Available online at http://www.journalcra.com

International Journal of Current Research
Vol. 5, Issue, 10, pp.2817-2820, October, 2013

INTERNATIONAL JOURNAL OF CURRENT RESEARCH

ISSN: 0975-833X

RESEARCH ARTICLE

EVALUATION OF CLASSICAL AND ROBUST DISCRIMINANT METHODS UNDER APPARENT ERROR RATE

Muthukrishnan, R. and *Mahesh, K.

Department of Statistics, Bharathiar University, Coimbatore – 46

ARTICLE INFO

Article History:
Received 25th July, 2013
Received in revised form
20th August, 2013
Accepted 14th September 2013
Published online 10th October, 2013

Key words:
Discriminant Analysis –
Robust Discriminant Analysis –
Apparent Error Rate– R.

ABSTRACT

All classical statistical methods rely explicitly or implicitly on parametric models based on number of assumptions. The most widely used assumption is that the observed data have normal distribution. This assumption about the structural and the stochastic parts of the model have been present in statistics for two centuries, and have been the framework for all the classical methods. Classical methods perform well if the data obeys the assumptions. Now-a-days data collected and stored at enormous speed (GB/TB/hr) and pressure to provide better customized service for an edge. The data does not follow the so-called assumptions then the result using classical methods get affected. In this context traditional techniques are infeasible due to enormity of data, high dimensionality of data and heterogeneous of data. The robust methods can be seen as extensions to the classical ones which can cope with deviations from the stochastic assumptions. Classification and data reduction techniques play an important role while handling large data. A reliable and precise classification aspect is essential in analyzing multivariate data. This paper presents the evaluation aspects such as apparent error rate of various classical and robust discriminant methods on a simulation study using R package.

INTRODUCTION

Classical statistical methods try to give a better classification result as well as possible. The usual criterion is discriminant analysis, is classified or predicted in the data. If the data set contains vulnerable, the classification may deviate strongly from those obtained from the non – vulnerable data. For instance, outliers can deviate discrimination classification analysis. Since all data have same weight in the discriminant criterion, large deviation is distributed over all the residuals, often making them hard to detect. Robust statistics is to reduce the impact of outliers, robust discriminant analysis methods try to give correct classification the bulk data, which assumes that the good observations outnumber the vulnerable data. Outliers can then be identified by looking at the residuals, which are large in the robust analysis. In this paper we have discussed the classical and robust discriminant analysis. The necessary preliminaries concerning this paper are briefly furnished in this section. The comparisons of classical and robust discriminant analysis are discussed in the forthcoming sections.

Linear Discriminant analaysis

Discriminant Analysis usually means classification by linear functions. The objective of discriminant analysis is to classify the sample objects into two or more groups. This is done with the help of linear combination of predictors or explanatory variables. The basic principle to determine the group with which an object is identified is that the misclassification error of that object is not maximum. The optimality of classification depends on the assumption of data, these are the predictor variables follow multivariate normal distribution and the covariance matrices of different groups of data are homogeneous. As the mathematical methods used in discriminant analysis are complex, they are described here only in general terms. The tendency of an individual to become a high performer can be written as a linear equation is,

$$D = \beta_0 + \beta_1 X_1 + \beta_2 X_2 +,...,+\beta_p X_p \qquad (2.1)$$

where D is the discriminant function score, βi is the discriminant function coefficient relating, independent variable i to the discriminant function score, xi is the value of independent variable i.

Maximum Likelihood Estimator (MLE)

The principle of maximum likelihood estimation (MLE), originally developed by R.A Fisher in the (1920). The standard estimates are the maximum likelihood estimates or their unbiased variants. The sample mean vectors for the provision of the estimates of $\hat{\mu}_i$ are given by

$$\hat{\mu}_i = \overline{Y}_i = \sum_{j=1}^{n_i} Y_{ij} / n_i \ , \qquad (i=1,\ldots,g)$$

for the hetroscedastic case, each $\sum_I$ is estimated by its training sample analogue usually after correction for bias, to give

$$\hat{\Sigma}_i = S_i = \sum_{j=1}^{n_i} (Y_{ij} - \overline{Y}_{ij})(Y_{ij} - \overline{Y}_{ij})' / (n_i - 1). \qquad (i=1,\ldots,g)$$

For the homoscedastic discriminant analysis model the standard estimate of the common covariance matrix $\sum$ is the pooled within – group sample covariance matrix

$$\hat{\Sigma} = S = \sum_{i=1}^{g} \sum_{j=1}^{n_i} (Y_{ij} - \overline{Y}i.)(Y_{ij} - \overline{Y}i.)' / (n - g),$$

where $n=\sum_{ni}$ is the total sample size across a groups.

***Corresponding author:** Mahesh, K.*
Department of Statistics, Bharathiar University, Coimbatore – 46

Robust Linear Discriminant Analysis

The robust linear discriminant analysis towards outliers depends completely on the estimator for the group center and common covariance matrix. There are two very popular ways to measure the robustness of an estimator is by mean of its breakdown value (Rousseeuw and Leroy, 1987) and its influence function. It is well known that sample averages and sample covariance matrix are very sensitive to outlying in the training sample may have an unduly large influence on the LDA. The influence of outliers on the discriminant warrants careful scrutiny by researchers, fortunately Huber (1977 b) and Maronna (1976) have developed the basic tools for identifying outliers and modifying the estimators to reduce their effects. Their theory deals with the robust estimation of a mean vector and covariance matrix of a single multivariate population. Randles et,al. (1978) used the robust estimates from each of two groups to define the two group discriminant function. Harner and Billings (1983) reviewed Huber's techniques for determining robust M-estimates of multivariate location and dispersion and also discussed robust discriminant analysis.

Minimum Covariance Determination Estimators

The multivariate location and diffusion estimation in high breakdown principles is based on the determinant of the covariance matrix. If the covariance matrix nxn positive semi-definite matrix, p eigenvalues are positive, the determinant of covariance matrix equals the product of eigenvalues. Thus, a small value in the determinant reflects some linear patterns in the data. Consider all C_h^n subsets, and compute the determinant of the covariance matrix for each subset. The subset with smallest determinant is used to calculate the usual $px1$ mean vector, and corresponding pxp covariance matrix, these estimators are called minimum covariance determinant estimators. The Fast MCD algorithm was developed by Rousseeuw P.J. Driessen K. (1999), sample of size h. we can evaluate the similarity between data points in the full set and our randomly simulated sample subset. In particular the Mahalanobis distance is used here, let M be the mean of the random subset and S be the standard covariance of random subset.

$$D = [(x - M)S^{-1}(x - M)]^{1/2}.$$

The algorithm is as follows:

1. choose a random subset of H_0 of x, with size h,
2. repeat
   a. Determine covariance S and mean M of the subset H_0
   b. Determine distances $d(x_i)$ or all x_i relative to H with the Mahalanobis distances
   c. Choose the h smallest distances and create a new subset H_1
   d. Repeat with $H_0 \leq H_1$ until H_0 and H_1 are equal or 0.
3. Evaluate from 1 for k times and determine the selection that had the smallest volume.

Minimum Covariance Determinant in all Group (MCD – A)

Minimum covariance determinant procedures in all groups was developed Todorov et al. (1990), Rousseeuw et al. (1992), Croux and Dehon (2001) all are applied this procedure for robustifying linear discriminant analysis based on S estimates. A drawback of this method is that the same trimming proportions are applied to all groups which could lead to a loss of efficiency if some groups are outliers free.

Minimum Covariance Determinant in each one of the Group (MCD – B)

Minimum covariance determinant analysis in each one of the group was proposed by He and Fung (2000) for the S estimates and modified algorithm was developed in Hubert and VanDriessen (2004), as a replacement for pooling the group covariance matrices, the observations are centered and collective to obtaining the each one of the individual group location estimates as the reweighted minimum covariance determinant location estimates of each group.

Minimum covariance Determinant in individual group (MCD – C)

This method is generally modifying algorithm of high breakdown point estimation in order to contain the individual sample or group. In this method was discussed in Valentin Todorov (2007), this algorithm taking advantages of the FAST-MCD, but it is still to compute the MCD for each individual group.

Feasible Solution Algorithm (FSA) using Minimum Covariance Determination (MCD)

The FSA for MCD method is using by Douglas M. Hawkins. (1994), it is clear from the definition that the MCD estimator satisfies the following necessary condition. The MCD estimator for μ and Σ is the sample mean vector and covariance matrix of a subset of size $n-h$. The determinant of $\hat{\Sigma}$ cannot be decreased by any case wise exchange exchanging one of the trimmed cases for one of the retained cases. This condition is satisfies the following algorithm. Start from a trial subset some random subset of $n-h$ of the cases retaining these cases and trimming the other h and find their mean vector and covariance matrix. Investigate all possible pairwise exchanges in which one of the determinants of the covariance matrix of the resulting new subset is smaller than that of the old subset could not be the MCD solution and may be replaced by the new one. Implement this exchange, getting a new trial subset and its mean vector and covariance matrix. If there is more than one exchange that will lead to a reduction in the determinant, than make that exchange that will lead to the greatest reduction. Repeat the process of investigating pariwise case exchange with the new trial subset. If there is no possible pairwise exchange that would lead to a reduction in the covariance determinant, then the current trial subset satisfies the necessary condition for the MCD optimum, this subset and its mean vector and covariance matrix a feasible set and a feasible solution respectively. Repeat this process with many different random trial subsets, following each to a feasible solution. Take as the estimated global optimum MCD the feasible solution with smallest determinant. The FSA properties consider the all values satisfy the necessary condition for the MCD. The properties of this implementation of the algorithm discussed as follows. The kernel of the algorithm is testing whether the determinant would be reduced by a pairwise case exchange. This test is conveniently implemented by augmenting each X_i with a 1, defining

$$z_i = (1 : X_i^T)^T.$$

Consider a trial partition of the cases into trimmed and retained cases. Let $J = \{i_1, i_2, ..., i_{n-h}\}$ be the set of indices of the currently retained cases and write the partitioned matrix

$$Z_J = (z_{i_1}\ z_{i_2} \ldots z_{i_{n-h}}).$$

Let $A = Z_J Z_J^T$. Then clearly A can be written in partitioned form as

$$A = \begin{bmatrix} n-h & (n-h)\overline{X}_J^T \\ (n-h)\overline{X}_J & X_J X_J^T \end{bmatrix}$$

where $\overline{X}_J$ represents the mean vector of the retained cases in the partition J. The determinant of A can be written using partitioned matrix results as

Table 1. Performance of classical and Robust Discriminant Analysis

	Classical Method	Robust Methods						Classical Method	Robust Methods					
Methods	MLE	MVE	MCD	MCD A	MCD B	MCD C	FSA	MLE	MVE	MCD	MCDA	MCD B	MCDC	FSA
ε			L.C $n_1=n_2=n_3=50$, g=2							L. C $n_1=n_2=n_3=100$, g=2				
0.0	0.03	0.03	0.03	0.02	0.03	0.02	0.02	0.030	0.06	0.030	0.050	0.030	0.030	0.030
0.05	0.03	0.02	0.02	0.00	0.02	0.02	0.01	0.070	0.10	0.020	0.020	0.020	0.020	0.020
0.10	0.04	0.02	0.02	0.00	0.02	0.02	0.02	0.080	0.11	0.025	0.040	0.025	0.025	0.025
0.15	0.06	0.02	0.02	0.01	0.02	0.02	0.02	0.080	0.13	0.020	0.040	0.020	0.025	0.020
0.20	0.06	0.03	0.03	0.02	0.02	0.02	0.02	0.090	0.15	0.020	0.020	0.020	0.020	0.025
ε			S. C $n_1=n_2=n_3=50$, g=2							S.C $n_1=n_2=n_3=100$, g=2				
0.0	0.03	0.03	0.03	0.02	0.03	0.02	0.02	0.060	0.07	0.060	0.050	0.060	0.030	0.025
0.05	0.02	0.03	0.03	0.00	0.03	0.02	0.01	0.140	0.11	0.065	0.065	0.070	0.050	0.050
0.10	0.02	0.01	0.02	0.00	0.03	0.02	0.01	0.155	0.16	0.070	0.065	0.070	0.060	0.055
0.15	0.02	0.04	0.02	0.00	0.03	0.02	0.02	0.220	0.27	0.070	0.085	0.075	0.065	0.065
0.20	0.02	0.05	0.01	0.01	0.01	0.01	0.01	0.225	0.37	0.085	0.100	0.085	0.080	0.075
ε			L&S C $n_1=n_2=n_3=50$, g=2							L&S C $n_1=n_2=n_3=100$, g=2				
0.0	0.03	0.14	0.03	0.02	0.02	0.02	0.02	0.02	0.03	0.015	0.015	0.020	0.010	0.010
0.05	0.05	0.14	0.02	0.02	0.02	0.02	0.02	0.20	0.07	0.020	0.020	0.02	0.02	0.020
0.10	0.09	0.14	0.02	0.02	0.02	0.02	0.02	0.23	0.08	0.020	0.025	0.020	0.030	0.020
0.15	0.09	0.14	0.02	0.02	0.02	0.02	0.02	0.23	0.10	0.020	0.020	0.020	0.030	0.020
0.20	0.12	0.12	0.02	0.02	0.02	0.02	0.02	024	0.32	0.020	0.020	0.025	0.020	0.020

L.C - Location Contamination, S.C – Scale Contamination, L&S C – Location and Scale Contamination

$$|A| = (n-h)\left|X_J X_X^T - (n-h)\overline{X}_J \overline{X}_J^T\right| = (n-h)|(n-h)S_J|,$$

where S_J is the sample covariance matrix of the retained cases. It follows at once that the minimum of the MCD criterion is given by

$$|S_J| = |A|/(n-h)^{p+1},$$

which is just a rescaling of $|A|$ by a constant. Thus $|S_J|$ may be minimized by minimizing $|A|$. The FSA involves evaluating pairwise exchange between a retained case and a trimmed case. This can be facilitated the determinate identify that for column vector u and v

$$\left|A + vv^T - uu^T\right| = |A|\left[(1-u^T A^{-1}u)(1+v^T A^{-1}v) + (u^T A^{-1}v)^2\right].$$

Here u and v represents the Z vector of the cases to be respectively trimmed and restored. The pairwise swap will lead to a reduction in the covariance determinant if the term in brackets is less than 1, and the swap to make is that for which the bracketed term is a minimum.

Simulation Study

For this paper purpose we have simulated multivariate random variable in different mean, and covariance matrix with and without contaminated data presented in the simulation study. The different proposals through a simulation study as in He and Fung (2000), there are three different type of comparison made in this simulation study, first one is location contamination μ_1= (0,0,0) , μ_2(1,1,1) and covariance matrix $\sum$ is identity matrix (1,0,0,0,1,0,0,0,1) with three variables X_1,X_2,X_3 , and also we simulate the contaminated mean vector is as follows $\hat{\mu}_1$ = (-4,-4,-4) , $\hat{\mu}_2$ (5,5,5), but the $\sum$ is same and number of group is 2. The second type is scale contamination in this case we simulate the same with and without contamination mean vector but the covariance matrix is differ identity matrix multiple (0.025^2) in contamination case, the last one is location and scale contamination in this type we mixed the location and scale contamination, with their study comparison the no of sample size in the all type of method is 50 and 100. There are many method comparisons in V.Todorov and Ana M. Pries (2007), but we have study some of the classical and robust discriminate methods and their apparent error rates. Table 1 is display the estimated over all apparent error rates as a function of the contamination proportion ε in different types of contamination. We will regard as all type of comparison the classical method is affected in the contamination situation but the robust methods are correctly classified when the data is affected in 20% of outliers

Conclusion

The classical and robust discriminant analysis produces the same result when the data contains no contaminated observations called vulnerable. The contaminated data encountered, it is observed from the table 1, the apparent error rate (or) misclassification of data is not more affected in the case of Robust Linear Discriminant analysis (RLDA). The robust discriminant procedures perform well when the data contains extreme observations. Robust procedure are not affected that much when the data contains outliers.

Acknowledgements

This research was partially supported by the Department of Science and Technology (DST), Promotion of University Research and Scientific Excellence (PURSE) programme in New Delhi, at Bharathiar University, Tamil Nadu, India.

REFERENCE

1. Anderson, T.W., Bahadur, R.R. (1962). Classification Into Two Multivariate Normal Distributions With Different Covariance Matrices, *The Annals of MathameticalStatistics* 33:420-431.
2. Anderson, T.W. (1984). An Introduction to Multivariate Statistical Analysis.*New York: John Wiley and Sons lnc.*
3. Pires Ana, M., Branco Joao.A. (2010). Projection-pursuit approach to robust linear discriminant analysis. *Journal of Multivariate Analysis* 101:2464-2485.
4. Croux, C., Dehon, C. (2001). Robust linear discriminant analysis using s-estimators. *The Canadian Journal of Statistics* 29: 473-492.
5. Cutillo, L., Amato.U. (2008). Localized empirical discriminant analysis. Computational Statistics and Data Analysis 52:4966-4978.
6. Douglas M. Hawkins (1994). The feasible solution algorithm for minimum covariance determinant estimator in multivariate data. *Computational Statistics & Data analysis 17:* 197-210.
7. Fisher, R.A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics*.
8. Fienberg, S.E. (1936).Graphical Methods in Statistics. *The amer.Statistician* 33:165-178.

9. Harner, E.J., Billings, A.A. (1983). Robust Discriminant Analysis. *Proceedings of the Eighth* Annual SUGI Conference: 873-878.
10. He, X. and FUNG, W. (2000). High breakdown estimation for multiple populations with applications to discriminant analysis, Journal of Multivariate Analysis, 72, 151-162.
11. Hubert, M., Rousseeuw, P.J., van Aelst,. S. (2008). High-Breakdown Robust Multivariate Methods. *Statistical Science* 23:92-119.
12. Huber, P.J. (1977 b). Robust Covariances. New York: In Gupta. S.S. and D.S. Moore. eds., Statistical Decision Theory and Related Topics II.Academic Press. 165-191.
13. Hubert M., Van Driessen .K. (2004). Fast and Robust Discriminant Analysis. *Computational Statistics & Data Analysis* 45: 301-320.
14. Maranna, R.C. (1976). Robust M estimators Multivariate Location and Scatter. *Ann Statist* 4:51-67.
15. Narayan, C.Giri. (2009). Multivariate Statistical Analysis .Second Edition, Revised and Expanded. Randles, R.H. et a1. (1978). Generalized Linear and Quadratic Discriminant Functions Using
16. Robust Estimates, *J. of the Amer. Statist. Assoc* 73:563-578
17. Rousseeuw, PJ., Leroy. AM. (1987). Robust Regression and Outlier Detection.*New York: John Wiley & Sons.*
18. *Rousseeuw PJ. Van Driessen K (1999).* A Fast Algorithm for minimum Covariance Determinant Estimator. Technometrics, 41, 212-223.
19. VALENTIN TODOROV and ANA M.PIRES (2007). Comparative Performance of Several Robust Linear Discriminant Analysis methods. *REVSTAT – Statistical Journal,* volume 5, Number 1: 63-83.

****