A NONPARAMETRIC DISCRIMINANT VARIABLE-ELIMINATION ALGORITHM

FOR CLASSIFICATION TO TWO POPULATIONS

S. PADMANABAN[1] & MARTIN L. WILLIAM[2]

[1]NIRRH Field Unit, Indian Council of Medical Research, KMC Hospital, Chennai, India

[2]Department of Statistics, Loyola College, Chennai, India

ABSTRACT

This paper provides a nonparametric discriminant variable-elimination algorithm to discriminate two multivariate populations and an associated optimal decision rule for membership-prediction. This is an alternative to the 'forward-stepwise' approach recently proposed for the same classification problem by Padmanaban and William (2016). As in the referred work, the present work relaxes the 'equal variance-covariance matrices' condition traditionally imposed and develops a discrimination-classification procedure by excluding variables that do not contribute to the 'discrimination', one-by-one in a backward-stepwise manner. The exclusion of variables in the discriminant is determined on the basis of least 'discriminating ability' as reflected in 'difference' between the distributions of the discriminant in the two populations. A decision-rule for classification or membership-prediction with a view to maximize correct predictions, balancing between 'sensitivity' and 'specificity', is provided. The proposed alogorithm is applied to develop an optimal discriminant for predicting preterm labour among expecting mothers in the city of Chennai, India and its performance is compared with logistic regression and also with the forward-stepwise discriminant algorithm of the same authors.

KEYWORDS: Classification, Discriminant, Variable-Elimination, Kolmogorov-Smirnov Statistic

1. INTRODUCTION

The problem of discriminating the objects belonging to two populations and the related issue of effectively classifying members to the two has existed for many decades now. It is a known fact that, applying the technique under a non-parametric setting needs the variance-covariance matrices of the two populations to be equal, even though this condition is not required for multivariate normal populations. Variables are included in the discriminant based on a comparison of the means in the two populations. Also, classification of a member to one of the two populations is based on the distances of the member's discriminant value from the means of the discriminant in the two populations.

The aim of the present work is to develop an algorithm for obtaining an ideal discriminant, starting with a 'large' set of candidate variables and pruning the variables one-by-one to obtain a parsimonious model having 'good' ability to 'classify' objects to the two populations. This is a modification to the 'variable selection algorithm' for constructing the ideal discriminant, introduced by Padmanaban and William (2016). Like the 'variable selection algorithm' referred, the 'variable elimination algorithm' being proposed in this paper also has a wider scope of application than the traditional discriminant analysis.

In practical situations where observations of multiple variables are involved, joint normality or equality of variance-covariance matrices is not assured. For multivariate normal datasets, the equality of the variance-covariance can

be tested and if affirmed, one can apply the traditional linear discriminant function; if equality is negated, quadratic discriminant function can be used. If the data are not from multivariate normal populations, the distribution-free Fisher's linear discriminant function can be used, but there is no easy procedure available for testing the equality of variance-covariance matrices. Many practitioners 'assume' equality and proceed. This gap between theory and practice has remained unfilled for long.

However, recently Padmanaban and William (2016) proposed a discrimination-classification procedure in a distribution-free context without imposing the condition of equal variance-covariance matrices. Specifically, Padmanaban and William (2016) provided a model-building algorithm for selecting variables that are to be fed into the discriminant function. That algorithm may be termed 'forward-model-building' process in line with the term usually employed in building predictive models. In this paper, we propose a 'backward' process of obtaining the ideal discriminant by starting with a number of candidate variables and 'eliminating' the variables with least predictive capacity one-by-one until we are left out with only those variables that are well-capable of discriminating the two populations. The theoretical framework for this work has been developed by Padmanaban and William (2016). Just as the 'forward' process, the proposed 'backward' process can be applied without the conditions that restrict the traditional approaches.

Interesting developments to the classical theory of discriminant analysis have been made by a number of authors in the past many decades. Different approaches to develop discriminant models focus on identifying the important variables for discriminating the populations. Some of the early contributions in this area include those of Chang (1983) who proposed using principal components for separating a mixture of two multivariate normal distributions and that of Bensmail and Celeux (1996) who considered Gaussian discriminant analysis through eigen-value decomposition. A stepwise algorithm using 'Bayesian Information Criterion' was developed by Murphy *et al.* (2010) following Raftery and Dean (2006) who proposed a similar approach for model-based clustering. The above approaches are parametric and are restricted in their scope of applications.

Other contributions in this area extended discriminant analysis to non-parametric settings in different directions. Nonparametric discriminant analysis with nonlinear classifiers was proposed by Hastie *et al.* (1994) to handle situations with a large number of input variables. Nonlinear discriminant analysis via kernel approach, theoretically close to support vector machines, was given by Baudat and Anouar (2000).Nonparametric discriminant analysis with adaptation to nearest-neighbour classification was developed by Bressan and Vitria (2003).Chiang and Pell (2004) proposed combining genetic algorithms with discriminant analysis for identifying key variables. In these works, a matter of major concern was to identify the variables that would be effective in discriminating the populations under consideration.

This paper takes a different approach from that of the above-mentioned works present in the literature on two-population discriminant analysis while adhering to the basic spirit and mathematical objective of classical discriminant analysis. We attempt to provide a backward process as an alternative method to the forward process developed by Padmanaban and William (2016) for building an effective discriminant model. A variable-elimination algorithm is proposed to obtain the discriminant model by removing variables that least contribute to the discrimination-ability one-by-one in a backward-stepwise manner. For a discussion on the 'model performance' measure to evaluate the classification ability of the discriminant model and the decision rule for identifying the optimal cut-off point for classification, reference is made to the paper of Padmanaban and William (2016).

Hence, the Objectives of the Present Work Are:

- To present a variable-elimination algorithm for discriminating two populations and an easy-to-apply procedure for classification of objects.

- To apply the algorithm to a biomedical phenomenon and compare its classification-performance with that of logistic regression and the forward-stepwise discriminant algorithm.

This paper is organized as follows: Following this introductory section, a review of the basic theoretical framework recently introduced in Padmanaban and William (2016) is given in Section 2. The new variable-elimination algorithm to build an efficient discriminant model is outlined in Section 3. As an application of this methodology, the prediction of 'pre-term labour' in pregnant women is considered in Section4. This is based on a sample of 200 women who delivered babies in the Department of Obstetrics and Gynaecology, Government Kilpauk Medical College and Hospital, Chennai, India, during the five-month period of 7th May, 2015 to 7th October, 2015.

2. A REVIEW OF THE RECENTLY INTRODUCED PROCEDURE

Consider two populations π_1 and π_2 whose relative sizes are given by the proportions p_1 and p_2.The objects in the two classes are to be discriminated using multidimensional data on a random vector, say, $X = (X_1, X_2,..., X_p)^T$. When there is a '*significant'* difference between the distributions, classification or membership-prediction of objects becomes pertinent and the 'correctness' or 'incorrectness' of classifications turns out to be a matter of concern.

Denote the mean-vectors of X in the two populations as $\mu_1 = E_1(X)$ and $\mu_2 = E_2(X)$ and the variance-covariance matrices of X in the two populations be Σ_1 and Σ_2. From Padmanaban and William (2016), we have the following theoretical results:

- For a random vector X and another random object W, the relationship between the unconditional and conditional mean vectors and variance-covariance matrices is given by

$$E(X) = E_W[E_{X|W}(X)] \text{ and } V(X) = E_W\{V_{X|W}(X)\} + V_W\{E_{X|W}(X)\} \tag{2.1}$$

- The overall variance-covariance matrix of the combined population is given by

$$\Sigma = p_1\Sigma_1 + p_2\Sigma_2 + p_1(1-p_1)\,\mu_1.\,\mu_1^T + p_2(1-p_2)\,\mu_2.\,\mu_2^T - p_1\,p_2(\mu_1\,\mu_2^T + \mu_2\,\mu_1^T) \tag{2.2}$$

In Discriminant Analysis, the multivariate observations (X) are transformed to univariate observations (Y) by considering linear combinations of the X_i's. For any linear combination $Y = \ell^T X$, where ℓ is a p x 1 vector of constants, the means of Y in the two populations are $\mu_{1Y} = \ell^T\mu_1$ and $\mu_{2Y} = \ell^T\mu_2$ and in the combined population it is given by $\mu_Y = p_1\ell^T\mu_1 + p_2\ell^T\mu_2$. And, the variance of Y in the combined population is given by $V(Y) = \ell^T \Sigma \ell$.

The linear combination which maximizes the (squared) distance between μ_{1Y} and μ_{2Y} relative to the variability of Y in the combined population helps in discriminating the two groups in the most 'optimal' manner. The 'distance-maximizing' linear combination of the X_i's is the 'optimum discriminant function' based on X. We call it 'X-based optimal discriminant' and is given by

$$Y = (\mu_1 - \mu_2)^T \Sigma^{-1} X \tag{2.3}$$

Suppose $X_{(s)}$ be a subset of the variables used to build the optimal discriminant. Denote the mean vectors of $X_{(s)}$ in

the two populations as $\mu_{1(s)}$ and $\mu_{2(s)}$ and the 'overall' variance-covariance matrix of $X_{(s)}$ as $\Sigma_{(s)}$. The $X_{(s)}$-based optimal discriminant is

$$Y_{(s)} = (\mu_{1(s)} - \mu_{2(s)})^T \Sigma_{(s)}^{-1} X_{(s)} \qquad (2.4)$$

Typically, these parameters are replaced by the sample estimates in practice. Computing the variable $Y_{(s)}$ for all members in both the samples, the performance of the $X_{(s)}$-based optimal discriminant is measured by the two sample Kolmogorov-Smirnov Statistic based on the $Y_{(s)}$ measurements. Denoting the (empirical) cumulative distribution functions of $Y_{(s)}$ for the two populations as $F_{1(s)}(\cdot)$ and $F_{2(s)}(\cdot)$, the performance measure is given by

$$KS_{(s)} = \max_{y} \left(| F_{1(s)}(y) - F_{2(s)}(y)| \right) \qquad (2.5)$$

Given two subvectors $X_{(s1)}$ and $X_{(s2)}$, the optimal $X_{(s1)}$-based discriminant is said to be 'more efficient' than the optimal $X_{(s2)}$-based discriminant if $KS_{(s1)} > KS_{(s2)}$. If there exists a random subvector $X_{(s*)}$ for which $KS_{(s*)} > KS_{(s)}$ for every other random subvector $X_{(s)}$, then the corresponding optimal discriminant $Y_{(s*)}$ is the 'most efficient' discriminant.

However, obtaining the 'most efficient' discriminant is computationally prohibitive in the presence of a very large number of predictor variables (i.e.) in case of very high dimension of the underlying random vector X. This is true of every model-building situation involving a large number of predictor variables and different algorithms are therefore suggested to 'build' improved models sequentially instead of considering 'all possible' models or identifying the 'most efficient'.

With this view, Padmanaban and William (2016) introduced a 'forward model-building' algorithm to build a 'sequence' of discriminant models, starting with a single variable and 'select' variables one-by-one evaluating their ability to 'add' to the discriminatory ability of the model. In the same spirit, the next section presents a model building algorithm to build a 'sequence' of discriminant models starting with the full set of observables, and 'eliminating' one-by-one those variables that do not contribute substantially to the discriminatory ability of the discriminant, ultimately leading to an efficient discriminant model.

3. THE PROPOSED VARIABLE-ELIMINATION ALGORITHM

The proposed algorithm evaluates each candidate 'input' variable in a sequential manner towards constructing the optimal discriminant function by 'pruning' the variables that do not contribute adequately to the discriminatory ability of the model. This 'backward' process of variable-elimination is a 'reversal' of the 'forward' process of variable-selection introduced by Padmanaban and William (2016).Variable-selection for discriminating between two populations has been addressed by Habbema and Hermans (1977) who considered selection of variables for Gaussian discriminant analysis on the basis of F-Statistics and error rates and by Pfeiffer (1985) who considered smoothing factors of kernel functions for nonparametric discriminant analysis with different criteria like distances, error rates and density-ratios.

The present work proposes a different process of 'eliminating' variables in a backward-stepwise manner. The algorithm starts by considering 'all' the candidate variables initially and proceeds by removing one input variable at a time on the basis of 'least' differentiation between the distributions of the discriminant scores in the two populations, as measured by the two sample Kolmogrov-Smirnov (KS) statistic used for comparison of two distributions. The exact backward process is described below.

Let $X_1, X_2,.., X_p$ be the candidate input variables and denote the vector $(X_1, X_2,.., X_p)$ as **X**.

Step 0: With all the candidate variables **X**, the **X**-based discriminant and the corresponding scores are obtained for each individual record in the data. Let the value of the Kolmogorov-Smirnov Statistic for this 'full' model be denoted $KS_{(0)}$. The significance of this statistic is evaluated at a desired level of significance.

Step 1: Removing one variable at a time, 'p' discriminants $Y_{(\sim 1)}$, $Y_{(\sim 2)}$,..., $Y_{(\sim p)}$, (where $Y_{(\sim i)}$ is the discriminant based on all variables except X_i), and their corresponding scores are obtained for each record in the data. Let the Kolmogorov-Smirnov Statistic for $Y_{(\sim i)}$ be denoted as $KS_{(\sim i)}$.

If

$$KS_{(\sim i)} > KS_{(\sim j)} \text{ for every } j \neq i \text{ and } KS_{(\sim i)} \geq KS_{(0)}$$

Then among the individual variables considered for elimination one a one-at-time, X_i is the least effective discriminator between the two populations. That is, leaving out X_i leads to a higher discrimination ability (or) at least does not reduce the ability of the model compared to the stage where X_i is part of the discriminant function. So, at the end of Step 1, X_i gets eliminated. In contrast, if

$$KS_{(\sim i)} > KS_{(\sim j)} \text{ for every } j \neq i \text{ but } KS_{(\sim i)} < KS_{(0)}$$

Then X_i does not leave the model, nor any of the remaining X_j's leave as its exit leads to reduced discriminatory ability and the model building stops with all the 'p' candidate variables present.

Step 2: If X_i were eliminated in Step 1, we remove one additional variable at a time and obtain (p–1) discriminants, in which the removed variables are (X_1, X_i),...., (X_{i-1}, X_i), (X_{i+1}, X_i),...., (X_p, X_i). Denote the discriminants as $Y_{(\sim 1 \sim, i)}$, $Y_{(\sim 2, \sim i)}$,..., $Y_{(\sim i-1, \sim i)}$, $Y_{(\sim i+1, \sim i)}$,... $Y_{(\sim p, \sim i)}$ and the corresponding Kolmogorov-Smirnov statistics as $KS_{(\sim 1 \sim, i)}$, $KS_{(\sim 2, \sim i)}$,..., $KS_{(\sim i-1, \sim i)}$, $KS_{(\sim i+1, \sim i)}$,..., $KS_{(\sim p, \sim i)}$. If for some 'm',

$$KS_{(\sim m \sim, i)} > KS_{(\sim j \sim, i)} \text{ for every } j \neq m, \text{ and } KS_{(\sim m, \sim i)} \geq KS_{(\sim i)},$$

Then X_m leaves the model in Step 2. In contrast, if

$$KS_{(\sim m, \sim i)} > KS_{(\sim j, \sim i)} \text{ for every } j \neq m, \text{ but } KS_{(\sim m, \sim i)} < KS_{(\sim i)},$$

Then X_m does not leave the model, nor any of the remaining X_j's leave, as its exit leads to reduced discriminatory ability and the model building stops with (p – 1) input variables present. Clearly no other variable can leave further.

At every subsequent step that is considered, one more additional variable leaves provided the maximum KS value at that step exceeds or equals the maximum KS value of the previous step. If it is less than the previous maximum, the process stops. When the process stops at the (k+1)[th] step, the optimal discriminant function is the one obtained in the kth step with the maximum KS value, leading to significant and maximum discrimination between the two populations. We denote the final subset of variables reached in this process as $X_{(S*)}$ and the 'final' efficient discriminant as $Y_{(s*)}$.

The classification or prediction rule, the 'explanation' to the KS statistic and also the suggestion to use the 'Reliability Function' for computing the KS Statistic are provided in the paper of Padmanaban and William (2016) wherein the proposal for forward model-building process was given.

4. APPLICATION OF THE ALGORITHM TO PREDICT PRETERM LABOUR

Preterm Labour

The recent lifestyle changes, nature of jobs and food habits have resulted in many health-related disorders among youngsters. In the case of married women, this leads to pregnancy-related issues and complications at the time of delivery. The birth of a baby ahead of the normal delivery time is a serious issue affecting the growth milestones of the child and possibly creating other life-long physical disabilities. For definitions and statistical information related to the phenomenon of preterm labour, details of the potential associated factors – lipid profiles, study design and sample size considered for the study, we refer to Padmanaban and William (2016).

Objective: This study aims to relate the above factors to preterm labour through the 'backward' discriminant model building algorithm developed in this paper. We wish to identify the significant factors that are associated to the risk of preterm labour.

A sample of the data on the six variables listed under 'Potential factors' along with the birth outcome (Term labour = 1, sPTB = 2) is given below:

Table 1

Record #	X_1 (BMI)	X_2 (AFI)	X_3 (TC)	X_4 (TGL)	X_5 (HDL)	X_6 (LDL)	Outcome
1	12.6	14.2	274	168	76	114	1
2	19.3	9.5	276	288	89	186	2
3	12.6	14.2	235	168	76	114	1
4	12.6	14.2	274	168	76	114	1
5	19.7	9.6	310	298	89	186	2

We apply the variable-elimination algorithm developed in this paper and get the following results.

Step 0: The KS statistic for the 'full' model with all the variables considered is found to be

$KS_{(0)} = 0.980$. The KS value of **0.980** is found to be statistically significant.

Step 1: The KS statistics for models leaving out one variable at a time are

$KS_{(\sim X1)} = \textbf{0.980}$, $KS_{(\sim X2)} = 0.960$, $KS_{(\sim X3)} = 0.960$, $KS_{(\sim X4)} = 0.920$, $KS_{(\sim X5)} = 0.960$, $KS_{(\sim X6)} = \textbf{0.980}$

X_1 and X_6 are found to be the least effective discriminators. As there is a tie on which one to leave out, we apply domain knowledge and decide to remove X_1 at the end of Step 1.

Step 2: In this step we get

$KS_{(\sim X1, \sim X2)} = 0.960$, $KS_{(\sim X1, \sim X3)} = 0.960$, $KS_{(\sim X1, \sim X4)} = 0.920$, $KS_{(\sim X1, \sim X5)} = 0.960$, $KS_{(\sim X1, \sim X6)} = \textbf{0.980}$

X_6 leaves the model in the second step. We note that X_6 was a competitor to X_1 for leaving at the end of Step 1.

Step 3: In this step we get

$KS_{(\sim X1, \sim X6, \sim X2)} = \textbf{0.980}$, $KS_{(\sim X1, \sim X6, \sim X3)} = 0.950$, $KS_{((\sim X1, \sim X6, \sim X4)} = 0.940$, $KS_{(\sim X1, \sim X6, \sim X5)} = \textbf{0.980}$

X_2 and X_5 are found least effective and as tie-breaker, we remove X_5 (HDL Cholestrol) as we have already found LDL Cholestrol (X_6) to be an ineffective discriminator.

Step 4: We get $KS_{(\sim X1, \sim X6, \sim X5, \sim X2)} = 0.960$, $KS_{(\sim X1, \sim X6, \sim X5, \sim X3)} = 0.930$, $KS_{((\sim X1, \sim X6, \sim X5, \sim X4)} = 0.940$

As all the KS statistics are less than the previous step KS maximum value, the variable-elimination algorithm stops with three variables being eliminated in the order of X_1, X_6 and X_5.This concludes that X_2, X_3, X_4 are the effective discriminators between pre-term labour population and the term labour population. The model has a KS value of **0.980**.

We have tried an alternative from Step 3 by removing X_2 instead of X_5. In this case, the Step 4 KS values are found to be

$KS_{(\sim X1,\sim X6,\sim X2,\sim X3)} = 0.910$, $KS_{(\sim X1,\sim X6,\sim X2,\sim X4)} = 0.880$, $KS_{((\sim X1,\sim X6,\sim X2,\sim X5)} = 0.960$ leading to a model with X_3, X_4 and X_5 with same KS value **0.980**.

However, we have found from domain experts that X_2 (AFI) is more associated with labour complications than X_5 (HDL Cholestrol) and moreover, our forward algorithm also selected X_2 but not X_5.

Thus, the 'Efficient Discriminant' obtained at the end of Step 3 of our algorithm is:

$$Y = 0.1853*AFI - 0.0343*TC - 0.0228*TGL \tag{4.1}$$

The estimated means of Y in the two populations are found to be

$\mu_{1Y} = -11.3522$, $\mu_{2Y} = -14.6097$

and the 'efficient cut-point' is $y_0 = -12.578$

Here, '1' denotes 'term labour group' and '2' denotes 'sPTB group'.

<u>Membership-Prediction Rule</u>: If '*y*' is the value of the 'Efficient Discriminant' Y of (4.1) for an individual, then the prediction rule is as follows:

Classify individual to: $\begin{cases} \textit{Term Labour Group} & \textit{if } y > -12.578 \\ \Pr \textit{eTerm Labour Group if } y \leq -12.578 \end{cases}$

We observe form (4.1) that, increased AFI, lower TC and lower TGL indicate the likelihood of normal term labour for a woman. Accordingly, we find that lower AFI, higher TC and higher TGL increase the risk for preterm labour for a woman. This result is same as the one we got in the 'forward' model building approach discussed by Padmanaban and William (2016).

Comparison with Logistic Regression Model:

Denoting 'preterm labour outcome' as the outcome of interest, when we built a logistic regression model using the stepwise method of model building, we get the following results.

Step 1: TC entered with very high significance and with a positive coefficient.

Step 2: TGL entered with very high significance and with a positive coefficient.

The model building process stopped in two steps and resulted in the following logit equation:

$$\log\left(\frac{p}{1-p}\right) = -59.154 + 0.119*TC + 0.108*TGL \tag{4.2}$$

where 'p' is the probability of preterm labour. From equation (4.2), we find that higher TC and higher TGL leads

to higher risk of preterm labour, which is in agreement with the conclusion of the discriminant model (4.1). However, the KS for the logit model is found to be 0.950 which is less than the KS obtained for the 'Efficient Discriminant Model'. Thus, the new method performs better than binary logistic regression method in predicting preterm labour among pregnant women.

We note that, while logistic regression identifies two factors TC and TGL, our model identifies one more important factor AFI. In this context we refer to the article of Weissmann-Brenner *et al.* (2009) in which it was stated that the mean AFI differs significantly between PPROM (PTB) cases and the normal cases. Our discriminant model confirms that AFI is an important discriminator between preterm and term labour cases and that lower AFI points to the risk of preterm labour. The finding here supports the discovery of the medical research team of Brenner *et al.*

Even if it may be argued that the new 'discriminant' approach needs to be applied to many more situations wherein logistic regression is applied to decide its 'higher' effectiveness in prediction of binary outcomes, the present work points out that this approach is a promising alternative to logistic regression model. It is possible that, in good many applications, this approach is also capable of performing better than logistic regression approach and could also discover some important discriminators that are not identified by the latter.

REFERENCES

1. Baudat, G. and Anouar, F. (2000). Generalized Discriminant Analysis using a Kernel Approach.   Neural Computation, 12 (10), 2385-2404

2. Bensmail, H. and Celeux, G. (1996). Regularized Gaussian discriminant analysis through eigen value decomposition. J. Amer. Statist. Assoc. 91, 1743-1748

3. Bressan, M. and Vitria, J. (2003). Nonparametric Discriminant Analysis and Nearest Neighbor   Classification. Pattern Recognition Letters. 24, 2743-2749

4. Catov, J.M., Bodnar, L.M., Ness, R.B., Barron, S.J. and Roberts, J.M. (2007). Inflammation and   Dyslipidemia related risk of Spontaneous Preterm Birth. Am. J. Epidemiol. 166, 1312-1319

5. Chang, W.-C. (1983). On using Principal Components before Separating a Mixture of two   Multivariate Normal Distributions. J. Roy. Statist. Soc. Ser C. 32, 267-275

6. Chiang, L.H. and Pell, R.J. (2004). Genetic algorithms combined with discriminant analysis for key variable identification. J. Process Control, 14, 143-155

7. Habbema, J.D.F. and Hermans, J. (1977). Selection of variables in discriminant analysis by F-statistic and error rate. Technometrics. 19, 487-493

8. Hastie, T., Tibshirani, R. and Buja, A. (1994). Flexible Discriminant Analysis by Optimal Scoring.Amer.   Statist. Assoc. 89, 1255-1270

9. Mudd, L.M., Holzman, C.B., Catov, J.M., Senagore, P.K. and Evans, R.W. (2012). Maternal lipids at midpregnancy and risk of preterm delivery. ActaObstet. Gynecol. Scand. 91, 726-735

10. Murphy, T.B., Dean, N. and Raftery, A.E. (2010). Variable Selection and updating in Model- Based Discriminant Analysisfor High Dimensional Data with Food AuthenticityApplications. The Annals of Applied Statistics, Vol.4,

No.1, 396-421

11. Padmanaban, S and Martin L. William (2016). A nonparametric discriminant variable-selection algorithm for classification to two populations. International Journal of Applied Mathematics and Statistical Sciences, Vol. 5, Issue 2, 87-98

12. Pfeiffer, K.P. (1985). Stepwise Variable Selection and Maximum Likelihood Estimation of Smoothing Factors of Kernel Functions for Nonparametric Dsicriminant Functions evaluated by Different Criteria. J. Biomed. Informatics. 18, 46-61

13. Raftery, A.E. and Dean, N. (2006). Variable Selection for Model-Based Clustering. J. Amer. Statist. Assoc. 101, 168-178

14. Weissmann-Brenner, A., O'Reilly-Green, C. and Ferber, A. (2009). Values of amniotic fluid index in cases of preterm premature rupture of membranes. J. Perinatal Medicine. 37, 232-235.