

# A New Approach for Classifying Large Number of Mixed Variables

Hashibah Hamid

**Abstract**—The issue of classifying objects into one of predefined groups when the measured variables are mixed with different types of variables has been part of interest among statisticians in many years. Some methods for dealing with such situation have been introduced that include parametric, semi-parametric and non-parametric approaches. This paper attempts to discuss on a problem in classifying a data when the number of measured mixed variables is larger than the size of the sample. A propose idea that integrates a dimensionality reduction technique via principal component analysis and a discriminant function based on the location model is discussed. The study aims in offering practitioners another potential tool in a classification problem that is possible to be considered when the observed variables are mixed and too large.

**Keywords**—classification, location model, mixed variables, principal component analysis.

## I. INTRODUCTION

**D**ISCRIMINANT analysis is a statistical technique that allows one in understanding the differences of objects between two or more groups with respect to several variables simultaneously [1]-[2]-[3]. It is the first multivariate statistical classification method used for decades by researchers and practitioners in developing classification models [4]. In general, discriminant analysis concerns with the development of a rule for allocating objects into one of some distinct groups. Then, the constructed classification rule will be used to determine a group of some future objects. Many different classification rules have been developed with various conditions such as rules that are suitable for either continuous variables or categorical variables (see [5]-[6]-[7]). However, real problems of classification sometimes show that the variables are mixed with continuous and categorical variables [8]-[9]-[10].

Many studies on dealing with mixed types of variables have been discusses and they can be generally grouped into three common strategies. The simplest strategy is to transform all the variables into the same type, either continuous or categorical, then construct a classification rule that is suitable to this type. However, it leads to loss of information due to the transformation process [11]-[12]-[13]. Alternatively, one may construct separate classification rules for each type of variables and then combine the results for summarising as

Hashibah Hamid is with the Statistics Department, College of Arts and Sciences, Universiti Utara Malaysia, 06010 Sintok Kedah, MALAYSIA and now is on study leave (e-mail: hashibah@uum.edu.my).

Author thank to the Malaysia Ministry of Higher Education (MOHE) and Universiti Utara Malaysia for the financial support.

overall classification. Such strategy however needs more discussions on tackling the issue of combining the results as few studies on this strategy have been done [14]-[15]. If these two strategies may not offer a good solution, then one may develop a model that can manage different types of variable simultaneously, after that derives a classification rule from this model. This strategy has been discussed in depth by [9]-[11]-[16]-[17]-[18] from the statistical point of view.

Some researchers have advocated a classification rule for mixed variables. To name few: reference [19] proposed a method based on the kernel approach in which the classification procedure is based on the estimation of groups' density, [20] introduced the logistic discrimination and [11] developed the location model where the estimators are obtained from the likelihood approach. It is beyond the scope of this paper to discuss in details of each existing rule, but readers are encouraged to have further information from [13]. Although there are many classification rules that can be considered, but most of them are not permitted to be constructed when the number of observed variables is larger than the number of the objects in the sample. The rules that depend on the covariance matrix may suffer from the occurrence of singularity of the matrix and others may obtain biased estimators if the likelihood approach is used.

Therefore, this paper is aiming on discussing the issue but the intention is based on the location model. It will be focusing on introducing the idea for constructing the model when the number of variables typically larger than the number of objects in the dataset. The discussion is limited to the two groups problem and the variables engage are the combination of continuous and binary. Section II is giving the idea of the location model and dimensionality reduction approaches. It is followed by Section III that outlines the proposed idea for constructing the location model on large dimensional space. Finally, Section IV summarizes the findings that expected to obtain from this study which is a new approach of classification model when dealing with large number of mixed variables.

## II. THE LOCATION MODEL AND THE ISSUE OF LARGE NUMBER OF VARIABLES

Location Model (LM) allows one to allocate a new object into one of the two-group based on the series of measurements taken on that object [21]. Its simplicity and reasonably efficient in classification procedures based on normal populations are essential in statistical practice. When the model assumes equal costs of misclassification and equal a priori probabilities of group membership, then it leads to a linear function of the observations as the appropriate

classification statistic [22]. Although it is almost impossible to guarantee the populations to be exactly normal distribution, the model has been proved feasible if the data are nearly normal distributed [23] or not much skewed from normality assumption [24].

Let a vector  $\mathbf{z}^T = (\mathbf{x}^T, \mathbf{y}^T)$  is observed on each object where  $\mathbf{x}^T = (x_1, x_2, \dots, x_b)$  is a vector of  $b$  binary variables and  $\mathbf{y}^T = (y_1, y_2, \dots, y_c)$  is a vector of  $c$  continuous variables. The binary variables are treated as a single multinomial variable having  $s = 2^b$  cells. The location model assumes the probabilities of obtaining an object in cell  $m$  of the table for population  $\pi_i$  is  $p_{im}$  for  $i=1,2$  and  $m = 1, 2, \dots, s$ . We also assume the vector of continuous variables to have a multivariate normal distribution with mean  $\boldsymbol{\mu}_i^{(m)}$  in cell  $m$  of  $\pi_i$  and a homogeneous covariance matrix across cells and populations. Hence, we have  $\mathbf{Y}_{im} \sim N(\boldsymbol{\mu}_i^{(m)}, \boldsymbol{\Sigma})$  for the combination of  $m$  cell on the  $i$ th population. If we assume all population parameters are known then the optimal allocation rule for an object  $\mathbf{z}^t = (\mathbf{x}^t, \mathbf{y}^t)$  can be derived easily from the general theory of classification [25]. Then, we assign a new object  $\mathbf{z}^t = (\mathbf{x}^t, \mathbf{y}^t)$  to  $\pi_i$  if  $\mathbf{x}$  falls in cell  $m$  of the multinomial variable and

$$(\boldsymbol{\mu}_1^{(m)} - \boldsymbol{\mu}_2^{(m)})^T \boldsymbol{\Sigma}^{-1} \left\{ \mathbf{y} - \frac{1}{2}(\boldsymbol{\mu}_1^{(m)} + \boldsymbol{\mu}_2^{(m)}) \right\} \geq \log\left(\frac{p_{2m}}{p_{1m}}\right) + \log(a) \quad (1)$$

otherwise to  $\pi_2$  [11]-[12]. The constant  $a$  depends on the costs due to misclassification and prior probabilities for the two groups and is equivalent to zero for the case of equal costs and equal priors. In practice, however, the parameter values are unknown and the common procedure is to replace all parameters in equation (1) with a sample-based allocation rule.

Often the maximum likelihood is used to estimate the unknown parameters [26] but estimators may be biased if one is dealing with a sparse data problem [27] thus limits the feasibility of the linear model approach [28]. Alternative approach which is based on a non-parametric smoothing is possible to be considered since it gives a marked reduction in the number of parameters that need to be estimated and hence alleviates the over-parametrization problems of the classical location model [9]. However, [18] discovered that this alternative is sometimes infeasible as the smoothing approach that based on the nearest neighbour may take negative or zero values. Such phenomenon lead to a problem in allocating future objects. Further modifications were done where a smoothed location model was proposed to tackle such problem but it may be not practically applicable when the number of binary variables is too large as the model will suffering from the singularity of the covariance matrix.

Following this, some adjustments towards reducing the number of variables need to be done to allow classification rule to be computed, it is the case the corresponding number of variables typically larger than the number of objects in the samples [29]-[30]-[31]. As issued by [32], it may provide a promising approach to deal with a very large dimensional and low sample size data. Additionally, according to [33] and [34], dimension reduction step is an important, thus ignoring them would definitely underestimate the prediction error. Therefore, common adjustment procedures are either (i) choosing some

important variables through feature/variable selection process or (ii) projecting the data onto a low dimensional subspace via linear combinations of those variables, this is known as a feature extraction technique.

#### A. The Classification with Dimensionality Reduction Approaches

According to [35], an important query in application of classification is whether all the variables on which measurements are made contain useful information or only some of them may be sufficient for the principle of classification. Therefore, the first step is to choose the best set of variates in order to construct and to obtain a superior classification rule which is able to bound and explain the nature of the study problem as originate as possible. But, selecting the most useful variables in discriminant analysis involves big challenges [36] as it is entailing with a large number of variables [37]. In such situation, it is desirable to select a subset of variables and hoping that not too much information has been ignored [38]. However, omitting the most useless or the least important variables would result in several negative aspects. For instance, stepwise regression methods (stepwise method, forward selection and backward elimination) can be used for this purpose due to their simplicity but some of the researchers are often not satisfied with its results. Their performance is poor when it is believe that multicollinearity exists among variables [39]. Besides, omitting some of the original variables may appear to be relevant and contributed to the study later [40] and the omitting the ones that are appear to have highly correlated with those retained will give a major influence on the efficiency measures [41]. The results of omitting different variables are extremely difficult to predict. Unfortunately, it is vary greatly according to which highly correlated variables are included or omitted even when the scientific or managerial justification for the omission or inclusion of certain variables is reasonable.

On the contrary, feature extraction has been approved as an important method when facing with large number of variables compared to the size of sample [42]. A considerable amount of research has been assigned for solving small sample and high dimensional problems [43]-[44]-[45]-[46]-[47]-[48]. It plays an important role in many applications due to a large dataset such as data mining, machine learning and bioinformatics [49]. One necessary part of multivariate statistical analysis in such applications is by applying dimensional reduction [30]. As quoted by [31], feature extraction is superior in getting a very low dimensional representation. He further stated that, by this procedure, it can effectively encode by a low dimensional vector which can significantly reduce the recognition cost.

Several methods exist for that particular purpose and some researchers have been applied the principal component analysis [50]-[51], latent semantic indexing [52]-[53], Partial least squares [54]-[55]-[56]-[57]-[58] and sliced inverse regression [59]-[60]-[61]-[62] as preprocessor for dimensionality reduction [63]-[64]. Over the past ten years, however, the principal component analysis (PCA) as data

reduction has received significant attention among all methods [65]. PCA is a well-known method for dimensional reduction [66]. Based on [67], the goal of PCA is to reduce its dimension into a low dimensional space which most of the essential information contained in a high dimensional space. Reference [68] and [69] further stated that, the most common technique which dimensionality can be reduce easily without disturbing the overall features of the sample is PCA. From its process, the singularity problem would be solved [70] and could avoid an *over-fitting* [71]. Furthermore, the new variables produced from PCA show no correlation among variables which is typically required particularly in high dimensional space [72]. When predictor variables are highly correlated, the model may encounter instability crisis because the existing of multicollinearity [73]. This is due to the arithmetic problem occurs from such collinearity issue [74]. As quoted by [75] that, even our model gives a good discrimination performance, there still exist some weaknesses as it is too flexible in situations if there are many highly correlated among variables. PCA can avoid and remedy this problem of multicollinearity [38] because PCA is a typical case that decorrelated and extracts the variables [76]. It is the case where the independent variables are linearly or near-linearly depend on each other, it is common practice to transform these variables into principal component scores (PCs) which are orthogonal and then regress the dependent variable on these new variables (the PCs) [77].

PCA yields a low dimensional subspace which minimizes the mean square error by finding the directions of major variations in the whole learning set [71]. In addition, the almost information in reduced components as there is in the original variables [23] with the idea of minimizing information loss is fundamental [78]. Throughout PCA, we will get a lower dimensional representation of the data which moves some of the *noisy directions* [79]. In fact, procedures involve in selecting variables using PCA will retain overall multivariate structure of the complete data as much as possible [77]. Furthermore, according to [67], PCA is a mathematically appealing statistical tool suitable for examining variables in high dimensional spaces. Nowadays, even with readily available compute power and sophisticated statistical algorithms, PCA is still a popular tool for dimensionality reduction or exploratory analysis [80]. He added that, the application of PCA to a very high dimension with small data has only increased. With many advantages of PCA as described, thus we fix on PCA to reduce the data dimension as the main tool to achieve the objectives of this research paper.

### III. METHODOLOGY

#### A. Data

We will use simulated data where the two groups are having multivariate normal distributions;  $MVN(\mu_1, \Sigma)$  and  $MVN(\mu_2, \Sigma)$  with  $p = 50$  continuous variables and  $q = 50$  binary variables. Common procedures for separately generating either continuous or binary variables are available in most statistical packages such as Minitab, S-Plus and R. However,

there is no package so far is able to generate and provide a set of data for various types of variables simultaneously.

Following Guo et al. (2007), a set of multivariate continuous variables will be generated by setting all entries of  $\mu_1$  to be 0 and for  $\mu_2$ , the first 30 entries are 0.5 and the rest are all 0, i.e.  $\mu_1 = \{0\}_{100}$  and  $\mu_2 = \{0.5, \dots, 0.5_{30}, 0_{31}, \dots, 0_{100}\}$ . Then, the first 50 variables in each group will be remained as they are generated while the last 50 observations will be discretised to form binary variables. We will generate  $n = 30$  with 15 objects for each group. This study will examine the model performance from some typical data distributions which are normal and non-normal cases. Finally, some investigations on real datasets also will be considered.

#### B. Model Construction and Evaluation

The objects in the sample will be divided into two parts: learning set and test set. The learning set will be used for constructing the model and the test set will be used for evaluation purposes. This mechanism will give more reliable and unbiased estimate of classification error as different objects are used for model construction and model evaluation following [81].

The proposed location model will be evaluated using the common leave-one-out error rate. It gives the proportion of objects that are misclassified hence is able to represent the performance of the constructed rule. In general, some procedures involve sequentially in order to accomplish the final algorithm of location model which can be summarized in several steps specifically:

- i) Omit object  $i$  from the sample  $n$  where  $i = 1, 2, \dots, n$ .
- ii) Perform PCA for the continuous variables from the remaining objects ( $n_1+n_2-i$ ) to choose the best combination of components or to reduce its dimensions.
- iii) Repeat step (ii) for conducting PCA purposely for binary variables, then combine the results from phase (ii) and (iii) to produce 2PCA.
- iv) Compute and estimate  $\mu_i$ ,  $\Sigma$  and  $\rho_i$  using the new components resulting from 2PCA, further construct the location model function.
- v) Pool and run step (ii-iv) together to produce a new algorithm of 2PCA plus LM.
- vi) Predict the group of the omitted object  $i$  using a new constructed model, if the prediction made is correct then assign error ( $\epsilon_{ij}$ ) = 0 otherwise  $\epsilon_{ij} = 1$ .
- vii) Repeat step (i) - (vi) for all objects in turn.
- viii) Compute the leave-one-out error rate using  $(\sum \epsilon_{ij} / n) \times 100$ .

Fig. 1 demonstrates the study design that integrates the two standard methods in producing a new classification algorithm of the location model when dealing with large number of mixed variables namely 2PCA plus LM or simply called 2PCALM.

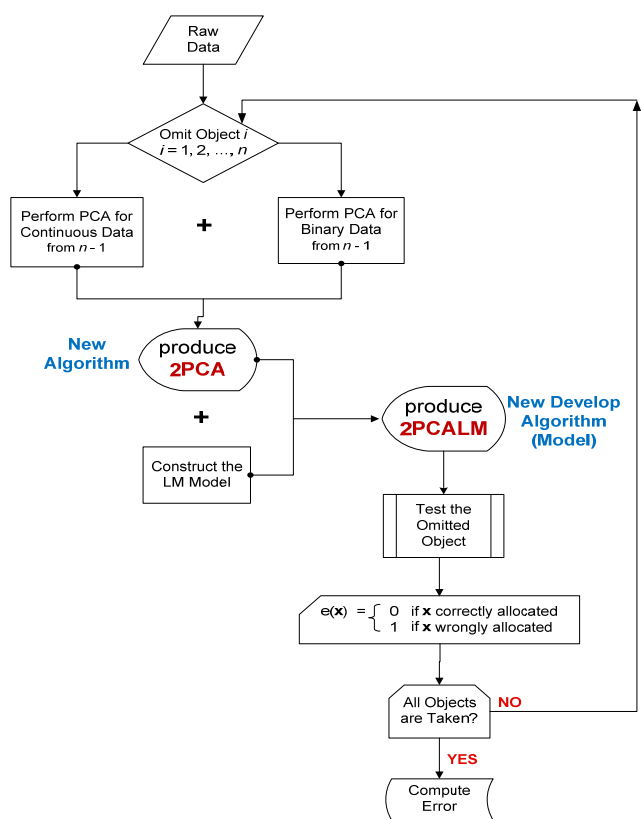


Fig. 1 The new location model for large number of mixed variables

#### IV. EXPECTED FINDINGS

A new mechanism of discriminant analysis of the location model for handling large number of mixed variables will be produced. We also attempt to obtain a new algorithm of data reduction that is able to manage different types of variables (binary and continuous) simultaneously.

#### REFERENCES

[1] W. R. Klecka, *Discriminant Analysis. Series: Quantitative Applications in the Social Sciences*. A Sage University Paper. Beverly Hills, CA: Sage, 1980.

[2] M. Núñez, A. Villarroya and J. M. Oller, "Minimum Distance Probability Discriminant Analysis for Mixed Variables," *Biometrics*, vol. 59, pp. 248-253, 2003.

[3] Y. H. Chan, "Biostatistics 303: Discriminant Analysis," *Singapore Medical Journal*, vol. 46, no. 2, pp. 54-61, 2005.

[4] M. Doumpos and C. Zopounidis, *Multicriteria Decision Aid Classification Methods*. Kluwer Academic Publishers, 2002.

[5] R. A. Fisher, "The Use of Multiple Measurements in Taxonomic Problems," *Annals of Eugenics*, vol. 7, no. 2, pp. 179-188, 1936.

[6] S. B. Bull and A. Donner, "The Efficiency of Multinomial Logistic Regression compared with Multiple Group Discriminant Analysis," *Journal of the American Statistical Association*, vol. 82, pp. 1118-1122, 1987.

[7] J. A. Anderson, *Logistic Discrimination*. In *Handbook of Statistics (Vol. 2)* P. R. Krishnaiah and L. N. Kanal (Eds.). Amsterdam: North-Holland, pp. 169-191, 1992.

[8] J. J. Daudin, "Selection of Variables in Mixed-variable Discriminant Analysis," *Biometrics*, vol. 42, no. 3, pp. 473-481, 1986.

[9] O. Asparoukhov and W. J. Krzanowski, "Non-parametric Smoothing of the Location Model in Mixed Variable Discrimination," *Statistics and Computing*, vol. 10, pp. 289-297, 2000.

[10] A. Merbouha and A. Mkhadri, "Regularization of the Location Model in Discrimination with Mixed Discrete and Continuous Variables," *Computational Statistics and Data Analysis*, vol. 45, pp. 563-576, 2004.

[11] W. J. Krzanowski, "Discrimination and Classification using Both Binary and Continuous Variables," *Journal of the American Statistical Association*, vol. 70, no. 352, pp. 782-790, 1975.

[12] W. J. Krzanowski, "The Location Model for Mixtures of Categorical and Continuous Variables," *Journal of Classification*, vol. 10, pp. 25-49, 1993.

[13] D. J. Hand, *Construction and Assessment of Classification Rules*. Chichester: John Wiley & Son, 1997.

[14] K. D. Wernecke, "A Coupling Procedure for the Discrimination of Mixed Data," *Biometrics*, vol. 48, no. 2, pp. 497-506, 1992.

[15] L. Xu, A. Krzyzak and C. Y. Suen, "Methods of Combining Multiple Classifiers and Their Applications to Handwriting Recognition," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 22, no. 3, pp. 418-435, 1992.

[16] P. C. Chang and A. A. Afifi, "Classification based on Dichotomous and Continuous Variables," *Journal of the American Statistical Association*, vol. 69, no. 346, pp. 336-339, 1974.

[17] W. J. Krzanowski, "Mixtures of Continuous and Categorical Variables in Discriminant Analysis," *Biometrics*, vol. 36, pp. 493-499, 1980.

[18] N. I. Mahat, W. J. Krzanowski and A. Hernandez, "Strategies for Non-Parametric Smoothing of the Location Model in Mixed-Variable Discriminant Analysis," *Modern Applied Science*, vol. 3, no. 1, pp. 151-163, 2009.

[19] J. Aitchison and C. G. G. Aitken, "Multivariate Binary Discrimination by the Kernel Method," *Biometrika*, vol. 63, pp. 413-420, 1976.

[20] J. A. Anderson, "Separate Sample Logistic Discrimination," *Biometrika*, vol. 59, no. 1, pp. 19-35, 1972.

[21] D. J. Hand, J. J. Oliver and A. D. Lunn, "Discriminant Analysis when the Classes Arise from a Continuum," *Pattern Recognition*, vol. 31, no. 5, pp. 641-650, 1998.

[22] P. A. Lachenbruch, C. Sneeringer and L. T. Revo, "Robustness of the Linear and Quadratic Discriminant Function to Certain Types of Non-normality," *Communications in Statistics*, vol. 1, pp. 39-56, 1973.

[23] R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis (3rd edition)*. New Jersey, Englewood Cliffs: Prentice Hall, 1992.

[24] A. Das, *Discriminant Analysis and Its Applications. BDM&DM Term Paper*, 2009.

[25] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis (2nd edition)*. New York: John Wiley & Sons, 1984.

[26] Guo, T. Hastie and R. Tibshirani, "Regularized Linear Discriminant Analysis and Its Application in Microarrays," *Biostatistics*, vol. 8, no. 1, pp. 86-100, 2007.

[27] I. G. Vlachonikolis and F. H. C. Marriott, "Discrimination with Mixed Binary and Continuous Data," *Applied Statistics*, vol. 31, no. 1, pp. 23-31, 1982.

[28] N. I. Mahat, W. J. Krzanowski and A. Hernandez, "Variable Selection in Discriminant Analysis Based on the Location Model for Mixed Variables," *Advances in Data Analysis and Classification*, vol. 1, no. 2, pp. 105-122, 2007.

[29] H. Ping, "Classification Methods and Applications to Mass Spectral Data," unpublished *PhD Thesis*. Hong Kong: Baptist University, Department of Mathematics, 2005.

[30] J. J. Dai, L. Lieu and D. Roocke, "Dimension Reduction for Classification with Gene Expression Microarray Data," *Statistical Applications in Genetics and Molecular Biology*, vol. 5, no. 1, article 6, 20 pages, 2006.

[31] Q. Li, "An Integrated Framework of Feature Selection and Extraction for Appearance-based Recognition," unpublished *PhD Thesis*. University of Delaware: Faculty of Computer Science, 2006.

[32] J. S. Marron, M. J. Todd and J. Ahn, "Distance Weighted Discrimination," *Journal of the American Statistical Association*, vol. 480, pp. 1267-1271, 2007.

[33] C. Ambroise and G. J. McLachlan, "Selection Bias in Gene Extraction on the Basis of Microarray Gene-Expression Data," *Proceedings of the National Academy of Sciences*, 2002, vol. 99, no. 10, pp. 6562-6566.

- [34] R. Simon, M. D. Radmacher, K. Dobbin and L. M. McShane, "Pitfalls in the Use of DNA Microarray Data for Diagnostic and Prognostic Classification," *Journal of the National Cancer Institute*, vol. 95, no. 1, pp. 14-18, 2003.
- [35] Z. Qiao, L. Zhou and J. Z. Huang, "Effective Linear Discriminant Analysis for High Dimensional, Low Sample Size Data," *Proceedings of the World Congress on Engineering (WCE)*, 2008, vol. 2, pp. 1070-1075.
- [36] G. J. McLachlan, "A Criterion for Selecting Variables for the Linear Discriminant Function," *Biometrics*, vol. 32, no. 3, pp. 529-534, 1976.
- [37] P. Xu, G. N. Brock and R. S. Parrish, "Modified Linear Discriminant Analysis Approaches for Classification of High-Dimensional Microarray Data," *Computational Statistics and Data Analysis*, vol. 53, no. 5, pp. 1674-1687, 2009.
- [38] Y. Lu, Q. Tian, M. Sanchez, J. Neary, F. Liu and Y. Wang, "Learning Microarray Gene Expression Data by Hybrid Discriminant Analysis," *IEEE Multimedia Magazine, Special Issue on Multimedia Signal Processing and Systems in Health Care and Life Science*, vol. 14, no. 4, pp. 22-31, 2007.
- [39] I. G. Chong and C. H. Jun, "Performance of Some Variable Selection Methods when Multicollinearity is Present," *Chemometrics and Intelligent Laboratory Systems*, vol. 78, pp. 103-112, 2005.
- [40] J. M. Weiner and O. J. Dunn, "Elimination of Variate in Linear Discrimination Problems," *Biometrics*, vol. 22, no. 2, pp. 268-275, 1966.
- [41] L. Jenkins and M. Anderson, "A Multivariate Statistical Approach to Reducing the Number of Variables in Data Envelopment Analysis," *European Journal of Operational Research*, vol. 147, no. 1, pp. 51-61, 2003.
- [42] F. Nie, S. Xiang, Y. Song and C. Zhang, "Extracting the Optimal Dimensionality for Discriminant Analysis," *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 2, pp. 617-620, 2007.
- [43] P. N. Bellhumeur, J. P. Hespanha and D. J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition using Class Specific Linear Projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711-720, 1997.
- [44] W. Zhao, R. Chellappa and N. Nandhakumar, "Empirical Performance Analysis of Linear Discriminant Classifiers," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1998, pp. 164-169.
- [45] L. F. Chen, H. Y. M. Liao, M. T. Ko, J. C. Lin and G. J. Yu, "A New LDA-based Face Recognition System which can Solve the Small Sample Size Problem," *Pattern Recognition*, vol. 33, no. 10, pp. 1713-1726, 2000.
- [46] J. Yang and J. Y. Yang, "Optimal FLD Algorithm for Facial Feature Extraction," *SPIE Proceedings of the Intelligent Robots and Computer Vision XX: Algorithms, Techniques, and Active Vision*, 2001, vol. 4572, pp. 438-444.
- [47] H. Yu and J. Yang, "A Direct LDA Algorithm for High-Dimensional Data with Application to Face Recognition," *Pattern Recognition*, vol. 34, no. 10, pp. 2067-2070, 2001.
- [48] J. Yang and J. Y. Yang, "Why Can LDA be Performed in PCA Transformed Space? Rapid and Brief Communication," *Pattern Recognition*, vol. 36, no. 2, pp. 563-566, 2003.
- [49] J. Ye and T. Xiong, "Computational and Theoretical Analysis of Null Space and Orthogonal Linear Discriminant Analysis," *Journal of Machine Learning Research*, vol. 7, pp. 1183-1204, 2006.
- [50] S. Wold, K. Esbensen and P. Geladi, "Principal Component Analysis," *Chemometrics Intelligent Laboratory Systems*, vol. 2, pp. 37-52, 1987.
- [51] D. Ghosh, "Singular Value Decomposition Regression Modeling for Classification of Tumors from Microarray Experiments," *Proceedings of the Pacific Symposium on Biocomputing*, 2002, pp. 11462-11467.
- [52] S. Deerwester, S. T. Dumais, G. W. Furnas, T. K. Landauer and R. Harshman, "Indexing by Latent Semantic Analysis," *Journal of the American Society for Information Science*, vol. 41, no. 6, pp. 391-407, 1990.
- [53] A. Kabán and M. A. Girolami, "Fast Extraction of Semantic Features from a Latent Semantic Indexed Corpus," *Neural Processing Letters*, vol. 15, no. 1, pp. 31-43, 2002.
- [54] P. M. Garthwaite, "An Interpretation of Partial Least Squares," *Journal of the American Statistical Association*, vol. 89, no. 425, pp. 122-127, 1994.
- [55] D. V. Nguyen and D. M. Rocke, "Tumor Classification by Partial Least Squares using Microarray Gene Expression Data," *Bioinformatics*, vol. 18, no. 1, pp. 39-50, 2002a.
- [56] D. V. Nguyen and D. M. Rocke, "Multi-class Cancer Classification via Partial Least Squares with Gene Expression Profiles," *Bioinformatics*, vol. 18, pp. 1216-1226, 2002b.
- [57] X. Huang and W. Pan, "Linear Regression and Two-class Classification with Gene Expression Data," *Bioinformatics*, vol. 19, pp. 2072-2078, 2003.
- [58] A. Boulesteix, "PLS Dimension Reduction for Classification with Microarray Data," *Statistical Applications in Genetics and Molecular Biology*, vol. 3, pp. 1-33, 2004.
- [59] R. D. Cook, *Regression Graphics*. New York: John Wiley & Sons, 1998.
- [60] F. Chiaromonte and J. Martinelli, "Dimension Reduction Strategies for Analyzing Global Gene Expression Data with a Response," *Mathematical Biosciences*, vol. 176, pp. 123-144, 2002.
- [61] A. Antoniadis, S. Lambert-Lacroix and F. Leblanc, "Effective Dimension Reduction Methods for Tumor Classification using Gene Expression Data," *Bioinformatics*, vol. 19, pp. 563-570, 2003.
- [62] E. Bura and R. M. Pfeiffer, "Graphical Methods for Class Prediction using Dimension Reduction Techniques on DNA Microarray Data," *Bioinformatics*, vol. 19, pp. 1252-1258, 2003.
- [63] W. Zhao, R. Chellappa and P. J. Phillips, Subspace Linear Discriminant Analysis for Face Recognition. *Technical Report CAR-TR-914*. University of Maryland, College Park, 1999.
- [64] J. Baeka and M. Kimb, "Face Recognition using Partial Least Squares Components," *Pattern Recognition*, vol. 37, no. 6, pp. 1303-1306, 2004.
- [65] W. Zuo, D. Zhang, J. Yang and K. Wang, "BDPCA plus LDA: A Novel Fast Feature Extraction Technique for Face Recognition," *IEEE Transactions on Systems, Man, and Cybernetics: Part B-Cybernetics*, vol. 36, no. 4, pp. 946-953, 2006.
- [66] A. Caprihan, G. D. Pearlson and V. D. Calhoun, "Application of Principal Component Analysis to Distinguish Patients with Schizophrenia from Healthy Controls based on Fractional Anisotropy Measurements," *Neuroimage*, vol. 42, no. 2, pp. 675-682, 2008.
- [67] G. P. McCabe, *Principal Variables*. Technical Report. West Lafayette: Purdue University, 1982.
- [68] G. A. F. Seber, *Multivariate Observations*. New York: John Wiley & Sons, 1984.
- [69] I. T. Jolliffe, *Principal Component Analysis*. New York: Springer-Verlag, 1986.
- [70] H. B. Deng, L. W. Jin, L. X. Zhen and J. C. Huang, "A New Facial Expression Recognition Method on Local Gabor Filter Bank and PCA plus LDA," *International Journal of Information Technology*, vol. 11, no. 11, pp. 86-96, 2005.
- [71] W. Hwang, T. K. Kim and S. C. Kee, "LDA with Subgroup PCA Method for Facial Image Retrieval," *The 5th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS)*, Portugal: Lisbon, April 2004, pp. 21-23.
- [72] I. T. Jolliffe, *Principal Component Analysis* (2<sup>nd</sup> edition). New York: Springer-Verlag, 2002.
- [73] B. G. Amidan and D. N. Hagedorn, Logistic Regression Applied to Seismic Discrimination. *Technical Report*. Pacific Northwest National Laboratory (PNNL): Washington (US), Richland, 1998.
- [74] A. P. Worth and M. T. D. Cronin, "The Use of Discriminant Analysis, Logistic Regression and Classification Tree Analysis in the Development of Classification Models for Human Health Effects," *Theochem*, vol. 622, pp. 97-111, 2003.
- [75] H. C. Kim, D. Kim and S. Y. Bang, "Extensions of LDA by PCA Mixture Model and Class-wise Features," *Journal of the Pattern Recognition Society*, vol. 36, pp. 1095-1105, 2003.
- [76] Y. Liang, C. Li, W. Gong and Y. Pan, "Uncorrelated Linear Discriminant Analysis based on Weighted Pairwise Fisher Criterion," *The Journal of the Pattern Recognition Society*, vol. 40, pp. 3606-3615, 2007.
- [77] M. M. Sithole, "Variable Selection in Principal Component Analysis: Using Measures of Multivariate Association," unpublished *Master*

*Thesis*. Curtin University of Technology, School of Mathematics and Statistics, 1992.

- [78] C. R. Rao, "The Use and Interpretation of Principal Component Analysis in Applied Research," *Sankhyā: The Indian Journal of Statistics, Series A*, vol. 26, no. 4, pp. 329-358, 1964.
- [79] W. J. Krzanowski, "Selection of Variables to Preserve Multivariate Data Structure using Principal Components," *Applied Statistics*, vol. 36, no. 1, pp. 22-33, 1987.
- [80] D. C. Hoyle, "Automatic PCA Dimension Selection for High Dimensional Data and Small Sample Sizes," *Journal of Machine Learning Research*, vol. 9, pp. 2733-2759, 2008.
- [81] W. K. Härdle and Z. Hlávka, *Multivariate Statistics: Exercises and Solutions*. New York: Springer, 2007.