Linear Fuzzy Clustering Based on Least Absolute Deviations *

Katsuhiro Honda, Nobuhiro Togo, Taro Fujii and Hidetomo Ichihashi
Graduate School of Engineering, Osaka Prefecture University
1-1 Gakuen-cho Sakai Osaka JAPAN
honda@ie.osakafu-u.ac.jp

Abstract

This paper proposes a technique of linear fuzzy clustering based on least absolute deviations. The novel method partitions a data set into several linear clusters by extracting local minor components. Using the least absolute deviations, the method provides robust clustering that is free from the influences of outliers.

1 Introduction

Fuzzy c-Varieties (FCV) proposed by Bezdek *et al.* [1] [2] is the fuzzy clustering method that partitions a data set into several linear clusters by using linear varieties as the prototypes of the clusters. Because the FCV algorithm estimates the vectors spanning the prototypical linear varieties by solving the eigenvalue problems of the fuzzy scatter matrices, it can be said that the vectors are equivalent to local principal component vectors derived in each cluster and the algorithm performs a simultaneous application of fuzzy clustering and Principal Component Analysis (PCA). In the algorithm, the eigenvectors that correspond to largest eigenvalues of the fuzzy scatter matrix span the prototypical linear varieties.

On the other hand, Minor Component Analysis (MCA) contrasts with PCA and tries to extract an eigenvector corresponding to the least eigenvalue. The first minor component is the normalized linear combination with minimum variance and the minor component vector is useful to estimate the orthonormal basis of subspace or noise subspace. For the extraction of minor components, several algorithms that are mainly associated to neural networks have been proposed [3] [4] [5].

In this paper, we propose a modified linear clustering method that cannot be easily influenced by outliers using minor components. The objective function of FCV can be expressed in a simpler form by considering the extraction of

*Proc. of 2002 IEEE International Conference on Fuzzy Systems, 1444-1449 (2002)

local minor components when the prototypical linear varieties are hyperplanes. The prototypical hyperplane is defined by using the eigenvector corresponding to the least eigenvalue of the fuzzy scatter matrix in each cluster. Then, the squared distances between data points and prototypical hyperplanes are replaced with the absolute deviations. Because the objective function is expressed in the simpler form, the squared distances can be easily replaced with the absolute deviations. In the novel algorithm, the calculation of the optimal local minor component vectors is reduced to the constrained minimization problem that can be solved by a certain optimization technique and the optimal cluster centers are estimated using a simple linear search algorithm. Generally, the absolute deviations are useful to estimate the robust model in Regression Analysis [6] [7]. Miyamoto *et al.* [8] proposed to use the absolute deviations in Fuzzy c-Regression Models (FCRM) [9] that can be regarded as a simultaneous application of Regression Analysis and fuzzy clustering. Replacing the squared deviations with the absolute deviations, the calculation of the regression coefficients is reduced to the solution of a set of linear programming problems and the regression model derived in each cluster is robust even though the observations include outliers.

The simplicity of the proposed objective function also makes it possible to handle other difficulties. In real world applications, data sets often include not only outliers but also missing values. While it is difficult to deal with missing values in the application of the original FCV algorithm, the proposed objective function can be extended into incomplete data sets by ignoring the missing coordinates.

Finally, the characteristic properties of the proposed method are shown in numerical examples.

2 Linear Fuzzy Clustering with Minor Components

Because the Fuzzy c-Varieties (FCV) clustering [1] [2] can be regarded as the simultaneous application of Principal Component Analysis (PCA) and fuzzy clustering, it is equivalent to the problem of extracting local minor components considering the memberships of the data points under a certain condition. The goal of the linear fuzzy clustering is to partition the data set into several clusters using local principal component vectors to express local linear structures. FCV is a clustering method that partitions a data set into C linear clusters. Let $X = (x_{ij})$ denotes a $(n \times m)$ data matrix consisting of m dimensional observations of n samples. The objective function of FCV consists of squared distances from data points to p dimensional prototypical linear varieties spanned by the linearly independent unit vectors $\boldsymbol{a}_{cj}$'s and is defined as follows:

$$L_{fcv} = \sum_{c=1}^{C}\sum_{i=1}^{n} u_{ci}^{\theta}\Big\{(\boldsymbol{x}_i - \boldsymbol{b}_c)^\top(\boldsymbol{x}_i - \boldsymbol{b}_c)$$

$$
\begin{aligned}
& -\sum_{j=1}^{p} \boldsymbol{a}_{cj}^{\top} R_{ci} \boldsymbol{a}_{cj} \Big\}, & (1)\\
R_{ci} &= (\boldsymbol{x}_i - \boldsymbol{b}_c)(\boldsymbol{x}_i - \boldsymbol{b}_c)^{\top}, & (2)
\end{aligned}
$$

where $\boldsymbol{b}_c = (b_{c1}, \cdots, b_{cm})$ is the center of the cth cluster and u_{ci} denotes the membership degree of the data point $\boldsymbol{x}_i$ to the cth cluster with the constraint

$$\sum_{c=1}^{C} u_{ci} = 1 \quad ; i = 1, \cdots, n. \tag{3}$$

The weighting exponent θ is added for fuzzification. The larger θ is, the fuzzier the membership assignments are. Because we can derive $\boldsymbol{a}_{cj}$'s as the eigenvectors corresponding to large eigenvalues of the fuzzy scatter matrix, $\boldsymbol{a}_{cj}$'s are local principal component vectors.

In this paper, we capture the local linear structures by using minor components. Minor Component Analysis (MCA) [3] [4] [5] is a useful technique to estimate the frequencies of sinusoidal signals buried in white noise and tries to determine the normalized linear combination with minimum variance. The problem is equivalent to finding the unit normal vector to the hyperplane spanned by the principal component vectors and to finding the eigenvector corresponding to the smallest eigenvalue of covariance matrix. In the following, linear combination $\boldsymbol{a}_{cm}^{\top}(\boldsymbol{x}_i - \boldsymbol{b}_c)$, where $\boldsymbol{a}_{cm}$ is associated with the smallest eigenvalue of fuzzy scatter matrix is called a local minor component. We define the distances between data points and prototypical linear varieties by using the local minor components. In the case of that the prototypes of clusters are hyperplanes ($p = m-1$), the deviations can be calculated by considering the projections onto the minor component vector $\boldsymbol{a}_{cm}$ and are equivalent to the minor component scores of the data points.

$$||\boldsymbol{x}_i - \boldsymbol{b}_c||^2 - \sum_{j=1}^{m-1} |\boldsymbol{a}_{cj}^{\top}(\boldsymbol{x}_i - \boldsymbol{b}_c)|^2 = |\boldsymbol{a}_{cm}^{\top}(\boldsymbol{x}_i - \boldsymbol{b}_c)|^2. \tag{4}$$

Therefore, the objective function of FCV can be transformed as follows:

$$L_{fcv'} = \sum_{c=1}^{C} \sum_{i=1}^{n} u_{ci}^{\theta} |\boldsymbol{a}_{cm}^{\top}(\boldsymbol{x}_i - \boldsymbol{b}_c)|^2. \tag{5}$$

The optimal minor component vector is estimated by using Lagrangean multiplier method, and the problem is reduced to the following eigenvalue problem.

$$S_{fc} \boldsymbol{a}_{cm} = \gamma_c \boldsymbol{a}_{cm}, \tag{6}$$

where γ_c is the Lagrangean multiplier and S_{fc} denotes the generalized fuzzy scatter matrix in cth cluster,

$$S_{fc} = \sum_{i=1}^{n} u_{ci}^{\theta} R_{ci}. \tag{7}$$

Here, $\boldsymbol{a}_{cm}^{\top}\boldsymbol{a}_{cm} = 1$ derives

$$\begin{aligned}\gamma_c &= \boldsymbol{a}_{cm}^{\top} S_{fc} \boldsymbol{a}_{cm} \\ &= \sum_{i=1}^{n} u_{ci}^{\theta} |\boldsymbol{a}_{cm}^{\top}(\boldsymbol{x}_i - \boldsymbol{b}_c)|^2. \end{aligned} \tag{8}$$

Then, the optimal $\boldsymbol{a}_{cm}$ is equivalent to the eigenvector corresponding to the smallest eigenvalue of the generalized fuzzy scatter matrix. In this way, the FCV algorithm can be represented as the technique for the extraction of the local minor components.

3 Application of Least Absolute Deviations to Linear Fuzzy Clustering

Although the FCV algorithm is useful to capture the local linear structures of a multivariate data set, it is easily influenced by outliers because it uses the squared distances as the criterion. In this section, we propose a modified linear clustering method that cannot be easily influenced by outliers using the absolute deviations. Since the objective function composed of minor component scores is simpler than that of FCV, it isn't so difficult to introduce the absolute deviations into the objective function. By changing the definition of distances, the problem for obtaining linear clusters is defined as the following minimization problem.

Minimization Problem LAD

$$\min \sum_{c=1}^{C} \sum_{i=1}^{n} \{u_{ci} \frac{|\boldsymbol{a}_{cm}^{\top}(\boldsymbol{x}_i - \boldsymbol{b}_c)|}{\sqrt{\boldsymbol{a}_{cm}^{\top}\boldsymbol{a}_{cm}}} + \lambda u_{ci} \log u_{ci}\} \tag{9}$$

$$\text{s.t.} \quad \sum_{c=1}^{C} u_{ci} = 1 \qquad ; i = 1, \cdots, n \tag{10}$$

$$\sum_{j=1}^{m} |a_{cmj}| = 1 \qquad ; c = 1, \cdots, C \tag{11}$$

Where the first term in the bracket of Eq.(9) represents the absolute distance between the data point and the prototypical hyperplane. For fuzzification, we use entropy regularization [10] and add the second term instead of the weighting exponent in the standard FCV algorithm. The larger λ is, the fuzzier the membership assignments are. The fuzzification technique has several merits, e.g. "singularities" don't occur even if several points are on the prototypes. The constraints for the length of the minor component vectors are replaced with linear functions. The solution algorithm is a fixed-point iteration scheme as in many FCM-type algorithms.

3.1 Calculation of Optimal Local Minor Component Vectors

To obtain the optimal solution for the local minor component vector with fixed u_{ci}'s and $\boldsymbol{b}_c$, we consider the minimization of the objective function in each cluster. Introducing slack variables s_j^- and r_i^- and surplus variables s_j^+ and r_i^+, the minimization problem in cth cluster is represented as follows.

Minimization Problem ACM

$$\begin{aligned}
\min \quad & \left\{\sum_{j=1}^{m}(s_j^+ - s_j^-)^2\right\}^{-\frac{1}{2}} \sum_{i=1}^{n} u_{ci}(r_i^+ + r_i^-) && (12)\\
\text{s.t.} \quad & \sum_{j=1}^{m}(s_j^+ - s_j^-)(x_{ij} - b_{cj}) - r_i^+ + r_i^- = 0 && (13)\\
& \sum_{j=1}^{m}(s_j^+ + s_j^-) = 1 && (14)\\
& s_j^+, s_j^-, r_i^+, r_i^- > 0 && (15)\\
& i = 1, \cdots, n
\end{aligned}$$

Where

$$a_{cmj} = s_j^+ - s_j^-.$$

The goal of the problem is to minimize the absolute deviations $(r_i^+ + r_i^-)$ and maximize the length of the minor component vector $(\sum_{j=1}^{m}(s_j^+ - s_j^-)^2)$ simultaneously. Eq.(14) is the same constraint as Eq.(11) when the products between s_j^+ and s_j^- equal zero, i.e. the maximization of the length of the minor component vector is responsible for the constraint (11). The optimal local minor component vectors are derived by using a certain optimization technique such as interior point method.

3.2 Calculation of Optimal Cluster Centers

To derive the optimal cluster center $\boldsymbol{b}_c$ with fixed u_{ci}'s and $\boldsymbol{a}_{cm}$, **Minimization Problem LAD** is solved in each cluster. Here, the objective function to be minimized is represented as the weighted sum of the l_1 errors and has no constraint about $\boldsymbol{b}_c$.

$$J_c(\boldsymbol{b}_c) = \sum_{i=1}^{n} \frac{u_{ci}}{\sqrt{\boldsymbol{a}_{cm}^\top \boldsymbol{a}_{cm}}} |\boldsymbol{a}_{cm}^\top(\boldsymbol{x}_i - \boldsymbol{b}_c)| \qquad (16)$$

In the case of Fuzzy c-Means (FCM) [1], the l_1 norm based methods have been studied by several researchers [11] [12] and Miyamoto *et al.* [13] proposed to use a linear search algorithm for calculating cluster centers in the iteration

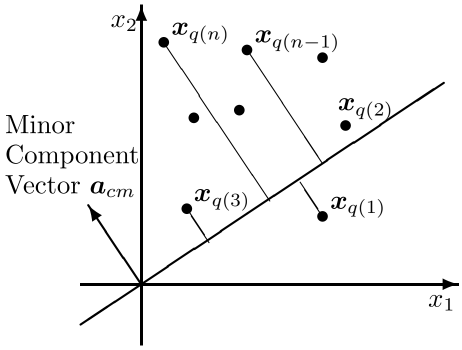

Figure 1: Ordering of sample data points

procedure. The algorithm is based on the property that at least one of the jth coordinates of the points $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n$ is the optimal solution.

In the same way, we can use the property that at least one of $\boldsymbol{x}_i$'s is on the optimal prototype. Assume that $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n$ is ordered and subscripts are changed using a permutation function $q(k), k = 1, \cdots, n$ as shown in Fig. 1 where

$$\boldsymbol{a}_{cm}^{\top}\boldsymbol{x}_{q(1)} \leq \boldsymbol{a}_{cm}^{\top}\boldsymbol{x}_{q(2)} \leq \cdots \leq \boldsymbol{a}_{cm}^{\top}\boldsymbol{x}_{q(n)}. \tag{17}$$

After the ordering, the optimization problem is reduced to finding the weighted median. Here, the cluster center has the arbitrariness like that of the FCV algorithm, i.e. the absolute deviations are invariant with respect to $\boldsymbol{b}_c$ so long as $\boldsymbol{b}_c$ is on the prototypical hyperplane. Because l_2 norm is invariant with respect to an orthonormal basis, the FCV algorithm adopted the weighted average as the cluster center [14]. However, l_1 norm isn't invariant with respect to an orthonormal basis. Then, we apply the following algorithm that derives a unique solution in a orthonormal system where $\boldsymbol{a}_{cm}$ is a orthonormal basis vector.

Algorithm BC

begin
  $\boldsymbol{b}_c := \boldsymbol{0}$;
  $\boldsymbol{v}_0 := \boldsymbol{a}_{cm}$;
  $j := 0$;
  while $(j < m)$ do begin
    ORTHONORMALIZING($\boldsymbol{v}_j$);
    ORDERING(X);
    $S := -\sum_{i=1}^{n} u_{ci}$;
    $r := 0$;

```
        while (S < 0) do begin
            r := r + 1;
            S := S + 2u_{cq(r)};
        end;
        b_c := b_c + (v_j^⊤ x_{q(r)}) v_j;
        j = j + 1;
        v_j = x_{q(r)};
    end;
    output b_c;
end.
```

Where ORTHONORMALIZING($\boldsymbol{v}_j$) denotes the subroutine that derives an orthonormal basis vector using the Gram-Schmidt's orthogonalization and ORDERING(X) performs the ordering of the subscripts $q(k)$'s so that

$$\boldsymbol{v}_j^\top \boldsymbol{x}_{q(1)} \leq \boldsymbol{v}_j^\top \boldsymbol{x}_{q(2)} \leq \cdots \leq \boldsymbol{v}_j^\top \boldsymbol{x}_{q(n)}. \tag{18}$$

Although the ordering of the sample data points is necessary in each iteration, the calculation of cluster centers is achieved by the simple and fast linear search algorithm.

3.3 Calculation of Memberships and Proposed Algorithm

The memberships can be derived by using Lagrangean multiplier method like the FCV algorithm and the new u_{ci} is given by

$$u_{ci} = \frac{\exp(-\lambda^{-1} E_{ci})}{\sum_{l=1}^{C} \exp(-\lambda^{-1} E_{li})}, \tag{19}$$

where

$$E_{li} = \frac{|\boldsymbol{a}_{lm}^\top (\boldsymbol{x}_i - \boldsymbol{b}_l)|}{\sqrt{\boldsymbol{a}_{lm}^\top \boldsymbol{a}_{lm}}}. \tag{20}$$

The following three-step algorithm is used to obtain the (local) optimal solution.

Step 1, Initialize the memberships u_{ci}'s and the local minor component vector $\boldsymbol{a}_{cm}$ randomly in each cluster and normalize them so that they satisfy the constraints.

Step 2, Initialize the cluster center $\boldsymbol{b}_c$ as follows:

$$\boldsymbol{b}_c = \frac{\sum_{i=1}^{n} u_{ci}\boldsymbol{x}_i}{\sum_{i=1}^{n} u_{ci}}.$$

Step 3, Solve **Optimization Problem ACM** to derive the optimal $\boldsymbol{a}_{cm}$ in each cluster.

Step 4, Calculate $\boldsymbol{b}_c$ by using **Algorithm BC** in each cluster.

Step 5, Update the memberships u_{ci}'s using Eq.(19).

Step 6, If

$$\max_{i,c} \mid u_{ci}^{NEW} - u_{ci}^{OLD} \mid < \epsilon,$$

then stop. Otherwise, return to **Step 3**.

4 Handling missing values

In many real world applications, data sets often include missing values. Miyamoto *et al.* [15] proposed some approaches that can handle missing values in FCM. A basic strategy is to replace a missing value by the weighted average of the corresponding attribution. Another simple approach is to ignore the missing values and calculate the distances from the remaining coordinates. Timm *et al.* [16] proposed similar techniques and reported that the simple approach ignoring the missing values gave fuzzier membership assignments than the basic strategy of replacing the missing values. However the FCV algorithm cannot deal with missing values without the transformation of the objective function. Then several techniques that can extract local principal components from incomplete data sets including missing values have been proposed [17] [18]. In these techniques, the modified objective functions are defined by ignoring missing coordinates. In this section, we consider the extraction of local minor components by ignoring missing coordinates. The simplicity of the proposed objective function also makes it possible to handle the difficulties.

Ignoring missing coordinates, **Minimization Problem LAD** is rewritten as follows.

Minimization Problem LADM

$$\min \qquad \sum_{c=1}^{C}\sum_{i=1}^{n}\{u_{ik}\frac{|\sum_{j=1}^{m} d_{ij}a_{cmj}(x_{ij}-b_{cj})|}{\sqrt{\sum_{j=1}^{m} d_{ij}a_{cmj}^{2}}}$$

$$+\lambda u_{ci} \log u_{ci}\} \tag{21}$$

$$\text{s.t.} \quad \sum_{c=1}^{C} u_{ci} = 1 \quad ; i = 1, \cdots, n \tag{22}$$

$$\sum_{j=1}^{m} |a_{cmj}| = 1 \quad ; c = 1, \cdots, C \tag{23}$$

Where

$$d_{ij} = \begin{cases} 1 & ; x_{ij} \text{ is observed.} \\ 0 & ; x_{ij} \text{ is missing.} \end{cases}$$

The optimal $\boldsymbol{a}_{cm}$'s, $\boldsymbol{b}_c$'s and u_{ci}'s are derived by **Minimization Problem ACM**, **Algorithm BC** and Eq.(19) ignoring missing coordinates respectively.

5 Numerical Experiments

We present two numerical examples to illustrate the characteristics of the proposed method.

5.1 Comparison with FCV Algorithm

First, we show the result of the comparison with the FCV algorithm. The artificial data set composed of 100 samples is distributed on the 2-D space forming two lines buried in noise data. Fig.2 shows the result of the analysis with the FCV algorithm where the data set is classified into two linear clusters ($\circ$ and $\times$) and their prototypical lines that pass through the centers ($\blacksquare$) are depicted by the dotted lines. Because the FCV algorithm uses the squared distances as the criterion, the result was influenced by the noise data and could not capture the local linear structures properly. Then we applied our proposed technique to the same data set. The clustering result is shown in Fig.3. In the figure, the local minor component vectors point in the direction depicted by the dotted lines intersecting perpendicularly with the prototypical lines. The prototypes represented the local linear structures properly by using the absolute deviations.

5.2 Handling Missing Values

Next, we analyzed an incomplete data set including missing values. The original data set shown in Fig.4 is composed of 100 samples and includes two local linear structures. Table I shows the clustering result with the original data set. Withholding 20 elements from data matrix, we made an incomplete sample data set where 20% of the samples include a missing value. The result of the analysis is shown in Table II. Even though the data set included missing values, the proposed method could capture the local structure properly.

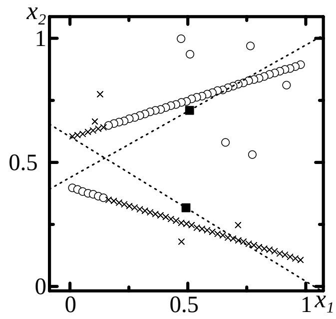

Figure 2: Clustering result of FCV

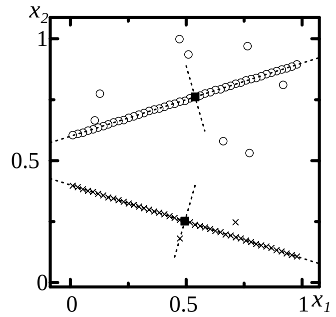

Figure 3: Clustering result of proposed method

6 Conclusions

In this paper we proposed a technique of linear fuzzy clustering that captures the local linear structures of a data set by extracting local minor components in the case of that the prototypes are hyperplanes. Considering minor component vectors instead of principal component vectors, the objective function was defined in the simpler form than that of the original FCV. This simplicity made it possible to introduce the absolute deviations into the objective function and the clustering results could not be easily influenced by outliers. The simplicity of the objective function also enabled us to handle missing values.

However, the proposed method is useful only when the prototypes are hyperplane, and the problems of initialization remain in our technique as well as the FCV algorithm. The future works include the introduction of the absolute deviation to more general situations.

Acknowledgement: This work was supported through the scientific research fund from the Ministry of Education, Culture, Sports, Science and Technology of Japan

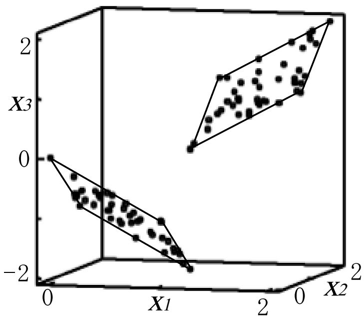

Figure 4: 3D plot of original sample data set

(#13680375)

References

[1] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, 1981

[2] J. C. Bezdek, C. Coray, R. Gunderson and J. Watson, "Detection and Characterization of Cluster Substructure 2. Fuzzy c-Varieties and Convex Combinations Thereof," *SIAM J. Appl. Math.*, Vol.40, No.2, pp 358-372, 1981

[3] E. Oja, "Principal Components, Minor Components and Liner Neural Networks," *Neural Networks*, Vol.5, pp.927-935, 1992

[4] F.-L. Luo, R. Unbehauen and A. Cichocki, "A Minor Component Analysis Algorithm," *Neural Networks*, Vol.10, pp.291-297, 1997

[5] H. Sakai and K. Shimizu, "A New Adaptive Algorithm for Minor Component Analysis," *Signal Processing*, Vol.71, pp.301-308, 1998

[6] F. Y. Edgeworth, "The Law of Error," *Phil. Magazine*, Vol.16, pp.360-375, 1883

[7] Z. Nahorski, "Regression Analysis: a Perspective," in J. Kacprzyk M. Fedrizzi, eds., *Fuzzy Regression Analysis*, Omnitech Press, pp.3-13, 1992

[8] S. Miyamoto, K. Umayahara, T. Nemoto and O. Takata, "Algorithm of Fuzzy c-Regression Based on Least Absolute Deviations," *J. of Japan Society for Fuzzy Theory and Systems*, Vol.12, pp.578-587, 2000 (in Japanese)

[9] 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

[10] S. Miyamoto and M. Mukaidono, "Fuzzy c-Means as a Regularization and Maximum Entropy Approach," *Proc. of the 7th International Fuzzy Systems Association World Congress*, Vol.2, pp.86-92, 1997

[11] K. Jajuga, "L_1 Norm Based Fuzzy Clustering," *Fuzzy Sets and Systems*, Vol.39, pp.43-50, 1991

Table 1: Cluster Centers and Local Minor Component Vectors (Without Missing Value)

$\boldsymbol{b}_c$			
	x_1	x_2	x_3
$c = 1$	0.51	0.47	-0.97
$c = 2$	1.51	1.31	0.82

$\boldsymbol{a}_{cm}$			
	x_1	x_2	x_3
$c = 1$	0.33	0.33	0.33
$c = 2$	0.33	0.33	-0.33

Table 2: Cluster Centers and Local Minor Component Vectors (With Missing Values)

$\boldsymbol{b}_c$			
	x_1	x_2	x_3
$c = 1$	0.47	0.58	-1.06
$c = 2$	1.54	1.31	0.85

$\boldsymbol{a}_{cm}$			
	x_1	x_2	x_3
$c = 1$	0.33	0.33	0.33
$c = 2$	0.33	0.33	-0.33

[12] L. Bobrowski and J. C. Bezdek, "c-Means Clustering with l_1 and l_∞ Norms," *IEEE Trans. on Systems, Man, and Cybernetics*, Vol.21, No.3, pp.545-554, 1991

[13] S. Miyamoto and Y. Agusta, "An Efficient Algorithm for l_1 Fuzzy c-Means and Its Termination," *Control and Cybernetics*, Vol.24, No.4, pp.421-436, 1995

[14] F. Höppner, F. Klawonn, R. Kruse and T. Runkler, *Fuzzy Cluster Analysis*, Jhon Wiley & Sons, 1999

[15] S. Miyamoto, O. Takata and K. Umayahara, "Handling Missing Values in Fuzzy c-Means," *Proc. of the Third Asian Fuzzy Systems Symposium*, pp 139-142, 1998

[16] H. Timm, and R. Kruse, "Fuzzy Cluster Analysis with Missing Values," *Proc. of 17th NAFIPS International conference*, pp 242-246, 1998

[17] K. Honda, N. Sugiura and H. Ichihashi, "Simultaneous Approach to Principal Component Analysis and Fuzzy Clustering with Missing Values," *Proc. of Joint 9th IFSA World Congress and 20th NAFIPS International Conference*, pp.1810-1815, 2001

[18] K. Honda, A. Yamakawa, A. Kanda and H. Ichihashi, "An Application of Fuzzy c-Means Clustering to PCA-like Method for Missing Value Estimation," *Proc. of 16th International Conference on Production Research*, 2001