ISSN 1985-7918

Strategies of Handling Different Variables Reduction for LDA

Strategi Pengendalian Pengurangan Pemboleh-ubah yang Berbeza untuk LDA

Hashibah Hamid and Nor Idayu Mahat
UUM College of Arts and Sciences, School of Quantitative Sciences
Universiti Utara Malaysia, 06010 Sintok, Kedah, Malaysia.
e-mail: hashibah@uum.edu.my, noridayu@uum.edu.my

Abstract

This paper discusses the strategy of conducting variable reduction processes such that they contribute to optimise the performance of linear discriminant analysis (LDA). The variables selection technique with local searching algorithm is manipulated. The technique is proposed to choose useful variables that give minimum error rate on LDA. Meanwhile, principal component analysis is used to extract important information from the original variables. The behaviour of eigenvalue and total variation explained is studied to understand how these two indicators may give optimum performance of LDA. Performance of the proposed strategy and LDA with all variables was assessed in leave-one-out fashion to avoid biasness. This study discovers that LDA with backward elimination is competitive to the full model, but extra concern needs to be given to the PCA.

Keywords leave-one-out error, linear discriminant analysis, principal component analysis, variables reduction

Abstrak

Kajian ini membincangkan strategi menjalankan proses pengurangan pemboleh-ubah sebagaimana mereka menyumbang untuk mengoptimumkan prestasi analisis diskriminan linear (LDA). Teknik pemilihan pemboleh-ubah dengan algoritma carian tempatan dimanipulasikan. Teknik ini dicadangkan untuk memilih pemboleh-ubah yang berguna yang memberikan kadar kesilapan yang minimum pada LDA. Sementara itu, analisis komponen utama digunakan untuk mendapatkan maklumat penting daripada pemboleh-ubah asal. Tingkah laku nilai eigen dan jumlah variasi dijelaskan dikaji untuk memahami bagaimana kedua-dua penunjuk ini boleh memberikan prestasi LDA yang optimum. Prestasi strategi yang dicadangkan dan LDA dengan semua pemboleh ubah telah dinilai dalam fesyen *leave-one-out* bagi mengelakkan bias. Kajian ini mendapati bahawa LDA dengan penghapusan ke belakang adalah kompetitif kepada model penuh, tetapi perhatian tambahan perlu diberikan kepada PCA.

Kata kunci ralat *leave-one-out*, analisis diskriminan linear, analisis komponen utama, pengurangan pemboleh-ubah

Introduction

Linear Discriminant Analysis (LDA) is a well established statistical technique for classification and discrimination which was originally developed in 1936 by R.A. Fisher.

It has been widely used in many classification problems (Croux *et al*., 2008) and performs well in various applications including medicine, engineering, psychology, computer sciences, education and finance (see Pasiouras *et al*., 2005; Guliashki, 2006; Lu *et al*., 2007). A related mathematical function in LDA called linear discriminant function has optimum performance when normality holds with homogeneous covariance matrix among groups. Also, this classification rule performs well even in situations where the underlying properties like normally distributed data with constant covariance matrices over all groups are not met (Czogiel *et al*., 2007). Such behaviour has made the LDA becomes trusted and chosen by many practitioners.

However, the LDA can be seriously degraded if singularity covariance matrix occurs which often due to the measured variables exceeds the number of data points (Friedman, 1989; Chen *et al*., 2000; Zhang & Jia, 2007) or limited sample size (Thomaz & Gillies, 2005). The classification in such circumstances is typically a critical issue (Liang *et al*., 2007), indicates an over-fitting rule (Kim *et al*., 2003) and makes LDA difficult to work (Nie *et al*., 2007). As a result, it generally shows poor classification performance (Qiao *et al*., 2008). A common approach to deal with the singularity problem is to apply an intermediate variables reduction prior to construction the LDA by either (i) selecting the variables that are best discriminating the groups (Murray, 1977; Bishop, 1995) or (ii) combining the variables (Zhu & Martinez, 2006; Zuo *et al*., 2006) in such a way that its combination optimizes some performance indicator, e.g. minimum error rate (Belhumeur *et al*., 1997; Li, 2006; Fearn, 2008) and give largest separation between the groups (Jeffers, 1967; Héberger & Andrade, 2004; Huang *et al*., 2005; Dai *et al*., 2006).

The selection of variables that best discriminating the groups means that researchers are dealing with the process of determining a subset of measured variables. Then, use the chosen one to construct the linear discriminant function. Meanwhile, the combination of variables needs a systematic mechanism to join all measured variables through a mathematical function. Then, the new variables produced from the mathematical function are used for classification purposes. Both techniques have been exercised in classification task when researchers prefer to keep the number of variables at minimum. Most existing studies perform variables reduction process prior to LDA. Such independent processes are questionable because the aim of variable reduction process (to reduce the original variables) does not match with the aim of LDA (to split the groups).

Thus, this paper attempts to propose the idea of joining the two processes such that they are working at the same aim. The investigation covers some common data sets with different sizes of variables for two-group problem where the groups are assumed to have homogeneous covariance matrix. Section 2 overviews the concept of LDA and variables reduction techniques. Then, Section 3 gives the details about the proposed idea and investigations that were carried out. Results of the investigations are summarised in Section 4 and the final section concludes the findings.

LDA with Many Variables

Linear Discriminant Analysis

Suppose there are two groups, π_1 and π_2, both consist of objects with large number of p continuous variables. We denote the vector of p continuous variables in group π_i for

i = 1, 2 as $x^T = (x_1, x_2,..., x_p)$ and the probability of being in π_i as p_i. The idea of linear discriminant function in LDA is to use a linear combination of x of the n objects as $f(x)$ and choosing the coefficients so that the difference of the means of the linear combination in the two groups to its variance is maximized. When both groups are multivariate normal distributions having means i_1 and i_2 with a common covariance matrix Ó, the density function of x in π_i is

$$f(x \mid \pi_i) = \frac{1}{(2\pi)^{p/2} \mid \text{Ó} \mid^{1/2}} e^{-\frac{1}{2}(x-\mu_i)^T \sum^{-1}(x-\mu_i)} \tag{1}$$

Taking the logarithm of the ratio between $f(x)|\pi_1$ and $f(x)|\pi_2$ will give optimal classification function and will assign x to π_1 if

$$(i_1 - i_2)^T \text{Ó}^{-1}\left[x - \tfrac{1}{2}(i_1 - i_2)\right] > log\left(\frac{p_2}{p_1}\right) \tag{2}$$

Commonly the parameters i_i, Ó and p_i are unknown and they have to be estimated from sample. The simplest estimation approach is based on the maximum likelihood but sometimes the estimation is not permissible if the size of observed variables (p) is bigger than the size of sample (n) relatively due to the occurrence of singular covariance matrix. In such a case, some adjustments need to be done to allow equation (2) to be computed. Common procedures that can be applied are either (i) choosing some important variables or (ii) projecting the data onto a low dimensional subspace by linear combination of its variables.

LDA with variables selection technique

Selection of the most useful variables in discriminant analysis is an important but difficult task (Urbakh, 1971). There are two concerns in the variable selection technique: (i) the indicator to determine the best variables and (ii) the searching process of the variables. Previous studies have introduced some indicators in order to choose the best possible variables such as rule performance criteria (Ganeshanandam & Krzanowski, 1989), group separation criteria (McKay & Campbell, 1982; Daudin & Bar-Hen, 1999), model goodness-of-fit criteria including AIC and BIC (Daudin, 1986) and other criteria such as R^2, Hotelling's T^2, Wilk's Λ (Rencher, 1993) and t-statistics (Weiner & Dunn, 1966). Different used of indicators may lead to different results hence the choice of indicator varies depending on the aim of the study and application.

The best searching process of the best set of variables is to seek for all possible subsets of combination variables and choose the best combination that gives the best performance (Krzanowski, 1987). However, this strategy is exhaustive for large number of variables. Alternatively, researchers use some systematic searching strategy based on local searches through the famous forward, backward and stepwise selection. The local search is easy to perform but the outcome may not contain all useful variables and may eliminate the useful ones.

LDA with variables extraction technique

Sometimes the number of useful variables is big and variables selection process may not suitable to be implemented. In such a case, variable dimensionality reduction can be performed by projecting the original data into a low dimensional subspace through an extraction process (Liang *et al*., 2007). The extraction process assists to reduce the burden of data management and facilitates more accurate estimation of statistics (Nenadic, 2007). Thus, it helps to improve the recognition of accuracy and efficiency of the constructed rule (Li, 2006). Many possible variables extraction techniques are feasible to use such as principal component analysis, factor analysis, corresponding analysis, multidimensional scaling, Fourier analysis and much more. Different technique has different strengths and weaknesses and it is not the intention of this paper to review them all.

Over the past ten years, the principal component analysis (PCA) which introduced by Karl Pearson (1901) has received great attention as an extraction technique (Zuo *et al*., 2006). PCA extracts the p original variables into q new uncorrelated components such that $q < p$ with little missing information (Rao, 1964; Johnson & Wichern, 1992). PCA has been widely used as an exploratory multivariate data analysis and predictive models. In some studies that are swamped with many observed variables such as image processing, voice recognition, graphical information system and microarray, PCA becomes as an important tool to extract most of the variation in the original data (see Sirovich & Kirby, 1987; Turk & Pentland, 1991; Belhumeur *et al*., 1997; Wu *et al*., 2003; Liu & Chen 2006; Xu *et al*., 2009. The new extracted components from PCA allow more analyses to be done at convenience time and computational tasks.

Classification problems sometimes are burden with many observed variables. The discussed two techniques namely variables selection technique and variables extraction technique have been implemented to reduce the burden. But, often researchers perform variable reduction process and construct classification rule independently. So, this paper takes an effort to investigate the combination of these two processes simultaneously with an attempt to reduce biasness of choosing useful variables for classification purpose.

Materials and Methods

Data sets

Three famous data sets were used in the investigation with vary sizes. The *iris* data set contains four variables with 50 random records of flowers from each species of *setosa*, *versicolor* and *virginica* (Anderson, 1935). The measured variables in this small sample size include sepal length, sepal width, petal length and petal width (all in centimetres). This paper limits the discussion for two species, *versicolor* and *virginica*, as the distribution of data of these two groups are overlapping.

The second data set is considered moderate sample size is based on the Pima Indian tribe by the intramural research program of the National Institute of Diabetes and Digestive and Kidney Diseases. The investigation aims to study the differences of patients who show a sign of diabetes based on criteria of World Health Organization. All patients are females at least 21 years old of Pima Indian heritage with eight measured variables: number of times pregnant, plasma glucose concentration a 2 hours in an oral glucose tolerance test, diastolic blood pressure (mm Hg), triceps skin fold thickness (mm), two-hour serum insulin (mu U/ml), body mass index (weight in kg/(height in m)^2), diabetes pedigree function and age of patients (years) (see Hanson *et al*., 2007).

The final data set which is considerably big size of sample concerns about *crime* (see Hand *et al*., 1994) with 13 variables. The variables include crime rate, number of males of age 14-24 years old, education level, 1960 per capita expenditure, 1959 per capita expenditure, labor force, number of males per 1000 females, population size, number of non-whites, unemployment rate of urban males of age 14-24, unemployment rate of urban males of age 35-39, family income and income inequality (the number of families per 1000 earnings below 1/2 the median income). The aim of the investigation on this data is to compare the crime rate between the southern parts of the United States with other regions, which consists of 19 states for each region.

Conceptual framework and assessment

In general, the process of classification with variables reduction can be performed with the following steps: (i) choose a set of useful variables from the original number of variables, (ii) use the chosen set of variables to construct a classification rule, (iii) assess the constructed rule and (iv) use the accepted rule to classify future objects into one of the two groups. This study investigated two techniques for reducing the original variables namely variables selection and variables extraction. Variables selection was performed via stepwise selection, forward selection or backward elimination. Although there are many indicators available for choosing useful variables from the variables selection technique, this paper preferred to choose variables that contribute to minimise the assessment criterion, percentage of error due to misclassifying objects to groups. Meanwhile, variables extraction was performed using principal component analysis. In practice, common indicators for choosing the number of components are based on eigenvalue of greater than 1, total variation explained by the components and scree plot. Even so, these indicators do not promise to contribute small error rate in classification process. Therefore, this paper examines error rate for each number of components in PCA.

In order to produce an unbiased rule, this study performed the classification process in a leave-one-out fashion. First, the first object from the sample was taken out as a test object. Then, the remaining $n - 1$ objects which act as training objects were used to choose useful variables either via (i) variables selection or (ii) extraction variables of PCA. Next, the useful variables were used to construct the linear discriminant function. The omitted object was classified into either group 1 or group 2 using the constructed rule. Then, the omitted object was returned back to the sample and the second object in the sample was taken out as a test object. Then, the process of choosing useful variables, constructing linear discriminant function and assessing the constructed rule were performed. These steps were repeated until each object in the sample was taken out in turn. Finally, the error rate was computed by comparing the actual group with the predicted group, and divided by n number of objects. The value of result obtains from these strategies is known as *leave-one-out error rate*. The framework of these steps is depicted in Figure 1.

This paper investigated the performance of LDA in three forms: (i) the construction of LDA without the process of reducing the original variables, (ii) the construction of LDA with variables selection and (ii) the construction of LDA with variables extraction.

Jurnal Sains dan Matematik Vol.4 No.2 (2012) 37-48
ISSN 1985-7918

Performance of the constructed models was assessed with leave-one-out error rate and the best model was the one with the lowest error rate. At the same time, this paper investigated the pattern of eigenvalues and total variation explains of PCA in relation to the error rate.

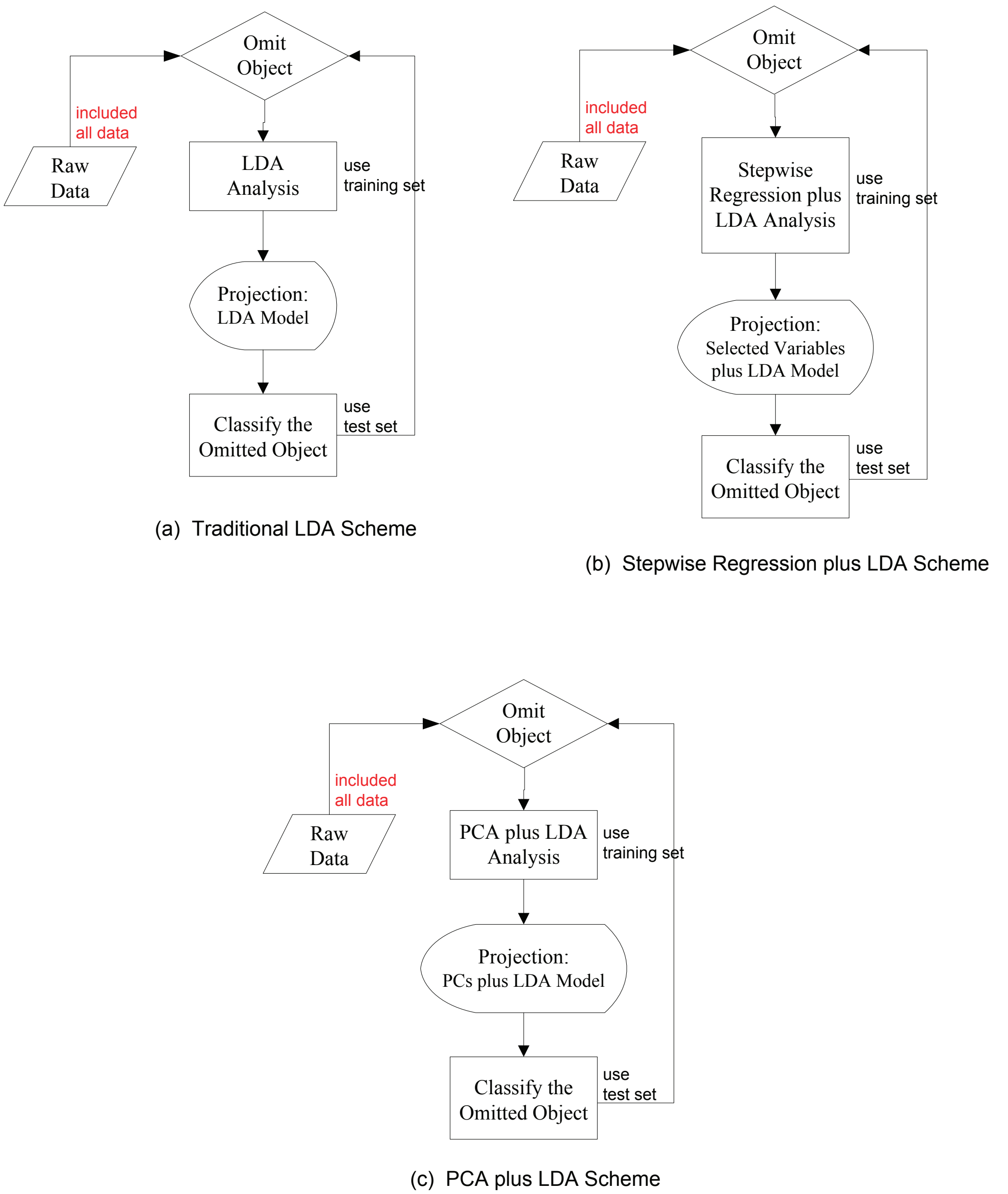

Figure 1 Three forms of investigation of LDA in leave-one-out fashion

Results and Discussion

Table 1 shows the performance of LDA of the three forms for *iris* data set. Among the LDA with variables selection, LDA with backward elimination performs the best and as good as the full model with three percent error rate. In fact, the backward elimination does not eliminate any variables at all hence it is an original *iris* data set. The stepwise selection and forward selection have greater error rate than the full LDA because they use one variable with little information to describe the classification. Meanwhile, the LDA with variables extraction scores shows perfect classification (0% error rate) except for LDA with two components. Such result indicates that LDA with principal component is the best compared to full LDA and LDA with variables selection. However, performance of LDA with four components is questionable as it performs better than the full model although the variations in the two models are the same.

Table 1 Percentage of misclassifications using Full-LDA and LDA with reduced variables for *iris* data set

Models	Selected number of variables or components	Error rate
Full LDA	4	0.03
LDA with variables selection		
1. Stepwise method	1 (Petal.Width*)	0.06
2. Forward selection	1 (Petal.Width*)	0.06
3. Backward elimination	4	0.03
LDA with variables extraction	1	0.00
	2	0.01
	3	0.00
	4	0.00

**selected variables*

Further analysis was carried on LDA with variables extraction. The scatter plot as in Figure 2 is used in order to visualise the differences between the original variables and the extracted components of PCA in discriminating the groups where "*" represents Group 1 and "0" represents Group 2. The figure demonstrates that the original variables of *iris* data are capable to show clear separation between the two groups in linear fashion but with some overlapping. However, the extracted components show great redundancy of the two groups in random behaviour. Such results occurs as the components are uncorrelated hence may influence the over performance of LDA. The investigations on the recorded eigenvalues and total variation explain of components (as tabulated in Table 2) show that the use of the first component is capable to achieve common point of selected component (often eigenvalues greater than 1 are considered useful) with almost 74% variation of the original variables explained by the component. However, it is hardly to relate the behaviour of the obtained error rate with either eigenvalue and total variation explains for this data set.

Table 2 Eigenvalue and total variance explained based on components for *iris* data set

Number of components	1	2	3	4
Eigenvalue	1.72	0.74	0.64	0.28
Percent of total variance explained	73.97	87.83	97.98	100.00

Jurnal Sains dan Matematik Vol.4 No.2 (2012) 37-48
ISSN 1985-7918

Plotting of Each Combinations of Original Variables

Plotting of Each Combinations of PCs Variables

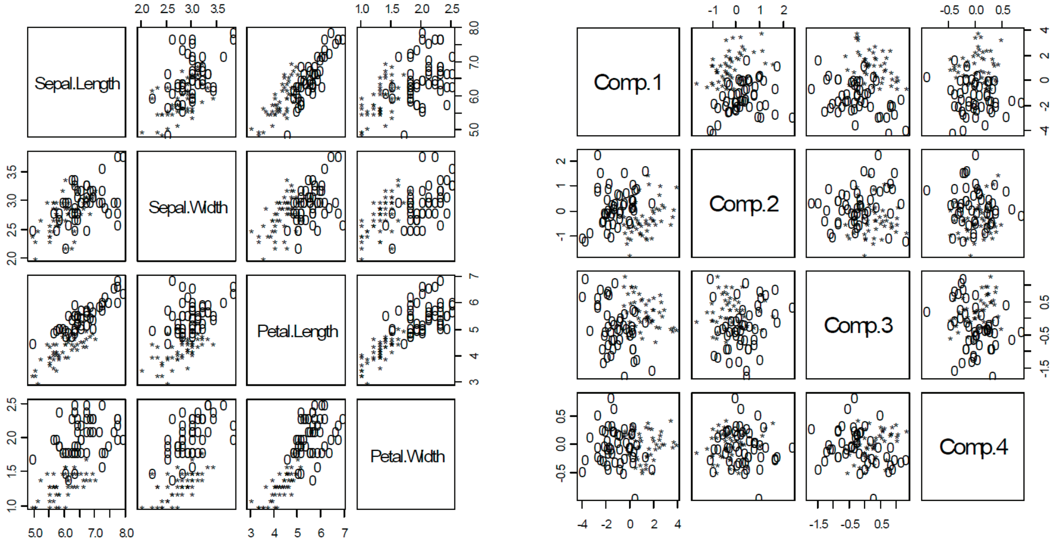

Figure 2 Scatter plot of original variables and PCA components scores of *iris* data set

The *Pima Indians diabetes* shows that the use of all variables in LDA contributes to perfect classification task. This is explained by the zero mistakes in full LDA and eight components of LDA with variables extraction (see Table 3). Among the LDA with variables selection, backward elimination gives better result than the other searching procedures. However, it is the worst compared to the full LDA and LDA with variables extraction. These results indicate that the *Pima Indians diabetes* is best explained by using the all measured variables. If a set of reduced variables is a concern on this data set, then the PCA suggests either three or four components to remain (see Table 4) based on eigenvalues greater than 1. At these points, the total variation explained is 61% and 72% respectively with error rate equal to zero percent.

Table 3 Percentage of misclassifications using Full-LDA and LDA with reduced variables for *Pima Indians Diabetes* data set

Models	Selected number of variables or components	Error rate
Full LDA	8	0.00
LDA with variables selection		
1. Stepwise method	2 (no. of times pregnant, diabetes pedigree*)	0.05
2. Forward selection	2 (no. of times pregnant, diabetes pedigree*)	0.05
3. Backward elimination	8	0.01
LDA with variables extraction	1	0.00
	2	0.01
	3	0.00
	4	0.00
	5	0.00
	6	0.00
	7	0.00
	8	0.00

**selected variables*

Table 4 Eigenvalue and total variance explained based on components for *Pima Indians Diabetes* data set

Number of components	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Eigenvalue	1.54	1.32	1.01	0.94	0.87	0.83	0.65	0.64
Percent of total variance explained	26.16	47.81	60.67	71.63	81.16	89.70	94.94	100.00

The performance on the *crime* data set is tabulated in Table 5. The full LDA makes no mistake in classifying the objects hence it tells us that all 13 variables are able to discriminate the objects correctly to their groups. The LDA with variables selection indicates that the best rule is to remain all the measured variables (see LDA with backward elimination) which it supports the result of full LDA. Finally, the LDA with extracted variables gives more choices. Nevertheless, keeping all 13 components gives slightly error rate. If a set of reduced variables is a concern, PCA suggests to remain four components (total variation explained = 86%) and it also give zero percent error rate.

Table 5 Percentage of misclassifications using Full-LDA and LDA with reduced variables for *crime* data set

Models	**Selected number of variables or components**	**Error rate**
Full LDA	13	0.00
LDA with variables selection		
1. Stepwise method	3 (crime, income inequality and education*)	0.06
2. Forward selection	3 (crime, income inequality and education*)	0.06
3. Backward elimination	13	0.00
LDA with variables extraction	1	0.01
	2	0.00
	3	0.01
	4	0.00
	5	0.00
	6	0.01
	7	0.02
	8	0.01
	9	0.00
	10	0.00
	11	0.00
	12	0.00
	13	0.01

**selected variables*

Table 6 Eigenvalue and total variance explained based on components for *crime* data set

Number of components	Eigenvalue	Percent of total variance explained
1	2.34	42.20
2	1.60	6.186
3	1.43	77.58
4	1.02	85.52

Note: Component with eigenvalue less than 1.0 is not presented

Conclusions and Recommendation

This paper is able to show that the proposed idea of performing variables reduction that contribute to minimise the error rate, either through variables selection or variables extraction, is competitive to the full LDA. Study on three data sets shows that LDA with backward elimination may replace the LDA with all measured variables. Also, if all measured variables are important, the backward elimination will not eliminate any of the variables hence it acts as a full model. The LDA with variables extraction can be considered as alternative if variables selection is not permissible. This study shows that LDAs with components which eigenvalue greater than 1 perform as good as the full model. The indicator which based on the total variation explained is too subjective as it is based on the choice of researchers. Therefore, there will be some obstacles to determine the best cutting point for this indicator.

The findings in this study cannot be used to generalise the behaviour of LDA in much wider context. But, these findings give a promising idea that the process of reducing variables needs to meet the overall aim of classification. In future investigation, focus can be given out to PCA in classification problems so that the chosen components from PCA are contributing to optimise the performance of LDA. Extensive investigations will be planned to study the LDA in much wider problems especially in dealing with variability of sample size.

References

Anderson, E. (1935). The Irises of the Gaspé Peninsula. *Bulletin of the American Iris Society*. 59: 2-5.

Belhumeur, P. N., Hespanha, J. P. and Kriegman, D. J. (1997). Eigenfaces vs. Fisherfaces: Recognition using Class Specific Linear Projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 19(7): 711-720.

Bishop, C. M. (1995). *Neural Networks for Pattern Recognition*. UK: Oxford University Press.

Chen, L. F., Liao, H. Y., Ko, M. T., Lin, J. C. and Yu, G. J. (2000). A New LDA-based Face Recognition System which can Solve the Small Size Problem. *Pattern Recognition*. 33(10): 1713-1726.

Croux, C., Filzmoser, P. and Joossens, K. (2008). Classification Efficiencies for Robust Linear Discriminant Analysis. *Statistica Sinica*. 18: 581-599.

Czogiel, I., Luebke, K., Zentgraf, M. and Weihs, C. (2007). Localized Linear Discriminant Analysis. *Advances in Data Analysis*. 133-140.

Dai, J. J., Lieu, L. and Rocke, D. (2006). Dimension Reduction for Classification with Gene Expression Microarray Data. *Statistical Applications in Genetics and Molecular Biology*. 5(1), Article 6.

Daudin, J. J. (1986). Selection of Variables in Mixed-variable Discriminant Analysis. *Biometrics*. 42(3): 473-481.

Daudin, J. J. and Bar-Hen, A. (1999). Selection in Discriminant Analysis with Continuous and Discrete Variables. *Comput. Stat. Data Anal*. 32: 161-175.

Dowling, E. M., Gestsdottir, S., Anderson, P. M., von Eye, A. and Lerner, R. M. (2003). Spirituality, Religiosity, and Thriving Among Adolescents: Identification and Confirmation of Factor Structures. *Applied Developmental Science*. 7(4): 253-260.

Fearn, T. (2008). Principal Component Discriminant Analysis: Competition on Clinical Mass Spectrometry based Proteomic Diagnosis. *Statistical Applications in Genetics and Molecular Biology*. 7(2): Article 6.

Fisher R. A. (1936). The Use of Multiple Measurements in Taxonomic Problems. *Annals of Eugenics*. 7(2): 179-188.

Friedman, J. H. (1989). Regularized Discriminant Analysis. *Journal of the American Statistical Association*. 84(405): 165-175.

Ganeshanandam, S. and Krzanowski, W. J. (1989). On Selecting Variables and Assessing Their Performance in Linear Discriminant Analysis. *Aust. J Stat*. 31: 433-447.

Guerreiro, P. M. C. (2008). Linear Discriminant Analysis Algorithms. *Master Thesis.* Universidade Técnica de Lisboa, Instituto Superior Tèchnico, Department of Electrical and Computer Engineering.

Guliashki, V. G. (2006). The Two-Group Classification Problem Challenge - Polynomial Heuristic Algorithms. *Cybernetics and Information Technologies*. 6(2): 43-53.

Hand, D. J., Daly, F., Lunn, A. D., McConway, K. and Ostrowski, E. (1994). *A Handbook of Small Data Sets*. Chapman & Hall.

Hanson, R. L., Clifton, B., David, D., Sayuko, K., Michele, K., Aniello M. I., Leslie M., Deb, B., Leslie, J. B. and William, C. K. (2007). A Search for Variants Associated With Young-Onset Type 2 Diabetes in American Indians in a 100K Genotyping Array. *Diabetes*. 56: 3045-3052.

Härdle, W. K. and Hlávka, Z. (2007). *Multivariate Statistics: Exercises and Solutions*. New York: Springer.

Héberger, K. and Andrade, J. M. (2004). Procrustes Rotation and Pair-wise Correlation: A Parametric and A Non-parametric Method for Variable Selection. *Croatica Chemica Acta*. 77(1-2): 117-125.

Huang, Z., Lui, H., McLean, D. I., Korbelik, M. and Zeng, H. (2005). Raman Spectroscopy in Combination with Background Near-infrared Autofluorescence Enhances the *In Vivo* Assessment of Malignant Tissues. *Photochemistry and Photobiology*. 81: 1219-1226.

Jeffers, J. N. R. (1967). Two Case Studies in the Application of Principal Component Analysis: Applied Statistics. *Journal of the Royal Statistical Society*. 16: 225-236.

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

Jolliffe, I. T. (2002). *Principal Component Analysis* (2nd edition). New York: Springer-Verlag. Kaiser, H. F. (1960). The Application of Electronic Computers to Factor Analysis. *Educational and Psychological Measurement*. 20: 141-151.

Kim, H. C., Kim, D. and Bang, S. Y. (2003). Extensions of LDA by PCA Mixture Model and Class-wise Features. *Journal of the Pattern Recognition Society*. 36: 1095-1105.

Krzanowski, W. J. (1987). Selection of Variables to Preserve Multivariate Data Structure using Principal Components. *Applied Statistics*. 36(1): 22-33.

Li, Q. (2006). An Integrated Framework of Feature Selection and Extraction for Appearance-based Recognition. *PhD Dissertation*. University Of Delaware: Faculty of Computer Science.

Liang, Y., Li, C., Gong, W. and Pan, Y. (2007). Uncorrelated Linear Discriminant Analysis based on Weighted Pairwise Fisher Criterion. *The Journal of the Pattern Recognition Society*. 40: 3606-3615.

Liu, J. and Chen, S. (2006). Discriminant Common Vectors versus Neighbourhood Components Analysis and Laplacianfaces: A Comparative Study in Small Sample Size Problem. *Image and Vision Computing*. 24: 249-262.

Lu, Y., Tian, Q., Sanchez, M., Neary, J., Liu, Fand Wang, Y. (2007). 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*. 14(4): 22-31.

McCabe, G. P. (1982). Principal Variables. *Technical Report*. West Lafayette: Purdue University, Department of Statistics.

McGarigal, K., Cushman, S. and Stafford, S. (2000). *Multivariate Statistics for Wildlife and Ecology Research*. New York: Springer-Verlag.

McKay, R. J. and Campbell, N. A. (1982). Variable Selection Techniques in Discriminant Analysis ii: Allocation. *British J. Math. Stat. Psychol*. 35: 30-41.

ISSN 1985-7918

Murray, G. (1977). A Cautionary Note on Selection of Variables in Discriminant Analysis. *Appl. Statist*. 26(3): 246-250.

Nenadic, Z. (2007). Information Discriminant Analysis: Feature Extraction with an Information-Theoretic Objective. *IEEE Transactions on Pattern Analysis*. 29(8): 1394-1407.

Nie, F., Xiang, S., Song, Y. and Zhang, C. (2007). Extracting the Optimal Dimensionality for Discriminant Analysis. *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2: 617-620.

Pasiouras, F., Tanna, S. and Zopounidis, C. (2005). *Application of Quantitative Techniques for the Prediction of Bank Acquisition Targets* (Vol. 5). World Scientific Publishing Company, Inc.

Pearson, K. (1901). On Lines and Planes of Closest Fit to Systems of Points in Space. *Philosophical Magazine*. 2(6): 559-572.

Qiao, Z., Zhou, L. and Huang, J. Z. (2008). Effective Linear Discriminant Analysis for High Dimensional, Low Sample Size Data. *Proceedings of the World Congress on Engineering (WCE)*. 2: 1070-1075.

Rao, C. R. (1964). The Use and Interpretation of Principal Component Analysis in Applied Research. *Sankhyā: The Indian Journal of Statistics, Series A*. 26(4): 329-358.

Rencher, A. C. (1993). The Contribution of Individual Variables to Hotelling's T^2, Wilk's Λ and R^2. *Biometrics*. 49: 479-489.

Sirovich, L. and Kirby, M. (1987). Low-dimensional Procedure for the Characterization of Human Faces. *Journal of the Optical Society America*. 4(3): 519-524.

Stevens, J. P. (2002). *Applied Multivariate Statistics for the Social Sciences* (4th edition). New Jersey: Lawrence Erlbaum Associates, Inc.

Thomaz, C. E. and Gillies, D. F. (2005). A Maximum Uncertainty LDA-based Approach for Limited Sample Size Problems - with Application to Face Recognition. In *Proceedings of the XVIII Brazilian Symposium on Computer Graphics and Image Processing (SIBGRAPI'05)*. 89-96.

Turk, M. A. and Pentland, A. P. (1991). Face Recognition using Eigenfaces. In *Proc. Computer Vision and Pattern Recognition Conference*. 586-591.

Urbakh, V. Y. (1971). Linear Discriminant Analysis: Loss of Discriminating Power when A Variate is Omitted. *Biometrics*. 27(3): 531-534.

Weiner, J. M. and Dunn, O. J. (1966). Elimination of Variate in Linear Discrimination Problems. *Biometrics*. 22(2): 268-275.

Wu, B., Abbott, T., Fishman, D., McMurray, W., Mor, G., Stone, K., Ward, D., William, K. and Zhao, H. (2003). Comparison of Statistical Methods for Classification of Ovarian Cancer using Mass Spectrometry Data. *Bioinformatics*. 19(13): 1636-1643.

Xu, P., Brock, G. N. and Parrish, R. S. (2009). Modified Linear Discriminant Analysis Approaches for Classification of High-Dimensional Microarray Data. *Computational Statistics and Data Analysis*. 53(5): 1674-1687.

Zhang, X. and Jia, Y. (2007). A Linear Discriminant Analysis Framework based on Random Subspace for Face Recognition. *Journal of the Pattern Recognition Society*. 40: 2585-2591.

Zhu, M. and Martinez, A. M. (2006). Subclass Discriminant Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 28(8): 1274-1286.

Zuo, W., Zhang, D., Yang, J. and Wang, K. (2006). BDPCA plus LDA: A Novel Fast Feature Extraction Technique for Face Recognition. *IEEE Transactions on Systems, Man, and Cybernetics - Part B: Cybernetics*. 36(4): 946-953.