Fuzzy c-Varieties with Different Dimensionalities

Kazutaka Umayahara and Yoshiteru Nakamori

December 25, 1997

ISE-TR-97-149

Fuzzy c-Varieties with Different Dimensionalities

Kazutaka Umayahara
Institute of Information Sciences and Electronics, University of Tsukuba
1-1-1 Tenno-dai, Tsukuba, Ibaraki, 305, Japan

Yoshiteru Nakamori
Department of Applied Mathematics, Konan University
8-9-1 Okamoto, Higashinada-ku, Kobe 658, Japan

Abstract

One of the recent interests in the field of fuzzy clustering is the simultaneous determination of a fuzzy partition of a given data set and parameters of assumed models of different shapes that explain respective partitioned data sets. This paper proposes a new objective function to improve existing approaches for detecting linear varieties with different dimensionalities. Since this is not an all-purpose method, some techniques will be suggested by using artificial examples to show how to implement the clustering successfully.

1 Introduction

This paper suggests a fuzzy clustering approach to finding linear varieties with *different dimensionalities.* The origin of clustering method to be considered is the fuzzy c-means (FCM) method developed by Bezdek[1], the ideas of which are traced back to Ruspini[2].

One of the recent interests in the field of fuzzy clustering is the simultaneous determination of a fuzzy partition of a given data set and parameters of assumed models of different shapes. Gustafson and Kessel[3] introduced a volume constraint to the FCM algorithm to obtain different local substructures. Bezdek et al.[4] introduced and analyzed a family of objective functions that replace the centers of clusters with linear varieties of arbitrary dimensions. This approach is called the fuzzy c-varieties (FCV). The combination of FCM and FCV was also discussed in Bezdek et al.[4], and later called the fuzzy c-elliptotypes (FCE) in Dave[5], in which the adaptive fuzzy c-elliptotype algorithm (AFC) is developed, which modifies the weights of combination adaptively in the process of clustering.

Hathaway and Bezdek[6] introduced a family of objective functions, called the fuzzy c-regression models (FCRM), to fit switching regression models to certain types of mixed data. Minimization of particular objective functions in the family yields simultaneous estimates for the parameters of regression models, together with a fuzzy partition of the data. Nakamori and Ryoke[7] modified the FCRM in such a way that the shapes of clusters are changed dynamically and adaptively in the clustering process in order to detect a better data partition for developing fuzzy prediction models, following the idea in Dave[5]. The FCRM approach has been extended to detect nonlinear substructures, for instance, in Dave and Bhaswan[8], Krishnapuram, Frigui and Nasraoui[9][10].

There is another school of fuzzy modeling originated in Takagi and Sugeno[11] that is interested in developing fuzzy implication inference models which consist of a number of if-then rules, where the premises are represented by fuzzy propositions and the consequences are given by linear equations. The main purpose of the above fuzzy clustering methods is explanation of the given data, whereas the fuzzy implication inference model is oriented to be used in prediction of the underlying system. The original idea for identification of a fuzzy implication inference model in Sugeno and Kang[12] is to obtain a fuzzy partition of the input space and linear equations simultaneously. After this work, a number of identification algorithms were developed, for instance, in Sugeno and Tanaka[13], Nakamori and Ryoke[14].

In all above works, one of the unsolved problems is the detection of clusters with *different dimensionalities*. This paper proposes a new objective function and an algorithm for detecting clusters with different dimensionalities. But, since the result by this algorithm depends on the clustering parameters and the initial condition, some ideas to obtain satisfactory results will be suggested by using artificial examples.

2 Preliminaries

Let $\{x_1, x_2, \cdots, x_m\}$ be the set of variables, $\{x_{1i}, x_{2i}, \cdots, x_{ni}\}$ the standardized data set of x_i, and $\boldsymbol{w}_k$ the k-th data vector of all variables:

$$\boldsymbol{w}_k = (x_{k1}, x_{k2}, \cdots, x_{km})^\top, \qquad k = 1, 2, \cdots, n \tag{1}$$

where $(\cdot)^\top$ indicates the transposition. The problem is to obtain c fuzzy clusters C_1, C_2, $\cdots$, C_c by partitioning the data set:

$$S = \{\boldsymbol{w}_1, \boldsymbol{w}_2, \cdots, \boldsymbol{w}_n\}. \tag{2}$$

Let u_{ik} be the membership value of $\boldsymbol{w}_k$ in the cluster C_i, satisfying the following conditions:

(i) $\quad 0 \le u_{ik} \le 1, \quad i = 1, 2, \cdots, c, \quad k = 1, 2, \cdots, n,$

(ii) $\quad \sum_{k=1}^{n} u_{ik} > 0, \quad i = 1, 2, \cdots, c,$

(iii) $\quad \sum_{i=1}^{c} u_{ik} = 1, \quad k = 1, 2, \cdots, n.$

The FCM algorithm[1] determines the membership matrix $U = (u_{ik})$ and the vectors of cluster centers $V = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_c\}$ simultaneously by minimizing the function:

$$J_{fcm}(U, V) = \sum_{k=1}^{n} \sum_{i=1}^{c} (u_{ik})^q \; d_{ik}(\boldsymbol{v}_i), \tag{3}$$

where

$$d_{ik}(\boldsymbol{v}_i) = \|\boldsymbol{w}_k - \boldsymbol{v}_i\|^2, \tag{4}$$

and q is the smoothing parameter, a real number greater than one.

The FCV algorithm[4] detects r-dimensional $(0 \leq r < m)$ linear varieties in $\boldsymbol{R}^m$:

$$V_i^r = \left\{ \boldsymbol{z} \in \boldsymbol{R}^m \mid \boldsymbol{z} = \boldsymbol{v}_i + \sum_{j=1}^{r} t_{ij}\, \boldsymbol{p}_{ij}\ ,\ t_{ij} \in \boldsymbol{R} \right\}, \quad i = 1, 2, \cdots, c \tag{5}$$

where $\boldsymbol{p}_{ij}$ $(j = 1, 2, \cdots, r)$ are linearly independent vectors in $\boldsymbol{R}^m$. The first r normalized eigenvectors $\{\boldsymbol{e}_{ij}\}$ of the fuzzy scatter matrix:

$$S_i = \sum_{k=1}^{n} (u_{ik})^q\ (\boldsymbol{w}_k - \boldsymbol{v}_i)(\boldsymbol{w}_k - \boldsymbol{v}_i)^\top, \qquad i = 1, 2, \cdots, c \tag{6}$$

are usually used for $\boldsymbol{p}_{ij}$ $(j = 1, 2, \cdots, r)$. Here, the normalized eigenvectors $\boldsymbol{e}_{i1}$, $\boldsymbol{e}_{i2}$, $\cdots$, $\boldsymbol{e}_{im}$ correspond to the ordered eigenvalues of S_i:

$$\lambda_{i1} \geq \lambda_{i2} \geq \cdots \geq \lambda_{im}. \tag{7}$$

The square distance between $\boldsymbol{w}_k$ and V_i^r is given by

$$D_{ik}^r(\boldsymbol{v}_i) = \|\boldsymbol{w}_k - \boldsymbol{v}_i\|^2 - \sum_{j=1}^{r} |< \boldsymbol{w}_k - \boldsymbol{v}_i,\ \boldsymbol{e}_{ij} >|^2, \qquad \|\boldsymbol{e}_{ij}\| = 1, \tag{8}$$

where $< \cdot, \cdot >$ denotes the inner product. The FCV algorithm searches the minimum of the function:

$$J_{fcv}^r(U, V) = \sum_{k=1}^{n} \sum_{i=1}^{c} (u_{ik})^q\ D_{ik}^r(\boldsymbol{v}_i). \tag{9}$$

Since a linear variety is extended to infinity, there is a possibility that some cluster contains two widely-separated groups of data points. To get over this problem, a convex combination of the FCM and the FCV is suggested in Bezdek et al.[4]:

$$J_{fce}^r(U, V) = (1 - \alpha)\, J_{fcm}(U, V) + \alpha\ J_{fcv}^r(U, V), \quad 0 \leq \alpha \leq 1, \quad 0 \leq r < m. \tag{10}$$

This is the criterion of the FCE which takes into account the continuity and linearity of the data distribution at the same time.

The AFC algorithm[5] determines the parameter α in (10) locally and adaptively. The local parameters are defined by

$$\alpha_i = 1 - \frac{\lambda_{im}}{\lambda_{i1}}, \qquad i = 1, 2, \cdots, c. \tag{11}$$

The objective function of AFC is then defined by

$$J_{afc}^r(U, V) = \sum_{k=1}^{n} \sum_{i=1}^{c} (u_{ik})^q \left[(1 - \alpha_i) d_{ik}(\boldsymbol{v}_i) + \alpha_i D_{ik}^r(\boldsymbol{v}_i) \right], \quad 0 \leq \alpha_i \leq 1, \quad 0 \leq r < m. \tag{12}$$

Thus, the AFC algorithm detects different shapes of clusters with the *same* dimensionality. However, the problem of finding clusters with *different* dimensionalities is left unsolved.

3 Objective Function

This paper proposes the following objective function:

$$J_{fvd}(U,V) = \sum_{k=1}^{n}\sum_{i=1}^{c} (u_{ik})^q \; E_{ik}(\boldsymbol{v}_i) \tag{13}$$

where

$$\text{(a)}\quad E_{ik}(\boldsymbol{v}_i) = \beta_i^0 \, \frac{G_{ik}^0(\boldsymbol{v}_i)}{m} + \sum_{r=1}^{m-1} \beta_i^r \, \frac{G_{ik}^r(\boldsymbol{v}_i)}{m-r}$$

$$\text{(b)}\quad G_{ik}^r(\boldsymbol{v}_i) = \begin{cases} \|\boldsymbol{w}_k - \boldsymbol{v}_i\|^2, & r = 0 \\ \|\boldsymbol{w}_k - \boldsymbol{v}_i\|^2 - \sum_{j=1}^{r} |< \boldsymbol{w}_k - \boldsymbol{v}_i, \boldsymbol{e}_{ij} >|^2, & r = 1, 2, \cdots, m-1 \end{cases}$$

$$\text{(c)}\quad \beta_i^r = \frac{\gamma_i^r}{\sum_{r=0}^{m-1} \gamma_i^r}$$

$$\text{(d)}\quad \gamma_i^r = \begin{cases} (\lambda_{im})^l, & l \geq 0, \quad r = 0 \\ (\lambda_{ir} - \lambda_{i,r+1})^l, & l \geq 0, \quad r = 1, 2, \cdots, m-1 \end{cases}$$

The ideas of using the function (13) are summarized in the following:

- In addition of $G_{ik}^0 (= d_{ik})$ and $G_{ik}^r (= D_{ik}^r)$ for a specified r, the square distances from the data point $\boldsymbol{w}_k$ to the linear varieties with all dimensionalities V_i^r $(r = 1, 2, \cdots, m-1)$ are considered as in (b) to detect clusters with different dimensionalities.

- In order to compare distances from a point to linear varieties with different dimensionalities, the square distance G_{ik}^r is divided by $m - r$ which is the dimensionality of the orthogonal complement of the linear variety V_i^r as shown in (a).

- The weight β_i^r in (c) is defined by using the difference between λ_{ir} and $\lambda_{i,r+1}$ as shown in (d). If β_i^r is relatively large, the possibility of the dimensionality of the cluster C_i being r is relatively high.

- The exponent l in (d) is called *the degree of linearity.* If $l = 0$, the shapes of all clusters become vague. On the other hand, if $l = 1$, the shapes of clusters are expected to reflect the data distribution. A real number greater than one can be used for l to stress the linearity in the data distribution.

4 Clustering Algorithm

The clustering is a process to determine

$$\{u_{ik}, \; \boldsymbol{v}_i; \; i = 1, 2, \cdots, c, \; k = 1, 2, \cdots, n\}$$

that minimize the objective function with assumed parameters c, q, and l. Traditionally, the fuzzy clustering assumes a group of parameters and searches the rest of parameters by the necessary conditions of optimality, which are given in the following for the present case:

1. When the membership matrix $U = (u_{ik})$, the eigenvalues $\{\lambda_{ij}\}$ and the eigenvectors $\{\boldsymbol{e}_{ij}\}$ of the fuzzy scatter matrices are given, the objective function becomes

$$J_1(V) = \sum_{k=1}^{n} \sum_{i=1}^{c} (u_{ik})^q \; E_{ik}(\boldsymbol{v}_i). \tag{14}$$

The necessary condition is

$$\frac{\partial J_1}{\partial \boldsymbol{v}_i} = \sum_{k=1}^{n} (u_{ik})^q \left[\frac{\beta_i^0}{m} \frac{\partial G_{ik}^0(\boldsymbol{v}_i)}{\partial \boldsymbol{v}_i} + \sum_{r=1}^{m-1} \frac{\beta_i^r}{m-r} \frac{\partial G_{ik}^r(\boldsymbol{v}_i)}{\partial \boldsymbol{v}_i} \right] = \boldsymbol{0}. \tag{15}$$

Letting

$$\frac{\beta_i^0}{m} = a_i, \qquad \sum_{r=1}^{m-1} \frac{\beta_i^r}{m-r} \left(I + \sum_{j=1}^{r} \boldsymbol{e}_{ij} \boldsymbol{e}_{ij}^\top \right) = A_i, \tag{16}$$

where I is the $m \times m$ unit matrix, we have

$$(a_i I + A_i) \sum_{k=1}^{n} (u_{ik})^q \; (\boldsymbol{w}_k - \boldsymbol{v}_i) = \boldsymbol{0}. \tag{17}$$

The matrix $a_i I + A_i$ is nonsingular because its orthogonal entries have the maximum absolute values in the respective columns. Then, we have

$$\boldsymbol{v}_i = \frac{\sum_{k=1}^{n} (u_{ik})^q \; \boldsymbol{w}_k}{\sum_{k=1}^{n} (u_{ik})^q}, \qquad i = 1, 2, \cdots, c. \tag{18}$$

2. When $\{\boldsymbol{v}_i\}$ and $\{\beta_i^r\}$ are given, the objective function becomes

$$J_2(U) = \sum_{k=1}^{n} \sum_{i=1}^{c} (u_{ik})^q \; E_{ik} + \sum_{k=1}^{n} \mu_k \left(\sum_{i=1}^{c} u_{ik} - 1 \right), \tag{19}$$

where $\mu_1, \mu_2, \cdots, \mu_n$ are the Lagrange multipliers. From the necessary condition of optimality, we have

$$u_{ik} = \left[\sum_{j=1}^{c} \left(\frac{E_{ik}}{E_{jk}} \right)^{\frac{1}{q-1}} \right]^{-1}, \qquad i = 1, 2, \cdots, c, \quad k = 1, 2, \cdots, n. \tag{20}$$

Here, if $E_{ik} = 0$ for some k,

$$u_{ik} = \begin{cases} \dfrac{1}{\#\{j \mid E_{jk} = 0\}}, & E_{ik} = 0 \\ 0, & E_{ik} > 0 \end{cases} \tag{21}$$

where $\#\{\cdot\}$ indicates the number of elements of a set.

The clustering algorithm is given in the following:

Step 1. Let $t = 0$. Assume the values of parameters c, q, l and the stopping parameter ε. Assume a set of initial membership values by randomization:

$$\left\{u_{ik}^{(t)};\ i = 1, 2, \cdots, c,\ k = 1, 2, \cdots, n\right\}.$$

Step 2. Compute the cluster centers by

$$\boldsymbol{v}_i^{(t)} = \frac{\sum_{k=1}^{n} \left(u_{ik}^{(t)}\right)^q \boldsymbol{w}_k}{\sum_{k=1}^{n} \left(u_{ik}^{(t)}\right)^q}, \qquad i = 1, 2, \cdots, c.$$

Step 3. Compute the eigenvalues $\left\{\lambda_{ij}^{(t)}\right\}$ and the corresponding eigenvectors $\{\boldsymbol{e}_{ij}\}$ of the fuzzy scatter matrix:

$$S_i^{(t)} = \sum_{k=1}^{n} \left(u_{ik}^{(t)}\right)^q \left(\boldsymbol{w}_k - \boldsymbol{v}_i^{(t)}\right)\left(\boldsymbol{w}_k - \boldsymbol{v}_i^{(t)}\right)^{\mathsf{T}}, \qquad i = 1, 2, \cdots, c.$$

Step 4. Renew the membership values by

$$u_{ik}^{(t+1)} = \left[\sum_{j=1}^{c} \left(\frac{E_{ik}^{(t)}}{E_{jk}^{(t)}}\right)^{\frac{1}{q-1}}\right]^{-1}, \qquad i = 1, 2, \cdots, c, \quad k = 1, 2, \cdots, n.$$

If $E_{ik}^{(t)} = 0$ for some k,

$$u_{ik}^{(t+1)} = \begin{cases} \dfrac{1}{\#\{j \mid E_{jk} = 0\}}, & E_{ik} = 0 \\ 0, & E_{ik} > 0. \end{cases}$$

Step 5. If the condition

$$\max_{i,k} \left\{\left|u_{ik}^{(t+1)} - u_{ik}^{(t)}\right|\right\} < \varepsilon$$

holds, then stop. Otherwise, let $t = t + 1$ and go to **Step 2.**

5 Numerical Examples

5.1 Example 1

The first example uses a two-dimensional data in Dave[5]. The result shown in Fig.1(a) is obtained by using the parameters:

$$c = 5,\ q = 1.8,\ l = 0.8,\ \varepsilon = 0.0001.$$

Every point is classified into one cluster in which it has the largest membership value. The behaviors of the function J_{fvd} and the value:

$$\delta = \max_{i,k} \left\{\left|u_{ik}^{(t+1)} - u_{ik}^{(t)}\right|\right\}$$

are shown in Fig.1(b), in which

- the figures in the top indicate the number of iterations,
- the figures in the left side correspond to the values of δ, and
- the figures in the right side correspond to the values of J_{fvd}.

Another result shown in Fig.2.(a) is obtained by changing the degree of linearity l form 0.8 to 0.5. Other parameters and the initial clusters are the same as previous ones. It is understandable that a smaller value of the degree of linearity l leads unsatisfactory clusters. The value of J_{fvd} is worse as shown in Fig.2(b).

When the number of clusters c is four, the result is out of the question as shown in Fig.3(a). The problem here is that the value of J_{fvd} in Fig.3(b) is almost the same as the first case shown in Fig.1(b). This means that for higher dimensional cases, a cluster validity criterion or a visual appealing technique should be developed, which is, however, left for future study.

5.2 Example 2

The second example uses a three-dimensional artificial data. The result shown in Fig.4(a) is obtained by using the parameters:

$$c = 4, \ q = 1.8, \ l = 0.8, \ \varepsilon = 0.0001.$$

Four line-like clusters are detected and the value of J_{fvd} is quite well as shown in Fig.4(b). Clusters seen from different angles are shown in Fig.4(c) and Fig.4(d). The cluster validity can be checked visually up to three dimensional cases.

Even when the degree of linearity l is 0.2 instead of 0.8, the clusters shown in Fig.5(a) are not bad visually, but the value of J_{fvd} is apparently worse as shown in Fig.5(b).

It is interesting to see what happens when the number of clusters c is assigned to other values. For the cases when $c = 5$, $c = 4$, $c = 3$, and $c = 2$, the clustering results are shown in Fig.6(a), Fig.6(b), Fig.6(c), and Fig.6(d), respectively. Here, starting with $c = 5$, we reduce the number of clusters one by one with the following strategy:

- For each cluster C_i, calculate

$$z_i = \sum_{k=1}^{n} u_{ik}, \quad i = 1, 2, \cdots, c. \tag{22}$$

- For the cluster C_i with the smallest z_i, let

$$u_{ik} = 0, \quad k = 1, 2, \cdots, n. \tag{23}$$

- Replace c with $c - 1$, and start again the clustering by using the current membership values instead of randomization.

This strategy tentatively spoils the conditions which the membership values should satisfy. But, of course, the next iteration recovers losses.

The behavior of the objective function J_{fvd} is shown in Fig.6(e). Since the value of J_{fvd} is changed in a large way from $c = 4$ to $c = 3$, we can conclude without seeing Fig.6(a) to Fig.6(d) that the appropriate number of clusters is four.

5.3 Example 3

The third example uses another three-dimensional artificial data: a line-like cluster and a plane-line cluster cross each other in the three dimensional space. The result shown in Fig.7(a) is obtained by setting:

$$c = 2,\ q = 1.5,\ l = 0.8,\ \varepsilon = 0.0001.$$

The behaviors of the function J_{fvd} and the value:

$$\delta = \max_{i,k} \left\{ \left| u_{ik}^{(t+1)} - u_{ik}^{(t)} \right| \right\}$$

are shown in Fig.7(b). Clusters seen from different angles are shown in Fig.7(c) and Fig.7(d). In these figures, data points included in the plane-like cluster are indicated by black points.

A failed example is shown in Fig.8(a) for which a different set of initial membership values is used. The algorithm finally detects two planes instead of one line and one plane. We can understand that this result is not acceptable by Fig.8(b) in which the value of J_{fvd} becomes worse. But, unfortunately, we cannot understand the behavior of J_{fvd} which goes well at first, but turns to the worse direction after all.

The strong dependence on the initial condition is unsurprising when using the fuzzy clustering algorithm. No one knows which initial state is connected with which final state. This requires a heuristics for successful application.

6 Parameters

Some suggestions for determining the clustering parameters and the initial membership values are summarized in the following:

1. The number of clusters c. There exist some proposals to determine an optimal value of c by defining cluster validity criteria. But, in the present study, we repeat the clustering from a larger value of c to a smaller one, and compare results by the membership values and the value of the objective function. One strategy for reducing the number of clusters was suggested in Section 5.2.

2. The smoothing parameter q. This parameter should be close to one from the viewpoint of data classification. But, when it is close to one, the result heavily depends on the initial condition, and the shapes of clusters do not change variously. It is recommended to start with a larger q, then make it smaller gradually. An automatic modification of this parameter is left for future study.

3. The degree of linearity l. This parameter determines the shapes of clusters. When $l = 0$, the shapes of clusters become vague. On the other hand, when $l = 1$, the shapes depend on the data distribution. It is recommended to repeat the clustering by changing this parameter from a smaller value to a larger one.

4. The stopping parameter ε. The algorithm does not guarantee a monotonous convergence in membership values because the shapes of clusters are changing in the process of clustering. A smaller value of ε is recommended for this reason.

5. The initial membership values. They are determined by randomization in the present study. It is unavoidable that different initial values usually lead different clustering results. We have no choice but to repeat the whole clustering process with different initial values. If the results are unstable for initialization, one should change other parameters.

7 Conclusion

This paper proposed a new objective function and a fuzzy clustering algorithm for detecting clusters with *different dimensionalities.* The new proposal is, however, not a magic. It takes a step forward, we believe, but does not always work well. Some techniques for determining the clustering parameters were suggested in order to implement the clustering successfully.

Acknowledgments

This research has partly been supported by TARA(Tsukuba Advanced Research Alliance), University of Tsukuba.

References

[1] J. C. Bezdek: *Fuzzy Mathematics in Pattern Classification.* PhD Thesis, Cornell University, 1973.

[2] E. H. Ruspini: A new approach to clustering. *Information and Control,* Vol. 15, No. 1, pp. 22-32, 1969.

[3] D. E. Gustafson and W. C. Kessel: Fuzzy clustering with a fuzzy covariance matrix. *In:* Proc. of IEEE CDC, pp. 761-766, San Diago, CA, 1979.

[4] J. C. Bezdek et al.: Detection and characterization of cluster substructure II. fuzzy c-varieties and convex combinations thereof. *SIAM J. Appl. Math.,* Vol. 40, No. 2, pp. 358-372, 1981.

[5] R. N. Dave: An adaptive fuzzy c-elliptotype clustering algorithm. *In:* Proc. of NAFIPS 90: Quarter Century of Fuzziness, Vol. I, pp. 9-12, 1990.

[6] R. J. Hathaway and J. C. Bezdek: Switching regression models and fuzzy clustering. *IEEE Trans. on Fuzzy Systems*, Vol. 1, No. 3, pp. 195-204, 1993.

[7] Y. Nakamori and M. Ryoke: Adaptive fuzzy clustering for fuzzy modeling. *In*: Proc. of IFSA'95, Vol. II, pp. 65-68, Sao Paulo, Brazil, July 22-28, 1995.

[8] R. N. Dave and K. Bhaswan: Adaptive fuzzy c-shells clustering and detection of ellipses. *IEEE Trans. on Neural Networks*, Vol. 3, No. 5, pp. 643-662, 1992.

[9] R. Krishnapuram, H. Frigui and O. Nasraoui: Quadratic shell clustering algorithms and their applications. *Pattern Recognition Letters*, Vol. 14, No. 7, pp. 545-552, 1993.

[10] R. Krishnapuram, H. Frigui and O. Nasraoui: Fuzzy and possibilistic shell clustering algorithms and their application to boundary detection and surface approximation: Part I and II. *IEEE Trans. on Fuzzy Systems*, Vol. 3, No. 1, pp. 44-60, 1995.

[11] T. Takagi and M. Sugeno: Fuzzy identification of systems and its applications to modeling and control. *IEEE Trans. on Systems, Man and Cybernetics*, Vol. 15, No. 1, pp. 116-132, 1985.

[12] M. Sugeno and G. T. Kang: Structure identification of fuzzy model. *Fuzzy Sets and Systems*, Vol. 28, pp. 15-33, 1988.

[13] M. Sugeno and K. Tanaka: Successive identification of a fuzzy model and its applications to prediction of a complex system. *Fuzzy Sets and Systems*, Vol. 42, pp. 315-334, 1991.

[14] Y. Nakamori and M. Ryoke: Identification of fuzzy prediction models through hyperellipsoidal clustering. *IEEE Trans. on Systems, Man and Cybernetics*, Vol. 24, No. 8, pp. 1153-1173, 1994.

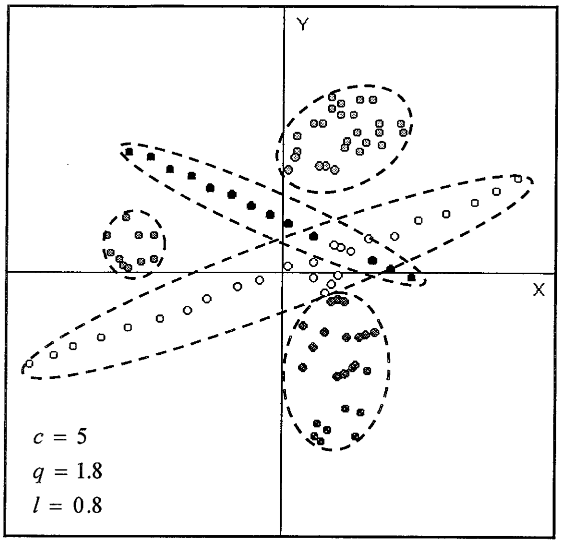

Fig.1(a) A successful result for the data in [5].

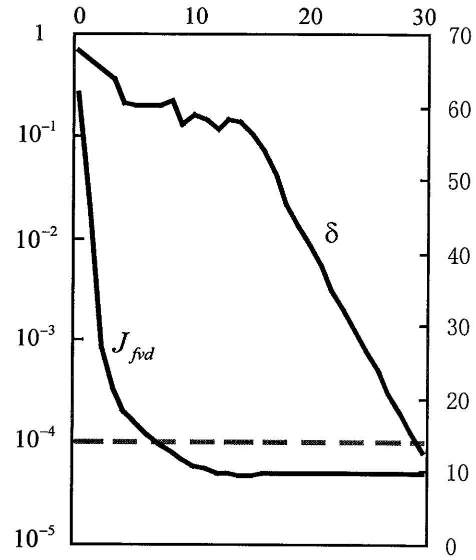

Fig.1(b): Behaviors of J_{fvd} and δ in detecting clusters in Fig.1(a).

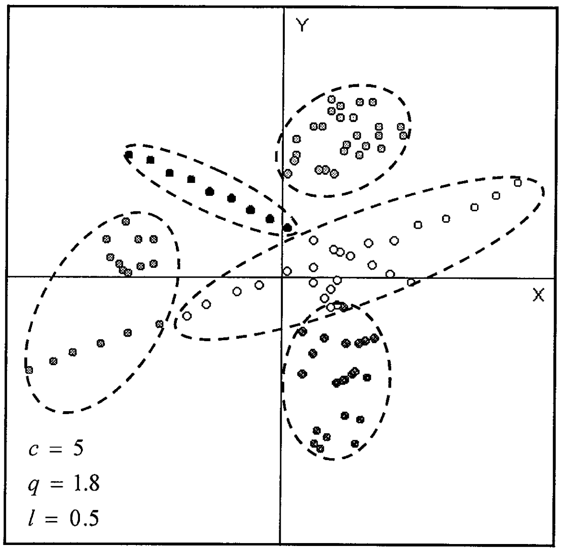

Fig.2(a) A failed example for the same data in Fig.1(a).

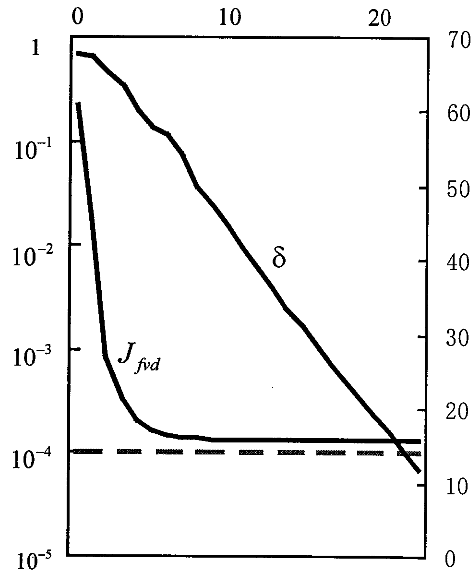

Fig.2(b) Behaviors of J_{fvd} and δ in detecting clusters in Fig.2(a).

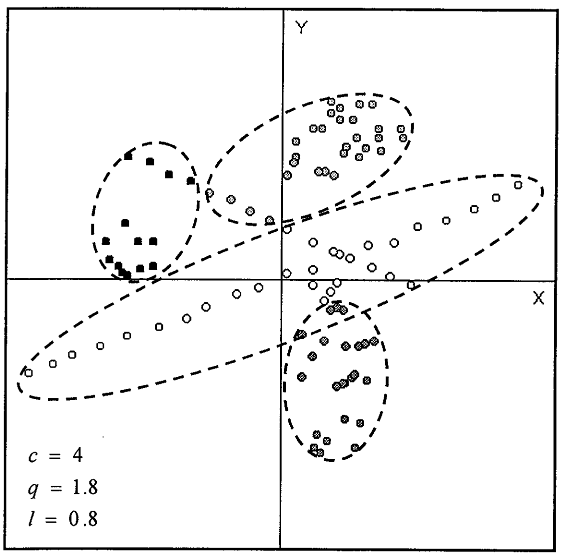

Fig.3(a) Another failed example for the same data in Fig.1(a).

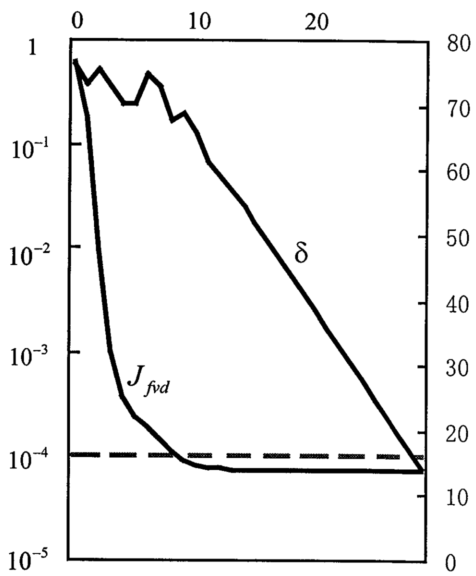

Fig.3(b) Behaviors of J_{fvd} and δ in detecting clusters in Fig.3(a).

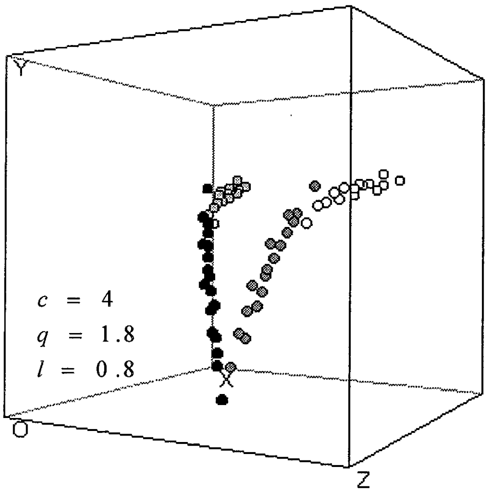

Fig.4(a) A successful result for a set of artificial data.

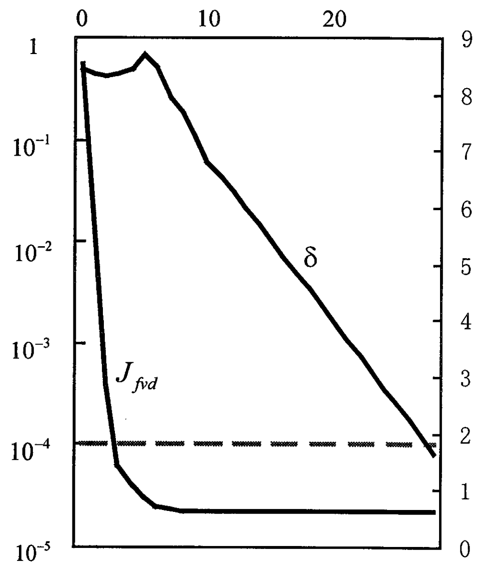

Fig.4(b) Behaviors of J_{fvd} and δ in detecting clusters in Fig.4(a).

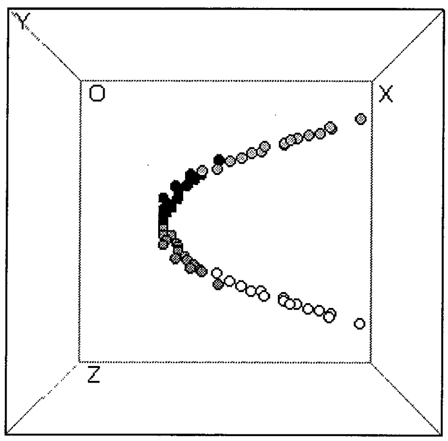

Fig.4(c) The same clusters in Fig. 4(a) seen from a different angle.

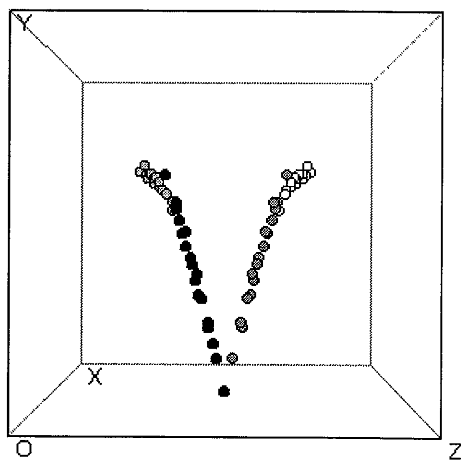

Fig.4(d) The same clusters in Fig. 4(a) seen from another different angle.

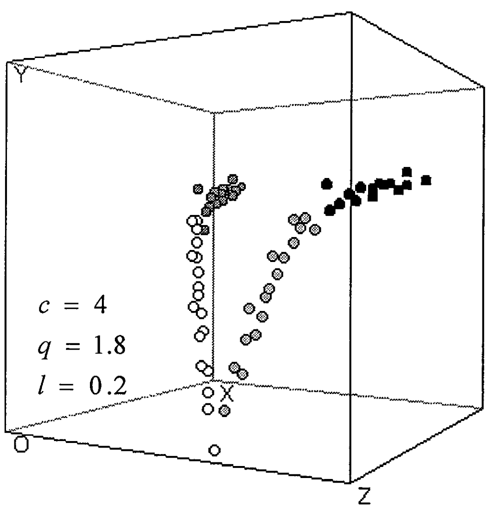

Fig.5(a) Another successful result for the same data in Fig. 4(a).

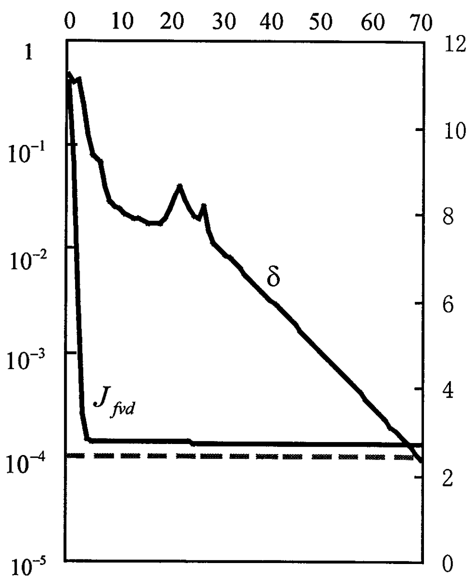

Fig.5(b) Behaviors of J_{fvd} and δ in detecting clusters in Fig.5(a).

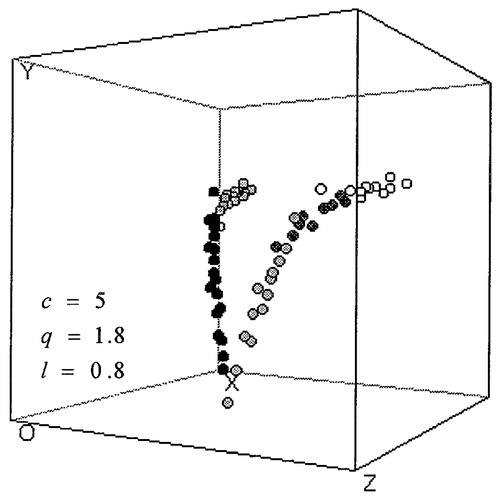

Fig.6(a) Five clusters for the same data in Fig.4(a).

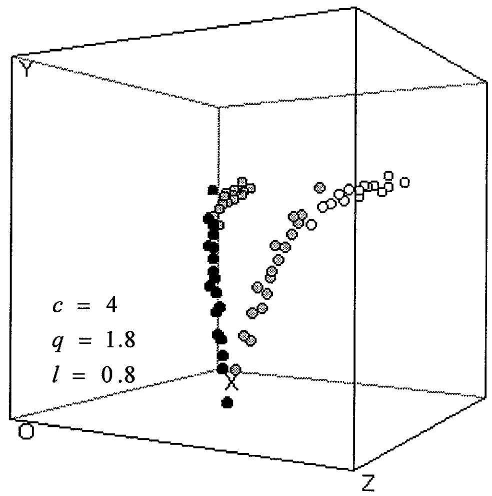

Fig.6(b) The same clusters in Fig.4(a).

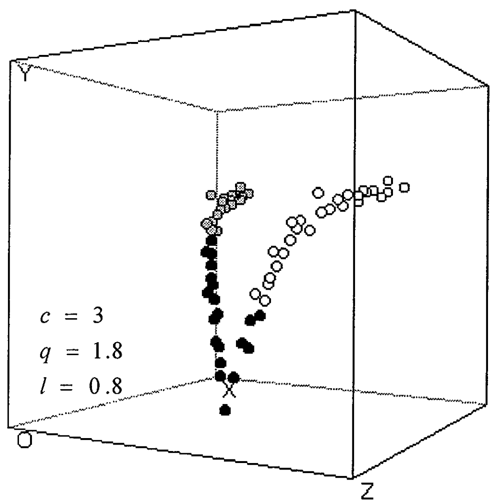

Fig.6(c) Three clusters for the same data in Fig.6(a).

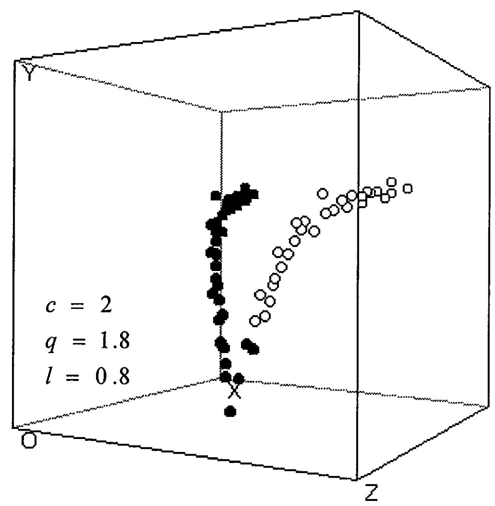

Fig.6(d) Two clusters for the same data in Fig.6(a).

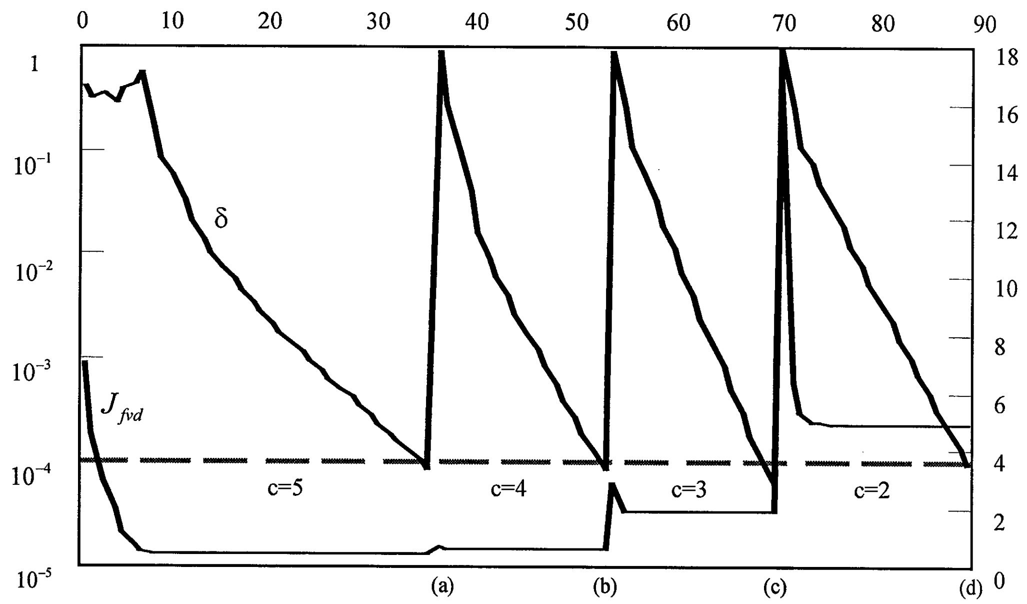

Fig.6(e) Behaviors of J_{fvd} and δ in detecting clusters in Fig.6 (a), (b), (c) and (d).

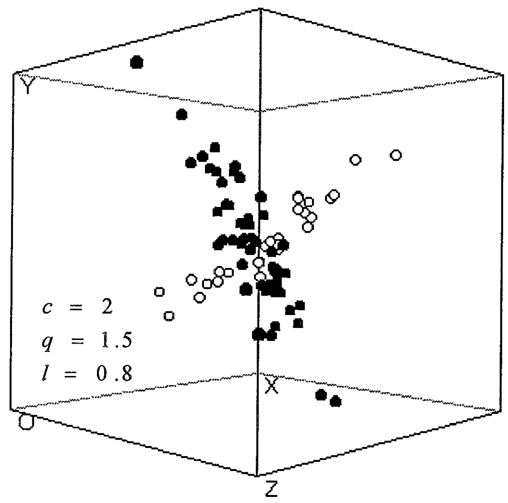

Fig.7(a) A successful result for another artificial data.

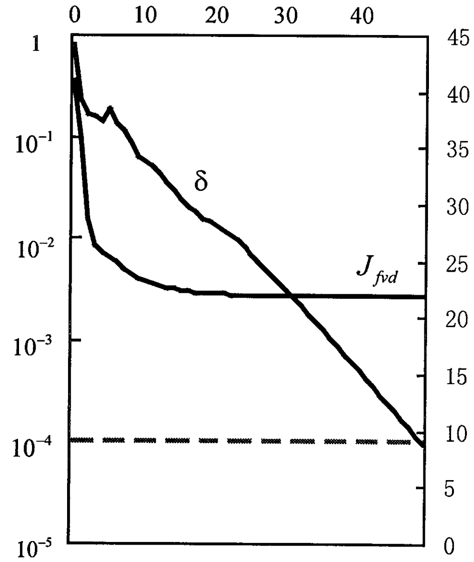

Fig.7(b) Behaviors of J_{fvd} and δ in detecting clusters in Fig.7(a).

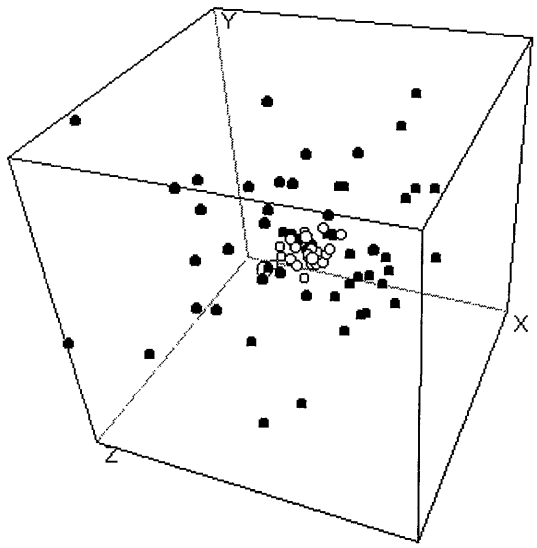

Fig.7(c) The same clusters in Fig.7 (a) seen from a different angle.

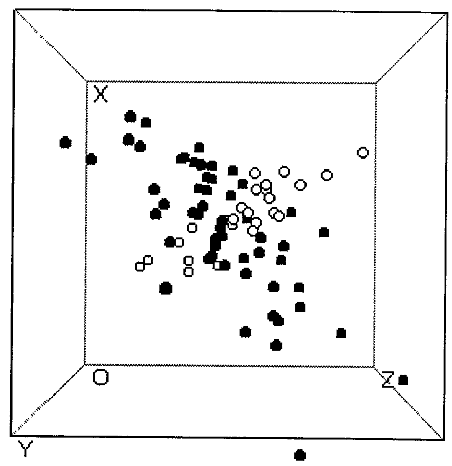

Fig.7(d) The same clusters in Fig.7 (a) seen from another different angle.

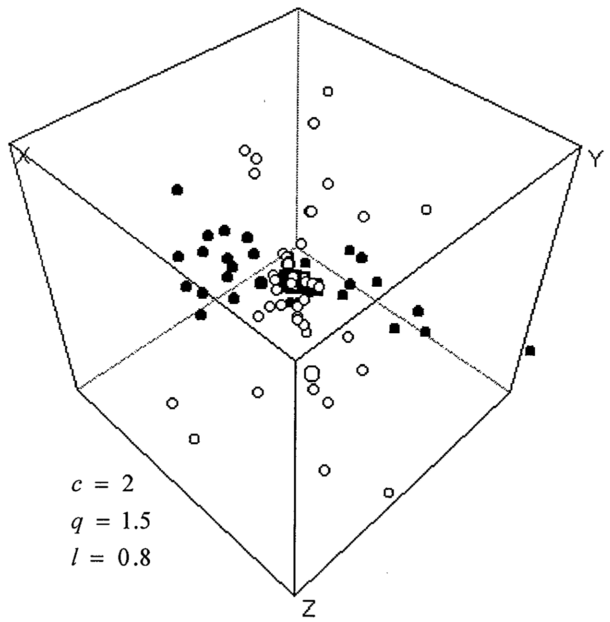

Fig.8(a) A failed result for the same data in Fig.7 (a) obtained by a different initial condition.

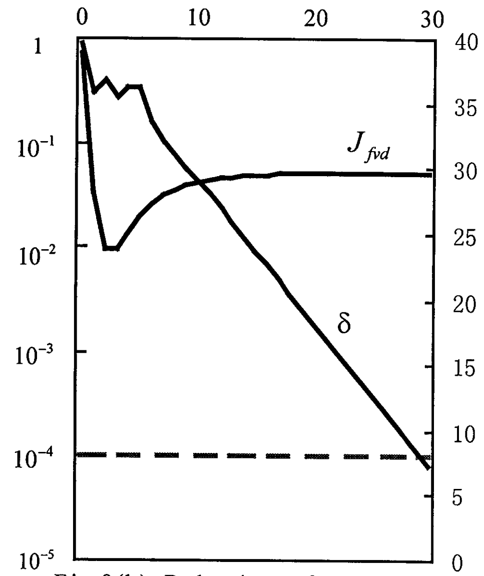

Fig.8(b) Behaviors of J_{fvd} and δ in detecting clusters in Fig.8(a).