ISAS

Available online at www.isas.org.in/jisas

JOURNAL OF THE INDIAN SOCIETY OF AGRICULTURAL STATISTICS 69(3) 2015 271-279

A Comparative Study of Various Classification Techniques in Multivariate Skew-Normal Data

Samarendra Das[1], Amrit Kumar Paul[1], S.D. Wahi[1] and Rohan Kumar Raman[2]

[1]ICAR-Indian Agricultural Statistics Research Institute, New Delhi

[2]ICAR-Central Inland Fisheries Research Institute, Kolkata

Received 10 November 2015; Accepted 25 December 2015

SUMMARY

The assumption of normality in data has been considered in the field of statistical analysis for a long time. However, in many practical situations, this assumption is clearly unrealistic. It has recently been suggested to study the performance of various statistical techniques like classification by using the data from distributions indexed by skewness/ shape parameters. In this study, four different classification techniques, namely linear discriminant analysis, quadratic discriminant analysis, *k*-th nearest neighbor and oblique axes method are considered for classification of observations. To assess the performance of the above techniques under non-normality caused by skewness, which is introduced in the ricebean data by using multivariate skew-normal distribution through simulation. Apparent error rate is used to study the classification performance of these techniques. The result of this study can be used for choosing the best method of classification for skewed-normal situation. The results indicate that *k*-th nearest neighbour followed by oblique axes methodand linear discriminant analysisperform better in skew-normal situations than normal condition and quadratic discriminant analysis performed better in normal data. For maximum consistency and accuracy of classification of skew-normal data, *k*-th nearest neighbor is best among the four classification techniques.

Keywords: Classification, Linear discriminant analysis, Quadratic discriminant analysis, *k*-th nearest neighbor, Oblique axes method, Apparent error rate, Multivariate skew normal distribution.

1. INTRODUCTION

Classification is of broad interest in science because it permeates many scientific studies and also arises in the contexts of many applications. For example in agriculture, crop varieties are classified into different groups which are suitable for different agroclimatic zones of a region and in the biological and medical sciences, applications of classification procedures include identifying patients with chronic heart failure, detecting lung cancer *etc*. The primary goal of classification is to correctly sort out objects into two or more mutually exclusive groups. Classification is often categorized into two subtypes: supervised and unsupervised (Rausch and Kelley 2009). Supervised classification, also known as discriminant analysis is used to correctly assign future objects to groups that are already known to exist (Johnson and Wichern 2002). Unsupervised classification is used to assign objects to groups that are not known a priori. Several researchers have utilised assumptions of normality in the data for the classification of observations in to two or multiple groups (McLachlan 1992). However, these studies have prolonged this practice for many years without using the flexible and modern distributions that have been introduced

Corresponding author: Samarendra Das
E-mail address: sam@iasri.res.in

recently. Each classification technique has its own strengths and limitations based on its mild assumptions. For example, the typical discriminant function method used is the linear discriminant function (LDF) when the data followed normal distribution and dispersion matrices are equal (Bobrowski 1986), and the quadratic discriminant function (QDF) (Hubert and Van der Veeken 2010), when normality of the data is satisfied. The real world data in general and crop morphological data in particular do not satisfy the assumptions like normality, equality of dispersion matrices (Wahi and Bhatia 2005). The selection of appropriate classification techniques for the grouping of crop genotypes, when the non-normality in the data is due to the shape parameters of the distribution is of paramount importance.

Simulation studies are extensively used by the researchers to study the performance of the classification techniques when some assumptions about the data characteristics are violated (Kiang 2003). For example, Whal and Kronmal (1977) used the simulation techniques to compare the performance of three discriminant functions *i.e.* the quadratic, best linear and Fisher's linear function in classifying individuals into two multivariate normally distributed populations when the dispersion matrices are unequal and the results indicated that for large samples from multivariate normal distributions, the quadratic is much better than Fisher's function and for small samples, the former performs worse than later. Wahi *et al*. (1986) used the best LDF for comparing the different grades of sheep in cross breeding programme and the instances state that 80% of comparisons among the different grades of sheep the probability of misclassification by the best LDF were found to be either lower or equal to the probability of misclassification obtained by Fishers linear discriminant function.An alternate classification technique was given by Das (1998) which is based on the classification of observations using distances in oblique co-ordinate system. Another case study conducted by Erimafa *et al*. (2009) suggested that LDF has hit ratio 88.2% and a valid tool for classifying fresh students of unknown origin into predefined groups (poor class degree and better class of degree). The effect of non-normality on LDA was conducted by Rausch and Kelly (2009) using Monte-Carlo simulation and the results showed that LDA is less robust than the logistic regression. Soni *et al.* (2010) applied different classification techniques like classification and regression tree, linear and quadratic discriminant analysis to the Indian stock market data with an aim for maximising profit of market analyst and investors to make decision for selling or purchasing stock a particular company. The relative efficiency of linear classification rule in multi group discriminant analysis through Monte-Carlo study was conducted by Glele *et al*. (2010). The effect of different combinations of dimensions and sample size was undertaken and they showed that for normal and homoscedastic populations, the efficiency of the rule is better for large number of groups. In this study, we tried to understand the strength and limitation of four different classification techniques when non-normality in the data is introduced by the skewness/shape parameters in a controlled setting. The intention here was to investigate how the classification techniques perform when certain assumptions are violated. Further, we studied the effect of skewness levels, sample size and dimension as well as their combinations on the performance of these techniques. The finding from this study would helpful in choosing the proper classification techniques for normal and skewed-normal situations.

2. MATERIALS AND METHODS

The secondary data on 131 genotypes of ricebean (*Vigna umbellate* L.) grown at Bhubaneswar, Odisha are used. The data is available in the Annual Report for the year 2007-08 of All India Coordinated Research Network on Underutilized Crops, NBPGR, New Delhi. The data consists of 9 morphological quantitative characters such asdays to 50% flowering, days to maturity, plant height (cm), pods per plant, pod length (cm.), seeds per pod, 1000 seed wt. (gm), seed yield/plant (gm), plot yield (gm). In the present investigation, we considered the ricebean data with nine morphological characters comprises of three groups. First group consists of 52 genotypes, second group consists of 38 genotypes and third group consists of 41 genotypes. The normality of the data set is tested by using the Mardia's test (1980). It is found that both the skewness and kurtosis components are coming highly significant.

2.1 Linear Discriminant Analysis

Linear Discriminant Analysis (LDA) is one of the most popular methods of supervised classification. This procedure can be conceptualized as a nonparametric method (i.e., distributional assumptions are not explicitly made) because it maximizes between group variability relative to within-group variability. However, it can also be conceptualized as a parametric procedure for classification. In particular, LDA is optimal (i.e., it maximizes classification accuracy) under the assumptions that the within-group predictors follow multivariate normal distributions and that the population covariance matrices are equal across groups.

2.2 Quadratic Discriminant Analysis

Quadratic Discriminant Analysis (QDA) is closely related to LDA and commonly used techniques for multi-group classification. Unlike LDA however, in QDA there is no assumption that the covariance matrices of each the groups are identical. When the assumption is true, the best possible test for the hypothesis that a given measurement is from a given group is the likelihood ratio test.

2.3 *k*-th Nearest Neighbor

The non-parametric (or distribution free) method, KNN (kiang, 2003; Wu, *et al.*, 2010) is used for classifying observations into multiple groups based on a set of quantitative variables. It relaxes the normality assumption and does not require a functional form as required in LDA and QDA. The distance, $d(x,y)$, between any two observations is usually defined by Mahalanobis distance between x and y. Using the nearest neighbor rule, an observation is classified to one of the groups to which a majority of its *k*-th nearest neighbors belong. The sample distribution approximation is accomplished by dividing the variable space in to arbitrary number of decision regions.

2.4 Oblique Axes Method

The Oblique Axes Method (OAM) is considered to be a non-parametric (or distribution free) method as it does not assume the distributional form of the population. The method classifies the observations into one of the several groups based on the square of distances between points corresponding to observation vectors using the oblique co-ordinate system. Some weight factors are associated with the distances are known as compounding values (Rao, 1946). These weights can be calculated by maximizing the ratio between average squared distances of all possible pairs of the group mean vectors to the pooled average squared distances within groups.

Next, to classify an observation vector $(z_1, z_2, z_3. . ., z_p)$ in to one of the several groups, the distance square of a point of the given observation vector from that of each of the points of the mean vectors of different groups is obtained. The observation vector belongs to that population whose mean vector has least distance from the observation vector point.

2.5 Criterion Used for Assessing the Performance

The performance of the classification procedures are assessed by using the Apparent classification Error Rate (APER) (Pohar *et al.*, 2004), which is discussed as follows

Table 1. Schematic Representation of g×g Confusion Matrix

	Π_1	Π_2	. . .	Π_{g-1}	Π_g	Total
Π_1	N_{11}	N_{12}	. . .	N_{1g-1}	N_{1g}	N_1
Π_2	N_{21}	N_{22}	. . .	N_{2g-1}	N_{2g}	N_2
Π_g	N_{g1}	N_{g2}	. . .	$N_{g,g-1}$	$N_{g,g}$	N_g

Where, Π_i (i, j=1, 2, . . .,g) are the groups

N_{ii} = number of Π_i items correctly classified as Π_i items

N_{ij}= number of Π_i items misclassified as Π_j items The apparent error rate is given by:

$$\text{APER}=1-\frac{N_{11}+N_{22}+\ldots+N_{gg}}{N}$$

or, in other words, the proportion of items in the training sample that are misclassified, where $N = N_1 + N_2 + \ldots + N_g$. The method which has least APER is considered to be best method for that situation.

2.5 Simulation of Experiment

The skew-normal data was generated by using the multivariate skew normal distribution, a member of a new family of asymmetric normal distributions (Azzalini 1985, 1996 and 2005). These classes of distributions also include multivariate normal distribution, when the skewness component is null. The simulation is programmed in R (*v* 2.13.0) and a program was written to generate skew-normal random number for different combination of sample size, skewness levels and dimensions. The following algorithm is used to generate skew-normal data.

[I] The Algorithm

1. Generate $z^* = (Z_0\ Z_1\ Z_2\ \ldots\ Z_P) \sim N_{p+1}(0, \Omega^*)$.
2. If $Z_0 > 0$, let $z = (Z_1, \ldots, Z_p)'$, otherwise let $z = -(Z_1, \ldots, Z_p)'$. Then,

z is an observation from a *p*-dimensional skew-normal distribution with

$$E(\mathbf{z}) = (2/\pi)^{1/2}\boldsymbol{\delta} \quad \text{and} \quad \text{cov}(\mathbf{z}) = \boldsymbol{\Omega} - (2/\pi)\boldsymbol{\delta}\boldsymbol{\delta}'$$

Let $\mathbf{x} = \mathbf{z} + \boldsymbol{\mu} - (2/\pi)^{1/2}\boldsymbol{\delta}$

As noted in Chapter 3, we then have that x is a multivariate skew-normal vector with

$$E(\mathbf{x}) = \boldsymbol{\mu} \quad \text{and} \quad \text{cov}(\mathbf{x}) = \boldsymbol{\Omega} - (2/\pi)\boldsymbol{\delta}\boldsymbol{\delta}'$$

3. Set sample size *n*, dimension ***p***, **μ**, **β** and Ω.
4. Using [1-3] independently generate multivariate skew-normal observations {xi, i = 1, 2,...,*n*}.

Here we took three levels of the sample size (thirty, sixty and hundred), two levels of dimensions (three and five) and five levels of skewness (-8, -4, 0, 4, 8) as well as their combinations. For each parameter setting the simulation is conducted 100 times independently for each group.

3. RESULTS

The different distances between the groups (D^2) are calculated from the real ricebean data and given in Table 2, which showed that out of three cases, D^2_b is greater than the D^2_M.

The four different methods of classification described in the *methods section* was applied on the simulated data sets generated by using multivariate skewed normal distribution. For the application of oblique axes method of classification, a SAS/IML code was been developed based on Das's method (Das 1988). Further, for the application of LDA, QDA and KNN method, we used the SAS (9.2) and SPSS (16.0) program. The Apparent Error Rates (APER) was calculated for each dataset and the results are averaged over 100simulations and represented in Table 3.1 to 3.6 for various combinations sample size, dimensions and skewness levels.

Table 2. Two distance measures obtained from real ricebean data

Groups	D^2_M	D^2_b
(1, 2)	35.376	33.237
(1, 3)	32.766	47.384
(2, 3)	62.017	70.849

D^2_M, Mahalanobis distance; D^2_b distance obtained from OAM

Table 3.1. APER of classification methods based on n=30, p=3 for different skewness levels

Methods	Beta=-8	Beta=-4	Beta=0	Beta=4	Beta=8
LDA	0.0883	0.0987	0.2107	0.0976	0.0881
QDA	0.5533	0.4578	0.2667	0.401	0.4333
KNN	0.0653	0.0768	0.0895	0.0766	0.0733
OAM	0.0786	0.0843	0.2685	0.0924	0.0833

n = 30, small sample size, *p* = 5, number of characters five, beta, different skewness levels

Table 3.2. APER of classification methods based on n=60, p=3 for different skewness levels

Methods	Beta=-8	Beta=-4	Beta=0	Beta=4	Beta=8
LDA	0.0713	0.0617	0.1456	0.1117	0.0787
QDA	0.2564	0.2447	0.2083	0.2138	0.2316
KNN	0.0615	0.0501	0.0671	0.0501	0.0519
OAM	0.0665	0.0589	0.1946	0.0856	0.0678

n = 60, moderate sample size, *p* = 3, number of characters three, beta, different skewness levels

Table 3.3. APER of classification methods based on n=150, p=3 for different skewness levels

Methods	Beta=-8	Beta=-4	Beta=0	Beta=4	Beta=8
LDA	0.0167	0.0323	0.0667	0.0215	0.0273
QDA	0.0333	0.0434	0.0344	0.0396	0.0367
KNN	0.0051	0.0067	0.0133	0.0192	0.0261
OAM	0.2213	0.1712	0.3613	0.1568	0.1347

n = 150, large sample size, *p* = 3, number of characters three, beta, different skewness levels

Table 3.4. APER of classification methods based on n=30, p=5 for different skewness levels

Methods	Beta=-8	Beta=-4	Beta=0	Beta=4	Beta=8
LDA	0.0878	0.0831	0.1604	0.0937	0.0871
QDA	0.2473	0.2363	0.2123	0.2245	0.2335
KNN	0.0333	0.0636	0.0667	0.0556	0.0333
OAM	0.0567	0.0756	0.2333	0.0890	0.0654

n = 30, small sample size, *p* = 5, number of characters five, beta, different skewness levels

Table 3.5. APER of classification methods based on n=60, p=5 for different skewness levels

Methods	Beta=-8	Beta=-4	Beta=0	Beta=4	Beta=8
LDA	0.0483	0.0516	0.1346	0.0927	0.03123
QDA	0.1976	0.1897	0.1827	0.2167	0.2447
KNN	0.0334	0.0317	0.0167	0.0325	0.0159
OAM	0.0434	0.0447	0.2007	0.0734	0.0217

n = 30, moderate sample size, *p* = 5, number of characters five, beta, different skewness levels

Table 3.5. APER of classification methods based on n=150, p=5 for different skewness levels

Methods	Beta=-8	Beta=-4	Beta=0	Beta=4	Beta=8
LDA	0.0143	0.0123	0.0278	0.0139	0.0112
QDA	0.0203	0.0203	0.0111	0.0203	0.0203
KNN	0.002	0.0108	0.0108	0.0108	0.0025
OAM	0.1568	0.0924	0.313	0.1342	0.1256

n = 150, large sample size, *p* = 5, number of characters five, beta, different skewness levels

To study the effect of sample size, dimensions and various skewness levels on APER of classification methods, we used analysis of variance (ANOVA) technique. The ANOVA was performed by considering the mean APER (defined as the mean value of classification error rate over 100 simulations) as dependent variable and sample size, levels of skewness and dimension as independent variable for each classification method. To meet the assumptions of ANOVA, the mean APER is transformed to normal scores by first ranking the mean APER values and then applying Bloom's transformation (implemented in SPSS). The obtained results are represented in tabular form and given in Table 4.1. 4.4.

Further, to study the behaviour of the classification methods in real crop data scenarios, we applied them in the ricebean data and the APER are calculated for each classification method by using confusion matrix (Table 1) described in methods section.

Table 4.1. Analysis of Variance for LDA

Sources of Variation	Degrees of Freedom	Sum of Square	Mean Square	F-Value	Pr (>F)	Significant
Dimension	1	0.0054	0.0054	6.83	< 0.01	* *
Sample Size	2	0.0387	0.019	51.98	< 0.01	* *
Skewness	4	0.0206	0.005	13.78	< 0.01	* *
Error	22	0.0083	0.00038			
Total	29					

Table 4.2. Analysis of Variance for QDA

Sources of Variation	Degrees of Freedom	Sum of Square	Mean Square	F-Value	Pr (>F)	Significant
Dimension	1	0.049	0.049	13.36	< 0.01	* *
Sample Size	2	0.4624	0.2312	63.03	< 0.01	* *
Skewness	4	0.0138	0.0034	6.94	< 0.01	* *
Error	22	0.08071				
Total	29	0.6060				

Table 4.3. Analysis of Variance for KNN

Sources of Variation	Degrees of Freedom	Sum of Square	Mean Square	F-Value	Pr (>F)	Significant
Dimension	1	0.0032	0.0032	27.93	< 0.01	* *
Sample Size	2	0.014	0.0069	59.77	< 0.01	* *
Skewness	4	0.0005	0.0001	1.09		NS
Error	22	0.026				
Total	29					

Table 4.4. Analysis of Variance for OAM

Source of Variation	Degrees of Freedom	Sum of Square	Mean Square	F-Value	Pr (>F)	Significant
Dimension	1	0.0053	0.0053	10.67	< 0.01	* *
Sample Size	2	0.054	0.027	54.75	< 0.01	* *
Skewness	4	0.1325	0.03312	66.32	< 0.01	* *
Error	22					
Total	29					

** represents values significant at 1% level of significance and NS represents not significant.

4. DISCUSSION

The Table 3.1 summarised the APER of four classification techniques when the sample size is thirty and dimension is five. The result showed that for small sample size the performance of LDA is affected due to introduction of skewness in the data. Under normality condition, the APER for LDA is0.1604 and as skewness increased the APER of LDA decreaseed, which is supported by the results given by Server *et al.* (2005). The result of QDA is that reverse of LDA. For QDA, the APER under normal condition is less than that of Skew normal condition *i.e.* when beta is zero (normal), the APER is 0.2123 and for beta is eight (highly skewed), the APER is 0.2335. Hence error rate increased with increase in beta value and this is true for negative beta also. For KNN method the APER is 0.0667, when beta is zero and 0.0333 in case beta is eight. When beta changes from zero to eight, the APER decreases, and this rate of change is gradual. The OAM has error 0.2333 in case beta is zero and when beta is eight the error rate is 0.0654. Hence the APER decreased from 0.2333 to 0.0654, when beta increased from zero to eight and similar result could be seen when beta decreased from zero to minus eight. Similar interpretation could be made for the effect of skewness levels on classification methods for different combination sample size and dimensions. To study the behaviour of classification techniques under varying condition of sample size can be well interpreted from the the Figs. 1.1, 1.3 and 1.5. Fig. 1.1 depicted that, when the sample size is thirty, the APER of LDA is given by 0.1604 (for beta is zero) and for sample size is hundred fifty, the APER is 0.0178. For LDA, the robustness increases with increase in sample size from 30 to 150. Similar type of trend is experienced for QDA, OAM and KNN methods for other levels of skewness (beta= -8, -4, 4 and 8). Fig. 2.2 showed the behaviour of QDA for different sample size, the APER is maximum (0.2123) when the n=30 and minimum (0.0111) when n=150. Similar type of pattern can be seen for all other levels of beta. For QDA, the APER decreased with increase in sample size. For large sample size (n=150), under normal condition (for beta is zero) the APER of QDA (0.0111) is less than that of LDA (0.0278). The results are motivated by the valid statement of Wahl and Kronomal (1977).Similar interpretation can be made for KNN method of classification from the Fig. 2.3. The result obtained from OAM (Fig. 2.4) is quite different from all other methods *i.e.* for OAM, the APER is 0.2333, when the sample size is thirty, 0.2007 (when n=60) and 0.3131 (n=150). The above result showed that when sample size changes from 30 to 60, the APER decreased from 0.2333 to 0.2007 and APER increased from 0.2007 to 0.3132 when sample size increased from 60 to 150. Similar interpretation can be made for the dimension equal to three with different skewness levels and sample sizes. Using ANOVA, the effect of sample size, dimension and skewness levels on the performance of classification methods was studied. The result showed that sample size has significant effect on the performance of each classification techniques. The result is as expected regarding the fact that with increase in sample size there is decrease in APER of classification methods. Classification performance for all the methods is also affected by dimension and the increase in dimension resulting in lower APER. The skewness had significant effect on APER but it has no effect on the performance of KNN method of classification. Further application of classification methods, namely LDA, QDA, KNN and OAM to the real ricebean data case yielded APER 0.1604, 0.3217, 0.0687 and 0.1221 respectively. The low APER of KNN technique made it efficient followed by OAM and LDA.

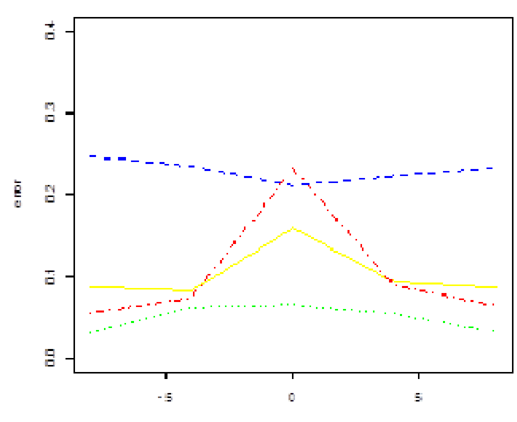

Fig. 1.1. Error Plot for n=30 and p=5

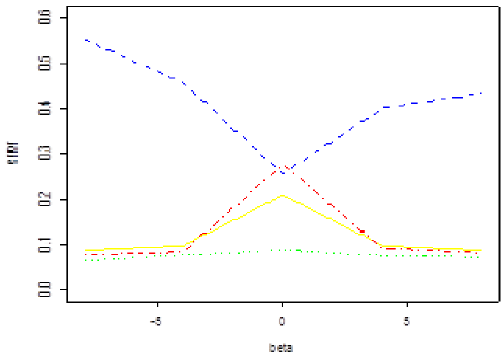

Fig. 1.2. Error Plot for n=30 and p=3

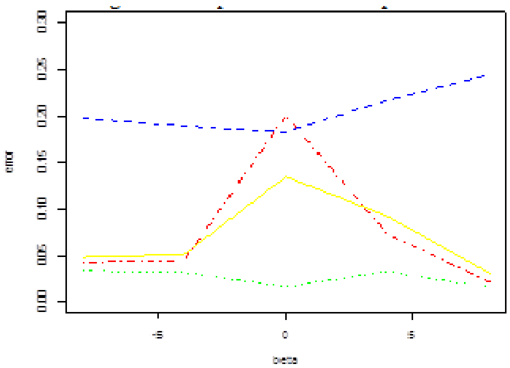

Fig. 1.3. Error Plot for n=60 and p=5

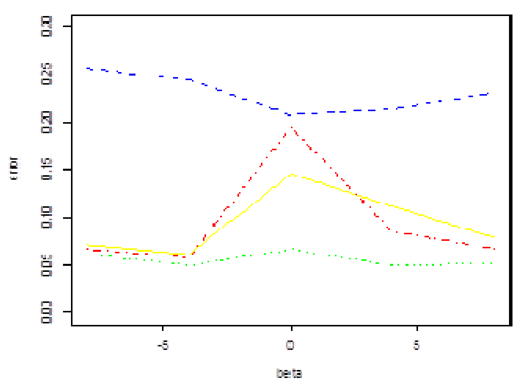

Fig. 1.4. Error Plot for n=60, p=5

Here, "Yellow line" stands for LDA, "Red dotted line" represented OAM, "Green dotted line" for KNN, "Blue dotted line" for QDA.

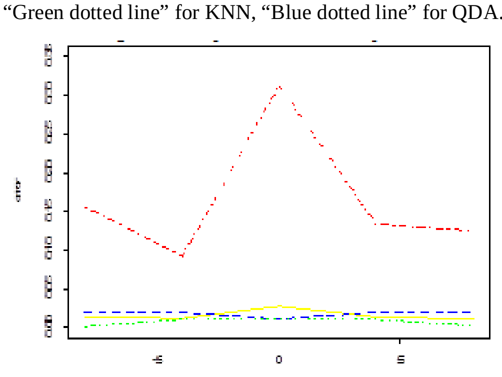

Fig. 1.5. Error Plot for n=150 and p=5

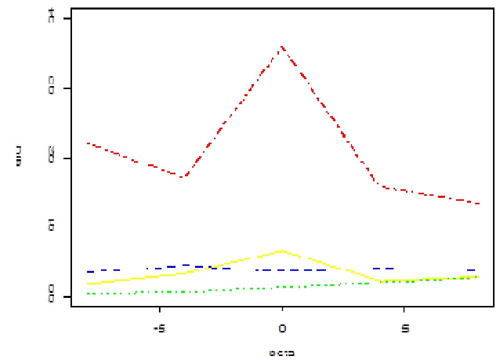

Fig. 1.6. Error Plot for n=150, p=3

Fig. 1.1-1.6. Performance of four Different Classification Techniques for Different Combinations of Sample Size and Dimensions

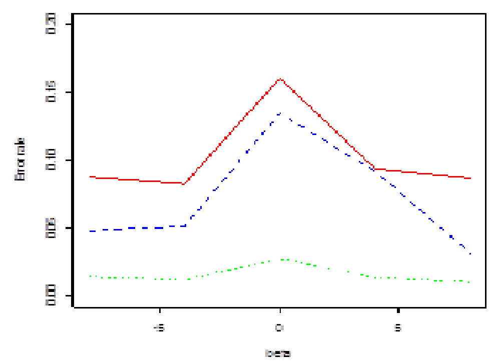

Fig. 2.1. Error Plot for LDA

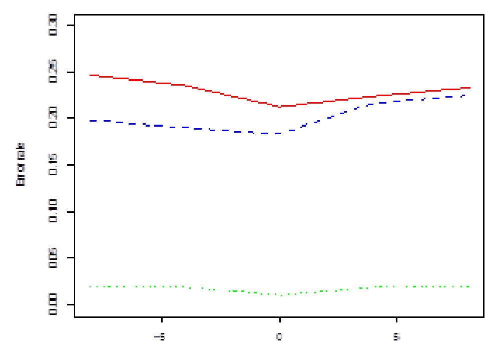

Fig. 2.2. Error Plot for QDA

Fig. 2.3. Error Plot for KNN

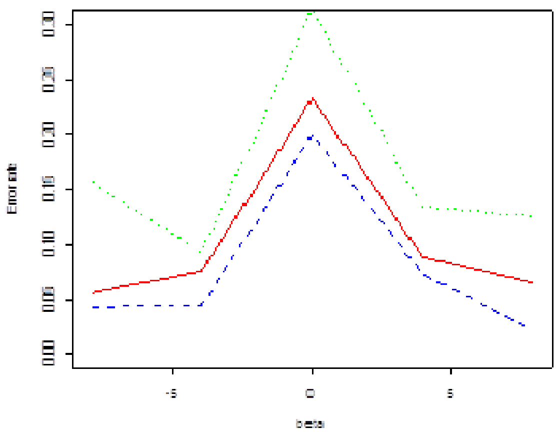

Fig. 2.4. Error Plot for QAM

Fig. 2.1-2.4. Performance of Classification Methods over Different Sample Sizes

Here, "red line" represented APER of classification techniques for n=30; "blue line' for n=60 and "green line" for n=150.

5. CONCLUSION

The purpose of the experiment described in this paper was to examine and compare the behaviour of four classification techniques (LDA, QDA, KNN, OAM) to sample size, dimension and skewness introduced to data by multivariate skew-normal distribution. From the obtained results, it can be concluded that the LDA is more robust to skew-normal data than normality condition and its error rate decreases with increase in sample size. QDA has more efficient in normal condition as compared to skew-normal situations and its error decreases with increase in sample size. For KNN method of classification, the APER decreases with increase in sample size and dimension. For OAM, the result is quite similar to LDA, as the APER in case of optimal situation is more than that of non-optimal (skewed) condition and its APER first decreases, when sample size changes from small to moderate and error rate is large when the size is large. The result is expected to be, due to the effect of central limit theorem. Among LDA and QDA, the APER of QDA is more than that of LDA in all skewness levels and sample sizes, hence LDA is considered to be more efficient than QDA. By comparing among LDA, QDA and OAM, the OAM has least APER. So, OAM is considered to best method for classification than LDA and QDA (except large sample). Further, KNN method provides the best result for classification with least APER among all the three methods of classification in case of simulated skew-normal data introduced by multivariate skew-normal distribution and as well as for the original data.

REFERENCES

Anderson, T.W. (2003). *An Introduction to Multivariate Statistical Analysis*, John Wiley and Sons, Inc., New York.

Azzalini, A. (2005). The Skew-Normal Distribution and Related Multivariate Families. *Scandinavian Journal of Statistics, Theory and Applications*, **32(2)**, 159-163.

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scand. J. Statist. Theo. Appl*, **12(2)**, 171-178.

Azzalini, A. and Dalla Valle, A. (1996).The Multivariate Skew-Normal Distribution. *Biometrika*, **83(4)**, 715-719.

Balbin, R.M., Safieva, R.Z. and Lomakina, E.I. (2010).Gasoline classification using near infrared spectroscopy data. *Analytica Chimica Acta*, **671**, 27-35.

Bobrowski, L. (1986). Linear discrimination with symmetrical models. *Pattern. Recogn.*, **19(1)**, 101−109.

Das, M.N. (1998). Classification of observations using distances using distances in oblique axes system. *J. Ind. Soc. Ag. Statist.*, **51**, 379-384.

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. Phy. Sci.*, **4(1)**, 16-21.

Genton, M. G. (2004). *Skew-elliptical distributions and their applications: a journey beyond normality*. Chapman & Hall/CRC.

Glele, K.R., Pelz, D., Palm, R. (2010). The efficiency of linear classification rule in multi-group discriminant analysis. *Afr. Jour. Math. Comp. Sci. Res.*, **3(1)**, 19-25.

Hubert, M., Van der Veeken, S. (2010). Robust classification for skewed data.*Adv. Data Anal.Classif.*, **4(4)**, 239−254.

Johnson, R.A. and Wichern, D.W. (2007).*Applied Multivariate Statistical Analysis*, 6th ed., Prentice Hall, Inc., New York.

Kiang, M.Y. (2003).A comparative assessment of classification methods. *Decision Support Sys.*, **35**, 441-454.

Mardia, K.V. (1980). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, **57**, 519-530.

McLachlan, G.J. (1992). *Discriminant Analysis and Statistical Pattern Recognition*. John Wiley and Sons, Inc., New York.

Pohar, M., Blas, M. and Turk, S. (2004). Comparison of logistic regression and linear discriminant analysis: A simulation study. *Metodoloskizvezki*, **1**, 143-161.

Rao, C.R. (1946). Tests with discriminant function in multivariate analysis. *Sankhya*, **7**, 407-410.

Rausch, R.J. and Kelly, A. (2009).A comparison of linear and non-linear models for discriminant analysis under non-normality.*Behave. Res. Methods*, **41(1)**, 85-98.

Sever, M., Lajovic, J. and Rajer, B. (2005). Robustness of Fisher's discriminant function to skewed normal distribution. *Metodoliskizvezki*, **2(2)**, 231-242.

Soni, S. and Shrivastava, S. (2010).Classification of Indian stock market data using machine learning algorithms. *Int. Jour. on comp. Sci. and Eng*., **2(9)**, 2942-2946.

Wahi, S.D., Narain, P. and Garg, L.K. (1986).Use of discriminant function for comparing different grades in crossbreeding programme with Sheep. IASRI Publication, New Delhi.

Wahi, S.D. and Bhatia, V.K. (1995). Use of bootstrap method in comparing the performance of linear discriminant function. *J. Ind. Soc. Ag. Statist.*, **47(1)**, 12-20.

Wahl, P.W. and Kronmal, R.A. (1977). Discriminant function when covariances are unequal and sample sizes are moderate. *Biometrics*, **33**, 479-484.

Wu, X., Kumar, V., Quinlan, J.R., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G.J., Liu, B., Yu, P.S. and Hand, D.J. (2008). Top 10 algorithims in data mining. *Know. Info. Sys.*, **14**, 1-37.