WEIGHTED PRINCIPAL COMPONENT ANALYSIS BASED ON FUZZY CLUSTERING

MIKA SATO-ILIC

Received May 12, 2003

Abstract. In this paper, we propose a weighted principal component analysis (WPCA) using the result of fuzzy clustering [4]. The principal component analysis (PCA) [1], [7] is one widely used and well-known data analysis method. However there is a problem, when the data does not have a structure that the PCA can capture we cannot obtain any satisfactory results. For the most part, this is due to the uniformity of the data structure, which means we cannot find any significant proportion or accumulated proportion for the obtained principal components.

In order to solve this problem, we use the cluster structure and degree of belongingness of objects to clusters, which is obtained as the fuzzy clustering result. By the introduction of the pre-classification and the degree of belongingness to the data, we can transform the data into a clearer structured data, so avoiding the noise in the data.

1 Introduction Recently, PCA is the method of multivariate analysis which is the most widely used. Due to the particular feature of PCA which can represent the main tendency of an observed data compactly, PCA has been used as a method to follow up a clue when we can not see any significant structure in the data. Beside the importance of PCA as an exploratory method, several extended methods of PCA have been proposed with the viewpoint that PCA can show the ability to the full, by bringing additional considerations of the observed data into the analysis. For example, constrained PCA (CPCA) [11], [12], nonlinear PCA [10], and the time dependent principal component analysis [2] are typical examples of this.

The proposed method in this paper also takes the position that the method is based on the idea that we try to bring PCA's ability into full play by introducing pre-information of the cluster structure of an observed data. The pre-information of the data is represented as weights based on a fuzzy clustering result. That is, the main difference between conventional PCA and WPCA based on fuzzy clustering is the introduction of cluster degree as weights on clusters of the observation space. The weights are represented by the product of the degree of belongingness of objects to fuzzy clusters with respect to the fuzzy clusters when an object is fixed. Due to a property of the algebraic product and the condition of the degree of belongingness, these weights can show how much the data has the clustering structure. According to weighted linear combination, we can show that the estimates of principal components for WPCA based on fuzzy clustering are obtained in a similar way as in conventional PCA. The time dependent principal component analysis [2] also introduces weights to PCA, however, the analysis dose not use the cluster structure of the data. Moreover, a possibilistic regression model [13], the switching regression model [6], and the weighted regression analysis based on fuzzy clustering [9] are among several examples of research that introduced the concept of fuzziness to the regression model.

2000 *Mathematics Subject Classification.* 62H25, 62H30.

Key words and phrases. Fuzzy cluster, Principal component, Weight.

This paper is organized as follows: in section 2, we explain conventional principal component analysis. A fuzzy clustering method used in this paper is mentioned in section 3. In section 4, we propose the weighted principal component analysis based on the fuzzy clustering. Section 5 states several numerical examples for showing the validity and the better performance of the proposed weighted principal component analysis. The conclusion is given in section 6.

2 Principal Component Analysis (PCA) When we obtain the observation of objects with respect to variables, PCA assumes that there are several main components (factors) caused by the relationship of the variables. The purpose of PCA is to obtain the components from the data and capture the feature of the objects. That is, PCA abstracts the data to the main components and finds the feature of the obtained data.

The observed data which are values of p variables with respect to n objects are denoted by the following:

$$X = (x_{ia}), \quad i = 1, \cdots, n, \ a = 1, \cdots, p. \tag{1}$$

Suppose the first principal component $\boldsymbol{z}_1$ which is defined as the following linear combination:

$$\boldsymbol{z}_1 = X\boldsymbol{l}_1, \tag{2}$$

where $\boldsymbol{l}_1^{'} = (l_{11}, l_{12}, \cdots, l_{1p})$.

The purpose of PCA is to estimate $\boldsymbol{l}_1$ which maximizes the variance of $\boldsymbol{z}_1$ under the condition of $\boldsymbol{l}_1^{'}\boldsymbol{l}_1 = 1$. The variance of $\boldsymbol{z}_1$ is as follows:

$$V\{\boldsymbol{z}_1\} = V\{X\boldsymbol{l}_1\} = \boldsymbol{l}_1^{'}V\{X\}\boldsymbol{l}_1 = \boldsymbol{l}_1^{'}\Sigma\boldsymbol{l}_1, \tag{3}$$

where, $V\{\cdot\}$ shows the variance of $\cdot$ and Σ is a variance-covariance matrix of X. Using the Lagrange's method of indeterminate multiplier, the following condition is needed for which $\boldsymbol{l}_1$ has non-trivial solution.

$$|\Sigma - \lambda I| = 0, \tag{4}$$

where λ is an indeterminate multiplier and I is a unit matrix. (4) is the characteristic equation, so λ is obtained as an eigen-value of Σ. Using (4) and the condition $\boldsymbol{l}_1^{'}\boldsymbol{l}_1 = 1$, we obtain

$$\boldsymbol{l}_1^{'}\Sigma\boldsymbol{l}_1 = \lambda\boldsymbol{l}_1^{'}\boldsymbol{l}_1 = \lambda. \tag{5}$$

From (3) and (5), $V\{\boldsymbol{z}_1\} = \lambda$. So, $\boldsymbol{l}_1$ is determined as the corresponding eigen-vector for the maximum eigen-value of Σ.

In order to obtain the second principal component $\boldsymbol{z}_2$, we define the following linear combination:

$$\boldsymbol{z}_2 = X\boldsymbol{l}_2, \tag{6}$$

where $\boldsymbol{l}_2^{'} = (l_{21}, l_{22}, \cdots, l_{2p})$. We need to estimate $\boldsymbol{l}_2$ which maximizes the variance of $\boldsymbol{z}_2$ under the condition of $\boldsymbol{l}_2^{'}\boldsymbol{l}_2 = 1$ and covariance of $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$ is 0, that is, $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$ are mutually uncorrelated. If we denote the covariance between $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$ as $\mathrm{cov}\{\boldsymbol{z}_1, \boldsymbol{z}_2\}$, then this condition is represented as follows:

$$\mathrm{cov}\{\boldsymbol{z}_1, \boldsymbol{z}_2\} = \mathrm{cov}\{X\boldsymbol{l}_1, X\boldsymbol{l}_2\} = \boldsymbol{l}_1^{'}\mathrm{cov}\{X, X\}\boldsymbol{l}_2 = \boldsymbol{l}_1^{'}\Sigma\boldsymbol{l}_2 = \lambda\boldsymbol{l}_1^{'}\boldsymbol{l}_2 = 0.$$

That is, we need to estimate $\boldsymbol{l}_2$ which satisfy $(\Sigma - \lambda I)\boldsymbol{l}_2 = 0$ under the conditions $\boldsymbol{l}_2'\boldsymbol{l}_2 = 1$ and $\boldsymbol{l}_1'\boldsymbol{l}_2 = 0$. From the fact that the eigen-vectors corresponding to the different eigen-values are mutually orthogonal, $\boldsymbol{l}_2$ is found as the eigen-vector corresponding to the second largest eigen-value of Σ. If we get k $(k \leq p)$ different eigen-values of Σ, according to the above, we find the k principal components.

As an indicator which can show how many principal components are needed to explain the data satisfactory, the proportion has been proposed. The proportion of the α-th principal component C_α is defined as:

$$C_\alpha = \frac{\lambda_\alpha}{\mathrm{tr}(\Sigma)}. \tag{7}$$

From $V\{\boldsymbol{z}_\alpha\} = \lambda_\alpha$ and $\sum_{\alpha=1}^{p} \lambda_\alpha = \mathrm{tr}(\Sigma)$, C_α can explain the importance of the α-th principal component. Also, the accumulated proportion until k-th principal components is defined as:

$$P = \sum_{\alpha=1}^{k} C_\alpha. \tag{8}$$

In order to get the interpretation of the obtained principal components, the factor loading $r_{\alpha,j}$ which is defined as a correlation coefficient between the α-th principal component $\boldsymbol{z}_\alpha$ and the j-th variable $\boldsymbol{x}_j$ has been proposed as following:

$$r_{\alpha,j} = \frac{\mathrm{cov}\{\boldsymbol{z}_\alpha, \boldsymbol{x}_j\}}{\sqrt{V\{\boldsymbol{z}_\alpha\}V\{\boldsymbol{x}_j\}}} = \frac{\sqrt{\lambda_\alpha}\, l_{\alpha j}}{\sqrt{\sigma_{jj}}}, \tag{9}$$

where σ_{jj} is variance of $\boldsymbol{x}_j$.

3 Fuzzy Clustering Conventional clustering means classifying the given observation into exclusive subsets (clusters). That is, we can discriminate clearly if an object belongs to a cluster or not. However, such a partition is hardly enough to represent many real situations. Then a fuzzy clustering method is offered to contract clusters with vague boundaries, namely this method allows that one object belongs to some overlapping clusters with some grades. In other words, the essence of fuzzy clustering is to consider not only the belonging status to the assumed clusters, but also to consider how much the objects belong to the clusters. So, there is a merit to representing the complex data situations of real data.

The state of fuzzy clustering is represented by a partition matrix $U = (u_{ik})$ whose elements show the grade of belongingness of the objects to the clusters, $u_{ik},\ i = 1, \cdots, n,\ k = 1, \cdots, K$, where n is number of objects and K is number of clusters. In general, u_{ik} satisfies the following conditions:

$$u_{ik} \in [0, 1], \ \ \sum_{k=1}^{K} u_{ik} = 1. \tag{10}$$

Fuzzy c-means (FCM) [3] is one of the methods of fuzzy clustering. FCM is the method which minimizes the weighted within-class sum of square:

$$J(U, \boldsymbol{v}_1, \cdots, \boldsymbol{v}_K) = \sum_{i=1}^{n}\sum_{k=1}^{K} (u_{ik})^m\, d^2(\boldsymbol{x}_i, \boldsymbol{v}_k), \tag{11}$$

where $\boldsymbol{v}_k = (v_{ka}),\ k = 1, \cdots, K,\ a = 1, \cdots, p$ denotes the value of the centroid of cluster k, $\boldsymbol{x}_i = (x_{ia}),\ i = 1, \cdots, n,\ a = 1, \cdots, p$ is i-th object, and $d^2(\boldsymbol{x}_i, \boldsymbol{v}_k)$ is the square Euclidean distance between $\boldsymbol{x}_i$ and $\boldsymbol{v}_k$. The exponent m which determines the degree of fuzziness of the clustering is chosen from $[1, \infty)$ in advance. The purpose is to obtain the solutions U and $\boldsymbol{v}_1, \cdots, \boldsymbol{v}_K$ which minimize (11), and the solutions are obtained by Picard iteration of the following expressions:

$$u_{ik} = \frac{1}{\sum_{j=1}^{K} \{d(\boldsymbol{x}_i, \boldsymbol{v}_k)/d(\boldsymbol{x}_i, \boldsymbol{v}_j)\}^{\frac{2}{m-1}}}, \quad \boldsymbol{v}_k = \frac{\sum_{i=1}^{n} (u_{ik})^m \boldsymbol{x}_i}{\sum_{i=1}^{n} (u_{ik})^m}, \quad i = 1, \cdots, n; k = 1, \cdots, K.$$

From (11), we can rewrite as:

$$\begin{aligned}
J(U, \boldsymbol{v}_1, \cdots, \boldsymbol{v}_K) &= \sum_{i=1}^{n} \sum_{k=1}^{K} (u_{ik})^m \, d^2(\boldsymbol{x}_i, \boldsymbol{v}_k) \\
&= \sum_{i=1}^{n} \sum_{k=1}^{K} (u_{ik})^m \, (\boldsymbol{x}_i - \boldsymbol{v}_k, \boldsymbol{x}_i - \boldsymbol{v}_k) \\
&= \sum_{i=1}^{n} \sum_{k=1}^{K} (u_{ik})^m \, (\boldsymbol{x}_i - \boldsymbol{h}_k + \boldsymbol{h}_k - \boldsymbol{v}_k, \boldsymbol{x}_i - \boldsymbol{h}_k + \boldsymbol{h}_k - \boldsymbol{v}_k) \\
&= \sum_{i=1}^{n} \sum_{k=1}^{K} (u_{ik})^m \, [(\boldsymbol{x}_i - \boldsymbol{h}_k, \boldsymbol{x}_i - \boldsymbol{h}_k) + 2(\boldsymbol{x}_i - \boldsymbol{h}_k, \boldsymbol{h}_k - \boldsymbol{v}_k) \\
&\qquad + (\boldsymbol{h}_k - \boldsymbol{v}_k, \boldsymbol{h}_k - \boldsymbol{v}_k)],
\end{aligned}$$

where $(\cdot, \cdot)$ denotes real inner product. If we assume

$$\boldsymbol{h}_k = \frac{\sum_{i=1}^{n} (u_{ik})^m \boldsymbol{x}_i}{\sum_{i=1}^{n} (u_{ik})^m},$$

then minimizer of (11) is shown as:

$$J(U) = \sum_{k=1}^{K} \left(\sum_{i=1}^{n} \sum_{j=1}^{n} ((u_{ik})^m (u_{jk})^m d_{ij}) / (2 \sum_{l=1}^{n} (u_{lk})^m) \right), \tag{12}$$

using

$$2 \sum_{l=1}^{n} (u_{lk})^m \sum_{i=1}^{n} (u_{ik})^m (\boldsymbol{x}_i - \boldsymbol{h}_k, \boldsymbol{x}_i - \boldsymbol{h}_k) = \sum_{i=1}^{n} \sum_{l=1}^{n} (u_{ik})^m (u_{lk})^m (\boldsymbol{x}_i - \boldsymbol{x}_l, \boldsymbol{x}_i - \boldsymbol{x}_l),$$

and $d_{ij} = d(\boldsymbol{x}_i, \boldsymbol{x}_j)$. When $m = 2$, (12) is the objective function of the FANNY algorithm. The detail of the FANNY algorithm is shown in [8].

4 Weighted Principal Component Analysis (WPCA) based on Fuzzy Clustering We apply a fuzzy clustering method to the data matrix X shown in (1) and obtain the degree of belongingness $U = (u_{ik}),\ i = 1, \cdots, n,\ k = 1, \cdots, K$. Using the obtained U, we define the weight matrix W as follows:

$$W = \begin{pmatrix} \prod_{k=1}^{K} u_{1k}^{-1} & 0 & \cdots & \cdots & 0 \\ 0 & \prod_{k=1}^{K} u_{2k}^{-1} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & \cdots & 0 & \prod_{k=1}^{K} u_{nk}^{-1} \end{pmatrix} \equiv \begin{pmatrix} w_1 & 0 & \cdots & \cdots & 0 \\ 0 & w_2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & \cdots & 0 & w_n \end{pmatrix}.$$

Then we introduce the following weighted matrix WX:
(13)

$$WX = \begin{pmatrix} w_1 & 0 & \cdots & \cdots & 0 \\ 0 & w_2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & \cdots & 0 & w_n \end{pmatrix} \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \vdots & \vdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix} = \begin{pmatrix} w_1 x_{11} & \cdots & w_1 x_{1p} \\ \vdots & \vdots & \vdots \\ w_n x_{n1} & \cdots & w_n x_{np} \end{pmatrix}.$$

In order to avoid 0^{-1}, we replace (10) as the following condition:

$$u_{ik} \in (0, 1),\ \ \sum_{k=1}^{K} u_{ik} = 1. \tag{14}$$

From the property of the algebraic product and the condition (14), we can see that if $u_{ik} = \dfrac{1}{K}$, for $\exists i,\ \forall k$, then $\prod_{k=1}^{K} u_{ik}^{-1} (= w_i),\ (\exists i)$ takes minimum value. And if u_{ik} is close to 1 for $\exists k, \exists i$, then $\prod_{k=1}^{K} u_{ik}^{-1} (= w_i),\ (\exists i)$ is close to maximum value. That is, the weight w_i shows that if the belonging status of the object i to the clusters is clearer, then the weight becomes larger. Otherwise, if the belonging status of the object i is more vague situation, that is, the cluster structure of the data is close to uniformity, then the weight becomes small. So, WX shows that clearer objects under the cluster structure have larger values and that the objects which are vaguely situated under the clustering are avoided such as those which do not have any significant relation to the clustering for example observations which are treated as noise.

Suppose $\tilde{\boldsymbol{z}}_1$ is the first principal component of the transformed data WX shown in (13). $\tilde{\boldsymbol{z}}_1$ is defined as:

$$\tilde{\boldsymbol{z}}_1 = WX\tilde{\boldsymbol{l}}_1, \tag{15}$$

where $\tilde{\boldsymbol{l}}_1' = (\tilde{l}_{11}, \tilde{l}_{12}, \cdots, \tilde{l}_{1p})$. The purpose of the WPCA based on fuzzy clustering is to estimate $\tilde{\boldsymbol{l}}_1$ which maximizes the variance of $\tilde{\boldsymbol{z}}_1$ under the condition of $\tilde{\boldsymbol{l}}_1'\tilde{\boldsymbol{l}}_1 = 1$. The variance of $\tilde{\boldsymbol{z}}_1$ is as follows:

$$V\{\tilde{\boldsymbol{z}}_1\} = V\{WX\tilde{\boldsymbol{l}}_1\} = V\{\tilde{X}\tilde{\boldsymbol{l}}_1\} = \tilde{\boldsymbol{l}}_1'\tilde{\Sigma}\tilde{\boldsymbol{l}}_1, \tag{16}$$

where, $WX \equiv \tilde{X},\ \tilde{\Sigma} = \tilde{X}'\tilde{X} = (WX)'(WX) = X'WWX$.

Using the Lagrange's method of indeterminate multiplier, the following condition is needed for which $\tilde{\boldsymbol{l}}_1$ has non-trivial solution.

$$|\tilde{\Sigma} - \tilde{\lambda} I| = 0, \tag{17}$$

where $\tilde{\lambda}$ is an indeterminate multiplier and I is a unit matrix. (17) is the characteristic equation, so $\tilde{\lambda}$ is obtained as an eigen-value of $\tilde{\Sigma}$. Using (17) and the condition $\tilde{\boldsymbol{l}}_1^{'}\tilde{\boldsymbol{l}}_1 = 1$, we obtain

$$\tilde{\boldsymbol{l}}_1^{'}\Sigma\tilde{\boldsymbol{l}}_1 = \tilde{\lambda}\tilde{\boldsymbol{l}}_1^{'}\tilde{\boldsymbol{l}}_1 = \tilde{\lambda}. \tag{18}$$

From (16) and (18), $V\{\tilde{\boldsymbol{z}}_1\} = \tilde{\lambda}$. So, $\tilde{\boldsymbol{l}}_1$ is determined as the corresponding eigen-vector for the maximum eigen-value of $\tilde{\Sigma}$. In order to obtain the second principal component $\tilde{\boldsymbol{z}}_2$ for $\tilde{X}$, we define the following linear combination:

$$\tilde{\boldsymbol{z}}_2 = \tilde{X}\tilde{\boldsymbol{l}}_2, \tag{19}$$

where $\tilde{\boldsymbol{l}}_2^{'} = (\tilde{l}_{21}, \tilde{l}_{22}, \cdots, \tilde{l}_{2p})$. We need to estimate $\tilde{\boldsymbol{l}}_2$ which maximizes the variance of $\tilde{\boldsymbol{z}}_2$ under the condition of $\tilde{\boldsymbol{l}}_2^{'}\tilde{\boldsymbol{l}}_2 = 1$ and covariance of $\tilde{\boldsymbol{z}}_1$ and $\tilde{\boldsymbol{z}}_2$ is 0, that is, $\tilde{\boldsymbol{z}}_1$ and $\tilde{\boldsymbol{z}}_2$ are mutually uncorrelated and which is represented as follows:

$$\mathrm{cov}\{\tilde{\boldsymbol{z}}_1, \tilde{\boldsymbol{z}}_2\} = \mathrm{cov}\{X\tilde{\boldsymbol{l}}_1, \tilde{X}\tilde{\boldsymbol{l}}_2\} = \tilde{\boldsymbol{l}}_1^{'}\mathrm{cov}\{\tilde{X}, \tilde{X}\}\tilde{\boldsymbol{l}}_2 = \tilde{\boldsymbol{l}}_1^{'}\tilde{\Sigma}\tilde{\boldsymbol{l}}_2 = \tilde{\lambda}\tilde{\boldsymbol{l}}_1^{'}\tilde{\boldsymbol{l}}_2 = 0,$$

where $\mathrm{cov}\{\tilde{\boldsymbol{z}}_1, \tilde{\boldsymbol{z}}_2\}$ shows covariance between $\tilde{\boldsymbol{z}}_1$ and $\tilde{\boldsymbol{z}}_2$. That is, we need to estimate $\tilde{\boldsymbol{l}}_2$ which satisfy $(\tilde{\Sigma} - \tilde{\lambda} I)\tilde{\boldsymbol{l}}_2 = 0$ under the conditions $\tilde{\boldsymbol{l}}_2^{'}\tilde{\boldsymbol{l}}_2 = 1$ and $\tilde{\boldsymbol{l}}_1^{'}\tilde{\boldsymbol{l}}_2 = 0$. From the fact that the eigen-vectors corresponding to the different eigen-values are mutually orthogonal, $\tilde{\boldsymbol{l}}_2$ is found as the eigen-vector corresponding to the second largest eigen-value of $\tilde{\Sigma}$. If we get k $(k \le p)$ different eigen-values of $\tilde{\Sigma}$, according to the above, we find the k principal components for $\tilde{X}$.

The proportion of α-th principal component for WPCA based on fuzzy clustering, $\tilde{C}_\alpha$, is proposed as follows:

$$\tilde{C}_\alpha = \frac{\tilde{\lambda}_\alpha}{\mathrm{tr}(\tilde{\Sigma})}, \tag{20}$$

where $V\{\tilde{\boldsymbol{z}}_\alpha\} = \tilde{\lambda}_\alpha$ and $\sum_{\alpha=1}^{p} \tilde{\lambda}_\alpha = \mathrm{tr}(\tilde{\Sigma})$. The accumulated proportion until k-th principal components for WPCA based on fuzzy clustering is defined as:

$$\tilde{P} = \sum_{\alpha=1}^{k} \tilde{C}_\alpha. \tag{21}$$

The factor loading $\tilde{r}_{\alpha,j}$ between the α-th principal component $\tilde{\boldsymbol{z}}_\alpha$ and the j-th variable $\tilde{\boldsymbol{x}}_j$ is proposed as the following:

$$\tilde{r}_{\alpha,j} = \frac{\mathrm{cov}\{\tilde{\boldsymbol{z}}_\alpha, \tilde{\boldsymbol{x}}_j\}}{\sqrt{V\{\tilde{\boldsymbol{z}}_\alpha\}V\{\tilde{\boldsymbol{x}}_j\}}} = \frac{\sqrt{\tilde{\lambda}_\alpha}\tilde{\boldsymbol{l}}_{\alpha j}}{\sqrt{\tilde{\sigma}_{jj}}}, \tag{22}$$

where $\tilde{\sigma}_{jj}$ is variance of $\tilde{\boldsymbol{x}}_j$.

5 Numerical Example The data is Fisher iris data [5]. The data consists of 150 samples of iris flowers with respect to four variables, sepal length, sepal width, petal length, and petal width. The samples are observed from three kinds of iris flowers, iris sestosa, iris versicolor, and iris virginica.

Figure 1 shows the result of PCA for the iris data. In this figure, the abscissa shows the values of first principal component shown in (2) and the ordinate is the values of the second principal component shown in (6). "s" means iris sestosa, "c" means iris versicolor, and "v" is iris virginica. From this figure, we can see iris sestosa (the symbol is "s") is clearly divided from iris versicolor and iris virginica (the symbols are "c" and "v").

Figure 2 shows the proportion (shown in (7)) and accumulated proportion (shown in (8)) of the four components. In this figure, the abscissa shows each component and the ordinate shows the values of the proportion of each component. Also the value on the barplot shows the accumulated proportion. From this figure, we can see that it can almost be explained by the first and the second principal components. However, we can not ignore the third principal component completely.

Figure 3 shows the result of WPCA based on fuzzy clustering. As a weight W in (13), we use a fuzzy clustering result which is obtained by using the FANNY algorithm whose objective function is (12) when $m = 2$. In figure 3, the abscissa shows the values of the first principal component and the ordinate is the values of the second principal component. From this figure, we can see a clear distinction between iris sestosa (the symbol is "s") and iris versicolor, iris virginica (the symbols are "c" and "v"), comparing the result shown in figure 1.

Figure 4 shows the proportion (shown in (20)) and accumulated proportion (shown in (21)) of WPCA based on fuzzy clustering result. In this figure, the abscissa shows each principal component and the ordinate is the values of proportion for each principal component. The value on the barplot is the accumulated proportion for each principal component. Through a comparison between figure 2 and figure 4, we can see the higher value of the accumulated proportion until the second principal component in WPCA based on fuzzy clustering (0.991) than the value of the accumulated proportion until the second principal component of PCA (0.958). This shows the higher discrimination ability with WPCA based on fuzzy clustering. Moreover, we can not see any significant meaning for the third principal component from the result of the proportion of WPCA based on fuzzy clustering shown in figure 4. So, we can avoid the noise of the data by the introduction of the weights based on the fuzzy clustering result.

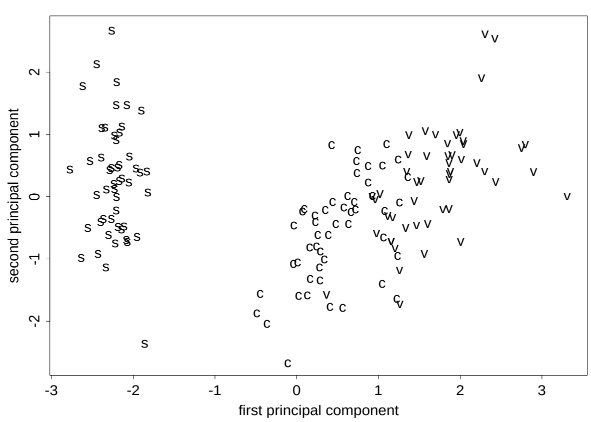

Figure 1 Result of PCA for Iris Data

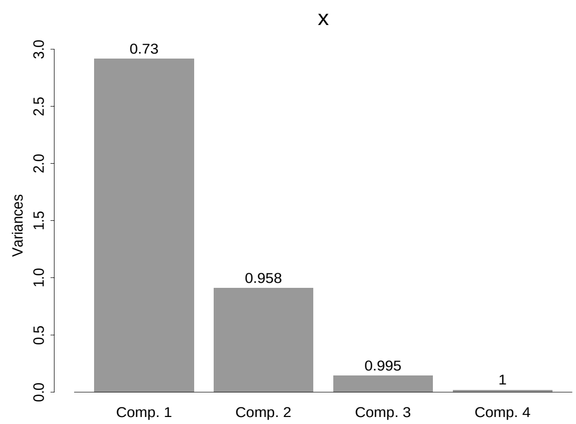

Figure 2 Barplot of Proportion for PCA

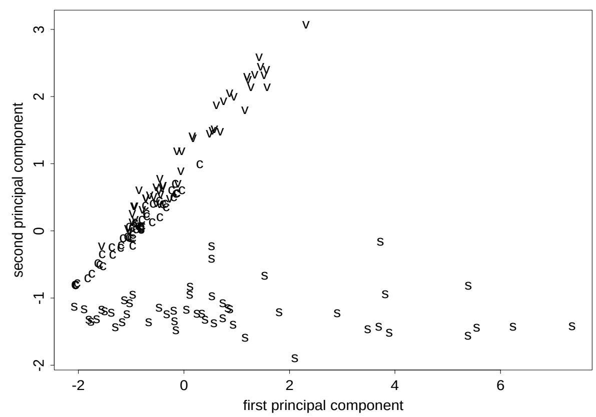

Figure 3 Result of WPCA based on Fuzzy Clustering for Iris Data

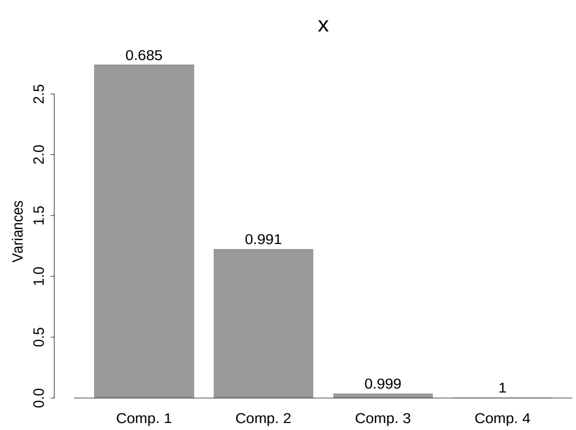

Figure 4 Result of Proportion for WPCA based on Fuzzy Clustering

Tables 1 and 2 show the squared value of λ_α shown in (5) and the squared value of $\tilde{\lambda}_\alpha$ shown in (18), respectively. In these tables comp.1 shows the first principal component, comp.2 shows the second principal component, comp. 3 is the third principal component, and comp. 4 is the forth principal component. From the comparison between the results of tables 1 and 2, we can see the higher value for the second principal component in table 2 than the value for the second principal component in table 1. Moreover, we can see the smaller values for the third and the forth principal components in table 2 as compared to the values for the third and the forth principal components in table 1. From this again, we can see that the WPCA based on fuzzy clustering has a high capability to capture the significance of the latent data structure clearly and tends to avoid the noise of the data, although the value for the first principal component becomes smaller in table 2.

Table 1 Standard Deviations for Principal Components in PCA

Comp.	Comp. 1	Comp. 2	Comp. 3	Comp. 4
SD	1.71	0.96	0.38	0.14

Table 2 Standard Deviations for Principal Components in WPCA based on Fuzzy Clustering

Comp.	Comp. 1	Comp. 2	Comp. 3	Comp. 4
SD	1.66	1.11	0.19	0.05

Moreover, tables 3 and 4 show the results of factor loadings which are shown in (9) and (22), respectively. The values of these tables show the values of the factor loadings which can show the relationship between each principal component and each variable. So, from these results, we can see how each component is explained by the variables. It is related with the interpretation of each component. In these tables, Sepal L. shows sepal length, Sepal W. shows sepal width, Petal L. shows petal length, and Petal W. is the petal width.

Table 3 Factor Loading in PCA

Comp.	Comp. 1	Comp. 2	Comp. 3	Comp. 4
Sepal L.	0.52	0.38	0.72	0.26
Sepal W.	-0.27	0.92	-0.24	-0.12
Petal L.	0.58	0.00	-0.14	-0.80
Petal W.	0.57	0.00	-0.63	0.52

Table 4 Factor Loading in WPCA based on Fuzzy Clustering

Comp.	Comp. 1	Comp. 2	Comp. 3	Comp. 4
Sepal L.	0.57	-0.32	-0.14	0.75
Sepal W.	0.54	-0.4	-0.37	-0.64
Petal L.	0.56	0.31	0.75	-0.15
Petal W.	0.27	0.80	-0.53	0.00

From the comparison between the results of tables 3 and 4, we can see quiet different results. For example, in table 3, the first principal component is mainly explained by the variables, sepal length, petal length, and petal width. However, in table 4, we can see that the first principal component is explained by the variables, sepal length, sepal width, and petal length. Moreover, for the second principal component, we can see the difference, that is, in table 3, the second principal component is represented by the strong relationship of the variable sepal width, but in table 4, we can see the high correlation between the second principal component and the variable petal width. From this, we can see that we can capture the components which have different meaning from conventional PCA by using WPCA based on fuzzy clustering. In other words, the ability to capture the different latent factors of the data is shown by introducing the cluster structure of the data.

6 Conclusion We proposed a weighted principal component analysis using the result of fuzzy clustering. If the cluster structure which are obtained by the fuzzy clustering and the principal component structure are the same, then the two results are essentially the same. However, normally both structures are different, we can obtain the different result from the result of conventional PCA by using the proposed method.

From several numerical examples, we showed higher discrimination ability of the proposed method compared with conventional PCA.

References

[1] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*, Second ed., John Wiley & Sons, (1984).

[2] Y. Baba and T. Nakamura, Time Dependent Principal Component Analysis. Measurement and Multivariate Analysis, S. Nishisato et al. eds., Springer, (2002), 155–162.

[3] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, (1987).

[4] J. C. Bezdek, J. Keller, R. Krisnapuram, and N. R. Pal, *Fuzzy Models and Algorithms for Pattern Recognition and Image Processing*, Kluwer Academic Publishers, (1999).

[5] R. A. Fisher, The Use of Multiple Measurements in Taxonomic Problems, *Ann. Eugenics*, **7(2)**, (1936), 179–188.

[6] R. J. Hathaway and J. C. Bezdek, Switching Regression Models and Fuzzy Clustering, *IEEE Trans. Fuzzy Syst.*, **1(3)**, (1993), 195–204.

[7] I. T. Jolliffe, *Principal Component Analysis*, Springer Verlag, (1986).

[8] L. Kaufman and P. J. Rousseeuw, *Finding Groups in Data*, John Wiley & Sons, (1990).

[9] M. Sato-Ilic and T. Matsuoka, On an Application of Fuzzy Clustering for Weighted Regression Analysis, *The 4th Conference of Asian Regional Section of the International Association for Statistical Computing*, (2002), 55–58.

[10] B. Scholkopf, A. Smola, and K. Muller, Nonlinear Component Analysis as a Kernel Eigenvalue Problem, *Neural Computation*, **10**, (1998), 1299–1319.

[11] Y. Takane and T. Shibayama, Principal Component Analysis with External Information on Both Subjects and Variables, *Psychometrika*, **56**, (1991), 97–120.

[12] Y. Takane, Seiyakutuki Syuseibun Bunseki, Asakura Syoten, (1995) (in Japanese).

[13] H. Tanaka and J. Watada, Possibilistic Linear Systems and Their Application to the Linear Regression Model, *Fuzzy Sets and Systems*, **27**, (1988), 275–289.

Faculty of Systems and Information Engineering
University of Tsukuba
Tennodai 1-1-1, Tsukuba, Ibaraki 305-8573, Japan
E-mail: mika@sk.tsukuba.ac.jp