboratory systems

Adaptive resonance theory based neural network for supervised chemical pattern recognition (FuzzyARTMAP) Part 1: Theory and network properties

Dietrich Wienke *, Lutgarde Buydens

Catholic University of Nijmegen, Laboratory for Analytical Chemistry, Toernooiveld 1, 6525 ED Nijmegen, Netherlands

Received 14 January 1995; accepted 24 July 1995

Abstract

The FuzzyARTMAP algorithm is studied with respect to its usefulness for supervised chemical pattern recognition. The theory of this relatively complex artificial neural classifier is presented in detail for chemists. An instructive data set of moderate size, describing male and female participants in courses of chemometrics by their body measures, is used to demonstrate how FuzzyARTMAP works and what its basic properties are.

Keywords: Artificial neural networks; Adaptive resonance theory (ART); Fuzzy set theory; Pattern recognition

1. Introduction

Adaptive resonance theory based artificial neural networks (ART) form a particular family of pattern recognition methods. They were introduced by Grossberg [1,2] to overcome the so-called stability–plasticity dilemma [3]. This dilemma describes the general difficulty for each classifier to be able to distinguish between two types of new knowledge: new knowledge that fits into the existing classifier structure and knowledge, that does not fit. The first type of new knowledge only requires a certain plasticity of the classifier, for example an adaptation of some fitting coefficients (network weights). The second type of new knowledge requires that the classifier not only adapts some weights but its structure to the new classification situation.

ART-1, ART-2, ART-2a, ART-3 and FuzzyART, for example, are methods for unsupervised grouping of n pattern vectors $\mathbf{X}$ (dimension $n \times m$) of length m into c clusters. These ART methods can be compared with respect of some algorithmic properties, for example, with the sequential leader algorithm [4]. Differences among the four ART methods exist in the type of processable input data (binary, real), data preprocessing (type of scaling, type of transfer function), speed of training and the type of similarity measure (Euclidian angle based, Fuzzy set theory based) [5,6]. The $l = 1...c$ clusters generated during the training, are described by $l = 1...c$ individual directed vectors $\mathbf{w}$ (dimension $1 \times m$) forming together a so-called weights matrix $\mathbf{W}$ of dimension $c \times m$. Caudell presented the use of ART-1 in combination

* Corresponding author. email: wienke@sci.kun.nl

SSDI 0169-7439(95)00084-4

with parallel operating opto-electronic multichannel detector arrays for commercial and military applications [7]. Optical, opto-electronical and electronical hardware implementations of ART-1 and ART-2 were reported by Kane and Paquin [8], Wunsch et al. [9,10] and Ho et al. [11]. Wienke and Kateman used ART-1 to classify UV/VIS and IR spectra and showed the quantitative chemical interpretability of the ART weights [12]. Lin and Wang fitted industrial time series data and classified the parameter vectors by ART-2 [13]. Whitley and Davis [14,15] proposed to use ART-2 for monitoring and control of chemical reactors by real-time interpretation of sensor data taken from the reactor. Wienke et al. [16,17] applied ART-2a in on-line monitoring of the environment by a classification of airborne particles using their scanning electron microscopy image data and their X-ray fluorescence emission pattern. Another application of ART-2 to pattern recognition with image data was reported recently by Resch and Szabo [18]. In Ref. [19] body measures of female and male persons were clustered by ART-2a. A comparison of ART-2a versus the principal component analysis based SIMCA classifier and backpropagation multilayer neural networks for rapid sorting of post-consumer plastics by remote NIR spectroscopy has been recently given [20]. Reported advantages for those four unsupervised ART methods compared to other types of classifiers are short training times (usually less than 20 epochs), built-in detector for outliers and for extrapolations, the quantitative interpretability of weights and the on-line applicability of ART in chemical apparatus.

A second group of ART algorithms (ARTMAP and FuzzyARTMAP) has been recently proposed by Carpenter et al. for supervised pattern recognition [21,22]. Their aim was to overcome the less powerful discrimination power of the former pure unsupervised working ART methods in case of very closely located clusters in the m-dimensional features space. Supervised learning by ARTMAP (resp. FuzzyARTMAP) means a controlled clustering of the **X**-space that is monitored by a corresponding desired output in a second space **Y**. Note, that in contrast to other supervised working classifiers (discriminant analysis, SIMCA, PLS, multilayer feedforward neural networks, etc.) the number, c, of clusters in the **X**-space and the number, b, of clusters in the **Y**-space in case of ARTMAP (resp. FuzzyARTMAP) are not priorly fixed. The correct number of clusters (groups of similar data vectors) will thus be formed during training. The network structure itself (number of required neurons) forms thus an important result of the training process and makes the design of the network structure less subjective due to different users. A further consequence of this approach is, that several distinct and distant clusters in the input space **X** can be related to one single (identical) desired output cluster. ARTMAP is restricted to binary input data making it less interesting for chemical applications. Therefore the present work concentrates on FuzzyARTMAP as the most advanced ART algorithm with no restriction with respect to the type of input data. It has been decided to split this study in two parts because of the complexity of the FuzzyARTMAP algorithm that includes aspects from adaptive resonance theory as well as from fuzzy set theory. By means of a moderate sized data set **X** of $n = 47$ human beings characterized by their individual $m = 5$ body measures and by their sex y ($p = 1$), the theory and the properties of FuzzyARTMAP are discussed in Part 1. The chosen data set allows a quantitative insight in the classification behavior of FuzzyARTMAP and a detailed presentation of the meaning of so-called network weights. Other moderate sized chemical data sets are presently under consideration of ART neural networks [28]. Part 2 of the present study [26] deals with the implementation of FuzzyARTMAP in a remote working near-infrared multisensor system for rapid on-line sorting of post-consumer plastics.

2. Theory and FuzzyARTMAP algorithm

The theory of FuzzyARTMAP follows the general ART paradigms developed by Grossberg [1,2]. According to these assumptions, a classifier should always first check if a currently offered input vector $\mathbf{x}_i$ fits into the actual model limits in the variables space covert by a set of training samples. If this check succeeds, the current input $\mathbf{x}_i$ 'resonates' with the ART network. Then the classifier model can be updated by upgrading corresponding coefficients (more specific: network weights). On the other hand, if the check of the current input fails, another action then fitting of

weights is required. Because this input deviates significant from the calibration space covered by the current classifier model, the structure of the classifier has to be extended towards the new area where the current deviating input sample is located. Structure adaptation during training plays thus in ART models an even important role as weights adaptation.

Technically considered, the FuzzyARTMAP algorithm consists of two independently operating, self-organizing FuzzyART neural networks (Fig. 1). One of them clusters the $i = 1...n$ pattern vectors in their m-dimensional **X**-space, the other one in their p-dimensional **Y**-space. m is the number of variables in the **X**-space and p the number of variables in the **Y**-space. A third algorithmic element, called 'mapfield', monitors and controls simultaneously both clustering processes.

2.1. Initial settings

Before training, two so-called weight matrices, $\mathbf{W}^x$ (dimension $m \times cmax$) and $\mathbf{W}^y$, (dimension $p \times bmax$) need to be initialized, whereby $cmax$ and $bmax$ are large constants, corresponding to the maximum expected number of clusters in the data sets (maximum is n). Usually the elements of $\mathbf{W}^x$ and $\mathbf{W}^y$ are initialized by $\mathbf{w}^x_{j,l} = 1/\sqrt{m}$ (resp. $\mathbf{w}^y_{k,h} = 1/\sqrt{p}$), whereby $j = 1...m$ and $k = 1...p$. The third weight matrix in a FuzzyARTMAP neural network, the so-called mapfield, $\mathbf{W}^{xy}$ (dimension $cmax \times bmax$), has to be initialized by zeros.

Training parameters such like learning rates η^x, $\eta^{x,y}$ and η^y and vigilance parameters $\rho^{x,max}$, $\rho^{x,y,max}$ and $\rho^{y,max}$ for all three networks have to be defined in advance (between 0 and 1, mostly around 0.1). A necessary scaling constant, $0 < \alpha < 1$ has also to be set.

2.2. Network training

After this initialization, the training is done by random offer of corresponding pairs $[\mathbf{x}_i, \mathbf{y}_i]$ of training pattern vectors to the initialized network. $\mathbf{x}_i$ is applied to the input of the network in the **X**-space and $\mathbf{y}_i$ to the input of the corresponding network in the **Y**-space. In the very beginning, when no cluster exists, $\mathbf{x}_i$ and $\mathbf{y}_i$ are copied as initial estimate of the first weight vectors $\mathbf{w}^x_1$ (resp. $\mathbf{w}^y_1$). Later on, a pairwise comparison is done between the current inputs, $\mathbf{x}_i$ and $\mathbf{y}_i$, with all active $l = 1...c$ weight vectors $\mathbf{W}^x$ (resp. $h = 1...b$ in $\mathbf{W}^y$) by calculating the fuzzy distance d^x_l (resp. d^y_h) for dissimilarity

$$d^x_{l,i} = |\mathbf{x}_i \wedge \mathbf{w}^x_l| / (\alpha + |\mathbf{w}^x_l|) \quad (1)$$

between the current inputs and all individual weight vectors with $|\mathbf{z}|$ = length of a vector $\mathbf{z}$. The fuzzy set theory based operator '$\wedge$' (intersection) provides a resulting vector $\mathbf{z} = min(\mathbf{x}_i, \mathbf{w}^x)$ (resp. $\mathbf{z} = min(\mathbf{y}_i, \mathbf{w}^y)$) which has as elements the corresponding minimum element values of the two considered vectors. Geometrically considered, this new resulting vector is always spatially located in between both vectors. Then, for each of both networks, a weight vector, $\mathbf{w}^x_{winner}$ (resp. $\mathbf{w}^y_{winner}$) can always be found having a minimum distance (highest similarity) to the present input vector. This weight vector is called winner. In the following step, called in ART 'resonance check', with the formula

$$\rho^x_i = |\mathbf{x}_i \wedge \mathbf{w}^x_{winner}| / |\mathbf{x}_i| \quad (2)$$

it is calculated if the present inputs are inside or outside winner's existing cluster. If the calculated distances are smaller then the so-called 'vigilance' parameters $\rho^{x,max}$ (resp. $\rho^{y,max}$) according to

$$\rho^x_i < \rho^{x,\max} \quad (3)$$

then the present input vector is inside an existing cluster (high similarity), or in other words, the input came into resonance with the network. If not, the secondly placed winner according (1) is checked by (2)–(3) and so on. In the following, 'network learning' will happen. In contrast to conventional neural networks, this step is called in ART 'adaptation'. Adaptation means more than simply fitting a fixed number of coefficients (network weights). Adaptation can also mean in ART an extension of the current network structure by additional weight vectors.

For sequential fitting its weights $\mathbf{W}^x$, FuzzyARTMAP uses the learning rule

$$\mathbf{w}^{new,x}_{winner} = \eta^x(\mathbf{x} \wedge \mathbf{w}^{old,x}_{winner}) + (1 - \eta^x)\mathbf{w}^{old,x}_{winner} \quad (4)$$

(corresponding formulae for fitting both other weight matrices $\mathbf{W}^y$ and $\mathbf{W}^{x,y}$ have simultaneously to be applied).

Expressed in classical terms of data fitting, the learning rates η are the step sizes for moving a current winning weight vector $\mathbf{w}^{x}_{winner}$ in its spatial direction towards the currently offered input vector $\mathbf{x}_i$ (resp. $\mathbf{w}^{y}_{winner}$ towards $\mathbf{y}_i$). For the influence of the size of η's on the training of an ART network, consult, for example, Ref. [12].

Additionally, the element in the lth row and the hth column in the mapfield $\mathbf{W}^{xy}$, corresponding to the two separate winners in both networks, is set to one in the learning step. Alternative ways of mapfield learning are the use of the binary coded output vector $\mathbf{y}^{out}$ of the Y-network or of the real-coded vector of $\mathbf{d}^{x}_{l,i}$ values (length c) as inputs to the mapfield. The mapfield as a real weight matrix is then trained by the corresponding training rule similar to Eq. (4).

If condition (3) is not fulfilled, the current input vectors are outside all currently existing clusters. In such a case FuzzyARTMAP will give a 'reset' signal and it will create a new additional cluster. In case of the network in the X-space, the number of formed clusters will grow from $l = 1...c$ to $l = 1...c + 1$ by increasing the dimension of the weight matrix $\mathbf{W}^{x}$ by one. The same happens for $\mathbf{W}^{y}$ with $h = 1...b$ weight vectors if a reset is given in the Y-space. That input vector that has caused a reset will directly be copied into the new formed weight vector. In this way a new additional cluster is always be initialized by its first offered pattern vector.

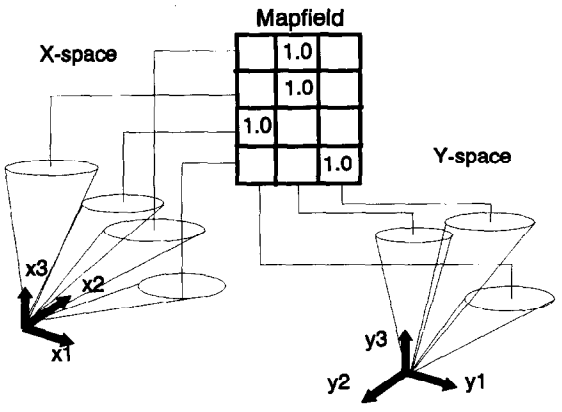

Fig. 1. The mapfield completes two FuzzyART neural networks to a FuzzyARTMAP network. The mapfield monitors and controls both simultaneous running clustering processes (network training). The number and the size of formed clusters depends for each single network from the size of the prior chosen vigilance parameter (see also Fig. 2).

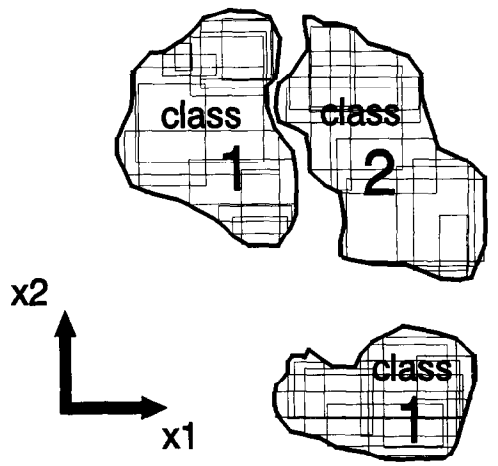

Fig. 2. Complement coding (see text) in combination with fuzzy distance measures provides hyperrectangular shaped clusters. Each cluster box grows during training to its maximum size. In this way, arbitrary shaped clouds of experimental data points are approximated by overlapping hyperrectangulars. FuzzyARTMAP's mapfield allows a link of very distant clusters in the X-space to one common cluster in the Y-space. On the other hand, it allows the dedication of spatially close located clusters in X to distinct clusters in Y.

The mapfield $\mathbf{W}^{xy}$ increases its dimensions simultaneously by closely following the cluster numbers c and b of $\mathbf{W}^{x}$ and $\mathbf{W}^{y}$. Another network reset happens, if a winner in the X-space does not correspond to a winner in the Y-space. This happens mostly for very close located clusters in variables space if they belong to different desired outputs. In such a case a reset is given by the mapfield, because the mapfield is that part of the entire network that covers the information which cluster in the X-space is linked to which cluster in the Y-space (Fig. 1). Two independent running neural networks that interact via a mapfield allow the link of distant clusters from X-space to an identical desired output cluster in the Y-space. As an example, NIR spectra of several grades of high density and low density polyethylene polymers ('PE') form distant clusters in the spectral space X. However, via the mapfield these distant clusters can be linked together to the same desired output 'PE'. On the other hand, close located clusters can be precisely separated (Fig. 2).

Table 1
Anonymized experimental data matrices **Y** and **X** of five body measures of 12 female (f) and 35 male (m) participants in training courses in chemometrics (for details, see Experimental)

Participant identifier	y = sex	x_1 = body length (cm)	x_2 = body weight (kg)	x_3 = shoe size (/a.u.)	x_4 = belly outline (cm)	x_5 = neck (cm)
m1	0	185	91	44.5	103	42
m2	0	196	98	43.5	100	39
m3	0	186	75	45.5	90	39
m4	0	169	59	41	83	38
f5	1	160	57	37	74	34
m6	0	178	72	44	85	37
f7	1	162	55	38	70	31
m8	0	183	65	41	75	34
m9	0	177	66.5	41	90	37
f10	1	170	63	39	72	35
m11	0	160	57	39	81	35
m12	0	190	80	46	83	36
m13	0	172	82	42	93	41
m14	0	183	70	43	91	38
m15	0	181	71	42	91	39
m16	0	185	78	42	89	39
m17	0	183	77	44	105	41
m18	0	183	90	42	106	46
m19	0	180	80	41	58	41
m20	0	187	81	44	87	41
f21	1	163	60	38	76	34
f22	1	164	55	39	76	31
m23	0	170	74	41	94	39
f24	1	170	75	37	99	37
f25	1	173	56	38.5	65	32
m26	0	168	75	41	99	39
m27	0	176	72	41.5	85	38
m28	0	172	73	41	92	41
m29	0	180	83	43	89	38
m30	0	172	82	42	100	41
f31	1	162	55	39	68	32
m32	0	186	105	45	115	43
m33	0	187	90	43	93	41
f34	1	165	55	37.5	76	33
m35	0	181	78	45	99	40
m36	0	188	72	43.5	77	33
m37	0	185	81	44.5	95	40
m38	0	188	75	44.5	87	37
m39	0	187	73	43	85	39
m40	0	178	84	43	98	41
f41	1	173	61	38	70	33
m42	0	189	75	45	87	38
m43	0	177	70	41	89	37
m44	0	181	76	42	87	39
f45	1	170	70	39	81	35
f46	1	170	57	39	70	35
m47	0	186	85	42	87	38
minimum	0	160	55	37	58	31
maximum	1	196	105	46	115	46

2.3. Complement coding as a special preprocessing procedure

Eqs. (1)–(2) require, that the elements of the input vectors $\mathbf{x}_i$ (resp. $\mathbf{y}_i$) are positive numbers in the range 0...1. This can be reached, for example, by data preprocessing using range scaling. A second condition has to be fulfilled. The input vectors have to be additionally 'complement coded'. This means that $\mathbf{x}_i$ is doubled in its length by appending its complement vector $1 - \mathbf{x}_i$. $\mathbf{y}_i$ is complement coded by its corresponding vector $1 - \mathbf{y}_i$. Calculation examples for complement coding are given, for example, in Refs. [27,28]. From an algebraic point of view, complement coding adds redundant information only. From the point of view of fuzzy set theory, a single crisp feature in a data vector is substituted by a function of this feature (here: a simple range between a lower and an upper feature limit). Thus, calculations (1)–(2) deal with functions of features instead of single crisp features. For two or more features, in this way, hyperrectangular shaped boxes for each cluster (Fig. 2) are obtained instead of single directed vectors. The boxes grow during training (network weights adaptation) from a small point vector to a large hyperrectangular. The mathematical proof for this effect can be found in Carpenter et al. [21,22]. Thus, an experimental data cluster will be approximated by FuzzyARTMAP by a sequence of overlapping hyperrectangulars. The outer limits of such a hyperrectangular are called in the present work 'upper fuzzy bounds' and 'lower fuzzy bounds'. Another effect of complement coding is, that it gives 'absent features', having low data values, even high contributions in the fuzzy similarity measures (1)–(2) as 'present features' that usually have higher values.

2.4. Qualification versus quantification

Each cluster formed by a FuzzyARTMAP artificial neural network describes a limited local region in the variables space. In this way, complicated shaped clouds of experimental data are approximated, in principle, by a sequence of overlapping hyperrectangulars. This is true for the data cloud in the **X**-space and in the **Y**-space. This approach of locally overlapping partial clusters the FuzzyARTMAP network is rather useful for supervised pattern recognition (qualification) than for function fitting (quantification). It has been shown [21–24] in studies with simulated and experimental data sets, that FuzzyARTMAP can successfully approximate data clouds that are located within each other and data clouds that are formed from combined spirals. This gives an impression of its discrimination power, for example, compared with MLF-BP neural networks [29]. However, the more a data set forms a continuous function, the more new single weight vectors are needed by FuzzyARTMAP. Function fitting is thus possible, but the costs one has to pay for this are a rapid proliferation of an enormous number of new single clusters.

3. Experimental

During the international COMETT Course in Chemometrics in Heyen (The Netherlands) that was organised in 1991 by the Laboratory of Analytical Chemistry of the Catholic University of Nijmegen the present author asked the participants for their individual body measures 'body length', 'body weight', 'foot size', 'belly' and 'neck'. The aim was to get an experimental data set that could be used in the exercises of multivariate data analysis and multivariate statistics. The same experiment has been repeated once more again during a chemometrics course for chemistry teachers at the technical college Utrecht (The Netherlands) in 1991. Both data sets were anonymized and merged providing a complete data matrix **X** (size 47×5) of $n = 47$ persons, characterized by $m = 5$ features, and their corresponding sex **Y** (dimension 47×1) as desired $p = 1$-dimensional output (Table 1).

4. Computations and software

A program 'FuzzyARTMAP.c' has been written by the author in ANSII-C computer language for UNIX computers (SUN) for the gcc-compiler. The alternative WINDOWS-version, now implemented in the master PC of an POLYTEC XDAP near-infrared diode array spectrometer, will be described in Part 2 of this publication series.

5. Results and discussion

First, the 47×5 data matrix (Table 1) has been analyzed in a classical way using multivariate statistical methods. This helps in understanding the following results provided by the FuzzyARTMAP neural network.

5.1. An introducing multivariate-statistical study

The present data set (Table 1) has been recently studied by hierarchical cluster analysis and principal component analysis [19]. The dendrograms for two different clustering methods of the columnwise autoscaled 47×5 data matrix showed some single outliers (m8, m18, m19, m32, m36). A group with a high contribution of 10 females and one single male participant (m11) has been found. Another large cluster concerns all remaining 34 male participants including two females f24 and f45 that were left. This large group splits into three subclusters whereby the classification result differs slightly due to Ward's and due to the average linkage method. The main result from hierarchical cluster analysis was, that obviously the five chosen body features provide combined information about differences between men and women. However, some few participants (m11, f24 and f45) were misclassified. Comparing their data with those ones from men in general makes these misclassifications understandable (Table 1). Principal component analysis over the 5×5 correlation matrix of **X** provided a distribution of the percentage contributions of the total variance over the five factors into 74.4%, 15.0%, 5.0%, 3.8% and 1.8%. Multivariate correlations were found between the five body measures. An example for this is that a high body weight can be caused by a large body length or/and a large belly. Thus, small men with a big belly can get an even high weight as tall but slender women. On the other hand, this does not necessary mean that a correlation between belly outline and body length exists. Fig. 3 summarizes quickly the results from the multivariate statistical study. So the reader will get somewhat feeling for the data before application of FuzzyARTMAP. The linear mapping result in Fig. 3 looks similar to the already published PCA scores plots [19]. The position of the 47 participants scores are in agreement with the cited clustering results. The two

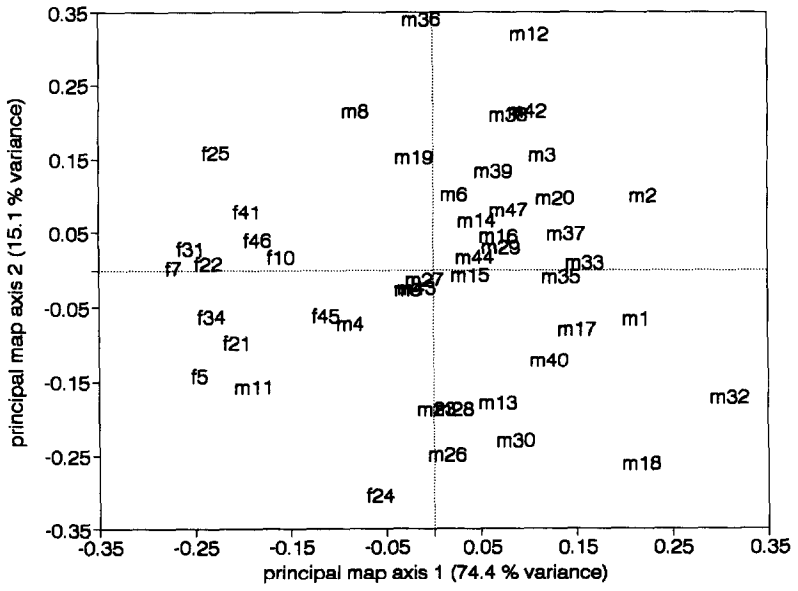

Fig. 3. Presentation of the two first principal axis obtained from a multidimensional scaling (MDS, Torgerson's method [30]) of data from Table 1. Female participants (f) are found in the left handed lower region that is mainly determined by low values in 'body size', 'body weight' and 'shoe measure'. Male participants (m) mostly can be found in the right handed region characterized by higher values in 'body length', 'belly', 'neck' and 'weight'.

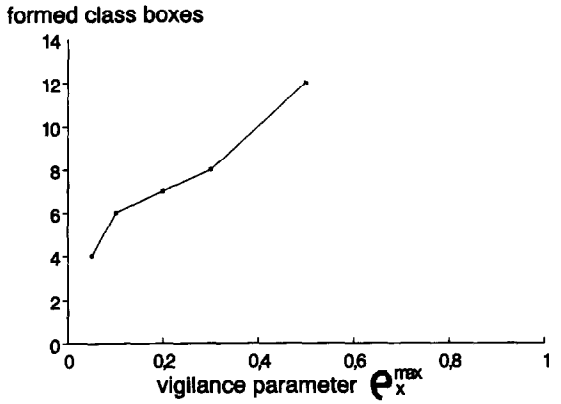

Fig. 4. Presentation of the growing resolution of a data set into finer subclusters by an increase of the vigilance parameter (here: ρ_x^{max}) in the X-space. The more compact the data are, the less they scatter. In such a case a change of ρ does not have nearly any effect (here: between 0.1 and 0.5). The more the data scatter or if they form a continuous function in place of a cluster the more sensitive the resolution reacts on small changes in ρ (here: 0.5–1.0).

large clusters of male and female participants occupy distinct areas in the space. Females show lower values in all five features compared to males. However, there are some exceptions and there is some overlap. Only combinations of several PC provided a perfect and distant separation of men and women. The reason is that special combinations of PC characterized the participants that took part in the courses of chemometrics coming from all over Europe. A few tall and gangling northern european women, for example, could compete with selected small european men according the combination of the three variable combination 'length + weight + foot size', for example. The closeness of the males m11, m4, m23 and m26 to the females group proofs this. On the other hand, men's combined measure ratios in belly outline and neck size allowed to distinguish them from female participants.

5.2. FuzzyARTMAP study

FuzzyARTMAP can be used to study data sets in three distinct directions. First, training runs with distinct chosen sizes of $\rho^{x,max}$, $\rho^{xy,max}$ and $\rho^{y,max}$ resolve the **X** and the **Y** matrix into distinct numbers of clusters and subclusters. In this way the raw structure and the fine structure of the data cloud can be explored. Second, after such a training, the obtained network weights can be decoded and rescaled again back to the original variables space for quantitative

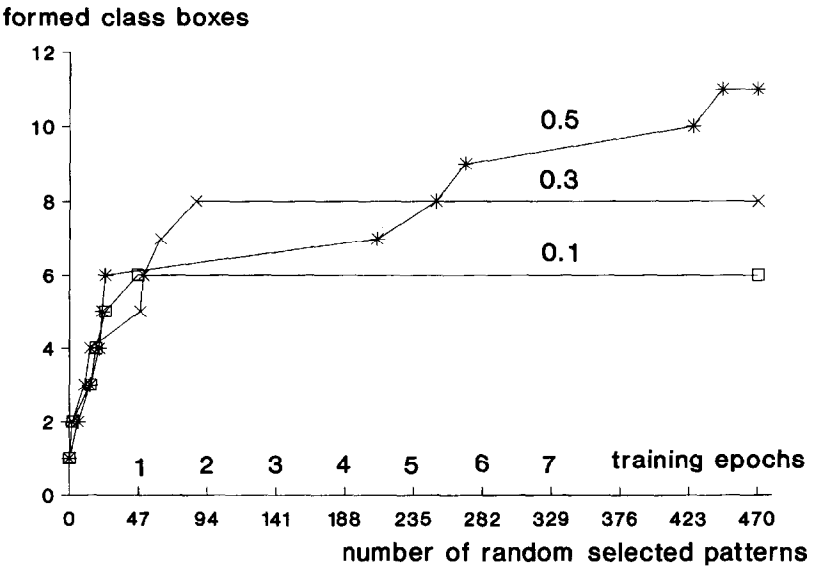

Fig. 5. Presentation of the formation of the final number of clusters as function of the number of training epochs for three distinct sizes of the vigilance parameter $\rho_x^{max} = 0.1$, 0.3, 0.5. In general, fast convergence of the training process can be observed for a FuzzyARTMAP network within a few epochs. The higher the desired cluster resolution is the more time the convergence takes (see also Fig. 6).

interpretation of the network structure. This helps to trace the nature of the formed clusters and it helps to understand why they were formed. Third, a trained FuzzyARTMAP network can be used to classify unknown pattern vectors.

Fig. 4 shows for increasing vigilance parameter $\rho^{x,max}$ (thus for decreasing cluster size), how the network needs additional single clusters (hyperrectangulars) to approximate the entire data cloud. This effect is also well known from the unsupervised working ART neural networks. Between 0.05 and 0.5 for the $\rho^{x,max}$ value the number of formed clusters does not change so much. That simply means that the more compact and focused to a single point a data cluster is, the less important the influence of the chosen size of $\rho^{x,max}$ is. On the other hand, it also indicates a disadvantage of all ART networks. The more a data cloud scatters (from noisy or wrong features) the more the FuzzyARTMAP network will tend to proliferation of single new clusters.

The next important property, that has been studied here, was the speed of the convergence of the training process. Carpenter et al. mentioned in their ART-2a study [5] the enormous gain in training speed by ART if it is compared, for example, with a multi-layer feedforward backpropagation neural network. Even for large data sets with many thousands of pattern vectors [21,22,24,25] a remarkable small number of training epochs were reported. For FuzzyARTMAP a significant speed has been observed, too, in the present work (Fig. 5). It can be seen, that FuzzyARTMAP converges for the training data from

Table 2
Weights matrix $\mathbf{W}^{xy}$ (mapfield) links seven clusters of the five-dimensional **X**-space (body measures of participants) to two clusters in the one-dimensional **Y**-space (sex of participants). This particular cluster structure was obtained for the data from Table 1 with following settings for the FuzzyARTMAP algorithm: learning rates: $\eta^x = 0.1$, $\eta^{x,y} = 0.1$, $\eta^y = 0.1$; vigilance parameters: $\rho^{x,max} = 0.2$, $\rho^{xy,max} = 0.6$, $\rho^{y,max} = 0.9$; scaling factor $\alpha = 0.01$. The network has been trained by 3 epochs

Cluster k/l	$\mathbf{Y}_{out,1}$	$\mathbf{Y}_{out,2}$
$\mathbf{X}_{out,1}$	1.0	0.0
$\mathbf{X}_{out,2}$	0.0	1.0
$\mathbf{X}_{out,3}$	1.0	0.0
$\mathbf{X}_{out,4}$	1.0	0.0
$\mathbf{X}_{out,5}$	0.0	1.0
$\mathbf{X}_{out,6}$	0.0	1.0
$\mathbf{X}_{out,7}$	1.0	0.0

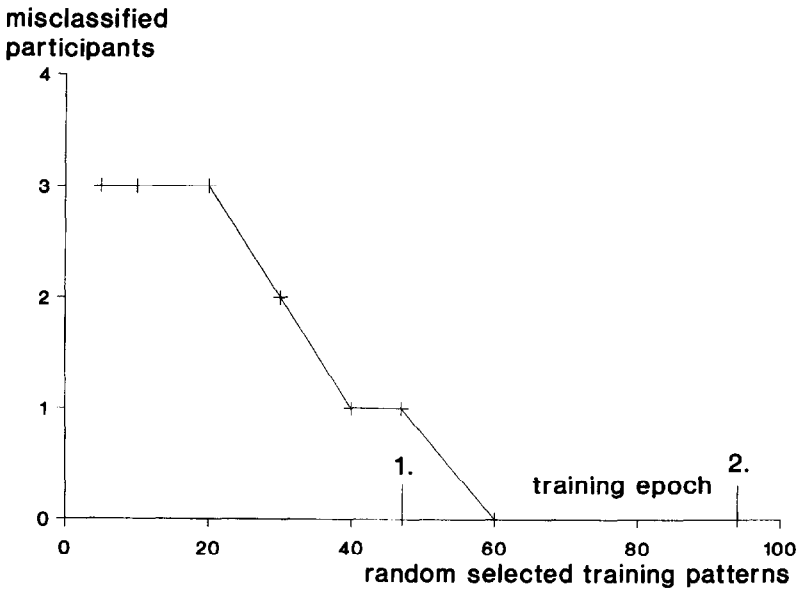

Fig. 6. Convergence of FuzzyARTMAP's training process shown by a presentation of the reclassification error (data set, Table 1) as function of the number of training epochs. In general, the reclassification error rapidly decreases within the first few epochs (here after less than two epochs).

Table 3
Weights matrix $\mathbf{W}^x$ describes seven rectangular shaped clusters among the $n = 47$ X pattern (Table 1). Each cluster is limited by the lower and upper 'fuzzy bounds' of the five contributing variables. $\mathbf{W}^x$ has been obtained after training the FuzzyARTMAP network with range scaled features (Table 1) using the network parameter given in Table 2. The sex identifier 'M' = male and 'F' = female are available via the mapfield (Table 2)

Scaled feature	Type of cluster						
	M1	F1	M2	M3	F2	F3	M4
lower fuzzy bounds							
length (a.u.)	0.30	0.00	0.51	0.00	0.07	0.27	0.25
weight (a.u.)	0.33	0.01	0.20	0.04	0.00	0.15	0.08
shoes (a.u.)	0.47	0.00	0.44	0.22	0.16	0.13	0.44
belly (a.u.)	0.47	0.24	0.15	0.40	0.19	0.28	0.43
neck (a.u.)	0.40	0.14	0.16	0.26	0.00	0.26	0.46
upper fuzzy bounds							
length (a.u.)	0.86	0.20	0.70	0.00	0.22	0.27	0.25
weight (a.u.)	0.87	0.10	0.36	0.04	0.00	0.34	0.08
shoes (a.u.)	0.93	0.08	0.57	0.22	0.22	0.22	0.44
belly (a.u.)	0.89	0.32	0.48	0.40	0.31	0.53	0.43
neck (a.u.)	0.86	0.22	0.46	0.26	0.04	0.32	0.46

Table 1 usually in less than three training epochs. With respect to a correct reclassification (Fig. 6) the convergence has been usually already reached after less than two epochs. The formation of the final fine-tuned cluster structure needed usually some additional training epochs (Fig. 5). This effect is caused by a few single patterns that are located at the border line of a formed cluster. Such single patterns have the same statistical chance to be randomly selected like the other ones that form together the center of a multi-pattern cluster. In this way the position of a cluster covering many of the training pattern is much faster formed than the position of a cluster with only a single pattern. The main clusters covering the majority of all training pattern were mostly formed within less than two training epochs. This is also almost independent from the chosen size of ρ and the number n of data set size (see Part 2 [26]). The split-

Table 4
Same weights matrix $\mathbf{W}^x$ from Table 3, but decoded and rescaled to original features space

Feature	Type of cluster						
	M1	F1	M2	M3	F2	F3	M4
lower fuzzy bounds							
length (cm)	171.12	160.00	178.40	160.00	162.62	170.00	169.00
weight (kg)	71.84	55.95	65.00	57.00	55.00	62.93	59.00
shoes (a.u.)	41.24	37.00	41.00	39.00	38.50	38.18	41.00
belly (cm)	85.13	71.82	67.03	81.00	68.91	74.06	83.00
neck (cm)	37.09	33.19	33.53	35.00	31.00	35.00	38.00
upper fuzzy bounds							
length (cm)	190.96	167.53	185.34	160.00	168.21	170.00	169.00
weight (kg)	98.66	60.26	73.22	57.00	55.46	72.04	59.00
shoes (a.u.)	45.43	37.80	42.19	39.00	39.00	39.00	41.00
belly (cm)	108.99	76.68	85.65	81.00	76.00	88.37	83.00
neck (cm)	43.92	34.39	37.95	35.00	31.71	35.81	38.00

Table 5
The rescaled and decoded weights matrix $\mathbf{W}^y$ describes two clusters among the $n = 47$ **Y** pattern (Table 1). Theoretically, each cluster should be limited by the lower and upper 'fuzzy bounds' of the two contributing variables. The binary type (0, 1) of the variable 'sex' causes a fuzzy range of zero. $\mathbf{W}^y$ has been obtained after training the FuzzyARTMAP network with range scaled features (Table 1) using the network parameter given in Table 2. The sex identifier 'm' = male and 'f' = female are available via the mapfield (Table 2)

Input feature	Clusters	
	M1	F1
lower fuzzy bounds		
m	1.000	0.000
f	0.000	1.000
upper fuzzy bounds		
m	1.000	0.000
f	0.000	1.000

ting of the remaining clusters with only single patterns in them takes a longer time based on their lower statistical formation chance.

After finishing the training, the weights of the FuzzyARTMAP network provide an insight into the structure of the obtained clusters (Tables 2–7). For such an insight one has to concentrate on the weight matrix of the mapfield (Table 2), on the weight matrix of the **X**-space (Tables 3 and 4) and on the weight matrix of the **Y**-space (Table 5).

For an understanding of the mapfield one can make use from the scheme shown in Fig. 1. The mapfield has been trained with binary output vectors of the **Y**-space network. This provides binary weights in the mapfield. Another way would be a use of the vector of all ρ_i^y values as input to the mapfield. This would provide real numbers in the mapfield.

Table 7
Similarity between the two **Y**-clusters given as vector angle between individual column vectors of weights matrix $\mathbf{W}^y$ (from Table 5)

Cluster index	Cluster index	
	m	f
m	1.000	0.000
f	–	1.000

The fuzzy bounds (Table 3) for the weights in the **X**-space reflect the individual range of a feature within a formed cluster. Lower and upper fuzzy bounds are obtained after decoding the complement coded weights. However, these weights have to be rescaled per feature, too, by using the scaling parameters of the original data (Table 1, below). After rescaling weights were obtained (Table 4) that can directly be interpreted in terms of the original data. As an example (Tables 3 and 4) one can see that FuzzyARTMAP formed the main groups 'small heavy man' (M1) and 'tall slender men' (M2). M3 and M4 were single patterns forming 'small men groups'. Women were classified into 'small sized women' (F1, F2) and 'large women' (F3). A similarity consideration between the formed clusters in the **X**-space (Table 6) proofs in a quantitative way the similarity between small men (M1, M4) and larger women (F3), in particular with respect to body length, shoe size and nek. In a similar way the cluster, formed in the **Y**-space, can be discussed (Table 7). However, for this particular binary coded **Y**-data set (Tables 1

Table 6
Similarity between the seven **X**-clusters given as cosine of the vector angle between individual column vectors of weights matrix $\mathbf{W}^x$ in Table 3 (note: only lower fuzzy bound taken into account)

Cluster index	Cluster index						
	M1	F1	M2	M3	F2	F3	M4
M1	1.000	0.704	0.859	0.865	0.802	0.941	0.954
F1	–	1.000	0.317	0.908	0.619	0.748	0.755
M2	–	–	1.000	0.544	0.722	0.833	0.801
M3	–	–	–	1.000	0.811	0.794	0.919
F2	–	–	–	–	1.000	0.700	0.811
F3	–	–	–	–	–	1.000	0.909
M4	–	–	–	–	–	–	1.000

and 5) the cluster results are obvious and not that much exciting. However, for future processing of spectral, environmental-analytical or process-analytical **Y**-data this consideration can provide interesting results, too.

5.3. Verification of the trained FuzzyARTMAP network

A test data set (Table 8) has been used to verify the trained network. This set contains body measures of as well male, female adults as of children. The children group (between 6 and 7 years old) forms with its specific body measures an extrapolation region in the features space, because the training set (Table 1) covered only data from adults. The test run has been performed in two different ways. First a validation run was done, for which the network has been undertrained with the training data from Table 1 using a growing number of training epochs. For each undertrained network the test adults (from Table 8) were predicted (Fig. 7). Similar to the decrease of the classification error it can be observed that the prediction error also rapidly decreases within a few training epochs (compare Figs. 6 and 7). Finally, all unknown adults have been correctly predicted by FuzzyARTMAP according their sex (male, female) based on their body measures if at least 2–3 training

Table 8
Anonymized test data set $\mathbf{X}^{test}$ of five body measures of female (f) and male (m) adults and of children (c) used for validation of trained FuzzyARTMAP neural network

Person	x_1 = body length (cm)	x_2 = body weight (kg)	x_3 = shoe size (a.u.)	x_4 = belly outline (cm)	x_5 = neck (cm)
f13	172	65	41	74	33
f14	160	68	36	88	38
m36	183	80	45	85	43
c1	121	24	32	54	25
f15	166	64	39	80	33
m37	186	80	43	93	42
m38	175	66	40	86	42
m39	183	75	43	79	38
m40	189	87	45	86	40
f16	158	51	36	74	40
m41	183	70	43	78	36
m42	193	72	45	80	38
m43	177	67	42	75	38
m44	170	89	43	102	41
m45	180	78	42	89	35
m46	197	84	45	88	36
m47	194	87	46	86	38
f17	173	63	39	69	32
f18	162	65	39	76	34
f19	170	70	38	79	36
f20	170	61	38	70	29
f21	171	61	40	71	30
f22	171	62	40	72	35
c2	114	22	29	47	22
c3	113	23	30	49	24
c4	124	28	33	59	26
c5	114	24	30	54	27
c6	120	27	31	55	26
c7	120	23	29	51	23
c8	116	24	30	51	25
c9	121	26	31	56	25
c10	113	20	28	46	20

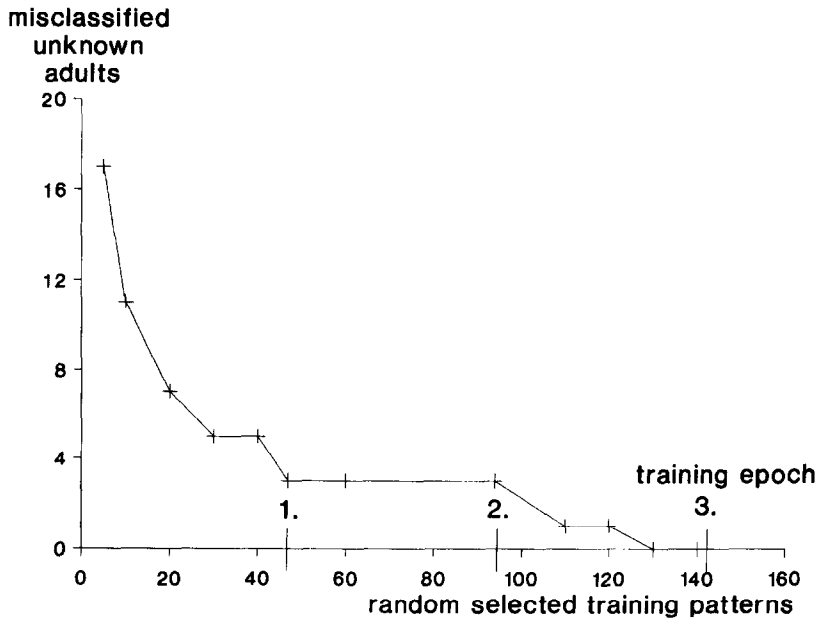

Fig. 7. Convergence of FuzzyARTMAP's training process, presented by the size of prediction error for a test data set (Table 8) as a function of the number of epochs for a training data set (Table 1).

epochs were applied. For this test run it was the network not allowed to go on with learning by modifying any weights of existing clusters (from training).

A second experiment has been performed with the test data set (Table 8) including now also the children's data. Now the network it was allowed to keep learning its weights as well its structure. Based on the random order of test vectors the weight vectors of the existing cluster (of adults) did only oscillate a little bit around their previous position. However, the first child's pattern vector from the test data set (Table 8) that was offered to the network, could not be classified into any existing adults's cluster. The network gave a reset and created in the **X**-space an additional new cluster and asked to declare a corresponding **Y**-pattern for this unknown. All following test vectors for children were classified then correctly by FuzzyARTMAP in this common new cluster.

6. Conclusions

FuzzyARTMAP is an alternative neural classifier for supervised pattern recognition with remarkable short training time and with full interpretability of its weights in terms of original variables. These two properties make the classifier interesting for quantitative pattern recognition studies. The built-in detector against outliers and extrapolations make FuzzyARTMAP also interesting for robust and save working technical applications in analytical-chemical process control and environmental monitoring tasks.

Acknowledgements

The authors thank all participants and teachers in the International COMETT Course in Chemometrics in Heyen (The Netherlands) in 1991 for their enthusiastic contribution to the experimental data processed in this work. Special thanks to Dr. Sijmen de Jong (Unilever Vlaardingen) and PhD student Eduard Derks (University of Nijmegen) for their excellent cooperation with the author in this experiment and in the course of multivariate statistics. Further we like to express thanks to all participants in the chemometrics course for chemistry teachers at the Technical College of Utrecht in 1991 for their exciting participation in the collection of data. Thanks to family members, friends, students and children for

their personal contribution to the test data set. Further we are grateful to both referees for their valuable hints.

References

[1] S. Grossberg, Biol. Cybernetics, 23 (1976) 121–134.
[2] S. Grossberg, Biol. Cybernetics, 23 (1976) 187–203.
[3] S. Grossberg, Studies of Mind and Brain, Reidel, Dordrecht, 1982.
[4] J.A. Hartigan, Clustering Algorithms, Wiley, New York, 1975.
[5] G.A. Carpenter, S. Grossberg and D.B. Rosen, Neural Networks, 4 (1991) 493–504.
[6] G.A. Carpenter, S. Grossberg and D.B. Rosen, Neural Networks, 4 (1991) 759–771.
[7] T.P. Caudell, Appl. Optics, 31 (1992) 6220–6229.
[8] J.S. Kane and M.J. Paquin, IEEE Trans. Neural Networks, 4 (1993) 695–702.
[9] D.C. Wunsch, D.J. Morris, R.L. McGann and T.P. Caudell, Appl. Optics, 32 (1993) 1399–1407.
[10] D.C. Wunsch, T.P. Caudell, C.D. Capps, R.J. Marks and R.A. Falk, IEEE Trans. Neural Networks, 4 (1993) 673–684.
[11] C.S. Ho, J.J. Liou, M. Georgiopoulos, G.L. Heileman and C. Christodoulou, Int. J. Electron., 76 (1994) 271–291.
[12] D. Wienke and G. Kateman, Chemom. Intell. Lab. Syst., 23 (1994) 309–329.
[13] C.C. Lin and H.P. Wang, Comput. Ind., 22 (1993) 143–152.
[14] J.R. Whitley and J.F. Davis, Comput. Chem. Eng., 18 (1994) 637–661.
[15] J.R. Whitley and J.F. Davis, IEEE Expert, 4 (1993) 54–63.
[16] D. Wienke, Y. Xie and P.K. Hopke, Chemom. Intell. Lab. Syst., 25 (1994) 367–387.
[17] Y. Xie, P.K. Hopke and D. Wienke, Environ. Sci. Technol., 28 (1994) 1921–1928.
[18] C. Resch and Z. Szabo, J. Nucl. Med., 5(35) (1994) 182–182 (abstract).
[19] D. Wienke, in: L. Buydens and W. Melssen (Eds.), Chemometrics: Exploring and Exploiting Chemical Information, University Press University of Nijmegen, Nijmegen, 1994.
[20] D. Wienke, W. van den Broek, R. Feldhoff, T. Huth-Fehre, T. Kantimm, L. Quick, W. Melssen, F. Winter, K. Cammann and L. Buydens, Anal. Chim. Acta, in press.
[21] G.A. Carpenter, S. Grossberg and J.H. Reynolds, Neural Networks, 4 (1991) 565–588.
[22] G.A. Carpenter, S. Grossberg, N. Markuzon, J.H. Reynolds and D.B. Rosen, IEEE Trans. Neural Networks, 3 (1992) 698–713.
[23] R. Hohenstein, Classification of Neuro-Magnetic Field Patterns Using a Fuzzy ARTMAP Network, Master Student Research Report, Institute Numerical and Instrumental Math/Informatics, Wilhelms-University Münster, 1994.
[24] P. Willems, The ART Neural Network Models Enlightend: Implementation on Sequential and Parallel Computer Systems, Master Student Research Report, Institute for Informatics, Catholic University of Nijmegen, 1994.
[25] D. Wienke, L. Buydens and K. Cammann, Fuzzy Adaptive Resonance Theory Based Neural Network for Supervised Chemical Pattern Recognition (FuzzyARTMAP), presented at International Chemometrics Meeting, Veldhoven, 3–5 July 1994.
[26] D. Wienke, W. van den Broek, L. Buydens, T. Huth-Fehre, R. Feldhoff, T. Kantimm and K. Camman, Chemom. Intell. Lab. Syst., 32 (1996).
[27] D. Wienke and L. Buydens, Trends Anal. Chem., 99(9) (1995) 1–18.
[28] D. Wienke, D. Domine, L. Buydens and J. Devillers, in J. Devillers (Ed.), Neural Networks and Genetic Algorithms in QSAR and Drug Design, Proceedings of 2nd International Congress on Neural Networks and Genetic Algorithms in QSAR and Drug Design, Lyon, 1995, Academic Press, Cambridge, in press.
[29] S.E. Fahlman and C. Lebiere, The Cascade Correlation Architecture, Carnegie Mellon University Report, CMU-CS-90-100, August 1991, pp. 1–13.
[30] W.R. Dillon and M. Goldstein, Multivariate Analysis, Wiley, New York, 1984.