# Missing Value Estimation by Local Minor Components *

Chi-Hyon Oh[†], Katsuhiro Honda[‡], and Hidetomo Ichihashi[‡]
[†] Faculty of Liberal Arts and Sciences, Osaka University of Economics and Law,
6-10 Gakuonji, Yao, Osaka, 581-8511 Japan
[‡] Graduate School of Engineering, Osaka Prefecture University,
1-1 Gakuen-cho, Sakai, Osaka, 599-8531 Japan

**Abstract**

In this paper, we propose a method of estimating missing values by local minor components. We assume that a data point, some attributes of which are missing, is on a hyper plane. The normal of the hyper plane is a local minor component vector. Thus, the data point is projected onto the hyper plane and its missing attributes are estimated. We extract the local minor components by executing fuzzy clustering and neural principal component analysis (PCA) approach simultaneously. The extracted local minor components take the substructures of the data set into consideration. Missing values are estimated simply and easily.

## 1    Introduction

Generally, principal components are extracted as eigen vectors associated with larger eigen values of a covariance matrix of a given data set. Oja [1] has proposed a technique of extracting principal components by neural networks which conduct Hebbian learning instead of the eigen value problem. He also has shown that the neural PCA approach can be applied for extracting minor components [2].

In this paper, we propose a method of estimating missing values by local minor components. To extract the local minor components, we employ the neural PCA approach and implement the fuzzy clustering [3] at the same time. Partitioning the data set into some clusters enables us to take its substructures into consideration. The local minor components well represent the distinctive structures of the data set. Hence, we can expect our method can achieve good missing value estimation by using the extracted local minor components. In the numerical example, we apply our method to a three dimensional artificial data set and examine its performance.

## 2    Extraction of Local Minor Coponents

Fuzzy $c$-Varieties (FCV) [4] proposed by Bezdek forms fuzzy clusters by minimizing the sum of distances from data points to prototypes. The prototypes of the FCV are multi dimensional linear varieties spanned by some vectors. We denote a data point and a vector spanning the prototypes

---

as follows:

$$\boldsymbol{x}_i = (x_{i1}, \cdots, x_{iK})^\top, \qquad i = 1, \cdots, N. \tag{1}$$

$$\boldsymbol{w}_{ck} = (w_{ck1}, \cdots, w_{ckK})^\top, \qquad c = 1, \cdots, C, \ k = 1, \cdots, K - m. \tag{2}$$

where $N$ and $C$ represents the number of data points and that of clusters respectively. $K$ is the dimensionality of the data points and $m$ is the number of minor component vectors. If we regard $\boldsymbol{w}_{ck}$ as the principal component vector, the FCV can be said to be the simultaneous analysis of the fuzzy clustering and the PCA.

Let us consider minor components now. The clustering criterion of the FCV can be written as follows by the minor component vectors.

$$\sum_{c=1}^{C} \sum_{i=1}^{N} \sum_{k=K-m+1}^{K} u_{ci} \boldsymbol{w}_{ck}^\top (\boldsymbol{x}_i - \boldsymbol{v}_c)(\boldsymbol{x}_i - \boldsymbol{v}_c)^\top \boldsymbol{w}_{ck}. \tag{3}$$

Note that $\boldsymbol{w}_{ck}$ in (3) represents the minor component vector and $\boldsymbol{v}_c$ does the cluster center of the cluster $c$. $u_{ci}$ is membership of the data point to the cluster which represents degree of belonging to the cluster. $u_{ci}$ satisfies the following condition.

$$\boldsymbol{u}_c \in \{(u_{ci}) | \sum_{c=1}^{C} u_{ci} = 1, u_{ci} \in [0, 1], c = 1, \cdots, C\}. \tag{4}$$

Optimization of (3) is resolved into minimization of the following objective function $L$ according to Lagrange's method of indeterminate multiplier.

$$
\begin{aligned}
L &= \sum_{c=1}^{C} \sum_{i=1}^{N} \sum_{k=K-m+1}^{K} u_{ci} \boldsymbol{w}_{ck}^\top (\boldsymbol{x}_i - \boldsymbol{v}_c)(\boldsymbol{x}_i - \boldsymbol{v}_c)^\top \boldsymbol{w}_{ck} \\
&\quad + T_0 \sum_{c=1}^{C} \sum_{i=1}^{N} u_{ci} \log u_{ci} + \sum_{i=1}^{N} T_i \left( \sum_{c=1}^{C} u_{ci} - 1 \right).
\end{aligned}
\tag{5}
$$

In (5), the first term represents the clustering criterion of the FCV. The second term represents entropy maximization as regularization which was introduced in Fuzzy $c$-Means (FCM) by Miyamoto *et al.* [5]. It enables us to obtain fuzzy clusters. The FCM is the most popular fuzzy clustering algorithms. $T_0 > 0$ is the weighting parameter which specifies degree of fuzziness. The remaining terms describe the constraint of the membership $u_{ci}$, *i.e.* (4). $T_i$ is the Lagrangian multiplier.

From the necessary condition $\partial L / \partial v_{ck} = 0$ and $\partial L / \partial u_{ci} = 0$ for the optimality of the objective function $L$, we can calculate cluster centers and memberships as follows:

$$v_{ck} = \sum_{i=1}^{N} u_{ci} \bigg/ x_{ik} \sum_{i=1}^{N} u_{ci}. \tag{6}$$

$$u_{ci} = \exp A_{ci} \bigg/ \sum_{l=1}^{C} \exp A_{li}. \tag{7}$$

where,

$$A_{li} = -\frac{1}{T_0} \sum_{k=K-m+1}^{K} \boldsymbol{w}_{lk}^{\top}(\boldsymbol{x}_i - \boldsymbol{v}_l)(\boldsymbol{x}_i - \boldsymbol{v}_l)^{\top} \boldsymbol{w}_{lk}. \tag{8}$$

We extract the local minor component vectors by the neural PCA approach. From the necessary condition $\partial J/\partial \boldsymbol{w}_{ck}$, we have the following equation.

$$\frac{\partial J}{\partial \boldsymbol{w}_{ck}} = 2 \sum_{i=1}^{N} u_{ci}(\boldsymbol{x}_i - \boldsymbol{v}_c)(\boldsymbol{x}_i - \boldsymbol{v}_c)^{\top} \boldsymbol{w}_{ck}. \tag{9}$$

Taking normalization and orthogonalization into account, we could derive the following learning rule to extract the local minor component vectors.

$$\begin{aligned} \Delta \boldsymbol{w}_{ck} = \gamma u_{ci} \bigg\{ &- y_{ck}(\boldsymbol{x}_i - \boldsymbol{v}_c) \\ &+ \left(y_{ck}^2 + 1 - (\boldsymbol{w}_{ck}^{old})^{\top}\boldsymbol{w}_{ck}^{old}\right)\boldsymbol{w}_{ck}^{old} - \alpha \sum_{l>k} y_{cl}y_{ck}\boldsymbol{w}_{cl}^{old} \bigg\}. \end{aligned} \tag{10}$$

where $\gamma > 0$ denotes the learning rate and $\alpha > 0$ does the coefficient of the orthogonalization. If $\alpha$ is sufficiently large, each local minor component vector is supposed to be orthogonalized.

The solution algorithms for our method are based on an iterative procedure. We remark them below.

*Extracting Algorithms of the Local Minor Components*
Step 1: Set each parameter and the terminal condition $\epsilon > 0$. Initialize memberships.
Step 2: Update cluster centers by using (6).
Step 3: Extract local minor components by using (10).
Step 4: Update memberships by using (7).
Step 5: If $\max |u_{ck}^{NEW} - u_{ck}^{OLD}| < \epsilon$, then stop. Otherwise, return to Step 2.

## 3 Missing Value Estimation by Local Minor Components

We exploit the extracted local minor components by the pre-mentioned algorithms to estimate the missing values. We assume that a data point some attributes of which are missing is on a hyper plane. The normal of the hyper plane is a local minor component vector. Thus, the data point is projected onto the hyper plane and its missing attributes are estimated. To begin with, the data points are decided which hyper plane they are to be projected onto since the data set is fuzzily partitioned. In this study, we employ the maximum membership rule which has the data points belong to the cluster with the maximum membership.

Suppose an attribute $x_{il}$ is missing in the data points $\boldsymbol{x}_i$. To estimate $x_{il}$, we consider a $K - 1$ dimensional hyper plane whose normal is a local minor component vector $\boldsymbol{w}_{cK}$. The linear equation for the hyper plane can be written as follows:

$$(\boldsymbol{x}_i - \boldsymbol{v}_c)^T \boldsymbol{w}_{cK} = 0. \tag{11}$$

If it is assumed that $x_{il}$ is on the hyper plane. From (11), we can derive it by the following equation.

$$x_{il} = -\frac{1}{w_{cl}} \sum_{m=1,m\neq l}^{M} (x_{im} - v_{cm})w_{cm} + v_{cl}. \tag{12}$$
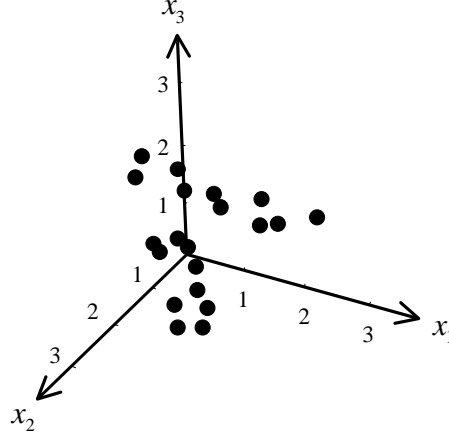
Figure 1: The three dimensional artificial data set.

If two attributes are missing, another hyper plane whose normal is the second minor component vector $\boldsymbol{w}_{cK-1}$ is derived.

$$(\boldsymbol{x}_i - \boldsymbol{v}_c)^T \boldsymbol{w}_{cK-1} = 0 \tag{13}$$

From (11) and (13), we can estimate two missing attributes. Thus, we can estimate any numbers of missing attributes if the same numbers of minor component vectors are extracted.

## 4  Numerical Example

In the numerical example, we apply our method to a three dimensional artificial data set and estimate the missing values. Figure 1 shows the data set used in the numerical example. The data set consists of 20 data points and two out of them are missing one attribute. Besides, the data set has a data point where two attributes are missing. From Figure 1, we can notice that the data set has two substructures, the upper part and lower one. The following parameters are employed.

The number of clusters $C$: 2
Entropy coefficient $T_0$: 0.2
Learning rate for the Hebbian learning $\gamma$: 0.000001
Coefficient of orthogonalization $\alpha$: 10
The terminal condition of iterative procedure $\varepsilon$: 0.001

The results of missing value estimation plotted on the two dimensional diagrams are shown in Figure 2 (a) and (b). The exact values by the estimations are also shown in Table 1. The estimated attributes are written in the bold numbers in Table 1. In Figure 2 (a) and (b), $\triangle$s represent the cluster center $\boldsymbol{v}_c$. $\circ$s and $\bullet$s depict the data points belonging to the two different clusters respectively. As we mentioned in the previous section, we have the data points belong to the cluster with the maximum membership. $\times$s are data points with missing attributes. The data points connected with $\times$s by the dotted lines are the estimated ones by our method. From Figure 2 (a), (b) and Table 1, we can see that good missing value estimation considering substructures of the data set is achieved.
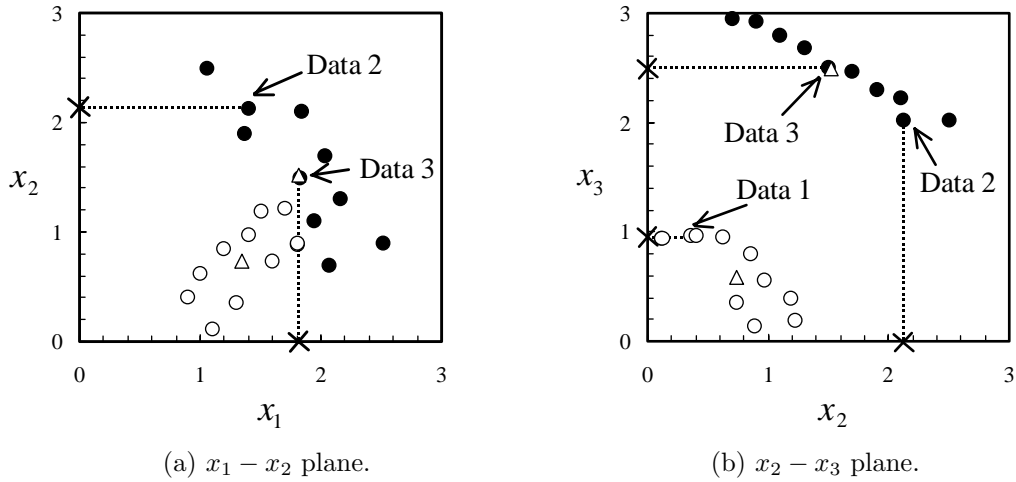
4

(a) $x_1 - x_2$ plane.    (b) $x_2 - x_3$ plane.

Figure 2: The results of missing value estimation.

Table 1: The results of missing value estimation.

|       | $x_1$ | $x_2$ | $x_3$ |
|-------|-------|-------|-------|
| Data1 | 1.30  | 0.36  | **0.97** |
| Data2 | 1.40  | **2.12** | 2.02 |
| Data3 | **1.82** | 1.50 | **2.51** |

# 5   Conclusion

In this paper, we have proposed a method estimating missing values by local minor components. We extracted the local minor components by executing the fuzzy clustering and the neural PCA approach simultaneously. In the numerical example, we applied our method to a three dimensional artificial data set and estimated its missing values. The results showed the good missing value estimation considering substructures of the data set has been done by our method.

# References

[1] E.Oja, A Simplified Neuron Model as a Principal Component Analyzer, Journal of Mathematical Biology 15 (1982) 267-273.

[2] E.Oja, Principal Components, Minor Components, and Linear Neural Networks, Neural Networks 5 (1992) 195-204.

[3] F. Höppner, F. Klawonn, R. Kruse and T. Runkler, *Fuzzy cluster analysis*, John Wiley & Sons, Chichester (1999).

[4] 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. 40, 2 (1981) 358-372.

[5] S. Miyamoto and M. Mukaidono, Fuzzy $c$-Means as a Regularization and Maximum Entropy Approach, Proc. IFSA'97 2 (1997) 86-92.