MATCH
Communications in Mathematical and in Computer Chemistry

MATCH Commun. Math. Comput. Chem. **69** (2013) 391-412

ISSN 0340 - 6253

A New Fuzzy Discriminant Analysis Method

Horia F. Pop

Babeş–Bolyai University
Faculty of Mathematics and Computer Science
1 M. Kogălniceanu St., 400084 Cluj–Napoca, Romania
E-mail: hfpop@cs.ubbcluj.ro

Costel Sârbu*

Babeş–Bolyai University
Faculty of Chemistry and Chemical Engineering
11 Arany János St., 400028 Cluj–Napoca, Romania
E-mail: csarbu@chem.ubbcluj.ro

(Received December 20, 2011)

Abstract

A new more informative and effective fuzzy discriminant analysis method based on fuzzy regression with point prototypes has been developed and applied on two relevant data sets (the classical Fisher's Iris data set and a clinical data set concerning different diseases). The proposed fuzzy method is consistent with the supervised character of the original discriminant analysis method. The classification and patterns obtained by membership degrees plot are in a very good agreement with the structure of data and the initial assignment of samples, which indicate that the new approach may be successfully employed in different fields. In addition, the graphical representation of fuzzy membership degrees to different classes provides a relevant way to visualize the relationships between the data items of the fuzzy classes.

1 Introduction

Discriminant function analysis or simply, discriminant analysis (DA), is based on the extraction of linear discriminant functions of the independent variables in a data set by

means of qualitative dependent variables and several quantitative independent variables [1, 2, 3, 4].

Discriminant analysis, in particular, has been extensively used in various fields of natural science (chemometrics, environmental sciences, biology, geology, etc.) [5, 6, 7, 8, 9, 10], as well as economy and humanistics [11, 12]. A very important issue in statistical analysis, fuzzy reasoning and similar theories is that their development goes hand in hand with their use in various domains. Because of the need for better, robust methods of data analysis coming from the application fields, often the theoretical developments are driven, as well, by the applicative research.

The fuzzy sets [13] represent a mathematical theory suitable for modeling imprecision and vagueness. Generally, vagueness is associated to the difficulty of making precise statements with respect to a certain topic. On the other side, in the *Fuzzy Sets Theory*, the hard alternative yes – no is indefinitely nuanceable. From this point of view, the fuzzy sets theory is not only a theory dealing with ambiguity; it is also a theory of fuzzy reasoning.

The fundamental fact that lies behind fuzzy logic is that any field and any theory may be fuzzified by replacing the concept of crisp set with the concept of fuzzy set. Thus, theoretic fields have appeared, such as fuzzy arithmetic, fuzzy topology, fuzzy graph theory, fuzzy probability theory, 'strict' fuzzy logic, etc. Similarly, applied fields that suffered generalizations are fuzzy neural network theory, fuzzy pattern recognition, fuzzy mathematical programming, etc. What is gained through fuzzification is *greater generality*, *higher expressivity*, an *enhanced ability to model* real-world problems, and a methodology for exploiting the *tolerance for imprecision* [14].

The first fuzzy DA method has been introduced in [15]. The original data is first fuzzified through a modified FCM method. Then, fuzzy within class and between class scatter matrices are computed and the resulted eigenvalue problem is then solved.

The paper [16] describes fuzzy discriminant analysis with kernel methods. The FDA approach is generalized by introducing kernel functions, in order to allow for nonlinear problems. A similar approach, using a kernel-based maximum a posteriori classification method, has been introduced in [17]. Robust methods are further proposed to estimate the probability densities.

A different approach of a fuzzy discriminant analysis is described in [18]. First, a

fourfold-objective model on the discriminant analysis is developed, by which a set of integrated subspaces derived from within-class and between-class scatter matrices are constructed, respectively. Second, an improved FDA algorithm based on the relaxed normalized condition is proposed to achieve the distribution information of each sample represented with fuzzy membership grade, which is incorporated into the rede?nition of Fisher's scatter matrices. A robust, non-fuzzy discriminant analysis method based on kernel functions, through nonlinear mapping in the feature space is described in [19].

A few other approaches have also been devised. Thus, the paper [20] describes a supervised iterative k-means like classification method using kernel functions. The paper [21] presents a fuzzy inverse FDA. The fuzzy membership degrees and each class center are obtained through Fuzzy k-Nearest Neighbor algorithm. Then, the membership degree of each sample is considered and the corresponding fuzzy within-class scatter matrix, fuzzy between-class scatter matrix and fuzzy total scatter matrix are computed. In case of small sample size problems, PCA is used in advance. A partial supervision in fuzzy clustering was also discussed [22] and the effect of various distance functions on the performance of the clustering mechanisms has also been investigated. In addition to the standard Euclidian distance being commonly exploited in fuzzy clustering, three more versatile and adaptive distance measures have been considered, such as its weighted version, a full adaptive distance, and a kernel-based distance.

In this paper a new fuzzy discriminant analysis method is developed and its efficient application is demonstrated using two data sets (the classical Fisher's Iris data set and a clinical data set concerning different diseases).

This paper is organized as follows. A theoretical introduction to the discriminant function analysis is given in Section 2. Section 3 follows with a description of our new fuzzy discriminant analysis approach. Then Section 4 presents two case studies, the former with the fuzzy discriminant analysis of the Fisher's Iris data set, and the latter with the analysis of a clinical data set.

2 Discriminant Function Analysis

DA can be formulated as follows: let $\mathbf{X} = \{x^1, \ldots, x^n\} \subset \mathbf{R}^s$ be a finite set of characteristic vectors, where n is the number of items and $s$is the number of the original variables (predictors), $\mathbf{x}^i = [x_1^i, x_2^i, \ldots, x_s^i]^T$ and y be a nominal characteristic (grouping

variable), with k values, each of which characterizes exactly one of the k sets composing the partition substructure of the given data set. The total variance/covariance matrix is first calculated according to the following expression

$$\mathbf{V} = \mathbf{X}^T\mathbf{M}\mathbf{X}, \tag{1}$$

where $\mathbf{X}$ is the centered data matrix, $\mathbf{X}^T$ is the transpose matrix, $\mathbf{M}$ is the diagonal matrix (in most cases is the unity matrix).

Considering a new characteristic defined as $\mathbf{c} = \mathbf{X}\mathbf{u}$, one can calculate its variance by applying the relation (2).

$$\|c\|^2 = \mathbf{c}^T\mathbf{M}\mathbf{c} = \mathbf{u}^T\mathbf{X}^T\mathbf{M}\mathbf{X}\mathbf{u} = \mathbf{u}^T\mathbf{V}\mathbf{u}, \tag{2}$$

The total variance $\mathbf{V}$ may be decomposed into two components: the between-group variance $\mathbf{B}$ and within-group variance $\mathbf{W}$, namely

$$\mathbf{V} = \mathbf{B} + \mathbf{W}, \tag{3}$$

and, as a consequence, the variance of the characteristic $\mathbf{c}$ becomes

$$\|c\|^2 = \mathbf{u}^T\mathbf{V}\mathbf{u} = \mathbf{u}^T\mathbf{B}\mathbf{u} + \mathbf{u}^T\mathbf{W}\mathbf{u}. \tag{4}$$

In this case, it is very easy to observe that eq. 4 can be rewritten in the following form

$$\frac{\mathbf{u}^T\mathbf{B}\mathbf{u}}{\mathbf{u}^T\mathbf{V}\mathbf{u}} + \frac{\mathbf{u}^T\mathbf{W}\mathbf{u}}{\mathbf{u}^T\mathbf{V}\mathbf{u}} = 1. \tag{5}$$

In practice the first ratio in eq. 5 is maximized

$$\lambda = \frac{\mathbf{u}^T\mathbf{B}\mathbf{u}}{\mathbf{u}^T\mathbf{V}\mathbf{u}} \quad (0 \leq \lambda < 1) \tag{6}$$

or, in a different form, of a generalized eigenvalue problem:

$$\mathbf{B}\mathbf{u} = \lambda\mathbf{V}\mathbf{u}, \tag{7}$$

Since matrix $\mathbf{V}$ of the total variance is symmetrical and positive definite, this equation may be rewritten to a matrix equation similar to that obtained in the case of principal component analysis,

$$\mathbf{V}^{-1}\mathbf{B}\mathbf{u} = \lambda\mathbf{u}, \tag{8}$$

where λ and $\mathbf{u}$ represent the eigenvalues (known, as well, as characteristic roots) and eigenvectors of the matrix $\mathbf{V}^{-1}\mathbf{B}$. The vector $\mathbf{u}^1$, named the first discriminant factor corresponds to the highest value of λ; the higher this value the higher will be the discriminant power of this factor. After obtaining the first discriminant characteristic $\mathbf{c}_1 = \mathbf{X}\mathbf{u}^1$, in a similar way can be obtained the discriminant characteristic $\mathbf{c}_2 = \mathbf{X}\mathbf{u}^2$, uncorrelated with the first and so on. It appears clearly that eigenvectors corresponding to the matrix $\mathbf{V}^{-1}\mathbf{B}$ namely $\mathbf{u}^1$, $\mathbf{u}^2$, ..., $\mathbf{u}^{k-1}$, ranked in decreasing order of the positive values λ_1, ..., λ_{k-1}, are successive solutions of the above matrix equation. The quality of discrimination and the selection of the most discriminant independent variable is given by the value of the largest eigenvalue, λ.

3 A New Fuzzy Discriminant Analysis Approach

Let us consider a data set $X = \{x^1, \dots x^n\} \subseteq \mathrm{R}^s$, and the predetermined classification matrix, denoted by A'. This matrix, produced by human experts, shows an *a priori* split of the n data items in k different classes. In such a case, the matrix A' is a Boolean matrix indicating the membership of a data item to one of the k classes.

One of the major issue human experts have is that they think in crisp terms. This means that the a-priori classes are defined in crisp terms. This is not a realistic decision, since in almost all real situations data is of a fuzzy nature. The given data classes most certainly have data items close to the central locations, but they have as well distant data items, also called outliers. As such, a preprocessing step must be done: for the crisp a-priory classes, suitable fuzzy regression sets will be determined. For each original class, a fuzzy regression with point prototypes is applied and fuzzy membership degrees are thus determined. We recall the main details here. The optimal fuzzy set A that best describes the given crisp set, and the associated point prototype $L \in \mathrm{R}^s$, are determined by minimizing the following fuzzy objective function:

$$J(A, L) = \sum_{j=1}^{n} A(x^j)^m \left\| x^j - L \right\|^2 + \sum_{j=1}^{n} \left(1 - A(x^j)\right)^m \left(\frac{\alpha}{1-\alpha}\right)^{m-1}, \tag{9}$$

where α is a positive subunit value set a-priory, identifying the fuzzy membership degree of the farthest outlier and $m > 1$ is the fuzziness index, set a-priori. The algorithm used

to solve this problem has been called Fuzzy Regression [23] and iterates by computing the prototype L that minimizes the function $J(A, \cdot)$ and by computing the fuzzy set A that minimizes the function $J(\cdot, L)$. As an improvement to this method, in order to ensure the independence of scale, we usually work with the relative dissimilarity when determining the fuzzy set above, i.e.

$$A(x^j) = \frac{\frac{\alpha}{1-\alpha}}{\frac{\alpha}{1-\alpha} + \left(\frac{\|x^j - L\|}{\max\|x^j - L\|}\right)^{\frac{2}{m-1}}}$$

Complete details of this fuzzy regression procedure and other variants thereof are given in [23, 24, 25].

Of course, this means that the result will be a sub-partition, i.e. the sum of membership degrees of a point to all classes is less than one. But, on the other side, this preprocessing step allows us to show light on the input data and the quality data items from each original cluster, as it has been a-priori proposed. As opposed to this, other methods use either an unsupervised clustering scheme here (which we find it is in principle un-appropriate), or use different mechanisms to set the membership degrees without any functional optimization.

The Fuzzy Discriminant Analysis problem is defined as follows: let $\mathbf{X} = \{x^1, \ldots, x^n\} \subset \mathbf{R}^s$ be a finite set of characteristic vectors, where $n$is the number of items and $s$is the number of the original variables (predictors), $\mathbf{x}^j = [x_1^j, x_2^j, \ldots, x_s^j]^T$ and let A_i (with $i = 1, \ldots, k$) be fuzzy sets on X, corresponding to the k a-priori sets composing the partition substructure of the given data set. A new vector (or characteristic) c is to be determined, that maximizes the fuzzy between-class variance of the projected data items, and minimizes the fuzzy within-class variance of the projected data items.

Considering this new characteristic defined as $\mathbf{c} = \mathbf{Xu}$, the fuzzy between-group variance $\mathbf{B}$ and fuzzy within-group variance $\mathbf{W}$, are defined as:

$$\mathbf{W} = \frac{1}{n-k} \sum_{i=1}^{k} \left(\sum_{j=1}^{n} A_i \left(x^j\right)^m \left(x^j - L^i\right)^T \left(x^j - L^i\right) \right), \tag{10}$$

$$\mathbf{B} = \frac{1}{k-1} \sum_{i=1}^{k} \left(\sum_{j=1}^{n} A_i \left(x^j\right) \right)^m \left(L^i - L\right)^T \left(L^i - L\right), \tag{11}$$

where the class means L^i are determined like the fuzzy point prototypes,

$$L^i = \frac{\sum_{j=1}^{n} A_i \left(x^j\right)^m x^j}{\sum_{j=1}^{n} A_i \left(x^j\right)^m},$$

and L is the central location for the whole data set.

As the fuzzy sets A_i form a sub-partition of the given data set, we formulate the problem of determining the optimal direction $\mathbf{u}$ as maximizing the ratio

$$\lambda = \frac{\mathbf{u}^T(\mathbf{V}-\mathbf{W})\mathbf{u}}{\mathbf{u}^T\mathbf{V}\mathbf{u}} \quad (0 \leq \lambda < 1) \tag{12}$$

or, in a different form, to solve the generalized eigenvalue problem

$$\mathbf{(V\text{-}W)u} = \lambda\mathbf{V}\mathbf{u}. \tag{13}$$

Since matrix $\mathbf{V}$ of the total variance is symmetrical and positive definite, this equation may be rewritten to a matrix equation similar to that obtained in the case of principal component analysis,

$$\mathbf{V}^{-1}\mathbf{(V\text{-}W)u} = \lambda\mathbf{u}, \tag{14}$$

where λ and $\mathbf{u}$ represent the eigenvalues (known, as well, as characteristic roots) and eigenvectors of the matrix $\mathbf{V}^{-1}\mathbf{(V\text{-}W)}$. The vector $\mathbf{u}^1$, named the first discriminant factor corresponds to the highest value of λ; the higher this value the higher will be the discriminant power of this factor. After obtaining the first discriminant characteristic $\mathbf{c}_1 = \mathbf{X}\mathbf{u}^1$, in a similar way can be obtained the discriminant characteristic $\mathbf{c}_2 = \mathbf{X}\mathbf{u}^2$, uncorrelated with the first and so on. It appears clearly that eigenvectors corresponding to the matrix $\mathbf{V}^{-1}\mathbf{(V\text{-}W)}$ namely $\mathbf{u}^1$, $\mathbf{u}^2$, ..., $\mathbf{u}^{k-1}$, ranked in decreasing order of the positive values λ_1, ..., λ_{k-1}, are successive solutions of the above matrix equation. The quality of discrimination and the selection of the most discriminant independent variable is given by the value of the largest eigenvalue, λ.

Finally, the original class means are projected in the new system of coordinates, and the final fuzzy membership degrees are determined from square-distances to the class means, using a relation similar to the Fuzzy C-Means case:

$$A_i\left(x^j\right) = \frac{1}{\sum\limits_{l=1}^{k}\left(\frac{\|x^j-L^i\|}{\|x^j-L^l\|}\right)^{1/(m-1)}}$$

The final fuzzy classification table is computed by counting cardinals of fuzzy sets: instead of counting the number of data items classified in a particular class, we are actually

computing an overall fuzzy membership degree. The fuzzy count of all items from the i-th original fuzzy set A'_i classified in the l-th fuzzy setA_l, denoted as C_{il}, is given by

$$C_{il} = \sum_{j=1}^{n} A'_i\left(x^j\right) \cdot A_l\left(x^j\right).$$

A friendlier version of this fuzzy classification matrix may be computed by scaling the fuzzy cardinal values and producing values representing the percentages of all items from the i-th original fuzzy set classified in the l-th fuzzy set:

$$C_{il}^{[\%]} = \frac{\sum\limits_{j=1}^{n} A'_i\left(x^j\right) \cdot A_l\left(x^j\right)}{\sum\limits_{j=1}^{n} A'_i\left(x^j\right)} \times 100.$$

A crisp classification matrix is as well determined by first defuzzifying the final fuzzy partition and then using the cardinals of the crisp classes.

After this learning phase, testing follows in various ways, including use of separate testing data, or by cross-validation.

The classical discriminant analysis method is known to provide maximum likelihood estimations under certain assumption (normality of the class distributions etc.). As the experiments will illustrate, and as previous research on data analysis methods based on fuzzy sets have also shown, the fuzzy discriminant analysis method is robust with respect to outliers and distribution of data.

We underline once again the robustness achieved by using fuzzy membership values. The main advantage of fuzzy sets over crisp sets and of fuzzy logic over binary logic is the availability of nuanced membership degrees. On one side, the classes input provided by the human expert is fuzzified, allowing robust treatment of outliers. On the other side, the output of the method is fuzzy as well, allowing a more detailed view of the relationships between data items and classes. These fuzzy membership degrees are not actually related to uncertainty, because there is nothing uncertain about the classification of a certain data item, but have to be regarded as a measure of 'typicality'.

The fuzzy discriminant analysis method presented here is a multiclass method by design, as no restriction with respect to the number of classes is introduced. This is a parameter to be set by the human experts as they establish the a-priori classes split.

4 Data sets

4.1 Data set 1 (Iris flower data set)

The first illustrative example uses the Iris flower data set, also known as Fisher's Iris data set [26]. The dataset consists of 50 samples from each of three species of Iris flowers (Iris setosa, Iris virginica and Iris versicolor). Four features were measured from each sample, namely the length and the width of sepal and petal, in centimeters.

4.2 Data set 2 (clinical data)

Fuzzy Discriminant Analysis was also applied as a method of disease identification, using data obtained from blood analysis of several patients. The compounds investigated by photometric methods in human blood samples were organic compounds of clinical interest (glucose, triglycerides, cholesterol, creatinine and urea), inorganic compounds (Na, K, Ca, Mg and Fe) and enzymes (Lactate Dehydrogenase (LDH), Alanine Transaminase (ALT), Aspartate Aminotransferase (AST), Alkaline Phosphatase (ALP) and Gamma Glutamyltransferase (GGT)). According to their concentration level the following diseases have been selected for study: hydroelectric disorders, hepatic diseases, lipid disorders, diabetes and renal disorders. Some patients resulted to be healthy. The training data set consisted of 100 patients, diagnosed by clinical evaluation as follows: 20 are healthy (marked 's'), 20 have lipid disorders ('l'), 20 hepatic diseases ('h'), 20 hydroelectric disorders ('d'), 10 diabetes ('z'), 10 renal disorders ('r').

5 Results and discussions

5.1 Data set 1

Figure 1 shows the graphical display of standardized canonical scores obtained with the classical DA algorithm for the Iris flower data set. Figure 2 shows the graphical display of standardized canonical scores obtained with our fuzzy DA algorithm. Both score values have been standardized in order to make any comparison possible. We remark a mirroring effect of the two images. This problem occurs with many implementations of multivariate analysis methods. PCA is another example. We are showing the graphics without any further alignment to illustrate that different implementations may align the variates differently, and that this is not an issue with the methods or their results.

By a careful visual examination of the two figures it is possible to observe that the three classes appear to be more compact and better separated in the fuzzy case. As well, the class outliers are better isolated. The linear structure of each class is clearly pointed out in the fuzzy representation.

Moreover, the graphical representation of fuzzy membership degrees to different classes provides a relevant way to visualize the relationships between the data items of the fuzzy classes. Figures 3, 4, and 5 show the graphical display of fuzzy membership degrees as obtained with our fuzzy DA algorithm. Figure 3 represents the fuzzy classes A1 and A2, Figure 4 represents the fuzzy classes A1 and A3, and Figure 5 represents the fuzzy classes A3 and A3, respectively.

We remark here quite a significant proportion of flowers in A2 and A3 that are classified in the other set. This is not a weakness of the method, but, the available data, given in terms of fuzzy membership degrees, show that there is actually an overlap between the two classes, with a few samples presenting hybrid features to the two classes. The overlap is very well seen in Figure 5. Figures 3, 4, 5 show a very clear separation between the class A1, on one side, and classes A2 and A3 on the other side, while Figure 5 shows three groups of samples for classes A2 and A3: a central group of A2, a central group of A3 and a group with hybrid samples. The representation of fuzzy membership degrees corresponding to the fuzzy partition leads to a clear discrimination of the three classes of iris data. In addition, the class overlaps are better distinguished, and the few data items with hybrid features are clearly isolated. The presence of hybrid data is not a weakness of the method, but an advantage, since it illustrates the natural hybrid character of the data items in question. An undisputable advantage of these fuzzy degrees representations over crisp sets representations is that the fuzzy membership degrees are already by design in the interval [0, 1], and, as such, all graphical representations are comparable without any scaling.

The eigenvalues, as determined by the Fuzzy DA method, are 0.986 and 0.728. As such, the discrimination quality is 0.986.

The classification matrix, produced using the classical DA method, is presented in Table 1. Comparatively, the classification matrix, produced using the Fuzzy DA method with maximum fuzzy membership defuzzification, is presented in Table 2. The extra data items that appear to be classified by FDA in a different class are described in Table 3. As

we see, excepting the three data items misclassified by the classical method (71, 78 and 107), all other data items have hybrid character, as they show membership degrees in the range 0.30-0.50, with five of them having second largest fuzzy membership degrees under 0.15 lower than the largest fuzzy membership degrees, all these increasing the correct classification rate for Fuzzy DA method.

	Correct [%]	**A1**	**A2**	**A3**
A1	**100**	**50**	0	0
A2	**96**	0	**48**	2
A3	**98**	0	1	**49**
Total	**98**	**50**	**49**	**51**

Table 1: The classification matrix produced using the classical DA method

	Correct [%]	**A1**	**A2**	**A3**
A1	**100**	**50**	0	0
A2	**88**	0	**44**	6
A3	**80**	0	10	**40**
Total	**89.33**	**50**	**54**	**46**

Table 2: The classification matrix produced using the Fuzzy DA method

A visual inspection of data clusters as depicted by the standardized canonical scores representations given in Figures 1, 2 shows that the classes in both figures are comparably compact and well separated and confirms the remark that the class switch for the data items indicated in Table 3 is justified, as there are essentially data items of hybrid membership degrees, situated closer to items of the class with slightly smaller membership degrees. As well, the graphical representation of fuzzy membership degrees in the data clusters obtained from the fuzzy discriminant analysis method clearly points out the hybrid character of the said data items.

We have tested our method using the cross-validation technique. We have repeatedly omitted one data item from the original set and we have determined the fuzzy membership degrees to the fuzzy classes learned by the Fuzzy DA method applied using all the other data items.

Figure 6 shows the cross-validation fuzzy membership degrees of the data items, against the fuzzy membership degrees as normally obtained by the Fuzzy DA method.

No.	A1	A2	A3	Initial class	Assigned class
51	0.0945	**0.3926**	**0.5130**	2	3
52	0.0789	**0.4358**	**0.4853**	2	3
71	0.0320	0.2191	**0.7489**	2	3
78	0.0226	0.2342	**0.7432**	2	3
86	0.1055	0.3465	**0.5480**	2	3
87	0.0568	**0.4266**	**0.5166**	2	3
102	0.0342	**0.4945**	**0.4713**	3	2
107	0.0687	**0.6749**	0.2565	3	2
114	0.0580	**0.5599**	0.3821	3	2
120	0.0951	**0.6041**	0.3008	3	2
124	0.0279	**0.6307**	0.3414	3	2
127	0.0235	**0.6121**	0.3644	3	2
134	0.0225	**0.6121**	0.3654	3	2
135	0.0353	**0.5599**	0.4048	3	2
143	0.0342	**0.4945**	**0.4713**	3	2
147	0.0490	**0.5960**	0.3550	3	2

Table 3: Membership degrees of the items classified by FDA in a different class

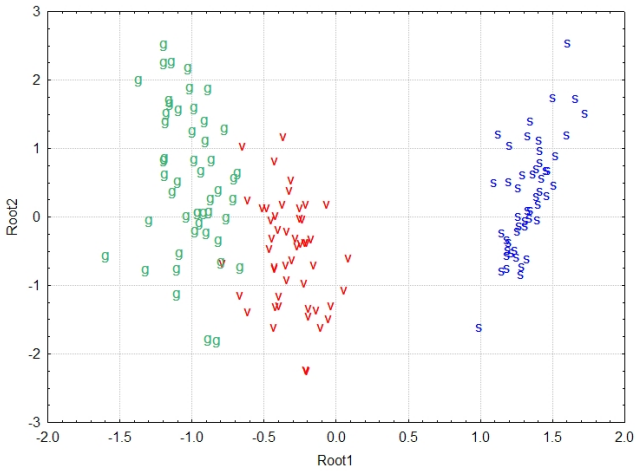

Figure 1: Standardized canonical scores obtained with the classical DA algorithm

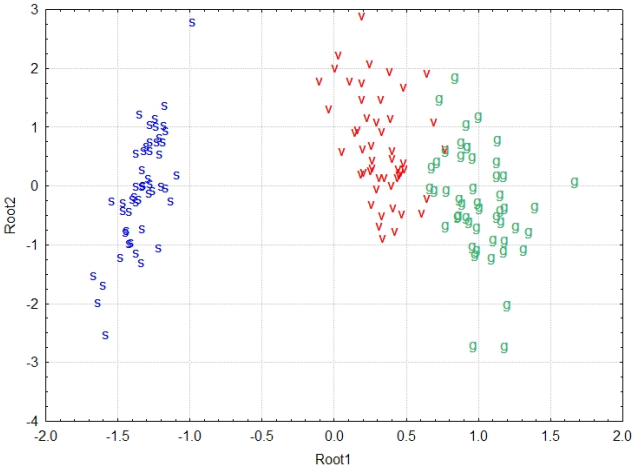

Figure 2: Standardized canonical scores obtained with the fuzzy DA algorithm

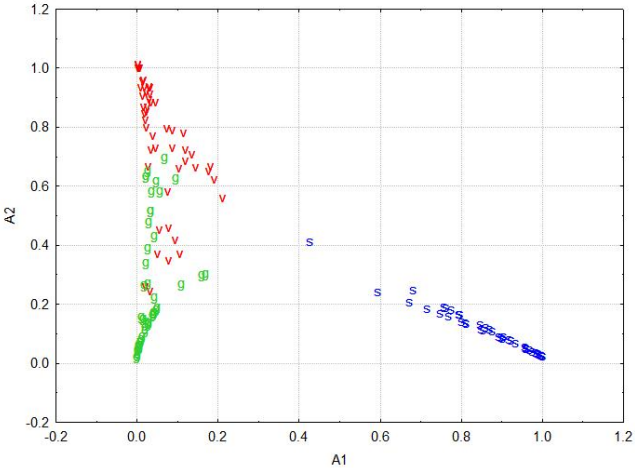

Figure 3: Representation of new fuzzy membership degrees to classes A1 and A2, as obtained with our fuzzy DA algorithm

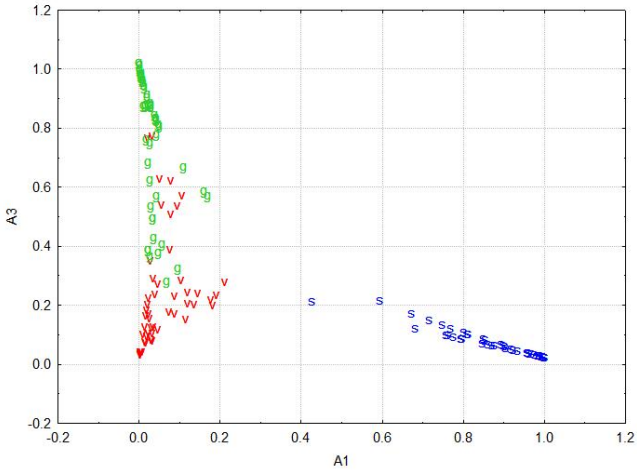

Figure 4: Representation of new fuzzy membership degrees to classes A1 and A3, as obtained with our fuzzy DA algorithm

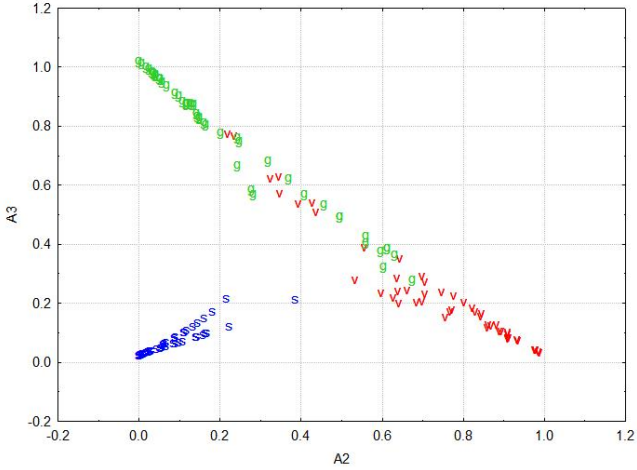

Figure 5: Representation of new fuzzy membership degrees to classes A2 and A3, as obtained with our fuzzy DA algorithm

We notice an almost perfect alignment across the y=x line, confirming the quality of the Fuzzy DA method.

Table 4 shows the cross-validation classification matrix for the classical DA method as compared with the proposed fuzzy DA method. We remark here quite comparable results.

Classic	**Correct [%]**	**A1**	**A2**	**A3**
A1	**100**	**50**	0	0
A2	**96**	0	**48**	2
A3	**94**	0	3	**47**
Total	**96.67**	**50**	**51**	**49**

Fuzzy	**Correct [%]**	**A1**	**A2**	**A3**
A1	**100**	**50**	0	0
A2	**84**	0	**42**	8
A3	**78**	0	11	**39**
Total	**87.33**	**50**	**53**	**47**

Table 4: The classification matrices produced using the classical and fuzzy DA method

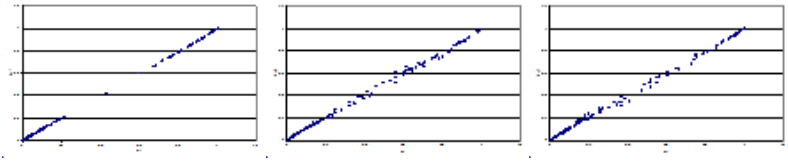

Figure 6: Fuzzy membership of the cross-validated items against the fuzzy DA output

A question may be raised with respect to the usefulness of a fuzzy sets based method for classification of crisp data. While the actual items of the Iris data set are flowers of one of four crisp types, there is an issue whether the data items as represented using the given four variables are indeed of a crisp nature. As any robust clustering method used on the Iris data shows, some of these data items demonstrate a class overlap. These lead to the conclusion that, while the actual items form crisp clusters, the data items, as characterized by the four variables, naturally form fuzzy clusters.

5.2 Data set 2

The values of the eigenvalues, determined by the Fuzzy DA method for the clinical data set, are 0.968, 0.955, 0.936, 0.895 and 0.877. As such, the discrimination quality is 0.968.

The classification matrix, produced using the classical DA method, is presented in Table 5. Comparatively, the classification matrix, produced using the Fuzzy DA method with maximum fuzzy membership defuzzification, is presented in Table 6. A direct examination shows slightly better results with the Fuzzy DA method against the classical

method, with 88% correctly classified items against 85% for the crisp method. The extra data items that appear to be classified by FDA in a different class are described in Table 7. As we see, excepting three data items (31, 38 and 47), all other data items have strong hybrid character, as they show membership degrees to two or more classes very closed to each other, with a difference of around 0.10, all these further increasing the correct classification rate for Fuzzy DA method. These may correspond to individuals suffering of more than one illness, even if one illness appears to be dominant.

	Correct [%]	**A1 (s)**	**A2 (l)**	**A3 (h)**	**A4 (d)**	**A5 (z)**	**A6 (r)**
A1 (s)	**90**	**18**	2	0	0	0	0
A2 (l)	**70**	6	**14**	0	0	0	0
A3 (h)	**80**	3	1	**16**	0	0	0
A4 (d)	**100**	0	0	0	**20**	0	0
A5 (z)	**80**	1	1	0	0	**8**	0
A6 (r)	**90**	1	0	0	0	0	**9**
Total	**85**	**29**	**18**	**16**	**20**	**8**	**9**

Table 5: The classification matrix produced using the classical DA method

	Correct [%]	**A1 (s)**	**A2 (l)**	**A3 (h)**	**A4 (d)**	**A5 (z)**	**A6 (r)**
A1 (s)	**95**	**19**	1	0	0	0	0
A2 (l)	**65**	7	**13**	0	0	0	0
A3 (h)	**86**	2	1	**17**	0	0	0
A4 (d)	**100**	0	0	0	**20**	0	0
A5 (z)	**90**	1	0	0	0	**9**	0
A6 (r)	**100**	0	0	0	0	0	**10**
Total	**88**	**30**	**15**	**17**	**19**	**9**	**10**

Table 6: The classification matrix produced using the fuzzy DA method

Figures 7, 8, and 9 show the graphical display of new fuzzy membership degrees as obtained with our fuzzy DA algorithm. Figure 7 represents the fuzzy classes A1 and A2, Figure 8 represents the fuzzy classes A3 and A4, and Figure 9 represents the fuzzy classes A5 and A6, respectively.

We have to remark a hepatitis item placed together with the disordered items. We assume this strange case is either a case of wrong human diagnosis, or a case of erroneous data collection.

We remark again the very good discrimination power of the fuzzy degrees representations. In this example, even the lower fuzzy values are very well separated; the spaces

No.	A1	A2	A3	A4	A5	A6	Initial class	Assigned class
6	**0.3314**	**0.4254**	0.0885	0.0477	0.0458	0.0612	1	2
27	**0.2910**	0.1525	0.0950	0.1501	**0.2633**	0.0481	2	1
28	**0.2639**	0.1158	0.1222	**0.1829**	0.1452	**0.1711**	2	1
29	**0.2696**	0.1168	0.1311	**0.2684**	0.0950	0.1192	2	1
32	**0.5684**	0.1740	0.0772	0.0638	0.0482	0.0685	2	1
38	**0.2719**	**0.2245**	0.1163	0.0848	0.1337	0.1688	2	1
39	**0.4100**	0.1795	0.0906	0.0675	0.2056	0.0468	2	1
40	**0.3208**	**0.1699**	0.1256	**0.1576**	**0.1625**	0.0636	2	1
47	0.1322	**0.8162**	0.0150	0.0124	0.0149	0.0093	3	2
50	**0.2718**	**0.2563**	**0.2045**	0.0785	0.1255	0.0634	3	1
51	**0.3532**	**0.2432**	**0.2363**	0.0661	0.0446	0.0567	3	1
81	**0.4261**	**0.3173**	0.0524	0.0638	0.1088	0.0315	5	1

Table 7: Membership degrees of the items classified by FDA in a different class

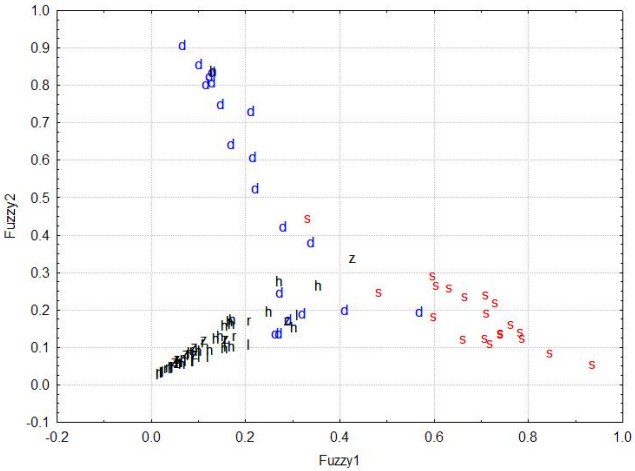

Figure 7: Representation of new fuzzy membership degrees to classes A1 and A2, as obtained with our fuzzy DA algorithm

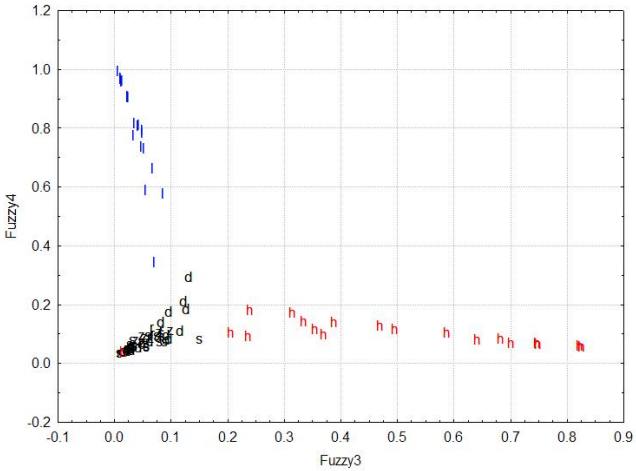

Figure 8: Representation of new fuzzy membership degrees to classes A3 and A4, as obtained with our fuzzy DA algorithm

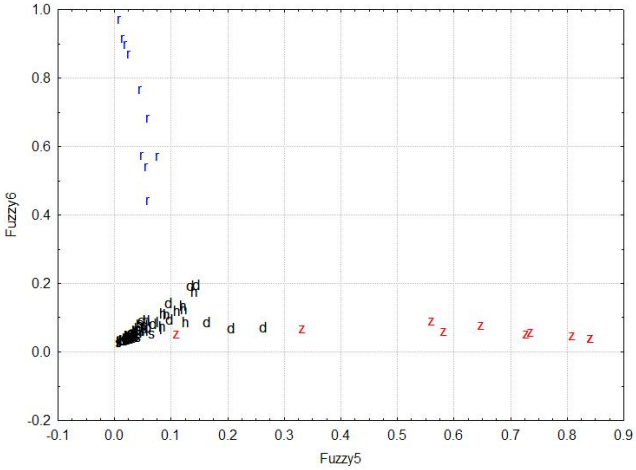

Figure 9: Representation of new fuzzy membership degrees to classes A5 and A6, as obtained with our fuzzy DA algorithm

between the separated classes and the rest of the data items are in many cases very large indeed.

On the other side, we observe that each graphic discriminates between two classes, and between these classes and the rest of the data set.

We notice, as well, the strong linear trend of each of the fuzzy classes depicted in Figures 7-9. As such, the fuzzy aspect is in good correlation with the status of the disease.

A particular remark is needed when observing Figure 7. Here, a few individuals with disorders are clustered with individuals marked healthy. This issue is real: such disorders are common to healthy people as well, and marking those individuals as having this disorder issue instead of being healthy is indeed the particular judgment of the medical doctor issuing the diagnostic.

A quality analysis similar to that performed for the first experiment may be done as well in this case. The studied data show hybrid character of some data items, situation found in most experimental data sets obtained from natural sciences.

Table 8 shows the cross-validation classification matrix for the classical DA method as compared with the proposed fuzzy DA method. We remark that the results in the fuzzy case are consistently better than the results using the traditional method.

Classic	**Correct [%]**	**A1 (s)**	**A2 (l)**	**A3 (h)**	**A4 (d)**	**A5 (z)**	**A6 (r)**
A1 (s)	**80**	**16**	4	0	0	0	0
A2 (l)	**50**	10	**10**	0	0	0	0
A3 (h)	**70**	4	2	**14**	0	0	0
A4 (d)	**95**	0	1	0	**19**	0	0
A5 (z)	**80**	1	1	0	0	**8**	0
A6 (r)	**80**	1	1	0	0	0	**8**
Total	**75**	**32**	**19**	**14**	**19**	**8**	**8**

Fuzzy	**Correct [%]**	**A1 (s)**	**A2 (l)**	**A3 (h)**	**A4 (d)**	**A5 (z)**	**A6 (r)**
A1 (s)	**95**	**19**	1	0	0	0	0
A2 (l)	**55**	6	**11**	0	1	2	0
A3 (h)	**75**	3	2	**15**	0	0	0
A4 (d)	**95**	1	0	0	**19**	0	0
A5 (z)	**80**	2	0	0	0	**8**	0
A6 (r)	**90**	1	0	0	0	0	**9**
Total	**81**	**32**	**14**	**15**	**20**	**10**	**9**

Table 8: The classification matrices produced using the classical and fuzzy DA method

Figure 10 shows the cross-validation fuzzy membership degrees of the data items,

against the fuzzy membership degrees as normally obtained by the Fuzzy DA method.

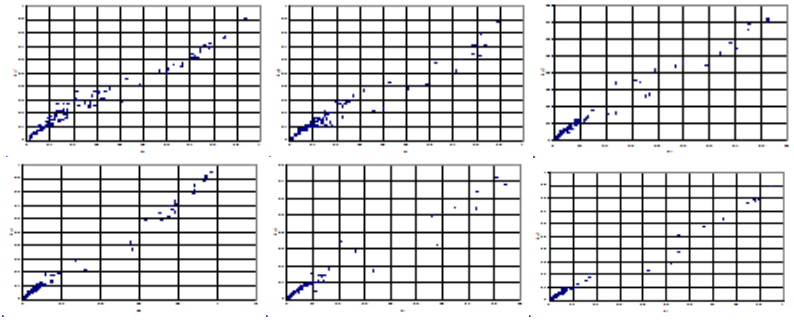

Figure 10: Fuzzy membership of the cross-validated items against the fuzzy DA output

The experiments have been run using our own software, called SADIC (System for Automatic Data Investigation and Classification), developed in C++ using the wxWidgets library. The program implements classical and fuzzy clustering methods of all kinds, classical and fuzzy linear and nonlinear regression, classical principal components analysis and a few fuzzy methods, classical and fuzzy factorial analysis and discriminant analysis, as well as other data analysis methods.

6 Concluding Remarks

The performances of a new fuzzy discriminant analysis method based on fuzzy regression with point prototypes have been evaluated on two relevant data sets (the classical Fisher's iris data set and a clinical data set concerning different diseases. The classification and patterns obtained by membership degrees plot were in a very good agreement with the structure and origin of data and illustrated the higher discrimination power of the new fuzzy discrimination approach.

We remark that both experiments show more or less similar numerical results, as comparing the performance of the fuzzy discriminant analysis method to the classical discriminant analysis method. This is actually expected, since it confirms that, on regular data sets, the fuzzy discriminant analysis method works closely to the classical one. The major issues to be noted here are the quality of results available for the fuzzy analysis. Having fuzzy membership degrees as output consistent to the classical analysis is a major information for the scientific analyst, since no two distinct items are the same, behave

exactly the same, or have identical properties and characteristics. Thus, the different fuzzy membership degrees of data items to the classes provided by human experts, allow the human experts to better distinguish between the different, otherwise similar, data. The cross-validation results confirm as well the quality of the fuzzy discriminant analysis method introduced in this paper.

Acknowledgment: This work was supported by a grant of the Romanian National Authority for Scientific Research, CNDIUEFISCDI, project number PN-II-PT-PCCA-2011-3.2-0917.

References

[1] Y. Kharin, *Robustness in Statistical Pattern Recognition*, Kluwer, Dordrecht 1996.

[2] G. H. McLachlan, *Discriminant Analysis and Statistical Pattern Recognition*, Wiley, New York, 2004.

[3] E. Micheli–Tzanakou, *Supervised and Unsupervised Pattern Recognition: Feature Extraction and Computational Intelligence*, CRC Press, Boca Raton, 2000.

[4] A. Webb, *Statistical Pattern Recognition*, Wiley, New York, 2002.

[5] R. G. Brereton, *Applied Chemometrics for Scientists*, Wiley, Chichester, 2007.

[6] B. Chanda, C. A. Murthy, *Advances in Intelligent Information Processing*, World Sci. Publishing, Singapore, 2008.

[7] J. C. Davis, *Statistics and Data Analysis in Geology*, Wiley, New York, 2002.

[8] J. W. Einax, H. W. Zwanziger, S. Geiss, *Chemometrics in Environmental Analysis*, Wiley–VCH, Weinheim, 1997.

[9] M. Otto, *Chemometrics: Statistics and Computer Applications in Analytical Chemistry*, Wiley–VCH, Weinheim, 1998.

[10] J. K. Percus, *Mathematics of Genome Analysis*, Cambridge Univ. Press, Cambridge, 2004.

[11] W. R. Klecka, *Discriminant Analysis*, Sage Publications, Iowa, 1980.

[12] M. Taniguchi, J. Hirukawa, K. Tamaki, *Optimal Statistical Inference in Financial Engineering*, Chapman, Boca Raton, 2008.

[13] L. A. Zadeh, Fuzzy sets, *Inf. Control* **8** (1965) 338–353.

[14] G. Klir, B. Yuan, *Fuzzy Sets and Fuzzy Logic: Theory and Applications*, Upper Saddle River, Prentice Hall, 1995.

[15] Z. P. Chen, J. H. Jiang, Y. Li, Y. Z. Liang, R.-Q. Yu, Fuzzy linear discriminant analysis for chemical data sets, *Chemom. Intell. Lab. Syst.* **45** (1999) 295–302.

[16] X. H. Wu, J. J. Zhou, Fuzzy discriminant analysis with kernel methods, *Pattern Recogn.* **39** (2006) 2236–2239.

[17] Z. Xu, K. Huang, J. Zhu, I. King, M. R. Lyu, A novel kernel–based maximum a posteriori classification method, *Neural Networks* **22** (2009) 977–987.

[18] X. Song, X. Yang, J. Yang, X. Wu, Y. Zheng, Discriminant analysis approach using fuzzy fourfold subspaces model, *Neurocomputing* **73** (2010) 2255–2265.

[19] Z. Liang, D. Zhang, P. Shi, Robust kernel discriminant analysis and its application to feature extraction and recognition, *Neurocomputing* **69** (2006) 928–933.

[20] C. Cifarelli, L. Nieddua, O. Seref, P. M. Pardalos, K-T.R.A.C.E: A kernel k-means procedure for classification, *Comput. Operations Res.* **34** (2007) 3154–3161.

[21] W. Yang, J. Wang, M. Ren, L. Zhang, J. Yang, Feature extraction using fuzzy inverse FDA, *Neurocomputing* **72** (2009) 3384–3390.

[22] A. Bouchachia, W. Pedrycz, Enhancement of fuzzy clustering by mechanisms of partial supervision, *Fuzzy Set. Syst.* **157** (2006) 1733–1759.

[23] H. F. Pop, C. Sârbu, A new fuzzy regression algorithm, *Anal. Chem.* **68** (1996) 771–778.

[24] H. F. Pop, A new regression technique based on fuzzy sets, *Studia Univ. Babes–Bolyai, Ser. Inform.* **43** (1998) 3–12.

[25] H. F. Pop, Development of robust fuzzy regression techniques using a fuzzy clustering approach, *Pure Math. Appl.* **14** (2004) 221–232.

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