Journal of the Korean Statistical Society (2002), **31**: *4, pp 519–532*

Projection Pursuit K-Means Visual Clustering

Mi-Kyung Kim[1] and Myung-Hoe Huh[2]

ABSTRACT

K-means clustering is a well-known partitioning method of multivariate observations. Recently, the method is implemented broadly in data mining softwares due to its computational efficiency in handling large data sets. However, it does not yield a suitable visual display of multivariate observations that is important especially in exploratory stage of data analysis. The aim of this study is to develop a K-means clustering method that enables visual display of multivariate observations in a low-dimensional space, for which the projection pursuit method is adopted. We propose a computationally inexpensive and reliable algorithm and provide two numerical examples.

Keywords. K-means clustering, projection pursuit method, visualization.
AMS 2000 subject classifications. Primary 62H30; Secondary 62H09.

1. Introduction

Clustering is a partitioning method of multivariate observations in such a way that the observations within the same group are similar each other and those belonging to different groups are dissimilar as much as possible. Among clustering methods, K-means clustering (Everitt, 1974; Hartigan, 1975) is known to be efficient and reliable (Milligan, 1980 and 1981), so that it is implemented broadly in most statistical and data mining softwares. However, the method does not provide any natural visual display of clustering results, which is very valuable in exploratory data analysis and/or data mining.

There appeared several studies on visual clustering. Kim (1999) and Kim, Kwon and Cook (2000), for instance, proposed a two-stage method that performs hierarchical/non-hierarchical clustering at the first stage and displays the results via multidimensional scaling on a suitable low-dimensional perceptual map at

Received February 2002; accepted September 2002.
[1]Kookmin Credit Card Ltd. Co., Seoul 110-719, Korea
[2]Department of Statistics, Korea University, Seoul 136-701, Korea

the second stage. Recently, Huh and Kim (2000) proposed an iterative procedure that repeats K-means clustering and the dimensional reduction by canonical discriminant analysis, to produce a meaningful visual display of K-means clustering results on the low-dimensional space. Even though it has very handy algorithm, the underlying rationale is indirect and the method lacks some theoretical back-up.

There have been proposed various visualization methods for multidimensional data, such as dynamic scatter plots and grand tours (Becker, Cleveland and Wilks, 1988; Young, Kent and Kuhfeld, 1988; Cook, Buja, Cabrera and Hurley, 1995; Cook and Buja, 1997). However, these tools are not specifically targeted to clustering purpose. On the theory side, Stute and Zhu (1995) studied K-means clustering based on projection pursuit and showed several asymptotic properties such as consistency. Recently, dozens of related papers were published in Korean journals. See Lee, Park and Kim (1995), Jhun and Jin (2000), Huh (2000) and Baek and Sim (2000) in clustering, Ahn and Rhee (1992) and Park, Choi and Koo (2000) in projection pursuit method, and Huh and Song (2001) in visualization.

The aim of this paper is to develop a K-means clustering method that enables more direct visual display of multivariate observations in the low-dimensional linear space, guided by projection pursuit method. We propose a computationally efficient and reliable algorithm, demonstrate it by two numerical examples, and provide several notes.

2. Dimensional Reduction for Visual Clustering

Let $\mathbf{X}$ be an $n \times p$ column-centered matrix of which the rows are observations and the columns are variables. When the p-variate observations are projected on the subspace spanned by two orthogonal unit-norm $p \times 1$ vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, $\mathbf{X}$ is reduced to $\mathbf{X}(\boldsymbol{\alpha}, \boldsymbol{\beta})$ or $\tilde{\mathbf{X}}$. Suppose that each observation is assigned to one of k groups. For the i^{th} observation belonging to the j^{th} group ($j = 1, \ldots, k$), we write $z_{ij} = 1$ and $\mathbf{Z} = (z_{ij})$ which is $n \times k$. Then, on the projected linear subspace, the overall coefficient of determination due to clustering can be expressed as

$$overall R^2_{(2)} = \frac{\text{trace}(\tilde{\mathbf{B}})}{\text{trace}(\tilde{\mathbf{T}})}$$

where

$$\tilde{\mathbf{T}} = \tilde{\mathbf{X}}^t\tilde{\mathbf{X}},\ \tilde{\mathbf{B}} = \tilde{\mathbf{X}}^t\mathbf{Z}(\mathbf{Z}^t\mathbf{Z})^{-1}\mathbf{Z}^t\tilde{\mathbf{X}},\ \tilde{\mathbf{W}} = \tilde{\mathbf{T}} - \tilde{\mathbf{B}} \quad (\tilde{\mathbf{T}} = \tilde{\mathbf{W}} + \tilde{\mathbf{B}}).$$

Ultimate goal for visual clustering on two-dimensional space is to find $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ so that the $overall R^2_{(2)}$ is maximized under the constraint that $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are orthonormal vectors in p-dimensional space. There are two subproblems to solve. First, for given basis determined by $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, how to partition observations into k-groups to maximize $overall R^2_{(2)}$? Second, how to choose $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ to achieve the ultimate goal?

The first problem poses a combinatorial computation, so that it is hardly possible to obtain the exact solution. However, we may rely on conventional K-means clustering for an approximate solution. For the second problem, we can make use of a variant of the project pursuit method originated from Friedman and Tukey (1974), as will be detailed in Section 3.

3. The Algorithm

Among several algorithms for general projection pursuit optimization, random search algorithm by Posse (1990, 1995) is known to be efficient and reliable, which can be described as follows:

Step 1. Set the two orthogonal unit-norm vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ in R^p. Moreover, specify m for the number of independent global trials, *half* for the maximal number of independent local trials, and c for the size of local random search, all of which are clarified shortly.

Step 2. Generate a unit-norm random vector $\boldsymbol{u}$ in R^p. Then replace $\boldsymbol{\alpha}$ by

$$\boldsymbol{\alpha}^* = \frac{\boldsymbol{\alpha} + c\boldsymbol{u}}{||\boldsymbol{\alpha} + c\boldsymbol{u}||},$$

a jittering of $\boldsymbol{\alpha}$ in the random direction specified by $\boldsymbol{u}$ with magnitude c. Accordingly, to guarantee the orthogonality of basis vectors, $\boldsymbol{\beta}$ is modified to

$$\boldsymbol{\beta}^* = \frac{\boldsymbol{\beta} - (\boldsymbol{\alpha}^{*t}\boldsymbol{\beta})\boldsymbol{\alpha}^*}{||\boldsymbol{\beta} - (\boldsymbol{\alpha}^{*t}\boldsymbol{\beta})\boldsymbol{\alpha}^*||}.$$

Also, to search the opposite side, set

$$\boldsymbol{\alpha}^{**} = \frac{\boldsymbol{\alpha} - c\boldsymbol{u}}{||\boldsymbol{\alpha} - c\boldsymbol{u}||},$$

$$\boldsymbol{\beta}^{**} = \frac{\boldsymbol{\beta} - (\boldsymbol{\alpha}^{**t}\boldsymbol{\beta})\boldsymbol{\alpha}^{**}}{||\boldsymbol{\beta} - (\boldsymbol{\alpha}^{**t}\boldsymbol{\beta})\boldsymbol{\alpha}^{**}||}.$$

Step 3. Compute projection indices $overall R^2_{(2)}$ for the data set $\mathbf{X}(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*)$ and $\mathbf{X}(\boldsymbol{\alpha}^{**}, \boldsymbol{\beta}^{**})$. If any improvement is found, renew $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ and initialize the counter. Otherwise, keep the old basis vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ and increase the counter by one.

Step 4. If the counter is less than equal to *half*, return to Step 2. Otherwise, reduce the value of c by half and check whether the new value of c is sufficiently small. If so, then proceed to Step 5. Otherwise, initialize the counter and go to Step 2.

Step 5. Repeat Steps 1 to 4 m times, to acquire a number of sample projection indices. Choose the maximum and keep the corresponding final vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$.

Posse's algorithm works wonderfully in many cases. In Step 2, however, we note that the perturbation scheme affects both $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, in such a way that the algorithm is more focused on $\boldsymbol{\alpha}$ than on $\boldsymbol{\beta}$. As the result, especially when one of basis vectors arrives near the optimal position, the algorithm may not keep it pivoted. Instead, the algorithm lets basis vectors wander some more time.

To overcome such deficiency, we propose modifications of Steps 2, 3 and 4:

New Step 2. Generate a unit-norm random vector $\boldsymbol{u}$ in R^p and make it orthogonal to current basis vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$:

$$\boldsymbol{u}_1 = \frac{\boldsymbol{u} - (\boldsymbol{\alpha}^t\boldsymbol{u})\boldsymbol{\alpha} - (\boldsymbol{\beta}^t\boldsymbol{u})\boldsymbol{\beta}}{||\boldsymbol{u} - (\boldsymbol{\alpha}^t\boldsymbol{u})\boldsymbol{\alpha} - (\boldsymbol{\beta}^t\boldsymbol{u})\boldsymbol{\beta}||}.$$

Set

$$\boldsymbol{\alpha}^* = \frac{\boldsymbol{\alpha} + c\boldsymbol{u}_1}{||\boldsymbol{\alpha} + c\boldsymbol{u}_1||}, \quad \boldsymbol{\alpha}^{**} = \frac{\boldsymbol{\alpha} - c\boldsymbol{u}_1}{||\boldsymbol{\alpha} - c\boldsymbol{u}_1||}$$

as jitterings of $\boldsymbol{\alpha}$ in the random direction specified by $\pm\boldsymbol{u}_1$ with magnitude c. We note that both $\boldsymbol{\alpha}^*$ and $\boldsymbol{\alpha}^{**}$ are orthogonal to the other basis vector $\boldsymbol{\beta}$. Similarly, generate independently another unit-norm random vector $\boldsymbol{v}$ in R^p, and make it orthogonal to current basis vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$:

$$\boldsymbol{v}_1 = \frac{\boldsymbol{v} - (\boldsymbol{\alpha}^t\boldsymbol{v})\boldsymbol{\alpha} - (\boldsymbol{\beta}^t\boldsymbol{v})\boldsymbol{\beta}}{||\boldsymbol{v} - (\boldsymbol{\alpha}^t\boldsymbol{v})\boldsymbol{\alpha} - (\boldsymbol{\beta}^t\boldsymbol{v})\boldsymbol{\beta}||}.$$

Subsequently, set

$$\boldsymbol{\beta}^* = \frac{\boldsymbol{\beta} + c\boldsymbol{v}_1}{||\boldsymbol{\beta} + c\boldsymbol{v}_1||}, \quad \boldsymbol{\beta}^{**} = \frac{\boldsymbol{\beta} - c\boldsymbol{v}_1}{||\boldsymbol{\beta} - c\boldsymbol{v}_1||}$$

as jitterings of β in the random direction specified by $\pm v_1$ with magnitude c. Note that both β^* and β^{**} are orthogonal to the other basis vector α.

New Step 3. Compute and compare projection indices $overall R^2_{(2)}$ for the data set $\mathbf{X}(\alpha^*, \beta)$, $\mathbf{X}(\alpha^{**}, \beta)$ and $\mathbf{X}(\alpha, \beta^*)$, $\mathbf{X}(\alpha, \beta^{**})$. If any improvement is found, renew α and β and initialize the counter. Otherwise, keep the old basis vectors α and β and increase the counter by one.

New Step 4. If the counter is less than equal to *half*/2, return to New Step 2. Otherwise, reduce the value of c by half and check whether the new value of c is sufficiently small. If so, then proceed to Step 5. Otherwise, initialize the counter and go to New Step 2.

The reason why we reduce the effective value of *half* by half is that the computing load is doubled in New Step 3 compared to that in Step 3. By such modification, the efficiency of the new algorithm relative to the existing algorithm by Posse can be assessed more fairly.

Since Mahalanobis distance between observations is more meaningful statistically than Euclidean distance, we may consider K-means clustering based on Mahalanobis distance as in Huh (2000). For this, one may consider further modification of New Step 4 as follows:

New Step 4 with Mahalanobis option. If the counter is less than equal to *half*/2, return to New Step 2. Otherwise, reduce the value of c by half, then check whether the new value of c is sufficiently small. If so, then proceed to Step 5. Otherwise, initialize the counter, transform the data set by

$$\mathbf{X}\mathbf{S}_P^{-1/2},$$

where $\mathbf{S}_P$ is the pooled sample covariance matrix resulting from the currently available grouping of observations. That is,

$$\mathbf{S}_P = [\mathbf{X}^t\mathbf{X} - \mathbf{X}^t\mathbf{Z}(\mathbf{Z}^t\mathbf{Z})^{-1}\mathbf{Z}^t\mathbf{X}]/(n-k).$$

Then, go to New Step 2.

In the next section, the computational efficiency of the algorithm is demonstrated with two numerical examples without using Mahalanobis option in New Step 4.

4. Numerical Demonstration of Computation Efficiency

We applied our algorithm to Fisher's iris data and Australian rock crab data (Campbell and Mahon, 1974; Ripley, 1996, p. 13). Fisher's iris data consists of 150 observations with four variables (x_1: sepal length, x_2: sepal width, x_3: petal length, x_4: petal width). For the moment, we ignore the fact that each observation belongs to one of three species of iris (1: setosa, 2: versicolor, 3: virginica) but we assume there are three groups ($k = 3$).

See Table 1 for the results with standardized (centered and scaled) Fisher's iris data. We executed the whole computation ten times by setting the starting/ending value, c_1 and c_0, of c equal to 1 and 0.001. The first column of Table 1 shows the result by Posse's algorithm with the *half* set to 10 that took 4 minutes 18 seconds for the computing time and achieved 0.9489 for the $overall R^2_{(2)}$ on average. In contrast, all the replications with proposed algorithm with the same parameters, with the same value of *half*, yielded the $overall R^2_{(2)}$ 0.9602 by consuming 1 minute 52 seconds on the average. We note here that the proposed algorithm is faster and more reliable. For another instance, we applied the proposed algorithm to Australian rock crab data, which consists of 200 observations of four gender $\times$ color groups with five variables (x_1: frontal lip, x_2: rear width, x_3: midline length, x_4: carapace width, x_5: body depth). Prior to main clustering analysis, the data are centered and sphered (covariance-adjusted) because of inherent high collinearity. Table 2 shows the result when the number k of clusters is set to four. In this case, the existing and proposed algorithms produce nearly equal computing time, while the proposed algorithm dominates the existing algorithm as for the projection index, the $overall R^2_{(2)}$.

We observed that the proposed algorithm is better in computing time and/or finding larger projection index, with considerable variation. Actually the numbers in Table 1 and Table 2 being produced by the proposed algorithm are accurate to the fifth decimal place. But these results depend on the data sets and so it needs further investigation on relevant conditions in the future. (All computations in this section are performed on Windows 98 IBM PC with Pentium III 450MHz processor, 64MB RAM and SAS/IML.)

5. Plotting Observations and Variables

It is very natural that the observations are represented on the two-dimensional display by the rows of $\mathbf{X}(\boldsymbol{\alpha}, \boldsymbol{\beta})$ where $\mathbf{X}$ is the $n \times p$ matrix of preprocessed

TABLE 1 *Two-dimensional projection pursuit K-means visual clustering applied to Fisher's iris data*

Argument	*Posse's algorithm* $m = 10$, *half* $=10$, $c_1 = 1$, $c_0 = 0.001$		*Proposed algorithm* $m = 10$, *half* $=10$, $c_1 = 1$, $c_0 = 0.001$	
Replication	*Time*	$overallR^2_{(2)}$	*Time*	$overallR^2_{(2)}$
1	3min. 20sec.	0.9595	1min. 51sec.	0.9602
2	3min. 33sec.	0.9538	1min. 52sec.	0.9602
3	3min. 35sec.	0.9578	1min. 46sec.	0.9602
4	4min. 36sec.	0.9531	1min. 56sec.	0.9602
5	3min. 20sec.	0.9086	1min. 44sec.	0.9602
6	10min. 10sec.	0.9530	1min. 43sec.	0.9602
7	3min. 26sec.	0.9550	1min. 55sec.	0.9602
8	3min. 15sec.	0.9398	1min. 55sec.	0.9602
9	3min. 50sec.	0.9509	2min. 1sec.	0.9602
10	3min. 53sec.	0.9578	1min. 59sec.	0.9602
Mean	4min. 18sec.	0.9489	1min. 52sec.	0.9602

TABLE 2 *Two-dimensional projection pursuit K-means visual clustering applied to Australian rock crab data*

Argument	*Posse's algorithm* $m = 10$, *half* $=10$, $c_1 = 1$, $c_0 = 0.001$		*Proposed algorithm* $m = 10$, *half* $=10$, $c_1 = 1$, $c_0 = 0.001$	
Replication	*Time*	$overallR^2_{(2)}$	*Time*	$overallR^2_{(2)}$
1	5min. 42sec.	0.7921	6min. 52sec.	0.8474
2	6min. 39sec.	0.8222	7min. 36sec.	0.8474
3	7min. 21sec.	0.7793	5min. 48sec.	0.8474
4	6min. 43sec.	0.7975	7min. 1sec.	0.8474
5	7min. 10sec.	0.7467	5min. 33sec.	0.8474
6	7min. 3sec.	0.7526	6min. 4sec.	0.8474
7	6min. 56sec.	0.7821	6min. 28sec.	0.8474
8	6min. 1sec.	0.7711	5min. 46sec.	0.8474
9	6min. 42sec.	0.7472	7min. 50sec.	0.8474
10	6min. 56sec	0.7735	7min. 30sec.	0.8474
Mean	6min. 43sec.	0.7764	6min. 39sec.	0.8474

(centered/scaled/sphered) measurements and $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are final basis vectors. Accordingly, the variable $\mathbf{X}_j$ can be plotted at $\boldsymbol{v}_j^t(\boldsymbol{\alpha}, \boldsymbol{\beta})$, if $\boldsymbol{v}_j^t$ represents the appropriate size and direction of the corresponding variable ($j = 1, \ldots, k$). We suggest

$$\begin{aligned}
\boldsymbol{v}_j &= \boldsymbol{e}_j, && \text{for the standardized (centered and scaled) data,} \\
&= s_j \boldsymbol{e}_j, && \text{for the centered data,} \\
&= s_j \mathbf{S}_T^{-1/2} \boldsymbol{e}_j, && \text{for the centered and sphered (covariance-adjusted) data.}
\end{aligned}$$

where s_j is the standard deviation of $\mathbf{X}_j$, $\boldsymbol{e}_j = (0, \ldots, 1, \ldots, 0)^t$ is the j^{th} elementary vector, and $\mathbf{S}_T$ is the $p \times p$ total covariance matrix.

For standardized Fisher's iris data when the algorithm is applied with the number k of cluster set to three and the parameters of optimization algorithm as $m = 10$, $half = 10$, $c_1 = 1$ and $c_0 = 0.001$, we obtained $overall R^2_{(2)} = 0.9602$ and the final basis vectors

$$\boldsymbol{\alpha} = \begin{pmatrix} 0.2322 \\ -0.1551 \\ -0.6571 \\ 0.7001 \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} 0.0221 \\ 0.2484 \\ -0.7295 \\ -0.6369 \end{pmatrix}.$$

Figure 1a and 1b show observations and variables on the two-dimensional plane obtained by the projection pursuit K-means visual clustering for Fisher's iris data. Overlaying Figure 1a to the top of Figure 1b, we summarize individual cluster's characteristics as follows:

Cluster 1. Being in positive direction of variable x_2 and in the negative direction of variables x_3, x_4, their sepals are wide and their petals are short and narrow.

Cluster 2. Being in negative direction of variable x_2 and in the positive direction of variables x_3, x_4, they have narrow sepals and long and wide petals.

Cluster 3. They are similar to Cluster 2, but they have narrower sepals and longer and wider petals.

By the way, this data set has the external classification code (iris species), and so it may be interesting to look at the classification table of clustering group and the classification code. It turns out, in Table 3, that the two-dimensional K-means

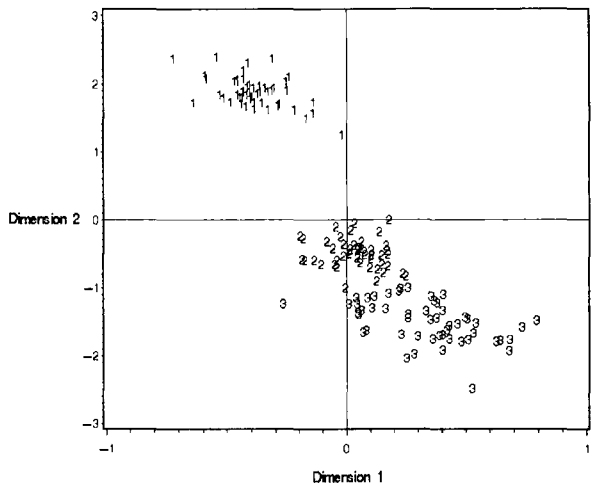

FIGURE 1a *K-means visual clustering display of Fisher's iris data: Observations are represented with cluster numbers.*

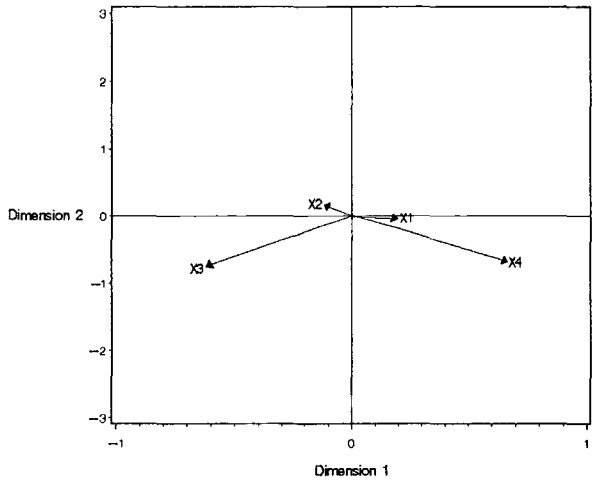

FIGURE 1b *K-means visual clustering display of Fisher's iris data: Variables are represented as arrows.*

TABLE 3 *Clustering result of Fisher's iris data*

(a) K-means clustering

	Cluster 1	*Cluster 2*	*Cluster 3*
Species 1	50	0	0
Species 2	0	39	11
Species 3	0	14	36

(b) Two-dimensional K-means visual clustering

	Cluster 1	*Cluster 2*	*Cluster 3*
Species 1	50	0	0
Species 2	0	46	4
Species 3	0	1	49

visual clustering is effective in reducing classification error. The conventional K-means clustering results in 25 classification errors out of 150 cases, while the proposed method yields 5 errors.

We applied the proposed algorithm to centered and sphered Australian rock crab data, the number of clusters $k = 4$, $m = 10$, $half = 10$, $c_1 = 1$, $c_0 = 0.001$, and obtained $overall R^2_{(2)} = 0.8474$ and the final basis vectors

$$\boldsymbol{\alpha} = \begin{pmatrix} 0.3680 \\ 0.8804 \\ -0.2505 \\ -0.1278 \\ 0.1024 \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} -0.2574 \\ 0.3311 \\ 0.8899 \\ -0.0740 \\ 0.1634 \end{pmatrix}.$$

Figure 2a and 2b show observations and variables on the two-dimensional space obtained by the projection pursuit K-means visual clustering. We may interpret the cluster characteristics as in the case of Fisher's iris data. When tabulated by cluster membership and natural groups by gender (male/female) and color (orange/blue), observations within each group are assigned to the modal cluster with smaller number of deviants, as seen in Table 4.

6. Concluding Remarks

The proposed algorithm depends on several parameters such as the starting and ending values (c_1 and c_0) of c, the number of inner repetitions $half$ and the number of independent global repetitions m. Kim (2000) recommended that

$$1 \leq c_1 \leq 2, \; c_0 = 0.001, \; half = 10 \text{ and } m \geq 10$$

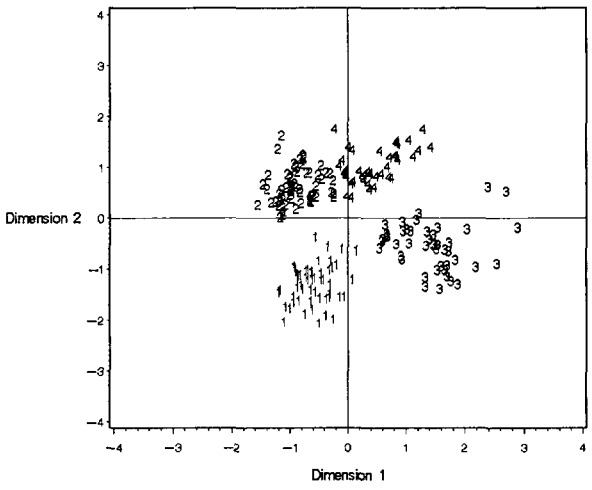

FIGURE 2a *K-means visual clustering display of Australian rock crab data: Observations are represented with cluster numbers.*

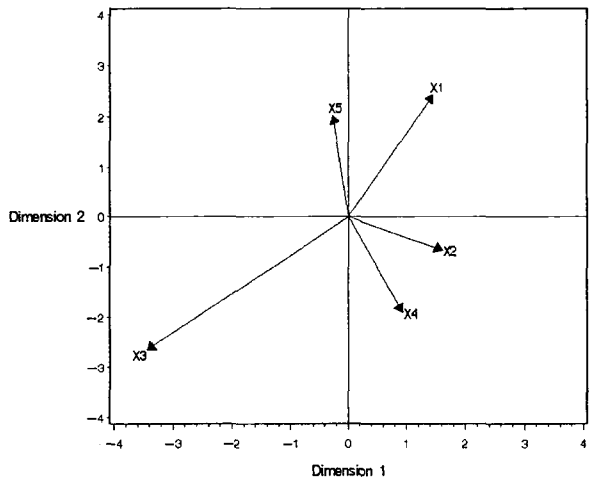

FIGURE 2b *K-means visual clustering display of Australian rock crab data: Variables are represented as arrows.*

TABLE 4 *Clustering result of Australian rock crab data*

(a) K-means clustering

	Cluster 1	*Cluster 2*	*Cluster 3*	*Cluster 4*
Species 1	48	2	0	0
Species 2	22	27	1	0
Species 3	1	0	0	49
Species 4	0	14	33	3

(b) Two-dimensional K-means visual clustering

	Cluster 1	*Cluster 2*	*Cluster 3*	*Cluster 4*
Species 1	0	48	0	2
Species 2	0	10	0	40
Species 3	50	0	0	0
Species 4	5	0	45	0

through several experimental analysis with optimal scaling.

This study proposed a practical algorithm for "Projection Pursuit K-Means Visual Clustering" which performs K-means clustering of multivariate observations by dimensional reduction. The improvement of computational efficiency and reliability can be achieved through the proposed algorithm. In applying the method, however, one needs to specify k, the number of clusters, and initialize the seeds of cluster, as in conventional K-means clustering. Huh (2002) is a helpful guidance for this.

REFERENCES

Aranda-Ordaz, F. J. (1981). "On two families of transformations to additivity for binary response data", *Biometrika*, **68**, 357–363.

Ahn, K. A. and Rhee, S. S. (1992). "A simulation study on projection pursuit discriminant analysis", *The Korean Journal of Applied Statistics*, **5**, 103–111.

Baek, J. S. and Sim, J. W. (2000). "Kernel pattern recognition using K-means clustering method", *The Korean Journal of Applied Statistics*, **13**, 447–455.

Becker, R. A., Cleveland, W. S. and Wilks, A. R. (1988). "Dynamic graphics for data analysis", In *Dynamic Graphics for Statistics* (W. S. Cleveland and M. E. McGill, eds.), Wadsworth Inc., California, 1–50.

Campbell, M. A. and Mahon, R. J. (1974). "A multivariate study of variation in two species of rock crab of genus Leptograpsus", *Australian Journal of Zoology*, **22**, 417–425.

Cook, D. and Buja, A. (1997). "Manual controls for high-dimensional data projections", *Journal of Computational and Graphical Statistics*, **6**, 464–480.

Cook, D., Buja, A., Cabrera, J. and Hurley, C. (1995). "Grand tour and projection pursuit", *Journal of Computational and Graphical Statistics*, **4**, 155–172.

Everitt, B. S. (1974). *Cluster Analysis*, Wiley, New York.

Friedman, J. H. and Tukey, J. W. (1974). "A projection pursuit algorithm for exploratory data analysis", *IEEE Transactions for Computers*, **23**, 881–890.

Hartigan, J. A. (1975). *Clustering Algorithms*, Wiley, New York.

Huh, M. H. (2000). "Double K-means clustering", *The Korean Journal of Applied Statistics*, **13**, 343–352.

Huh, M. H. (2002). "Setting the number of clusters in K-means clustering: Exploratory approach", In *Recent Advances in Statistical Research and Data Analysis* (Y. Baba, ed.), Springer-Verlag, Tokyo, 115–124.

Huh, M. H. and Kim, M. K. (2000). "Low-dimensional K-means clustering 1: Iterative canonical transforms method", *Journal of Data Science and Classification*, **4**, 1–15.

Huh, M. Y. and Song, K. R. (2001). "Variable arrangement for data visualization", *The Korean Communications in Statistics*, **8**, 643–650.

Jhun, M. S. and Jin, S. H. (2000). "On a modified k-spatial medians clustering", *Journal of the Korean Statistical Society*, **29**, 247–260.

Kim, M. K. (2000). *Low-dimensional K-means Clustering*, Ph.D. Dissertation, Korea University, Seoul.

Kim, S. S. (1999). "Interactive visualization of K-means and hierarchical clusters", *The Journal of Data Science and Classification*, **3**, 13–27.

Kim, S. S., Kwon, S. and Cook, D. (2000). "Interactive visualization of hierarchical clusters using MDS and MST", *Metrika*, **51**, 39–51.

Lee, S. H., Park, N. H. and Kim, Y. H. (1995). "A clustering method using the Coulomb energy network", *The Korean Journal of Applied Statistics*, **8**, 39–50.

Milligan, G. W. (1980). "An examination of the effect of six types of error perturbation of fifteen clustering algorithms", *Psychometrika*, **45**, 325–342.

Milligan, G. W. (1981). "A review of Monte Carlo tests of cluster analysis", *Multivariate Behavioral Research*, **16**, 379–407.

Park, H. J., Choi, D. W. and Koo, J. Y. (2000). "Prediction and classification using projection pursuit regression with automatic order selection", *The Korean Communications in Statistics*, **7**, 585–596.

Posse, C. (1990). "An effective two-dimensional projection pursuit algorithm", *Communications in Statistics-Simulation and Computation*, **19**, 1143–1164.

Posse, C. (1995). "Tools for two-dimensional exploratory projection pursuit", *Journal of Computational and Graphical Statistics*, **4**, 83–100.

Ripley, R. D. (1996). *Pattern Recognition and Neural Networks*, Cambridge University Press, Cambridge.

Stute, W. and Zhu, L. X. (1995). "Asymptotics of K-means clustering based on projection pursuit", *Sankhyā*, **A57**, 462–471.

Young, F. W., Kent, D. P. and Kuhfeld, W. F. (1988). "Dynamic graphics for data analysis", In *Dynamic Graphics for Statistics* (W. S. Cleveland and M. E. McGill, eds.), Wadsworth Inc., California, 391–423.