



ELSEVIER

Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

Computers & Operations Research 31 (2004) 877–888

computers &
operations
research

www.elsevier.com/locate/dsw

Fuzzy discriminant analysis with outlier detection by genetic algorithm

Chang-Chun Lin^{a,*}, An-Pin Chen^b

^a*Department of Information Management, Kun-Shan University of Technology, 949, Da-Wan Road, Yung-Kang, Tainan 710, Taiwan, ROC*

^b*Institute of Information Management, National Chiao-Tung University, Hsinchu 300, Taiwan, ROC*

Abstract

This paper proposes a method for performing fuzzy multiple discriminant analysis on groups of crisp data and determining the membership function of each group by minimizing the classification error using a genetic algorithm. Euclidean distance is used to measure the similarity between data points and defining membership functions. A numerical example is provided for illustration. The numerical example indicates that the classification obtained by fuzzy discriminant analysis is more satisfactory than that obtained by crisp discriminant analysis and is less fuzzy than that obtained by fuzzy cluster analysis. Moreover, the proposed fuzzy discriminant analysis is also a good approach to identifying outliers, of which the degree of membership to each group is zero.

© 2003 Elsevier Ltd. All rights reserved.

Keywords: Fuzzy sets; Fuzzy discriminant analysis; Genetic algorithms

1. Introduction

In multivariate analysis, interests often focus on accounting for the variation in one variable, the criterion or dependent variable, in terms of covariation with other variables, the predictor or independent variables. Methods for this purpose are classified as the dependent methods, and discriminant analysis is an important one. Conventionally, discriminant analysis is divided into two-group discriminant analysis and multiple discriminant analysis. Two-group discriminant analysis is generally conducted using Fisher's method [1]. The basic idea behind Fisher's method is to reduce a large set

* Corresponding author. Tel./fax: +886-6-273-2726.

E-mail address: chanclin@ms29.hinet.net (C.-C. Lin).

of multiple measurements to a single linear composite with values that can maximally distinguish between members of the two groups. In doing so, a set of multivariate data is transformed into a set of univariate data. The linear composite is called Fisher's discriminant function, and the univariate data are called the discriminant score, which is a projection of each point onto the discriminant axis.

Fisher's method can also be applied to multiple discriminant analysis. Based on Fisher's method, the objective of multiple discriminant analysis is to find linear composites of the original variables that maximize the ratio of among group variability to within group variability. This concern is one of the most important issues in multiple discriminant analysis. Notably, the number of discriminant axes, depending on the number of groups to be discriminated, will exceed one, and multiple discriminant scores will be then required.

Another approach to developing two-group or multiple discrimination is based on the distances between data points and centroids of groups. Consider, for example, Mahalanobis' D^2 method [2]. This method computes the distances from each individual point to the centroid of each group using the pooled within-group covariance matrix. Each point is then assigned to the group to which it is nearest. Clearly, other types of distance may also be used as discriminating tools. However, both Fisher's method and the distance method crisply discriminate data; that is, each point can only belong to a single group. Once two groups overlap each other, some points will be assigned to the other group rather than the one to which they were originally assigned. Besides, the closeness of a point to a group is unobtainable.

According to the idea behind fuzzy cluster analysis, originally proposed by Bezdek [3], data points do not necessarily belong only to one group. A point that belongs to one group may simultaneously belong to other groups with different membership degrees. This idea has been widely accepted by researchers and practitioners. As an extension of classical cluster analysis, fuzzy cluster analysis divides groups based on the similarity between the points in them, which is measured by distance. Notably, however, discriminant analysis differs from cluster analysis. Cluster analysis is an independent method applied when no criterion variable exists; that is, when the memberships of data points to each group are unknown beforehand. Instead, when the criterion variable exists, dependent methods such as discriminant analysis are usually used to classify data groups. Discriminant analysis distinguishes groups by minimizing the within group variability while maximizing the between group variability; that is, minimizing the error of classification. However, few methods for fuzzy discriminant analysis have been reported. Although Watada et al. [4] proposed a method for fuzzy discriminant analysis. Their proposed method is designed for discriminating among objects that are measured by fuzzy predictor variables and a fuzzy criterion variable. This method is simply not suitable for cases that involve crisp predictor variables and a crisp criterion variable.

This study paper presents a method for fuzzy discriminant analysis (FDA). The objective of FDA herein is to determine the membership functions of data groups, which functions can maximally, fuzzily distinguish the data groups. Accordingly, Fisher's method is inappropriate for situations that involve multiple discriminant variables and cannot conveniently define membership functions. For simplicity, Euclidean distance instead of Mahalanobis' D^2 is used as a measure of the similarity between data points. The degree to which a point belongs to a group can be obtained from the membership function that is obtained by minimizing the sum of squared errors in the classification. Consequently, a point can be considered to be an outlier if its degree of membership to each group is zero.

2. Fuzzy discriminant analysis

Suppose m groups of n points are to be distinguished. The membership degree that point j belongs to group i , determined from the group’s membership function μ_i , should be as close to the known membership value y_{ij} as possible, while retaining a specific degree of fuzziness. The difference between the membership degree and the known membership value is defined as the classification error at point j . Thus, the objective of FDA is defined herein to be to minimize the sum of squared classification errors for each group. That is,

$$\text{minimize } E_i = \sum_{j=1}^n (\mu_i(D_{ij}^2) - y_{ij})^2 \quad \text{for } i = 1, 2, \dots, m, \tag{1}$$

where y_{ij} is the criterion variable, $y_{ij} \in \{0, 1\}$. $y_{ij} = 1$ if point j originally belongs to group i ; otherwise $y_{ij} = 0$. D_{ij}^2 denotes the distance between point j and the centroid of group i . The distance from a point \mathbf{x}_j to the centroid of group i , \mathbf{c}_i , is defined as

$$D_{ij}^2 = (\mathbf{x}_j - \mathbf{c}_i)' \mathbf{G} (\mathbf{x}_j - \mathbf{c}_i), \tag{2}$$

where matrix \mathbf{G} is symmetric and positive-definite. \mathbf{G} is an identity matrix when Euclidean distance is used or is the inverse of the pooled within-group covariance matrix of the data when Mahalanobis’ D^2 is used. Mahalanobis’ D^2 is a generalization of the familiar Euclidean distance.

The simplest and most extensively used fuzzy sets, the ramp-type and Z-shaped fuzzy sets, are used to represent the membership function of each group. The membership function of a Z-shaped fuzzy set \tilde{A} is usually defined as

$$\mu_{\tilde{A}}(x) = \begin{cases} 1, & \text{if } x \leq a, \\ \frac{b-x}{b-a} & \text{if } a \leq x \leq b, \\ 0 & \text{if } b \leq x. \end{cases} \tag{3}$$

Thus, the membership function of group i can be expressed as

$$\mu_i(D_i^2) = \begin{cases} 1, & \text{if } D_i^2 \leq a_i, \\ \frac{b_i - D_i^2}{b_i - a_i} & \text{if } a_i < D_i^2 \leq b_i, \\ 0 & \text{if } b_i < D_i^2. \end{cases} \tag{4}$$

If $D_{ij}^2 \leq a_i$, then point j is considered to belong fully to group i , while if $b_i < D_{ij}^2$, point j is considered not to belong to group i . Furthermore, if $a_i < D_{ij}^2 \leq b_i$, point j is considered to belong partially to group i with a degree of membership between 0 and 1. Finally, for a point j , $\mu_i(D_{ij}^2) = 0$ for $i = 1, 2, \dots, m$, then point j does not belong to any group and can be considered to be an outlier.

Restated, the objective of FDA is to find each group a fuzzy boundary that surrounds the group, with a fuzzy radius $\tilde{R}_i = [a_i, b_i]$ from the centroid \mathbf{c}_i . The region between a_i and b_i represents the fuzzy boundary around group i . The ultimate objective of FDA is to find a_i and b_i that can maximally distinguish the points that belong to from those that do not belong to group i .

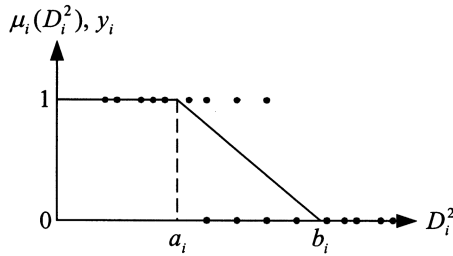


Fig. 1. Membership function for data group i .

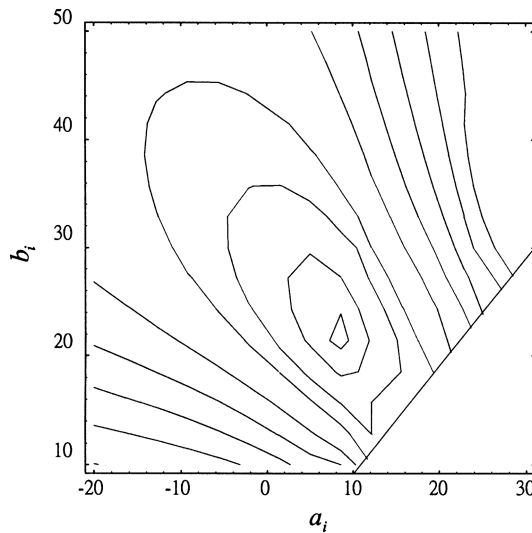


Fig. 2. Typical contour diagram of objective function $E_i(a_i, b_i)$.

Fig. 1 depicts the membership function of group i . The data points that belong to group i are those with criterion variable $y_i = 1$. Often, group i overlaps other groups. In such a case, the maximum distance from the points that belong to group i to the centroid may exceed some of the distances from the points that not belong to group i to the centroid, resulting in an overlapping area in Fig. 1. FDA then seeks to find the optimal a_i and b_i for group i that can minimize the classification errors, especially those that occur in the overlapping area.

The membership function expressed in Eq. (4) can be replaced by the following one:

$$\mu_i(D_i^2) = \frac{1}{2} \frac{|D_i^2 - b_i|}{b_i - a_i} - \frac{1}{2} \frac{|D_i^2 - a_i|}{b_i - a_i} + \frac{1}{2}. \tag{5}$$

Thus, from Eq. (1), the objective function associated with group i can be expressed as

$$E_i(a_i, b_i) = \sum_{j=1}^n \left(\frac{1}{2} \frac{|D_{ij}^2 - b_i|}{b_i - a_i} - \frac{1}{2} \frac{|D_{ij}^2 - a_i|}{b_i - a_i} + \frac{1}{2} - y_{ij} \right)^2. \tag{6}$$

Obviously, this is a bivariate, continuous but not differentiable function, and the optimization problem is a nonlinear one. Fig. 2 shows a typical contour diagram of $E_i(a_i, b_i)$, for example, the error function defined for the membership function in Fig. 1, under the condition $a_i < b_i$. Fig. 2 shows a valley in the direction $\mathbf{d} = (-1, 1)$. However, the error function may have several sub-optimal points where the gradient is zero and sometimes may be multi-modal. In such a case, fuzzy discriminant analysis becomes a global optimization problem, and genetic algorithms are an effective approach to finding the global optimal solution.

3. Genetic algorithm

Many optimization problems are very complex and quite hard to solve by conventional optimization techniques. Recently, genetic algorithms have received considerable attention for their potential in optimization technique for complex problems and have been successfully applied in many areas. Originally proposed and studied by John Holland [5], genetic algorithms are stochastic search techniques based on the mechanism of natural selection and genetics. Goldberg [6] described the typical form of genetic algorithms. Genetic algorithms start with an initial set of random solutions called population. Each individual in the population is a chromosome, and represents a solution to the problem to be optimized. A chromosome is a string of symbols that can be decoded into a solution. The chromosomes evolve through successive iterations called generations. During each generation, chromosomes are evaluated according to some measure of fitness, which is related to the objective function of the problem. Fitter chromosomes are more likely to be selected. Two operators are normally used to produce new chromosomes, called offspring, to form a new generation. One is the crossover operator, which merges two chromosomes selected from the current generation to form two new chromosomes. The other is the mutation operator, which modifies some genes of an old chromosome to generate a new chromosome. After several generations, the algorithms converge to the best chromosome, which is hopefully the optimal solution to the problem.

The genetic algorithm used in this study employs a different evolution process from conventional genetic algorithms. In Holland's original genetic algorithm, all parents are replaced by their offspring to form a new generation. This is called generational replacement. Offspring may be less fit than their parents because genetic algorithms are blind, such that some fitter chromosomes may be lost from the evolutionary process. Several replacement processes have been examined to solve this problem. Holland suggested that when each offspring is born, it replaces an arbitrarily selected chromosome of the current population. DeJong [7] proposed a crowding strategy that selects one parent to die when an offspring is born. The dying parent is the one that most closely resembles the new offspring, as measured by a simple bit-by-bit similarity count. In Gen and Cheng [8], both parents and their immediate offspring are all candidates for the new generation. The genetic algorithm used in this study follows a similar strategy to prevent the populations from degenerating. Each time an offspring is produced and is fitter better than the worst parent, a randomly chosen parent, but not the best one, is replaced. Consequently, the population evolves continually without being replaced by another population of the same size. Let $P(t)$ be the population at iteration t . The genetic algorithm is

described as follows.

```

begin
   $t \leftarrow 0$ ;
  initialize  $P(t)$ ;
  evaluate  $P(t)$ ;
  find the best and the worst chromosome of  $P(t)$ ;
  while  $t < a$  predetermined number of iteration do
    select two parents from  $P(t)$ ;
    generate two offspring by crossover and mutation;
    evaluate the offspring;
    if the offspring is fitter than the worst parent then
      select a parent except excluding the best one and replace it;
    find the best and the worst chromosome of  $P(t)$ ;
     $t \leftarrow t + 1$ ;
  end
end

```

Chromosome encoding is an essential part of every genetic algorithm. The genetic algorithm herein uses decimal encoding rather than binary coding. Decimal encoding can largely reduce the computation load of decoding the genotype into the phenotype. Each chromosome has a length of 12 genes. Each gene is a digit from 0 to 9. The first six digits and the last six digits each represent a real number of six significant figures, including 4 decimal places. The first real number is mapped onto a predetermined domain of a_i , and the second real number is mapped onto a domain of b_i . The domain of a_i is set to $[-k, 99.9999 - k]$ where k is a positive number since a_i may be negative. The domain of b_i is set to $[0, 99.9999]$ because b_i is not likely to be negative. The genetic algorithm in this study uses roulette wheel selection and one point crossover. The population size is set to 100. The crossover rate is set to 0.9 and the mutation rate to 0.05. Goldberg [6] or Gen and Cheng [8] detailed the algorithms for these operations.

4. Numerical example

Consider the data in Table 1, also depicted in Fig. 3. Fig. 3 shows that groups 1 and 2 do not overlap. Clearly dividing the two groups is impossible. Table 2 lists the primitive results of FDA, after the membership function for each group is obtained by applying the genetic algorithm described in Section 3. One point, point 12, is identified as an outlier because its degree of membership to every group is zero. Fig. 3 also indicates that point 12 is clearly an outlier. Outliers must be removed from the data to keep them from influencing the results of FDA. Notably, the membership functions must be determined again after the outliers are removed because this removal alters the centroid of the group. After the outlier is removed from the data, the FDA process is again performed on group 2. Table 4 compares the results of FDA and the results of crisp discrimination that is based on a comparison of distance.

Notably, point 6, originally belonging to group 1, is assigned to group 2 according to the distance method. According to FDA, the degrees of membership of point 6 to groups 1 and 2 are 0.36 and 1,

Table 1
Data of the numerical example

No. of sample	y_1	y_2	y_3	x_1	x_2
1	1	0	0	2	4
2	1	0	0	3	2
3	1	0	0	4	5
4	1	0	0	5	3
5	1	0	0	5	4
6	1	0	0	7	5
7	0	1	0	6	5
8	0	1	0	7	6
9	0	1	0	8	4
10	0	1	0	9	7
11	0	1	0	10	6
12	0	1	0	12	10
13	0	0	1	3	8
14	0	0	1	3	9
15	0	0	1	4	8
16	0	0	1	5	6
17	0	0	1	6	7
18	0	0	1	6	8

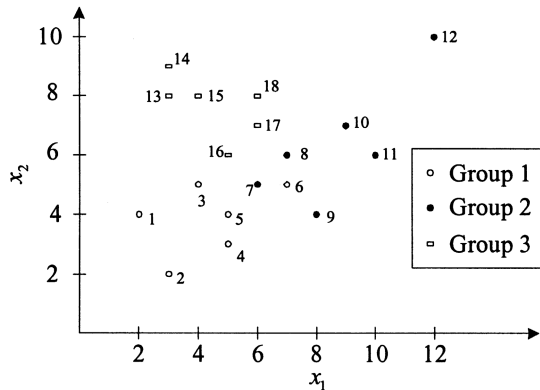


Fig. 3. Scatter plots of the numerical example.

respectively. These results seem more reasonable than those obtained by the distance method. Fig. 3 also shows that point 6 is closer to group 2 than to group 1. This result coincides with the results that the degree of membership of point 6 to group 2 exceeds its degree of membership to group 1.

The results for point 7 in Table 4 differ considerably from the results in Table 3. This point that originally belongs to group 2 is more likely to belong to group 1 than to group 2 before the outlier is removed. The degree of membership of point 7 to group 1 exceeds that to group 2 because the centroid of group 2 is farther from the point than that the centroid of group 1. After the outlier is removed, group 2 becomes more compact, and its centroid becomes closer to point 7. Thus,

Table 2
Primitive results of fuzzy discriminant analysis

No. of sample	μ_{1j}	μ_{2j}	μ_{3j}	Sum of membership degrees
1	0.55	0.00	0.00	0.55
2	0.57	0.00	0.00	0.57
3	0.92	0.00	0.00	0.92
4	0.97	0.00	0.00	0.97
5	1.00	0.00	0.00	1.00
6	0.36	0.65	0.00	1.01
7	0.65	0.32	0.00	0.97
8	0.19	0.82	0.00	1.01
9	0.11	0.54	0.00	0.65
10	0.00	1.00	0.00	1.00
11	0.00	0.95	0.00	0.95
12	0.00	0.00	0.00	0.00
13	0.00	0.00	1.00	1.00
14	0.00	0.00	1.00	1.00
15	0.00	0.00	1.00	1.00
16	0.57	0.05	1.00	1.62
17	0.14	0.41	1.00	1.55
18	0.00	0.26	1.00	1.26

Table 3
Parameters of membership functions

No. of group	a_i	b_i
1	0.9825	4.0000
2	2.0911	2.1503
3	2.3999	2.7099

the degree of membership of point 7 to group 2 becomes larger than its degree of membership to group 1.

The sum of the degrees of membership of a point to all the groups specifies the position of that particular point in relation to other points. The sum of membership degrees need not be unity because were that the case, the FDA method would be unable to identify outliers. Thus, a point is considered an outlier if the sum of its degrees of membership to all groups is 0. A higher sum of degrees of membership implies that the point is closer to the centroid of all points, and is more likely to belong to more than one group simultaneously.

Table 2 presents the parameters of each membership function. Based on these parameters, the membership function of each group can be determined. For example, the membership function of

Table 4
Discriminant results by comparing distance and fuzzy discriminant analysis

No. of sample	Distance comparison			Fuzzy discriminant analysis			Sum of membership degrees
	y'_1	y'_2	y'_3	μ_{1j}	μ_{2j}	μ_{3j}	
1	1	0	0	0.55	0.00	0.00	0.55
2	1	0	0	0.57	0.00	0.00	0.57
3	1	0	0	0.92	0.00	0.00	0.92
4	1	0	0	0.97	0.00	0.00	0.97
5	1	0	0	1.00	0.00	0.00	1.00
6	0	1	0	0.36	1.00	0.00	1.36
7	1	0	0	0.65	1.00	0.00	1.65
8	0	1	0	0.19	1.00	0.00	1.19
9	0	1	0	0.11	1.00	0.00	1.11
10	0	1	0	0.00	1.00	0.00	1.00
11	0	1	0	0.00	1.00	0.00	1.00
12	0	1	0	0.00	0.00	0.00	0.00
13	0	0	1	0.00	0.00	1.00	1.00
14	0	0	1	0.00	0.00	1.00	1.00
15	0	0	1	0.00	0.00	1.00	1.00
16	0	0	1	0.57	0.00	1.00	1.57
17	0	0	1	0.14	0.00	1.00	1.14
18	0	0	1	0.00	0.00	1.00	1.00

group 1 is

$$\mu_1(D_1^2) = \begin{cases} 1, & \text{if } D_1^2 \leq 0.98, \\ \frac{4.00 - D_1^2}{4.00 - 0.98} & \text{if } 0.98 < D_1^2 \leq 4.00, \\ 0 & \text{if } 4.00 < D_1^2. \end{cases} \tag{7}$$

In fact, the fuzzy boundary around group 3 is an empty zone. All points are either within or outside the fuzzy boundary. The minimal distance from the centroid of group 3 to the points that do not belong to group 3 is 2.713 while the maximal distance to the points that do belong to group 3 is 2.007. Consequently, group 3 does not overlap any other group. Notably, parameters a_3 and b_3 obtained by the genetic algorithm are 2.3999 and 2.7099, respectively, rather than 2.007 and 2.713 because the search of driven by the genetic algorithms tends to stall at an optimal point with zero gradient. Nevertheless, the value of the objective function is still the optimal value, 0. Table 3 also indicates that the results for group 3 obtained by fuzzy discriminant analysis are identical to those obtained by crisp discriminant analysis. These analytical results indicate that the proposed FDA method degenerates to a crisp discriminant analysis method, provided that the target group does not overlap the other groups.

Fuzzy c -means is still the most frequently used alternative to fuzzy discriminant analysis to classify fuzzily data groups, even when the memberships of samples are provided, because no method of fuzzy discriminant analysis for crisp data has yet been found in the literature. The sum of

Table 5
Clustering results by fuzzy c -means

No. of sample	μ_{1j}	μ_{2j}	μ_{3j}	Original group	Classification
1	0.73	0.10	0.17	1	1
2	0.80	0.10	0.10	1	1
3	0.78	0.09	0.13	1	1
4	0.88	0.08	0.04	1	1
5	0.89	0.07	0.04	1	1
6	0.09	0.87	0.05	1	2
7	0.35	0.52	0.13	2	2
8	0.04	0.92	0.04	2	2
9	0.16	0.77	0.07	2	2
10	0.08	0.82	0.10	2	2
11	0.11	0.79	0.10	2	2
13	0.03	0.02	0.94	3	3
14	0.05	0.05	0.90	3	3
15	0.00	0.00	1.00	3	3
16	0.35	0.27	0.37	3	3
17	0.15	0.47	0.38	3	2
18	0.12	0.32	0.56	3	3

memberships of a particular sample to all the groups in fuzzy c -means is constrained to be unity. These constrained memberships cause a serious problem that the performance is inadequate when the data is contaminated by noise and outliers. Robust fuzzy clustering methods have been proposed to overcome the problem of outliers, such as those of Ohashi [9], Dave [10,11], Krishnapuram and Keller [12], Frigui and Krishnapuram [13], and Ménard [14]. A review of robust clustering methods is given by Dave and Krishnapuram [15]. These methods depend on different criteria to determine whether a point is an outlier or not. Different criteria often lead to different outliers such that to compare the capabilities of detecting outliers of different methods is sometimes difficult. However, the outlier problem in fuzzy clustering is different from that in FDA because of the existence of the criterion variable. In a clustering problem where the number of clusters is unclear, outliers may either form another cluster or simply be regarded as outliers. But in a discriminant problem where the number of clusters is fixed, outliers that do not belong to any cluster can only be outliers. Thus, comparing the capabilities of detecting outliers between fuzzy clustering methods and FDA is not very appropriate.

Table 5 presents the results obtained by fuzzy c -means after the outlier, point 12, is removed from the data set. The purpose of so doing is to see what results will be obtained if fuzzy c -means is used instead of FDA. Generally, the results obtained by fuzzy c -means are reasonably consistent with those obtained by FDA, except for the results for points 6 and 17. It is understandable to classify point 6 to group 2, but it seems inappropriate to classify point 17 to group 2. Thus, the results obtained by FDA are comparatively more reasonable than those by fuzzy c -means. On the other hand, Table 3 also shows that the results obtained by fuzzy c -means are fuzzier than those obtained by FDA. This finding can be verified using entropy as a measure of fuzziness. Entropy thus applied is called partition entropy and was adopted by Bezdek [3] to measure cluster validity.

Partition entropy is a scalar defined as

$$H(\tilde{U}, m) = -\frac{1}{n} \sum_{j=1}^n \sum_{i=1}^m \mu_{ij} \ln(\mu_{ij}), \quad (8)$$

where \tilde{U} is a fuzzy partition matrix. Larger partition entropy corresponds to fuzzier partition. The partition entropy of fuzzy discriminant analysis is 0.1413 while the partition entropy of fuzzy c -means is 0.5754. The larger partition entropy obtained by fuzzy c -means indicates that the results obtained using fuzzy c -means are fuzzier than those obtained using fuzzy discriminant analysis. Less fuzzy partition results provide more information about the structure of clusters than do fuzzier partition results. Consequently, FDA outperforms fuzzy c -means provided that the criterion variable values of the data points are known beforehand since the former provide less fuzzy partition results.

5. Conclusions

This study proposed a novel approach to fuzzy discriminant analysis for groups of crisp data. The groups of data may overlap. Fuzzy discriminant analysis considers that a point that lies in an overlapping area and is classified as belonging to one group may also belong to another group, to a greater or lesser extent. The extent, to which a point belongs to a certain group, or its membership degree, is unobtainable by conventional crisp discriminant analysis but obtainable by fuzzy discriminant analysis. However, minimizing the classification error is difficult because the objective function is complicated nonlinear. This study shows that genetic algorithms can represent a good approach to solving this problem. The numerical example shows that the proposed fuzzy discriminant method can determine the membership function for each data group when it overlaps another group. However, when a group does not overlap any other group, the results of discrimination are identical to those obtained by crisp discriminant analysis.

Comparing the partition results obtained by fuzzy discriminant analysis to those obtained by fuzzy c -means reveals that fuzzy discriminant analysis provides clearer structural information concerning the clusters than does fuzzy c -means. Additionally, another advantage of fuzzy discriminant analysis is its capability of identifying outliers. When a point is sufficiently far from the others such that its degree of membership to each group is zero, the point is then regarded as an outlier. Notably, the membership functions must be determined again after the outliers are removed from a group. Crisp discriminant analysis is an important dependent method for multivariate analysis and has numerous applications. All the existing applications of crisp discriminant analysis can be potential applications of fuzzy discriminant analysis, which needs further investigation.

Acknowledgements

The authors would like to thank the anonymous reviewers for their constructive comments that help to improve the presentation of this paper.

References

- [1] Fisher RA. The use of multiple measurement in taxonomic problems. *Annals of Eugenics* 1936;7:179–88.
- [2] Mahalanobis PC. On the generalized distance in statistics. *Proceedings of the National Institute of Science, Calcutta*, 1936;12:49–55.
- [3] Bezdek JC. *Pattern recognition with fuzzy objective function algorithms*. New York: Plenum Press, 1981.
- [4] Watada J, Tanaka H, Asai K. Fuzzy discriminant analysis in fuzzy groups. *Fuzzy Sets and Systems* 1986;19: 261–71.
- [5] Holland JH. *Adaptation in natural and artificial systems*. Ann Arbor: University of Michigan Press, 1975.
- [6] Goldberg DE. *Genetic algorithms in search, optimization and machine learning*. Reading, MA: Addison-Wesley, 1989.
- [7] DeJong K. *An analysis of the behavior of a class of genetic adaptive systems*. Ph.D. thesis, University of Michigan, USA, 1975.
- [8] Gen M, Cheng R. *Genetic algorithms and engineering design*. New York: Wiley, 1997.
- [9] Ohashi Y. Fuzzy clustering and robust estimation. In: *Ninth Meeting of SAS Users Group International*, Florida, 1984.
- [10] Dave RN. Characterization and detection of noise in clustering. *Pattern Recognition Letters* 1991;12:657–64.
- [11] Dave RN. Robust fuzzy clustering algorithms. In: *Proceeding of the Second International Conference on Fuzzy Systems*, San Francisco, 1993; p. 1281–6.
- [12] Krishnapuram R, Keller JM. A possibilistic approach to clustering. *IEEE Transactions on Fuzzy Systems* 1993;1: 98–110.
- [13] Frigui H, Krishnapuram R. A robust algorithm for automatic extraction of an unknown number of clusters from noisy data. *Pattern Recognition Letters* 1996;17:1223–32.
- [14] Ménard M. Fuzzy clustering and switching regression models using ambiguity and distance rejects. *Fuzzy Sets and Systems* 2001;122:363–99.
- [15] Dave RN, Krishnapuram R. Robust clustering methods: a unified view. *IEEE Transactions on Fuzzy Systems* 1993;5:270–93.