Journal of the Nigerian Statistical Association, Vol. **29**, 2017, 71– 80

On the use of predictive discriminant analysis in academic prediction

A. Iduseri[a*] and J. E. Osemwenkhae[b]

[a,b]*Department of Mathematics, University of Benin, P.M.B. 1154, Benin City, Nigeria*

Selection of key predictor variables in classical statistical procedures such as predictive discriminant analysis (PDA) not only leads to identification of key predictor variables which separate the groups well, but also improves prediction or classification accuracy. Because the dependent variable in PDA is categorical, the technique lends itself to various uses in higher education predictions. But the predictive validity of the predictive discriminant function (PDF), in the context of academic prediction can best be evaluated based on the relevance of the selected key predictor variables to the PDF underlying construct when compared to what the PDF solution is intended to measure. This is not provided for by existing variable selection procedures used in PDA. Hence, final key predictor variables obtained by PDA in the context of academic prediction often lack the basic quality desired in a criterion measure such as 'relevance'. A simple approach that will validate the relevance of the key predictors to the PDF underlying construct when compared to what the PDF solution is intended to measure is proposed. The approach is based on modified splitting of historical sample and profiling of the selected key predictors. In application to four training samples from the same population, the final selected subset of key predictors' description was relevant when compared to what the PDF solution is intended to measure.

Keywords: predictive discriminant analysis; variable selection; key predictor variable; predictive validity.

1. Introduction

Since the early application of predictive discriminant analysis (PDA) in education by methodologists in Harvard University in the 1950s and 1960s, it has become a widely used analytical tool for academic prediction till date. Notable areas of successful application of PDA for academic predictions include predicting student placement (Bakari *et al.* 2016), predicting student performance or success in academic programs (Divjak and Oreski, 2009; Aluko *et al.* 2016), predicting student graduating class of degree (Erimafa *et al.* 2009), and predicting college student dropout or completion intention (Rai, 2014; Thomas, 2014). PDA is a predictive multivariate technique for classifying subjects into one of several groups. In PDA, the preprocessing steps prior to constructing a predictive discriminant function (PDF) involves the selection of useful list of predictors and selection of key predictors from the useful list of predictors. In order to select the key predictor variables, researchers often employ variable selection methods. Existing variable selection methods only search for best subset of key predictor variables, which separate the groups well in order to improve classification accuracy. A means of ascertaining whether the best subset of key predictor variables has a direct relationship with the characteristic of interest the PDF solution (or discriminant score) intends to measure is not provided for by any of the existing variable selection methods. However, in academic prediction, ascertaining whether the best subset of key predictor variables has a direct relationship with the characteristic of interest the PDF solution intends to measure (i.e., obtaining predictive validity of the PDF) can best be evaluated based on the 'relevance of the selected key variables description to the PDF underlying construct when compared to what the PDF solution is intended to measure'. The term 'construct' is used in PDA to refer to the characteristic or dimension the PDF is measuring. Therefore, the discriminant score when compared against a classification rule

*Corresponding author. Email: augustine.iduseri@uniben.edu

indicates whether an analysis unit possesses the characteristics of interest or dimension in high or low degree.

The first step of the preprocessing steps prior to constructing a PDF is the selection of a useful list of predictors. In the context of academic prediction, notable criteria that have been used are based on substantive theory and prior research (Huberty, 1974). This is an easier-said-than-done situation, of course. Limited knowledge and resources sometimes prelude the researcher from including all relevant predictors and from excluding all irrelevant predictors (Huberty and Olejnik, 2006). Also, most research finding on factors affecting academic success in various centers of learning vary from region to region and their results differ in various settings. Therefore, selection of useful list of predictors based on prior research may lead to the selection of irrelevant predictors. The second step of the preprocessing steps prior to constructing a PDF, is the selection of key predictor variables from the useful list of predictors. This second step is a critical step because of its direct impart on group assignment and predictive accuracy which are essential concerns in PDA. In addition, predictive accuracy of group membership prediction may be enhanced with fewer key predictors than the initial number of obtained useful predictors (Huberty and Olejnik, 2006). Over the past four decades, extensive research into key variables selection has been conducted. These include the stepwise methods (Draper and Smith, 1981), all possible subset method (Huberty and Olejnik, 2006), genetic search algorithms wrapped around Fisher discriminant analysis (Chiang and Pell, 2004), proposed alternatives to backward/forward/stepwise search wrapped around different discriminant functions (Pacheco *et al.* 2006), variable selection for kernel Fisher discriminant analysis (Louw and Steep, 2006), DALASS approach (Trendafilov and Jolliffe, 2007), shrinkage method (Tibshirani, 1996; Bertrand *et al.* 2009), the dimension reduction method used for deriving a low-dimensional set of features from a large set of variables (Chiang *et al.* 2001), and some recent representative methods which are based on dictionary learning (DL) for classification (Mairal *et al.* 2008; Yang *et al.* 2011). These variable selection methods search for best subset of key variables that can carry out classification task in an optimum way with the hope of reducing computational time, and not in terms of relevance of the selected key variables to the PDF underlying construct when compared to what the PDF solution is intended to measure. Whitaker (1997), earlier pointed out that regardless of the care taken in choosing predictors at the outset, some irrelevant predictor variables may still be chosen for inclusion in the final model using any of the variable selection methods. Since the discriminant weights are used both to assess the relative contribution to separation and to give a meaningful interpretation to the selected key predictors, this suggests that the interpretation of the PDF's solution will be misleading if irrelevant predictors are included in the final model. Herein lies the need for validation of selected key predictors, in terms of their relevance to PDF underlying construct when compared to what the PDF solution is intended to measure.

The concern for predictive validity in PDA has been a subject of discussion since the early 1970s (Huberty, 1974; Hand, 1997). This concern originally arises from the problem of generalizability due to instability of some PDFs solution or hit rates (Huberty, 1974). The view of predictive validity we consider herein relates directly to the question: Are the PDA solutions free from sample bias? For example, researchers are often disappointed when a PDF that predicts group membership well in the original data set becomes at best marginal when applied to fresh data drawn from the same population. The general solution is to evaluate the PDF by testing it on new data set distinct from the training sample using a validation procedure (Stone 1974), cross-validation procedure (Stone, 1974; Geisser, 1975), or bootstrap procedure (McLachlan, 1992). The use of either of these procedures will enable us to measure how predictive the model is. That is, is it useful in new situations? However, in the context of academic prediction, the view of predictive validity relates directly to providing answer to question like: What does the discriminant score tell us about the individual? This question becomes pertinent when different training samples, of the same size and number of predictors from the same population produces different subsets of key predictors with consistent high hit rates when used in constructing a PDF. From literature search, it is obvious that the main concern of most researchers for potential predictors is mainly on identifying the key predictors and their relative importance. The major indexes used include standard weight (Baggaley *et al.* 1970), variable-PDF correlation (Whellams, 1973), and group separation (Burnham and Hewitt, 1972; Iduseri and Osemwenkhae, 2015). In the context of academic prediction, these indexes lack the basic qualities desired in a criterion measure such as 'relevance' (Aggarwal, 2012). Ideally, researchers using PDA in academic prediction ought to ascertain the relevance of the key predictors to the PDF underlying construct when compared to what the PDF solution is intended to measure. This will in no doubt ensure meaningful interpretation of the discriminant score. In the context of academic prediction, this is yet to be achieved. Hence, knowing which subset of key variables are truly relevant to the PDF underlying construct when compared to what the PDF solution is intended to measure, will not only guarantee predictive validity, but will provide insight into the nature of the prediction/identification problem, which will in turn provide meaningful interpretation of the discriminant scores. In addition, uncertainty about the identified key predictor variables can be reduced, if their relevance to the PDF underlying construct when compared to what the PDF solution is intended to measure can be verified. This process is analogous to many trouble-shooting problems in process industries which relate to key variable identification for classification (Chiang and Pell, 2004). There is a dearth of literature on key variables validation in PDA. No procedure has yet been developed for validation of selected key variables obtained from existing variable selection methods.

Thus, the purpose of this research was, therefore, to demonstrate how a simple approach based on modified splitting of historical sample and profiling of the selected key predictors can be used to validate the relevance of key predictors to the PDF underlying construct when compared to what the PDF solution is intended to measure. This could form a major part of the preprocessing steps prior to constructing a PDF in the context of academic prediction. Towards this objective, the rest of the paper has been structured as follows: Section 2 describes the proposed approach of the study. Section 3 illustrates the proposed approach in the context of a major-prerequisite-identification problem, while section 4 presents the summary of findings with concluding remarks.

2. Methodology

The approach in this paper is based on modified splitting of historical sample and profiling of the selected key predictors. This approach will provide a formal means of validating the relevance of key predictors to the PDF underlying construct when compared to what the PDF solution is intended to measure. This could form a major part of the preprocessing steps prior to constructing a PDF in the context of academic prediction.

2.1 *Modified splitting of historical sample and profiling of the selected key predictors*

Suppose we have a data set (or a historical sample, D_N) that consists of N samples $\{(x_i, y_i)\}_{i=1}^{N}$, where $x_i \in \{1, 2, \ldots, P\}$ denote the corresponding predictor variable label, $y_i \in \{1, 2, ..., K\}$ denote the corresponding group label, P is the number of predictor variables, and K is the number of groups. Let $D_N = [x_1, x_2, \ldots, x_N] \in \mathbb{R}^{P\times N}$ be the historical sample data matrix. Also, let $D_n \in \mathbb{R}^{p\times n_k}$ be the historical data matrix of the kth group, where n_k is the sample size of the kth group, and $\sum_{k=1}^{K} n_k = n$. The outline of the proposed approach is described as follows:

Step 1: Draw a training sample, $D_n^{(t)}$ using simple random sampling and split $D_n^{(t)}$ into equal n-folds with $n \geq 20$ and $p \leq 6$ to obtain training set, $I^{(t)}$. The PDF will be robust to violations of normality if the smallest group has more than twenty cases and the number of independents is fewer than six (Tabachnick and Fidell, 2007). Splitting $D_n^{(t)}$ into n-folds will help detect inherent variability in the historical sample, D_N

Step 2: For each training sample, $D_n^{(t)}$ obtained in step 2, compute a PDF,Z using Stepwise option as a criterion by which the key predictors would be included in the PDF,Z defined as

$$Z = u_1 x_1^* + \ldots + u_p x_p^*$$

$$= \eta(D_n^t) \tag{2.1}$$

where u_i are the discriminant weights, x_i^* are the selected key predictor variables, and $\eta(D_n^t)$ indicates that the PDF is calibrated on a training sample. To obtain the hit rate or simply the percentage of correct classification which represents the overall correct classification within the confusion matrix, we let

$$d_j = \begin{cases} 1 \; if \hat{Z}_j = Z_j \\ 0 \; otherwise \end{cases}$$

where $\hat{Z}_j$ is the predicted response for the jth observation in the training sample, Z_j is the value for the jth observation in the training sample. The percentage of correct classification for the PDF, Z (2.1) is given as

$$P^{(a)} = \frac{1}{n} \sum_{i=1}^{n} (\hat{Z}_j = Z) \times 100 \tag{2.2}$$

where n is the total number of cases over all groups or size of the training sample

Step 3: If the linear combination of the selected key predictors for each computed PDF, Z are all the same, then the observed consistency serves as a criterion for validation of the selected subset of key variables relevance to what the PDF solution is intended to measure.

Step 4: If the linear combination of the selected key predictors for each computed PDF, Z are not the same, then a joint profiling of the selected subset of key predictors description for each computed PDF, Z is analyzed based on their intended goal in order to ascertain their relevance to what the PDF solution is intended to measure. This step serves as a criterion for choosing the subset of key predictors that is relevant to the PDF underlying construct when compared to what the PDF solution is intended to measure.

3. The data, computational results and discussion

A detailed illustration of how the proposed approach can be used to select key predictors that are relevant to the PDF underlying construct when compared to what the PDF solution is intended to measure is considered in this section. The performance of the proposed approach is investigated by analyzing a real life historical sample obtained from the department of Mathematics, Faculty of Physical Sciences, University of Benin, Nigeria. The academic records of some alumni serve as input data for the historical sample made up of two groups of students with the aim of identifying major prerequisites for success in Industrial Mathematics as a course of study. The first group (G1) is students who graduated with First Class Honours, Second Class Upper Division and Second Class Lower Division, while the second group (G2) is students who graduated with Third Class and Pass degrees. The designation procedure is tailored to the long-standing recognition that most students are of average achievement or extremely high, and a very few students have extremely low achievement levels. In Faculty of Physical Science, University of Benin, courses taken by students at first year are basically the same across individual courses of study or departments. Courses that constitute an integrated composition within any course of study are taken from second year and above. Hence, all the Industrial Mathematics second year courses from 2010/2011 to 2012/2013 academic sessions were treated as useful predictors. These three academic sessions were the sessions for which the Department throughput was significantly reduced when compared to previous sessions throughput.

For the *first step* of the proposed approach, we begin by first dividing the historical sample of size 160 (initially obtained using simple random sampling) into four folds of size 40 each to obtain four training samples with 20 students in each group. Therefore, a total of 40 students each, whose membership (in terms of graduating class of degree) was established were treated as training sample. All the four training samples contain the nine predictor variables or courses taken by students of Industrial Mathematics at 200 Level which includes: MTH212, MTH214, MTH218, MTH219, MTH230, MTH222, MTH227, MTH228, and MTH229.

The *second step* of the proposed approach involves constructing four PDFs from the four training samples $(D^t_{n1}, D^t_{n2}, D^t_{n3}, and D^t_{n4})$ obtained in step one using the DISCRIMINANT subprogram in SPSS 16 to determine the important predictors used to distinguish group membership. The METHOD=Stepwise option was chosen to specify the criteria by which the key predictor variables would be included in the analysis. At the end of the stepwise analysis, the obtained four PDFs had different linear combinations of the predictor variables out of the nine useful predictors. For the four PDFs obtained, their standardized coefficients, construct coefficients, and the overall prediction accuracy for the cross-validated group cases are shown in Tables 1 to 4.

Table 1: Canonical Variates and Hit Rate for Training Sample 1

Key Predictor Variables	**Standardized Coefficient**	**Construct Coefficient**	**LOOCV Hit-Rate**
MTH 214	0.463	0.521	87.5
MTH 222	0.604	0.649	
MTH229	0.547	0.670	

$$Z_1 = 0.463(MTH214) + 0.604(MTH222) + 0.547(MTH229)$$

Table 2: Canonical Variates and Hit Rate for Training Sample 2

Key Predictor Variables	**Standardized Coefficient**	**Construct Coefficient**	**LOOCV Hit-Rate**
MTH 227	0.894	0.878	90.0
MTH 229	0.478	0.229	

$$Z_2 = 0.894(MTH227) + 0.478(MTH229)$$

Table 3: Canonical Variates and Hit Rate for Training Sample 3

Key Predictor Variables	**Standardized Coefficient**	**Construct Coefficient**	**LOOCV Hit-Rate**
MTH 218	0.638	0.733	87.5
MTH 219	0.687	0.775	

$$Z_3 = 0.638(MTH218) + 0.687(MTH219)$$

Table 4: Canonical Variates and Hit Rate for Training Sample 4

Key Predictor Variables	**Standardized Coefficient**	**Construct Coefficient**	**LOOCV Hit-Rate**
MTH 212	-0.482	0.141	92.5
MTH 219	0.882	0.744	
MTH229	0.689	0.598	

$$Z_4 = -0.482(MTH212) + 0.882(MTH219) + 0.689(MTH229)$$

In Tables 1 to 4, Column 1 shows the best subset of key predictors that make up the linear combination of the four obtained PDFs (i.e., Z_1, Z_2, Z_3 and Z_4) at the end of the stepwise analysis, using the four training samples. Examination of Tables 1 and 4, shows that a best subset of three predictors were chosen as key predictors, while Tables 2 and 3 shows that a best subset of two predictors were chosen as key predictors. A cursory look at Tables 1 to 4 shows that both the subsets of two and three predictors are all different from each other in terms of the linear combination of the predictor variables. This suggests inherent variability among the four training samples used. Therefore, the observed inconsistency among the four subsets of key predictors obtained from the four training samples raises a question for consideration or solution. That is, which of these best subsets of key predictors are more relevant to the PDF underlying construct when compared to what the PDF solution is intended to measure?

Column two of Tables 1 to 4 shows the discriminant weights associated with each selected key predictor, while column 3 presents their respective structure coefficients. A cursory look at Tables 1 to 4 shows that the values of these estimates are all essentially equivalent. This implies that the key predictors in each of the subset of key predictors are essentially the same in terms of their order of importance (or unique contribution) in predicting group membership, as well as order of importance by total correlation with the PDF. Lets assume that the training sample one was the only available training sample. The consistency observed in the standardized coefficients and structure coefficients values among the three key predictor variables (Table 1), provide enough evidence to believe that MTH214, MTH222 and MTH229 are jointly effective, and useful major prerequisites that have relationship with performance in Industrial Mathematics as a course of study. Because this consistency was observed in Tables 2 to 4, also raises the same question for consideration or solution. That is, which of these subsets of key predictors are more relevant to the PDF underlying construct when compared to what the PDF solution is intended to measure?

A criterion for performance evaluation under PDA is based on correct classification or misclassification rate, obtained by applying the classification rules derived from the sample to a test data set or to leave-one-out cross-validation sample. Because the training sample size was small, the leave-one-out cross-validation procedure via the training sample was adopted to calculate the actual hit rate. The estimates of the actual hit rates for the four PDFs are shown in column four of Tables 1 to 4. Since equal training sample size will have a 50/50 chance, and most researchers would accept a hit ratio that is 25 % larger than that due to chance, therefore the overall prediction accuracy of the four PDFs (in Tables 1 to 4) is better than the expected percent. This obvious high degree of accuracy exhibited by the four PDFs gives enough reason to have accepted any of the PDFs if only one was available.

The *fourth step* of proposed approach involves providing answer to the questions raised above by employing a 'profiling' approach. This serves as a means of ascertaining the linear combination of key variables that are truly relevant to the PDF underlying construct when compared to what the PDF solution is intended to measure in order to provide meaningful interpretation for individual's discriminant score. In order words, we want to be able to provide answers to questions like: 'What does the discriminant score tell us about the individual'? To achieve this, we resorted to selected key predictors description using the University of Benin prospectus for Faculty of Physical Sciences. That is, the objective of each selected key predictor variables description was analyzed in terms of its relevance to Industrial Mathematics programme objective(s). Base on the prospectus, the main objective of Industrial Mathematics programme in the Department of Mathematics, University of Benin is to facilitate the education and training of Industrial Mathematicians who are able to effectively apply Mathematics and Statistics in the management and administration of industries and establishments.

After carrying out a careful study of all the key predictors' description with respect to identifying their intended goals, only MTH218 and MTH219 has a joint intended goal of providing the primary knowledge and skill needed for mathematical and statistical modeling. Base on the programme main objective, only the combination of MTH218 and MTH219 intended goals is best relevant in terms of achieving the programme objective. That is, the composition of the course outline for these two courses is fundamental with respect to achieving Industrial Mathematics programme main objective. The observed significant equivalent between the discriminant weights and structure coefficients values for PDF, Z_3 further reveal that MTH218 and MTH219 are the major prerequisite for success in Industrial Mathematics. In order words, the PDF, Z_3 identifies MTH218 and MTH219 as having a booster effect on final graduating Cumulative Grade Point Average (CGPA) for any student studying Industrial Mathematics.

Therefore, PDF, Z_3 underlying construct can be interpreted in terms of a discriminant score value. That is, a student having a discriminant score that is higher than the cutoff mark indicates a higher possession of mathematical and statistical modeling skills. In other word, having a discriminant score above the cutoff mark is a panacea for success in Industrial Mathematics as a course of study in Department of Mathematics, University of Benin. However, this may not be the case for other universities offering Industrial Mathematics as a course of study. This is because the course description for MTH218 and MTH219 may differ from one university to another. Hence, a different course or courses other than MTH218 and MTH219 may be a panacea for success in Industrial Mathematics using the profiling criterion.

4. Conclusion

This paper presents a simple approach for validating the relevance of key predictors to a PDF underlying construct when compared to what the PDF solution is intended to measure. This is aimed at providing for meaningful interpretation of the discriminant score in the context of academic prediction. The observed inconsistency among the four subsets of key predictors obtained from the four training samples as shown in Tables 1 to 4 confirms the fact that results from variable selection methods (in particular stepwise search) should only be considered descriptive for the training sample used. In the context of academic prediction, valid generalization of a PDF may be obtained only when a criterion for assessing the key predictors relevance to what the PDF solution is intended to measure had been incorporated into the preprocessing steps involved in the construction of the PDF. In addition, uncertainty about the predictive validity of the PDF can be reduced if the PDF solution or discriminant score is able to provide valid answer to question such as: what does the discriminant score tell us about the individual?

Although the findings from this research support the applicability and the merit of the proposed approach for validation of key predictor variables relevance, however, it is true that the proposed approach is limited to academic prediction. Therefore, this article believes that formal rules for validating key predictor relevance in other areas where PDA are applicable are still called for. That is another future research task.

Acknowledgement

The authors would like to thank the Editor-in-Chief and the anonymous reviewers for their valuable comments which have greatly improved the quality of the earlier manuscript.

References

Aggarwal, Y. P. (2012). Statistical methods: concepts, application and computation. Sterling, New Delhi.

Aluko, R.O., Adenuga, O.A , Kukoyi, P.O., Soyingbe, A.A and Oyedeji, J.O. (2016). Predicting the academic success of architecture students by pre-enrolment requirement: Using Machine-Learning Techniques. *Construction Economics and Building,* 16(4): 86–98.

Baggaley, A. R., Isard, E. S., and Sherwood, E. J. (1970). Discrimination of academic curricula by the Tunner studies of attitude patternscollege form. *Measurement and Evaluation in Guidance*, 16(4): 86–98.

Bakari, H.R., Isa, A.M. and Zannah, U. (2016). Application of discriminant analysis in modelling students placement in colleges of education. *Discovery*, 52(249): 1702–1712.

Bertrand, C., Ernest, F., and Hao, H. Z. (2009). Principles and theory for data mining and machine learning. Springer Series in Statistics doi:10.1007/978-0-387-98135-2, Springer Science+Business Media New York, 569–576.

Burnham, P. S., and Hewitt, B. A. (1972). Determining vocational characteristics versus predicting membership: A follow-up of male college graduates. *J. of Psych.*, 81: 73–84.

Chiang, L.H., Russell, E.L. and Braatz, R.D. (2001). Fault detection and diagnosis in industrial systems. Springer, New York.

Chiang, L. H. and Pell, R. J. (2004). Genetic algorithms combined with discriminant analysis for key variables identification. *Journal of Process Control*, 14: 143–155.

Draper, N. R. and Smith, H. (1981). *Applied regression analysis*. Wiley, New York.

Divjak, B. and Oreski, D. (2009). Prediction of academic performance using discriminant analysis. *Proceedings of the ITI 2009 31st International Conference on Information Technology Interfaces*. DOI: 10.1109/ITI.2009.5196084.

Erimafa, J. T., Iduseri, A., and Edokpa, I. W. (2009). Application of discriminant analysis to predict the class of degree for graduating students in a university system. *Int.*

J. Phys. Sci, 4(1): 016–021.

Geisser, S. (1975). The predictive sample reuse method with applications. *J. Amer. Statist. Associ*, 70: 320–328.

Hand, D. J. (1997). Constructon and assessment of classification rules, Chicester, UK, Wiley.

Huberty, C. J. (1974). Discriminant analysis, Paper presented at the annual meeting of the American Educational Research Association, Chicago, Illinois.

Huberty, C. J. and Olejnik, S. (2006). Applied manova and discriminant analysis. Wiley, Hoboken.

Iduseri, A. and Osemwenkhae, J. E. (2015). An efficient variable selection method for predictive discriminant analysis. *Ann. Data. Sci*, 2(4): 489–504 doi: 10.1007/s40745-015-0061-9.

Louw, W. and Steep, S. J. (2006). Variable selection in kernel Fisher discriminant analysis by means of recursive feature elimination. *Comput Stat Data Anali*, 51: 2043–2055.

McLachlan, G. J. (1992). Discriminant analysis and statistical pattern recognition, Wiley, New York.

Mairal, J., Bach, F., Ponce, J., Sapiro, G. and Zisserman, A. (2008). Supervised dictionary learning. *Advances in Neural Information Processing Systems (NIPS)*, 21. See http://papers.nips.cc/paper/3448-supervised-dictionary-learning.pdf

Pacheco, J., Casado, S., Nunez, L. and Gomez, O. (2006). Analysis of new variable selection methods for discriminant analysis. *Computational Statistics and Data Analysis*, 51: 1463–1478.

Rai, S. (2014). Students dropout risk assessment in undergraduate course at residential university. https://arxiv.org/pdf/1405.3727.

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *J Roy Stat Soc Ser Bi*, 36: 111–147.

Tabachnick, B. G. and Fidell, L. S. (2007). Using multivariate statistics, 5th ed. Pearson Education, Inc. USA, 382.

Thomas, D. (2014). Predicting student college completion intention: A discriminant analysis. *Asean Journal of Management and Innovationi*, journal.stamford.edu/index.php/ajmi/article/download/190/58.

Tibshirani, R. (1996). Regression shrinkage and selection via the LASSO. *J R Stat Soc, Ser B*, 58: 267–288.

Trendafilov, N. T. and Jolliffe, I. T. (2007). DALASS: Variable selection in discriminant analysis Via the LASSO. *Comput Stat Data Anal*, 51: 3718–3736.

Whellams, F. S. (1973). Musical abilities and sex differences in the analysis of aural-musical capacities. *Journal of Research in Music Education*, 21: 30–39.

Whitaker, J. S. (1997). Use of stepwise methodology in discriminant analysis. A paper presented at the annual meeting of the Southwest Educational Research Assoc, Austin, January, 1997.

Yang M., Zhang L., Feng X., Zhang D. (2011). Fisher discrimination dictionary learning for sparse representation. In: Proceedings of the IEEE international conference on computer vision (ICCV), 543550.