Journal of the Faculty of Environmental Science and Technology, Okayama University
Vol. 1, No. 1, pp.65-76, March 1996

Application of Partial Least Squares Linear Discriminant Function to Writer Identification in Pattern Recognition Analysis

Hyun Bin KIM * Yutaka TANAKA †

(Received November 17 , 1995)

Abstract

Partial least squares linear discriminant function (PLSD) as well as ordinary linear discriminant function (LDF) are used in pattern recognition analysis of writer identification based on arc patterns extracted from the writings written with Hangul letters by 20 Koreans. Also a simulation study is performed using the Monte Carlo method to compare the performances of PLSD and LDF. PLSD showed remarkably better performance than LDF in the Monte Calro study and slightly better performance in the analysis of the real pattern recognition data.

KEYWORDS: Writer identification, Arc patterns, Linear discriminant function, Partial least squares

1 Introduction

Pattern recognition has aroused researchers' interest in various fields such as image processing, speech recognition, character recognition and medical diagnosis as computer technologies have been developed rapidly in recent years. Great interest has been shown to the recognition of handwritten data in image processing and it can be put into practice in many fields of engineering. Of particular practice is writer identification in which writers of written texts are automatically identified by means of the analysis of handwriting.

The process of pattern recognition consists of feature extraction and classification, and for the latter part of pattern recognition there are three major approaches. One is statistical approach, another one is syntactic approach and last one, which has been developed recently, is neural network approach (see, e.g., Schalkoff, 1992).

In the present paper partial least squares linear discriminant function (PLSD), which was proposed by Kim and Tanaka (1995a), is used as a statistical method of classification in writer identification analysis of handwritten Hangul letters (Korean characters). As a feature extraction technique we use Yoshimura and Yoshimura (1991)'s method of arc pattern extraction, which extracts typical patterns of arc from a sequence of handwritten characters. However, using the arc patterns, we are often confronted with the problems of multicollinearity, which occur in the case where the variables are highly correlated with each other or in the case where the number of observations is smaller than that of variables. PLSD is a newly proposed linear discriminant function, which is devised to reduce the effects of multicollinearity in discriminant analysis and can be applied in the small-sample, high-dimensional setting. Also we try to compare the performance

*Graduate School of Natural Science and Technology, Okayama University, Tsushima, Okayama 700, Japan
†Department of Environmental and Mathematical Sciences, Faculty of Environmental Science and Technology, Okayama University, Okayama, 700 Japan.

of PLSD with that of linear discriminant function (LDF) by applying to artificial data generated with Monte Carlo method. The brief explanations of writer identification and arc pattern extraction are given in sections 2 and 3, respectively. Section 4 presents the algorithm of PLSD. The comparison of the performance of PLSD and that of LDF is given in section 5. Section 6 gives the results of analysis with PLSD of writer identification of handwritten Hangul letters. Section 7 provides a short discussion.

2 Writer identification

In pattern recognition analysis, writer recognition problems can generally be divided into two types, "writer identification" and "writer verification", depending on prior knowledge available about the classes under study and on the type of information extracted from the different objects. A writer identification system must establish a writer's identity by comparing some specific attributes of his writing with those of all the g writers enrolled in a reference data base. A writer verification system, on the other hand, decides on the claimed identity of a writer i by a one-to-one comparison. Signature or text can be acquired "on-line" in which the acquisition occurs during the writing process itself or "off-line" in which the acquisition is carried out from the paper on which text or signature appears after writing. Moreover, in case of writer identification, one needs to distinguish between two situations, "text-dependent" and "text-independent". In text-independent situation, the text of reference writings need not be the same as the one of the writing in question (see, e.g., Plamondon and Lorette, 1989).

The above problem of writer identification in off-line situation is essentially the same as that of classification in discriminant analysis, which is described as follows : An observation is taken from one of g groups. The observation is classified into one of g groups on the basis of the vector of measurements $\boldsymbol{x}^t = (x_1, \cdots, x_p)$,which can be expressed as a point in a p-dimensional space. This p-dimensional space is divided into g regions by some classification rule. If the observation falls in region R_i, it is classified as coming from the i-th group (see, e.g., Anderson, 1984). So, we can use ordinary discriminant functions, for example, linear discriminant function or quadratic discriminant function in writer identification.

3 Yoshimura and Yoshimura's method of arc pattern extraction

A method of arc pattern extraction was proposed by Yoshimura and Yoshimura (1991) to extract features from binary data. The basic idea can be summarized as follows : Consider seven arcs which connect two given points as shown in Figure 1. The radius of an arc is set to be the length obtained by multiplying the distance between two end points by one of four coefficients 2/3, 2/2, 2/1 and 2/0(= ∞), and use the locations of five equi-distant points on an arc to characterize the arc. Select all possible end points and count the frequencies of these arcs contained in hand written sequences of characters. Then adopt the relative frequencies of the arc patterns as the features of writers. Precisely speaking, all possible two end points are selected on 5×5 grids in such a way that the distances between these two end points are $0, 2, \text{or}, 4$ (even number) horizontally and vertically. Then, corresponding to the seven arc patterns we have 45 different types of patterns as shown in Figure 2. Some types of pattern such as No. 2 and No. 3 in Figure 2 are so similar that they are regarded as a same model pattern and as the result the number of model patterns become 37.

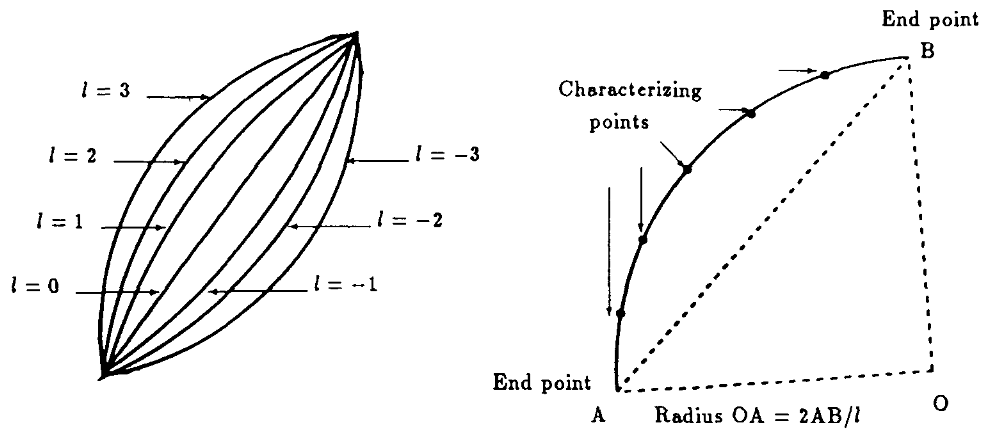

FIGURE 1. Arc patterns in their original forms. Seven arcs connect two given end points. It is assumed that five equi-distant points on an arc characterize the arc and the frequencies of these arcs contained in the writing characterize the feature for the inspection of a writing.

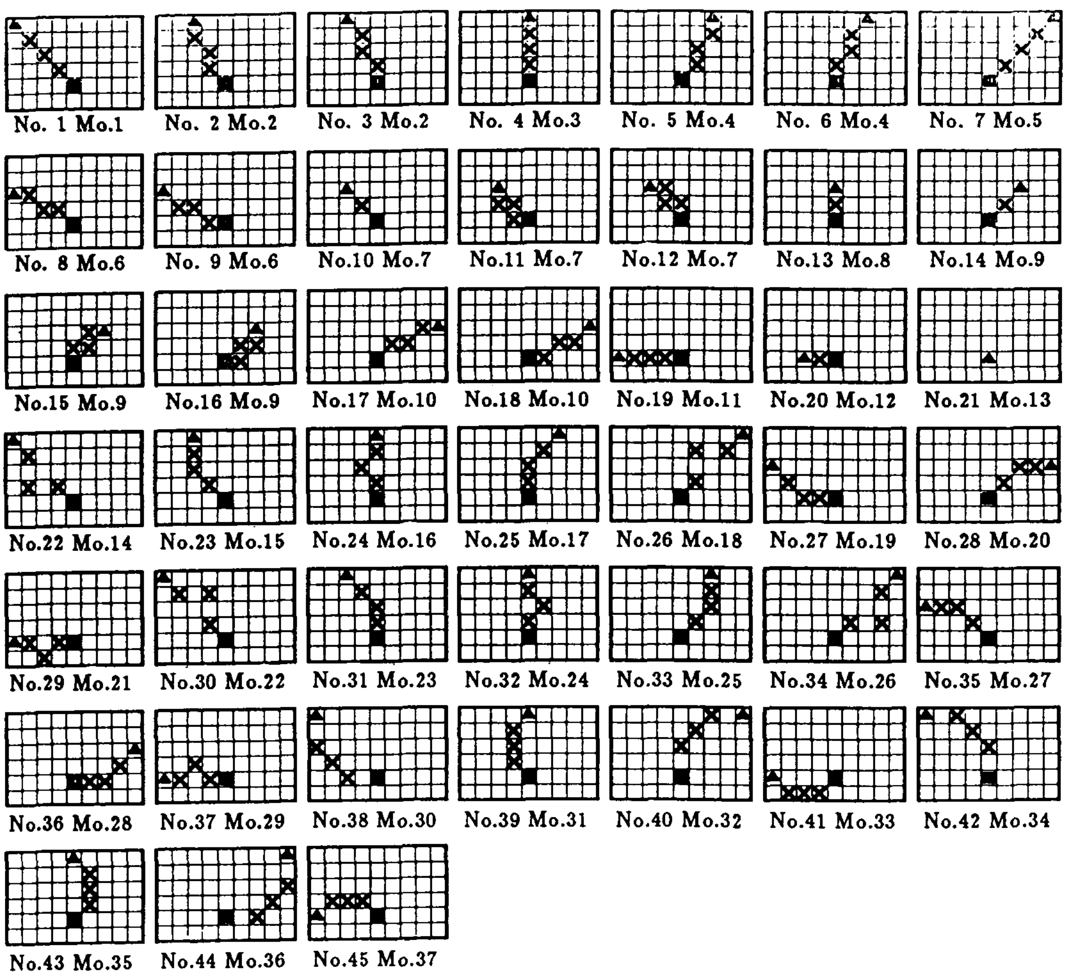

FIGURE 2. The models of the digitized arc pattern within 5 × 5 grids. The types of arc pattern in Figure 1 are reduced to the bellow listed patterns, where only oddly distant points are used as two end points and only one point is used in some models. Models (Mo.) with the same number are regarded as the same one in counting frequencies.

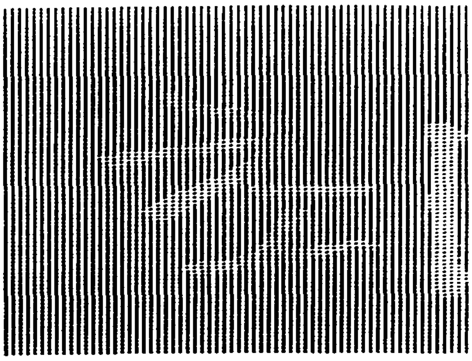

FIGURE 3. Example of the binary pattern of a Hangul letter. The area in which a writing is placed is meshed in 90 (vertically) $\times$ 1664 (horizontally) grids.

Frequencies of 37 model patterns are counted in each case where a rectangular-marked point is fixed at a point as in Figure 2 and they are summed up for all possible cases where the rectangular-marked point moves on all points of the area of the binary patterns. An example of binary patterns is given Figure 3. The relative frequencies thus obtained are referred as the feature to characterize a writer.

4 Partial least squares linear discriminant function (PLSD)

Partial least squares regression (PLSR) is one of regression methods which can be efficiently used even in the case where there exists multicollinearity among explanatory variables. Its performance has been compared with other regression methods using simulation (Frank and Friedman 1993; Kim and Tanaka 1994). It has worked well not only in the above simulation studies but in many chemical problems and has become one of the most popular regression methods in chemometrics.

Kim and Tanaka (1995a) proposed a new linear discriminant function, named partial least squares linear discriminant function (PLSD), using the algorithm of partial least squares (PLS) in regression analysis and showed its good performance by comparing with the ordinary LDF through numerical study.

4.1 Partial Least Squares Regression (PLSR)

An algorithm of PLSR proposed by Wold (1975) is as follows :

1. Initialize (Centering) :

$$\boldsymbol{X}_0 \leftarrow (\boldsymbol{X} - \boldsymbol{1}\bar{\boldsymbol{x}}^t), \quad \boldsymbol{y}_0 \leftarrow (\boldsymbol{y} - \bar{y}\boldsymbol{1})$$

2. For $k = 1, 2, \cdots$ to K :

   2.1 $\boldsymbol{w}_k = \boldsymbol{X}_{k-1}^t \boldsymbol{y}_{k-1}$

   2.2 $\boldsymbol{t}_k = \boldsymbol{X}_{k-1}\boldsymbol{w}_k$

   2.3 $\boldsymbol{p}_k = \boldsymbol{X}_{k-1}^t \boldsymbol{t}_k / \boldsymbol{t}_k^t \boldsymbol{t}_k \quad (= \boldsymbol{X}^t \boldsymbol{t}_k / \boldsymbol{t}_k^t \boldsymbol{t}_k)$

   2.4 $q_k = \boldsymbol{y}_{k-1}^t \boldsymbol{t}_k / \boldsymbol{t}_k^t \boldsymbol{t}_k \quad (= \boldsymbol{y}^t \boldsymbol{t}_k / \boldsymbol{t}_k^t \boldsymbol{t}_k)$

   2.5 $\boldsymbol{X}_k = \boldsymbol{X}_{k-1} - \boldsymbol{t}_k \boldsymbol{p}_k^t$

2.6 $\boldsymbol{y}_k = \boldsymbol{y}_{k-1} - \boldsymbol{t}_k q_k$

3. Calculate regression coefficients :

$$\widehat{\boldsymbol{\beta}}_K = \boldsymbol{W}_K(\boldsymbol{W}_K^t \boldsymbol{X}^t \boldsymbol{X} \boldsymbol{W}_K)^{-1} \boldsymbol{W}_K^t \boldsymbol{X}^t \boldsymbol{y}, \qquad \boldsymbol{W}_K = (\boldsymbol{w}_1, \cdots, \boldsymbol{w}_K)$$

4. Prediction equation :

$$\hat{y} = (\bar{y} - \widehat{\boldsymbol{\beta}}_K^t \bar{\boldsymbol{x}}) + \widehat{\boldsymbol{\beta}}_K^t \boldsymbol{x}.$$

In the above n and p indicate the numbers of observations and explanatory variables, respectively, $\boldsymbol{X}$ is an $n \times p$ matrix of explanatory variables, $\boldsymbol{y}$ is an $n \times 1$ vector of response variable, $\mathbf{1}$ is a unit vector with 1 in its all elements, $\bar{\boldsymbol{x}}$ is the mean vector of $\boldsymbol{X}$, $\bar{y}$ is the mean of $\boldsymbol{y}$, K ($\leq$ rank of $\boldsymbol{X}$) is the number of components employed in the model. PLS becomes equivalent to ordinary least squares (OLS) if it uses all possible components. So, it is important how to determine the number of components K in applying the PLS algorithm.

4.2 Linear Discriminant Function (LDF)

Suppose that there exist g groups $\pi_1, \pi_2, \cdots, \pi_g$ and that an observation $\boldsymbol{x}$ from group π_i follows a p-variate normal distribution $N(\boldsymbol{\mu}_i, \boldsymbol{\Sigma})$ with a common covariance matrix $\boldsymbol{\Sigma}$. Also suppose that the costs $c(j \mid i)$ due to misclassifying an observation from π_i to π_j for $i, j = 1, 2, \cdots, g$ are the same for all (i, j). Then LDF of the i-th group which minimizes the total cost is expressed as follows :

$$\delta_i(\boldsymbol{x}) = \boldsymbol{\mu}_i{}^t\, \boldsymbol{\Sigma}^{-1} \boldsymbol{x} - \frac{1}{2} \boldsymbol{\mu}_i{}^t\, \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_i + \ln(q_i), \tag{1}$$

where q_i is the prior probability of drawing an observation from group π_i. We have the following classification rule :

$$\text{assign } \boldsymbol{x} \text{ to } \pi_i \text{ if } \delta_i(\boldsymbol{x}) = \max_j \{\delta_j(\boldsymbol{x})\}.$$

In sample version $\boldsymbol{\mu}_i$, $\boldsymbol{\Sigma}$ and q_i are replaced by

$$\bar{\boldsymbol{x}}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \boldsymbol{x}_{ij}, \quad \widehat{\boldsymbol{\Sigma}} = \frac{1}{n-g} \sum_{i=1}^{g} \sum_{j=1}^{n_i} (\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_i)(\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_i)^t \text{ and } \hat{q}_i = \frac{n_i}{n},$$

respectively, where n_i is the number of observations from π_i for $i = 1, 2, \cdots, g$ and $n = \sum_{i=1}^{g} n_i$ (Anderson, 1984). Thus the sample LDF is obtained as

$$\mathrm{LDF}_i(\boldsymbol{x}) = \bar{\boldsymbol{x}}_i{}^t\, \widehat{\boldsymbol{\Sigma}}^{-1} \boldsymbol{x} - \frac{1}{2} \bar{\boldsymbol{x}}_i{}^t\, \widehat{\boldsymbol{\Sigma}}^{-1} \bar{\boldsymbol{x}}_i + \ln(\hat{q}_i). \tag{2}$$

Applying the above LDF to, say, the i-th and j-th groups, we have the classification rule as

$$\text{assign } \boldsymbol{x} \text{ to } \pi_i \text{ if } \mathrm{LDF}_{ij}(\boldsymbol{x}) \geq 0 \quad \text{or} \quad \pi_j \text{ if } \mathrm{LDF}_{ij}(\boldsymbol{x}) < 0,$$

where LDF_{ij} is defined by

$$\mathrm{LDF}_{ij}(\boldsymbol{x}) = (\bar{\boldsymbol{x}}_i - \bar{\boldsymbol{x}}_j)^t\, \widehat{\boldsymbol{\Sigma}}^{-1} \boldsymbol{x} - \frac{1}{2} (\bar{\boldsymbol{x}}_i - \bar{\boldsymbol{x}}_j)^t\, \widehat{\boldsymbol{\Sigma}}^{-1} (\bar{\boldsymbol{x}}_i + \bar{\boldsymbol{x}}_j) + \ln(\hat{q}_i) - \ln(\hat{q}_j). \tag{3}$$

4.3 Derivation of PLSD

PLSD was proposed by Kim and Tanaka (1995a) for the purpose of reducing the effects of the multicollinearity in data. The proposed PLSD is obtained by replacing $\widehat{\boldsymbol{\Sigma}}^{-1}$ in eq. (3) by $\ddot{\boldsymbol{H}}_{ij(K)}$ defined as

$$\ddot{\boldsymbol{H}}_{ij(K)} = \boldsymbol{W}_{ij(K)}(\boldsymbol{W}^t_{ij(K)}\widehat{\boldsymbol{\Sigma}}\boldsymbol{W}_{ij(K)})^{-1}\boldsymbol{W}^t_{ij(K)}. \tag{4}$$

Namely,

$$\mathrm{PLSD}_{ij}(\boldsymbol{x}) = (\bar{\boldsymbol{x}}_i - \bar{\boldsymbol{x}}_j)^t\, \ddot{\boldsymbol{H}}_{ij(K)}\boldsymbol{x} - \frac{1}{2}(\bar{\boldsymbol{x}}_i - \bar{\boldsymbol{x}}_j)^t\, \ddot{\boldsymbol{H}}_{ij(K)}(\bar{\boldsymbol{x}}_i + \bar{\boldsymbol{x}}_j) + \ln(\hat{q}_i) - \ln(\hat{q}_j). \tag{5}$$

Here $\boldsymbol{W}_{ij(K)}$ consists of K covariance vectors obtained by applying PLSR to the data of the i-th and j-th groups, where a dummy variable with the values $-\dfrac{n_j}{n_i + n_j}$ and $\dfrac{n_i}{n_i + n_j}$ is used as the response variable to indicate which group an observation belongs to. The basic idea behind this is that LDF is mathematically equivalent to the above regression of binary responses.

In case of more than three groups, ${}_gC_2$ PLSD$_{ij}$s are calculated and an observation $\boldsymbol{x}$ is assigned to the i-th or j-th group based on PLSD$_{ij}$ for all possible pairs. Finally, $\boldsymbol{x}$ is classified into the group to which $\boldsymbol{x}$ is most often assigned. The result classified by PLSD$_{ij}$ becomes equivalent to that classified by LDF if all possible components are used in PLSD$_{ij}$. This implies that PLSD has at least equal and possibly better performance than LDF if the number of components are properly determined. It is important, therefore, how to choose the number of components K. As a criterion to choose K, the correct discrimination rate (CDR) based on cross-validation can be used.

$$\mathrm{CDR} = 100 \times \sum_{i=1}^{g} \frac{c_i}{n_i}.$$

where n_i and c_i are the numbers of all observations and of correctly classified observations, respectively, in the i-th group. We have to search for a PLSD model with the largest value of cross-validated CDR. If two or more PLSD models have the largest CDR, the PLSD model with the smallest number of components is selected in our analysis. Refer to Kim and Tanaka (1995a) for details.

5 Monte Carlo study

This section presents a summary of results of a set of Monte Carlo experiments comparing the relative performance of PLSD with LDF. The two methods are compared in 24 different situations. The number of groups is fixed at three and the distances of centroids of each group are fixed with keeping the ratios of 3:4:5. The situations are differentiated by the training-sample size of each group (balance data ($n_1 = 20, n_2 = 20, n_3 = 20$) and unbalance data ($n_1 = 15, n_2 = 20, n_3 = 25$)), the structure of the (population) covariance matrix of each group (equal covariance ($\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}_3$), unequal covariance ($\boldsymbol{\Sigma}_1 \neq \boldsymbol{\Sigma}_2 \neq \boldsymbol{\Sigma}_3$)) and the number of variables ($p = 5, 15, 25, 35, 45, 55$). Here three groups in case of equal covariance type have the same (population) condition number κ, defined by $\lambda_{max}/\lambda_{min}$ of $\boldsymbol{\Sigma}$ to indicate the degree of multicollinearity, with the value 1.0. In case of unequal covariance type κ_1 is 900.0, κ_2 400.0 and κ_3 1.0. Generally, if the condition number is less than 100, there is no serious problem of multicollinearity. Condition numbers from 100 to 1000 indicate multicollinearity from weak degree to strong one, and if κ exceeds 1000, severe multicollinearity problems would take place (Belsley, Kuh and Welsch, 1980). Artificial data with the above degrees of multicollinearity of $\boldsymbol{\Sigma}$ in this experiment was generated by the method proposed by Kim and Tanaka (1995b). For each situation of all 24 (= 2 (sample size) ×

2 (covariance matrix) $\times$ 6 (variable number)) situations, 50 ($= N_t$) repetitions are performed. The evaluation criterion is the average value of CDRs obtained by 50 repetitions in which CDR is evaluated for the independently generated test-sample having the sizes of 50 times as large as the training-sample of each group.

$$\mathrm{CDR} = \frac{1}{N_t} \sum_{i=1}^{N_t} \mathrm{CDR}_i.$$

Results of the experiment are given in Figure 4. Note that, since the training-sample size is fixed at 60, the degree of multicollinearity becomes high as the number of variables increases.

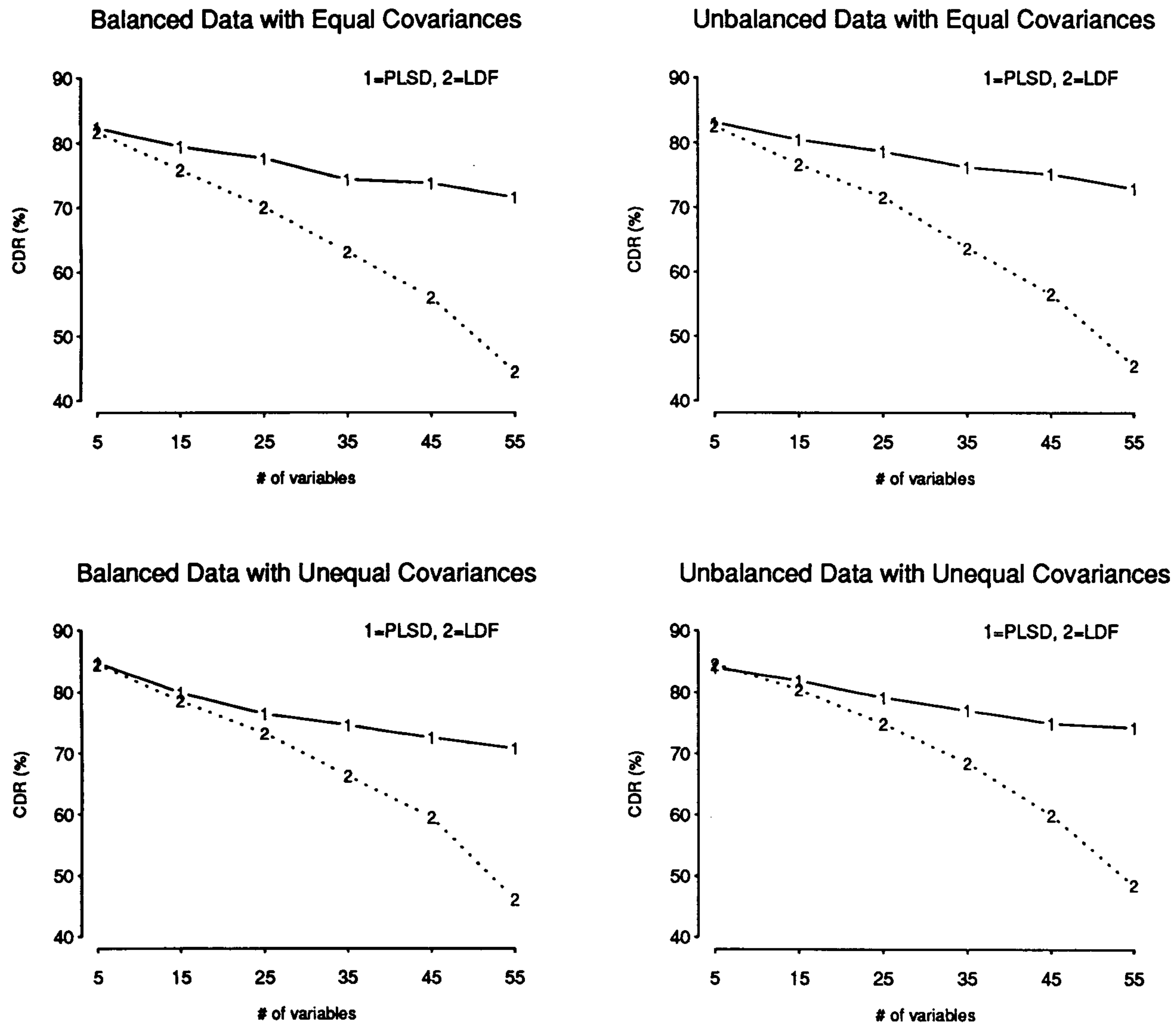

FIGURE 4. Average correct discrimination rates (CDRs) for PLSD (solid line) and LDF (dotted line) at 24 (2 (data type) $\times$ 2 (covariance type) $\times$ 6 (variable type)) situations. PLSD shows the same performance with LDF in case of no multicollinearity problem (X5) and better performances than LDF as the degree of multicollinearity becomes more serious in all situations except for X5.

There are no great differences between CDRs of PLSD and LDF in the case of $p = 5$. But differences between two methods increase as the number of variables increases. Such trends are shown in all the situations regardless of balance/unbalance data and of equal/unequal covariance matrices.

6 Application of PLSD to writer identification

In this section PLSD and LDF are applied to writer identification based on the arc pattern data for Korean text.

6.1 Data

We distinguish between two terms "text" and "writing". A set of letters as a semantic being is referred to as a text and a handwritten image of a text is referred to as a writing. Korean text was written with Hangul letters by 20 Koreans. A text was written with pencil on a A4 (20.0 cm $\times$ 29.5 cm) sheet with ruled horizontal lines 1 cm apart as shown in Figure 5. Two sheets for a text were repeatedly obtained three times with more than a week interval. As a result, six sheets of writings were obtained from one person. Three lines were selected from one sheet and

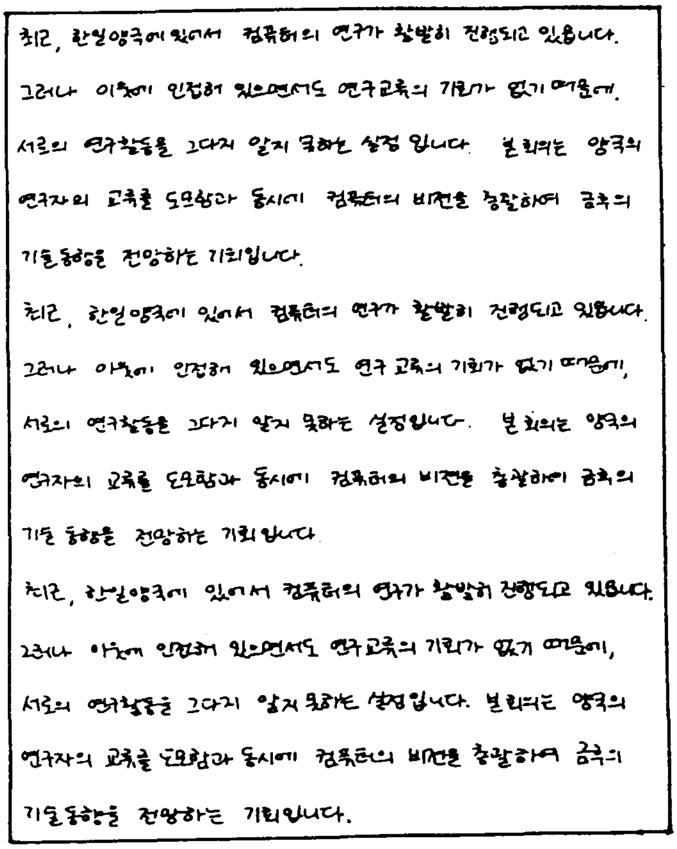

FIGURE 5. An example of text written with Hangul letters.

stored in a database as a set of lines of binary patterns, where the vertical and the horizontal sizes of grids are 90 and 1664, respectively for one line. From this database we get a set of data to be used for writer identification. The sample size n is 360 (= 20 (person) × 6 (sheet) × 3 (line)). The relative frequencies, obtained by dividing each frequency by the maximum frequency, of 37 arc patterns for each data unit were calculated from the database of binary patterns and used as explanatory variables to calculate PLSD and LDF.

The eigenvalues of the pooled within-group covariance matrix are shown in the descending order of magnitude in Table 1. The eigenvalues except for the largest nine are very close to zero. It indicates that the data show the phenomenon of singularity which is caused by the high degree of multicollinearity. So, it is impossible to calculate LDF of all variables because there does not exist the inverse of the covariance matrix. The above fact forces us to perform variable selection for obtaining LDF.

1	2	3	4	5	6	7	8
0.970688	0.182884	0.147894	0.044723	0.010319	0.005421	0.004145	0.002316
9	10	11	12	13	14	15	16
0.001089	0.000459	0.000435	0.000340	0.000206	0.000139	0.000111	0.000081
17	18	19	20	21	22	23	24
0.000071	0.000064	0.000058	0.000054	0.000041	0.000035	0.000031	0.000028
25	26	27	28	29	30	31	32
0.000025	0.000014	0.000010	0.000009	0.000009	0.000008	0.000006	0.000005
33	34	35	36	37			
0.000005	0.000003	0.000002	0.000002	0.000000			

TABLE 1. Eigenvalues of the pooled within-group covariance matrix in the descending order of magnitude.

Before applying discriminant analysis we performed principal component analysis to describe roughly the distribution of the observations of 20 persons. The first and second eigenvalues and their proportions are 1st = 0.9707 (70.83%) and 2nd = 0.1829 (13.34%), respectively. Figure 6 shows the scatter plot of the scores of the first and second principal components. 20 persons are marked as $0, 1, \cdots, 9, a, b, \cdots, j$ in the input order. From this scatter plot we can see that, though the same marks tend to locate relatively small regions, it is difficult to explain all the variations of 20 persons by these two principal components.

6.2 Results

As noted in the previous section, LDF with all of the 37 variables can not be calculated, while PLSD with the same variables can be obtained. At first, PLSD is applied to all variables of the arc pattern data. Next, we select variables by stepwise forward variable selection algorithm in LDF and compare the performance of PLSD and LDF calculated on the selected same variables.

Table 2 shows cross-validated classification matrix based on PLSD model with 29 components. Here the number of components is determined by the method of cross-validation. This classification matrix shows the numbers of correct and incorrect classifications where the discriminant functions are computed on a subset without one case and the removed one case is used for testing in turn. Correctly classified cases appear on the diagonal part of the matrix since the predicted person (row) and actual person (column) are the same. By using the PLSD model with 29 components, 351 cases out of 360 are classified correctly (CDR=97.50%).

In stepwise forward variable selection algorithm, the two selection criteria, Wilk's lambda and Mahalanobis' distance, are often used. So-called p-to-enter and p-to-remove are fixed at 0.15 and 0.20 for the two criteria. For details of the above stepwise method and the two criteria for variable selection, refer to, e.g., SPSS inc. (1993). One set selected by Wilk's lambda contains the following 23 variables:X1,X2,X3,X5,X6,X8,X9,X11,X12,X14,X16,X18,X22,X23,X26,X27,X29,X30,X33,X34, X35,X36,X37. The other set by Mahalanobis' distance contains the following 24 variables:X1,X2,X5, X9,X11,X12,X13,X14,X16,X18,X19,X20,X21,X22,X23,X25,X26,X27,X28,X29,X31,X33,X35,X37.

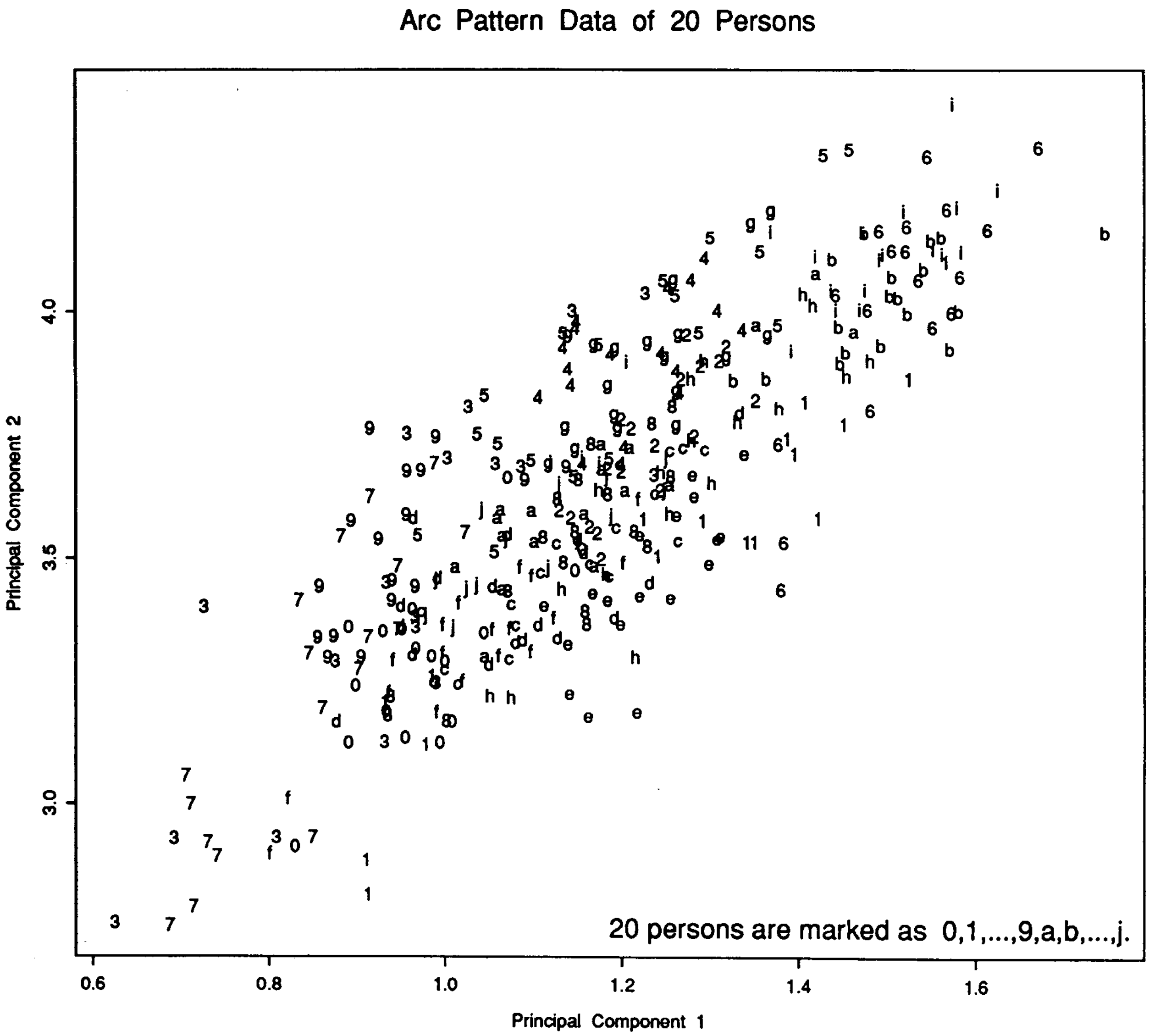

FIGURE 6. Scatter plot of the first and second principal components. The eigenvalues and their proportions of these two principal components are : 1st = 0.9707 (70.83%), 2nd = 0.1829 (13.34%). 20 persons are marked as $0, 1, \cdots, 9, a, b, \cdots, j$ in the input order.

Here symbol X corresponds to symbol Mo., for example, X1 is equivalent to Mo.1 in Figure 2. LDF and PLSD show the same cross-validated CDR, 97.78% for the set of 23 variables. For the set of 24 variables, LDF and PLSD show 95.83% and 96.11%, respectively. We can see that PLSD shows the same or better performance.

Person	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	18																			
2		18						1												
3			17																	
4				18					1											
5					18														1	
6						18														
7							18													
8								17												
9									17											
10										18										
11											18			1						
12												18								
13													17			2				
14														16						
15															18					
16													1			16				
17			1											1			18			
18																		18		
19																			17	
20																				18

TABLE 2. Cross-validated classification matrix by PLSD model with 29 components. (Note. Correctly classified cases appear on the diagonal part since the predicted person (row) and actual person (column) are the same. In total 351 cases out of 360 are classified correctly.)

7 Discussions

In the present paper artificial data in a simulation study and actual data in a writer identification problem were analyzed with LDF and PLSD. In the simulation study, PLSD showed almost the same performance with LDF in case of no multicollinearity problem and the performance superior to LDF in case of weak to severe multicollinearity situations. In the analysis of writer identification, LDF could not be obtained for all variables because of the nonexistance of the inverse matrix of the pooled within-group covariance matrix, while PLSD could be obtained and showed good performance. For the two sets of selected variables, both of LDF and PLSD showed good performance. This fact suggests that we could obtain a data set with no multicollinearity problem by applying the variable selection procedure in this case. But, even in this case we may say that PLSD has a merit that can be applied to the data with high degree of multicollinearity, which often occurs in pattern recognition analysis, without selecting variables.

It is our open and future task to compare PLSD not only with other kinds of discriminant functions such as Fisher's cannonical discriminant function and discriminant function using principal components but also with the recent computer-oriented methods like neural networks and genetic algorithm.

Acknowledgement

The authors would like to thank Professor Isao Yoshimura of Science University of Tokyo and Professor Mitsu Yoshimura of Chubu University for their kind permission of using their data and for their helpful advices.

References

[1] Anderson, T. W. (1984), An Introduction to Multivariate Statistical Analysis, 2nd Edition, John Wiley & Sons, Inc., New York, pp.195-243.

[2] Belsley, D. A., Kuh, E. and Welsch, R. E. (1980), Regression Diagnostics : Identifying Influential Data and Sources of Collinearity, Wiely, New York.

[3] Frank, I. E. and Friedman, J. H. (1993), A Statistical View of Some Chemometrics Regression Tools, Technometrics, Vol. 35, No. 2, pp.109-148.

[4] Kim, H. B. and Tanaka, Y. (1994), A Numerical Study of Partial Least Squares Regression with an Emphasis on the Comparison with Principal Component Regression, The Eighth Japan and Korea Joint Conference of Statistics, Okayama, pp.83-88.

[5] Kim, H. B. and Tanaka, Y. (1995a), Linear Discriminant Function Using Partial Least Squares Method, Proceedings of International Conference on Statistical Methods and Statistical Computing for Quality and Productivity Improvement (ICSQP '95), Vol 2, Seoul, pp.875-881.

[6] Kim, H. B. and Tanaka, Y. (1995b), Generating Artificial Data with Preassigned Degree of Multicollinearity by Using Singular Value Decomposition, The Journal of Japanese Society of Computational Statistics, Submitted for publication, September, 1995.

[7] SPSS inc. (1993), SPSS Professional Statistics, Release 6.x, SPSS inc.

[8] Plamondon, R. and Lorette, G. (1989), Automatic Signature Verification and Writer Identification – The State of the Art, Pattern Recognition, vol. 22, pp. 107-131.

[9] Schalkoff, R. (1992), Pattern Recognition: Statistical, Structural and Neural Approaches, John Wiley & Sons, Inc., New York.

[10] Wold, H. (1975), Soft Modeling by Latent Variables ; the Non-linear Iterative Partial Least Squares Approach, In Perspectives in Probability and Statistics, Papers in Honour of M. S. Bartlett, Edited J. Gani, Academic Press, Inc., London.

[11] Yoshimura, M. and Yoshimura, I. (1991), Writer Identification using Localized Arc Pattern Method, Trans. IEICE, vol. J74-D-II, pp. 230-238.

[12] Yoshimura, I., Yoshimura, M. and Kim H. B. (1993), A Text-Independent Off-Line Writer Identification Method for Japanese and Korean Sentences, Trans. IEICE, Inf. & Syst., vol. E76-D, No 4, pp. 454-461.