Research Journal of Mathematical and Statistical Sciences ________________________________ISSN 2320–6047
Vol. **3(12),** 1-4, December **(2015)** *Res. J. Mathematical and Statistical Sci.*

A Generalized Multi-Group Discriminant Procedure for Classification: A Comparative Study

Anyanwu Paul E.[1], Ekezie Dan D.[2], Onyeagu Sidney I.[3] and Mbaeyi, George C.[4]
[1]Department of Mathematics and Statistics, Federal Polytechnic, P. M. B. 55, Bida, Niger State, NIGERIA
[2]Department of Statistics, Imo State University, P.M.B. 2000, Owerri, Imo State, NIGERIA
[3]Department of Statistics, Nnamdi Azikiwe University, Awka, Anambra State, NIGERIA
[4]Department of Mathematics/Computer Science/Statistics/Informatics, Federal University, Ndufu-Alike Ikwo, P.M.B. 1010, Abakaliki, Ebonyi State, NIGERIA

Available online at: www.isca.in, www.isca.me
Received 17th October 2015, revised 27th October 2015, accepted 26th November 2015

Abstract

It has been earlier demonstrated that an alternative and dependable tool for discriminant analysis with many groups is obtainable by considering all possible pairs of group of the available multi-groups. An assessment of the performance of this procedure is therefore made by comparing its accuracy rate alongside other conventional and common procedures of classification, with a distributional data set under various sample sizes. The All Possible Pairs (APP) classification procedure performed better than its conventional counterparts under the various scenario. Thus, the All Possible Pairs procedure could still remain a better option in situations of any multivariate data structure with many groups.

Keywords: Discriminant function, classification, combination, accuracy rate, cross validation.

Introduction

Many areas of applications involve collection of multivariate data and the linear discriminant function has often been used in classifying observations that are multivariate in nature as coming from one of two populations. Multivariate analysis techniques, besides discriminant analysis also includes principal component analysis[1] and canonical correlation[2]. When more than one observation is available[3] provides a procedure for obtaining the best linear function for discriminating the population under study. A close look at the allocation rule associated with the Fisher's Linear Discriminant (FLD) procedure provides reasons to agree that the FLD procedure is important, easy and simple when applied to just two groups. Fisher R.A.[4] also pointed out that although the proposed technique have been applied in different field especially for the two group case, considerable work in theory remains to be done for the more than two group case. In classification generally, solutions to Multi-class (group) problems have been proposed by many researchers. Examples includes Linear Discriminant Analysis[5], Two-class linear discriminant analysis[3], Nearest Neighbour classifier[6], Aggregate Classifiers[7,8], Boosting[9], Multiple Group Logistic Model[10], the Super Vector Machines[11], One-versus-the-rest method[12], Pairwise Comparison[13,14], Direct Graph Traversal[15], Error Correcting Output Coding[16]. None of these methods seems entirely satisfactory.

According to[17], Fisher's approach to discriminant problem is parametric and relies on assumptions such as multivariate normality for optimality and therefore, may be less effective on more realistic classes of problems. The multiple group problems, however, has very rarely been addressed and most of the methods proposed for two groups do not generalize and the performance of the methods that can be used with several group is not generally reliable[18]. It is fair to say that there is probably no multi-class approach generally outperforms the others[19]. For practical problems, the choice of approach will depend on constraints on hand such as required accuracy, the time available for development and training, the nature of the classification problem, distributional assumptions of available data and the data structure. The simple, efficient and accurate discriminant analysis provides a good choice for practical multi-group classification problems. As multi-group classification problem is not confined to specific studies but it is rather faced by overall studies, verifying its general applicability is important.

The purpose of this paper, therefore, is to make a follow up to the previously described procedure[19] by comparing its performance with some conventional procedures commonly used in classification.

Methodology

Let us briefly recap the method for multi-group classification (hereafter referred to as All Possible Pairs (APP) classification procedure) as described previously[19]. Assuming we have a set of observation with attributes represented by variables $x_1, x_2, \dots, x_k$ coming from m-population (groups). Group I has n_1 observations, group II has n_2 observations and so on up to group m having n_m observations where $n_1 + n_2 + \cdots + n_m = N$. Our interest is to classify a future (or new) observation whose origin is unknown with same attributes as $x_1, x_2, \dots, x_k$ to the correct group. We desire to do this with so much caution so as to minimize the cost of misclassification. Fisher's procedure obtains a set of m-1 linear functions which represents the functional relationship between the discriminating attributes (or variables). The possible ways of arranging *n* objects in *r* ways (considering the order and without repetition) is given by

$$\binom{n}{r} = \frac{n!}{r!(n-r)!} \quad (1)$$

Applying (1) into our set up, with *n* replaced by *m*, the number of groups and *r* the possible pairs of group combination, thus

$$\binom{m}{r} = \frac{m!}{r!(m-r)!} = \lambda \quad (2)$$

λ is the number of functions arising from all possible pairs of combination with m-groups. λ would certainly be a non-negative integer. Given m-groups, evaluation of the number of all possible pairs would result to λ number of functions in the form of a linear functions arising from combining possible pairs of groups without repetition. Clearly, we would have a set of discriminant Functions (DF) representing every possible pairs of group combinations of the *m*-groups (in general) as

$$DF_{m,1}, DF_{m,2}, \dots, DF_{m,m-2}, DF_{m,m-1}. \quad (3)$$

[3]allocation rule is to Allocate to group I if

$$\alpha^T X > D(in\ multivariate\ case) \quad (4)$$

Otherwise allocate to group II

Where

$$Z = a^T X = a_1x_1 + a_2x_2 + \dots + a_kx_k \quad (5)$$

$$D = \frac{1}{2}(\bar{X}_1 + \bar{X}_2)^T \boldsymbol{S}^{-1}(\bar{X}_1 - \bar{X}_2) \quad (6)$$

It is worthy to note that D can only be computed for two groups at a time. Since our derivations are in pairs, it is also possible to obtain for each possible pair, a corresponding and appropriate D-value. Thus, for *m*-groups and λ number of discriminant functions, we would have a set of D-values (in general) in the form;

$D_{m,1}, D_{m,2}, \dots, D_{m,m-2}, D_{m,m-1}$

Table-1
Summary of Discriminating values for all Groups[19]

Group I	Group II	Group III	…	Group M
$D_{1,2}$	$D_{2,1}$	$D_{3,1}$	…	$D_{m,1}$
$D_{1,3}$	$D_{2,3}$	$D_{3,2}$	…	$D_{m,2}$
$D_{1,4}$	$D_{2,4}$	$D_{3,4}$	…	$D_{m,3}$
…	…	…	…	…
$D_{1,m-1}$	$D_{2,m-1}$	$D_{3,m-1}$		$D_{m,m-2}$
$D_{1,m}$	$D_{2,m}$	$D_{3,m}$	…	$D_{m,m-1}$

Since the combinatorial analysis so far has given us λ number of discriminant functions and λ number of D-values. It follows that λ number of rules would be required to conveniently allocate observations. Having stated this, the λ number of rules that can allocate future observation derived on the basis of our initial combinatorial concept would be such that each possible pair would have a corresponding allocation rule. This would clearly give rise to a set of λ independent rules for every possible pairs in each of the *m*-groups. In general, we would have set of rules for each of the *m*-groups, as

Allocate to

$$G_1\ if\ DF_{m,1} > D_{m,1}.$$

else

$$G_2\ if\ DF_{m,2} > D_{m,2}.$$

else

$$G_{m-1}\ if\ DF_{m,m-1} > D_{m,m-1}.$$

Otherwise

$$Allocate\ to\ G_m.$$

Table-2
Summary of the Allocation Rules[19]

All Original G_1	All Original G_2	All Original G_3	…	All original G_m
$G_1\ if\ DF_{1,2} > D_{1,2}$	$G_1\ if\ DF_{2,1} > D_{2,1}$	$G_1\ if\ DF_{3,1} > D_{3,1}$	…	$G_1\ if\ DF_{m,m-1} > D_{m,m-1}$
$G_2\ if\ DF_{1,3} > D_{1,3}$	$G_2\ if\ DF_{2,3} > D_{2,3}$	$G_2\ if\ DF_{3,2} > D_{3,2}$	…	$G_2\ if\ DF_{m,m-1} > D_{m,m-1}$
$G_3\ if\ DF_{1,4} > D_{1,4}$	$G_3\ if\ DF_{2,4} > D_{2,4}$	$G_3\ if\ DF_{3,4} > D_{3,4}$	…	$G_3\ if\ DF_{m,m-1} > D_{m,m-1}$
…	…	…	…	…
$G_{m-2}\ if\ DF_{1,m-1} > D_{1,m-1}$	$G_{m-2}\ if\ DF_{2,m-1} > D_{2,m-1}$	$G_{m-2}\ if\ DF_{3,m-1} > D_{3,m-1}$		$G_{m-2}\ if\ DF_{m,m-1} > D_{m,m-1}$
$G_{m-1}\ if\ DF_{1,m} > D_{1,m}$	$G_{m-1}\ if\ DF_{1,m} > D_{1,m}$	$G_{m-1}\ if\ DF_{1,m} > D_{1,m}$	…	$G_{m-1}\ if\ DF_{1,m} > D_{1,m}$
Otherwise G_m	Otherwise G_m	Otherwise G_m	…	Otherwise G_m

The initial work used a real data set to evaluate the feasibility and computational possibility of the APP procedure and compared it with the conventional FLD procedure. Here, multivariate normal distributed data was simulated under sample sizes n=30,50,100,250,500 using a U(1,10) to assign groups arbtrarily to them for ten groups. Accuracy rate (percentage of the number of observations each procedure correctly classified) for each procedure was obtained by cross validation for the various sample sizes and an Average Accuracy Rate (AAC) for each of the procedures was obtained accordingly.

Results and Discussion

The APP classification procedure gave an appreciable accuracy rate that varied with change in sample size especially as sample size increases. Though it reduced with n=250, it however did not end up on a poor note with an AAC of 60.68%. The next was the KCA which maintained a relative increase in accuracy rate as sample size increases with an AAC of 60.56% very close to that of APP. The FLD, a more conventional and often commonly implored procedure in this situation produced an accuracy rate that was a reverse of the others, in the sense that, its accuracy rate decreased with increase in sample sizes thereby producing a poor AAC of 23.54%. The CRT performed poorly than the FLD with an AAC as small as 11.48%. The logistic regression produced a seemingly outstanding accuracy rate with all the sample sizes with an AAC of 99.3%. This is perhaps as expected because of its inability to make strong distributional assumption concerning a data; the method has been adjudged weak and a last option if paradventure, other fair procedure fails. Its inclusion in this work is to justify the fact that some researchers and data analyst uses it for classification. The CRT's performance, though poor when compared with its non-parametric counterpart, could be attributed to the nature of data used, thus, it would be worth noting the nature and type of available data before considering the CRT as a method to be adopted in classification.

Conclusion

This study has so far considered and implemented the procedure suggested in this work, thereby observing that when available groups are many, it is better to consider and carry out evaluation in pairs. Evaluation in pairs makes sure that error resulting from combining the many groups at the same time is minimized. It also ensures that every possible pairs are considered appropriately since statistically accepted allocation rule makes provision for accommodating only two groups at a time. The FLD procedure might be considered for use when available data set is of a large sample size and the CRT procedure considered only after examining the nature and type of available data. This result agrees with that of Oyeyemi et.al. [20] in which it was observed that the performance of the FLD procedure is poor with few number of variables and a larger sample size. The KCA could have also been an alternative but since we are interested in discrimination and allocation and not forming of clusters, it therefore faces some limitation.

The procedure presented in this work has shown considerable and fairly appreciable performance when compared with its conventional parametric and non-parametric counterparts. It would also assuredly overcome the problem of sample size because even with a small sample, its performance was fairly outstanding. This procedure, as previously noted, is also based on mathematical acceptable concepts and has in no way violated or deviated from known and important statistical principles. The procedure though may look cumbersome and lengthy but carefully written computer programs using a user-friendly language would make the procedure more appreciable in terms of speed, time optimization and accuracy. Conclusively, when we have multiple groups, the conventional procedure only provides a method that exists in theory but contradictory in practice, thus, we observe and hence suggest that, with multiple groups, higher accuracy in discrimination and allocation of observation can be fairly achieved by adopting the procedure suggested in this work.

Table-3
Accuracy Rates (%) for each of the procedures for the various sample sizes

Sample size (n)	APP	FLD	Logistic Regression (LR)	Classification and Regression Tree (CRT)	K-means Cluster Analysis (KCA)
30	50.0	33.3	100.0	10.0	60.0
50	62.0	32.0	100.0	10.0	48.0
100	67.0	21.0	100.0	10.0	56.0
250	60.4	18.0	100.0	14.8	68.8
500	64.0	13.4	96.6	12.6	70.0
AAC	60.68	23.54	99.3	11.48	60.56

References

1. Ekezie D.D., Principal Component Analysis, an Aid to Interpretation of Data: A Case Study of Oil Palm (Elaeis guineensis jacq). *J. of Emerging Trends in Engr. and App. Sci*., **4(2)**, 237-241 **(2013)**

2. Onyeagu S.I., Osuji G.A. and Onyia O.M., Comparison of Canonical Correlation Analysis and the Generalized Canonical Correlation Analysis using the Lognormal and Cauchy Distributed Data, *J. of Math. Theory and Mod*., **5(4)**,177-185 **(2014)**

3. Fisher R.A., The Use of Multiple Measurements in Taxonomic Problems, *Ann. Eugenics*., **(7)**, 179-188 **(1936)**

4. Fisher R.A., The Precision of Discriminant Functions, *Ann. Eugenics*., **(10)**, 422-429 **(1940)**

5. Fukunaga F., Introduction to statistical pattern recognition, *Academic Press*, **(1990)**

6. Sandrine D., Jane F. and Terence P.S., Comparison of Discrimination Methods for the Classification of Tumors Using Gene Expression Data. Technical Report #576, Department of Statistics, University of California, **(2000)**

7. Breiman L., Bagging Predictors, *Machine Learning.*, **(24),**123-140 **(1996)**

8. Breiman L., Arcing Classifiers, *Ann. of Stat*., **(26),** 801-824 **(1998)**

9. Freund Y. and Schapire R.E., A Decision-Theoretic Generalization of On-line Learning and an Application to Boosting, *J. of Comp. and System Sci*., **(55),** 119-139 **(1997)**

10. Lesaffre E. and Albert A., Multiple-Group Logistic Regression Diagnostics, *J. of the Royal Stat. Soc*., **38(3)**, 425-440 **(1989)**

11. Vapnik V.N., Statistical Learning Theory, Wiley, New York., **(1998)**

12. Tao L., Shenghou Z. and Mitsunori O., Using Discriminant Analysis for Multi-class Classification: An Experimental Investigation, *Knowledge and Information Systems*., **10(4)**, 453-472 **(2006)**

13. Hastie T. and Tibshirani R., Classification by Pairwise Coupling, In M. I. Jordan, M. J. Kearns, and S. A. Solla (eds.), Advances in Neural Information Processing Systems, 10, **(1998)**

14. Kwon Y.S., Han I.G. and Lee K.C., Ordinal Pairwise Partitioning approach to Neural Networks, *Training in Bond Rating, Intelligent Systems in Accounting Finance and Management*., **(6),** 23-40 **(1997)**

15. Platt J.C., Cristianini N. and Taylor S.J., Large Margin DAGs for Multiclass Classification. In S. Solla., T. Leen, and K.R. Muller (eds.), Advances in Neural Information Processing Systems, **(2000)**

16. Dietterich T.G. and Bakiri G., Solving Multiclass Learning Problems via Error-Correcting Output Codes, *J. of Artif. Intell. Res*., **(2),** 263–286 **(1995)**

17. Anyanwu P.E., Ekezie D.D. and Onyeagu S.I., A Review of the limitations of some Discriminant Analysis procedures in Multiple Group Classification, *J. of Math Theory and Modeling*., **5(9)**,199-203 **(2015)**

18. David H., Error-Rate Estimation in Multiple-Group Linear Discriminant Analysis, *Technometrics*., **4(38)**, 389-399 **(1996)**

19. Anyanwu P.E., Ekezie D.D., Onyeagu. S.I. and Nwankwo I.I.M., A Generalized Multi-Group Discriminant Function procedure for Classification: An Application to Ten Groups of Yam Species, *J. of Math Theory and Modeling*., **5(10)**,115-123 **(2015)**

20. Oyeyemi G.M. and Mbaeyi G.C., On Discrimination procedure with mixtures of Continuous and Categorical Variables, Unpublished M.Sc Thesis, Department of Statistics, University of Ilorin, Ilorin, Nigeria, **(2013)**