Journal of Natural Sciences Research www.iiste.org
ISSN 2224-3186 (Paper) ISSN 2225-0921 (Online)
Vol.6, No.1, 2016

Bootstrap Method for Estimating Error Rate in Linear Discriminant Analysis (LDA)

Obiora-Ilouno Happiness O Nwoke Chidinma B. Uzuke C.A
Department of Statistics, Nnamdi Azikiwe University, Awka Anambra State.

Abstract
This paper proposes a Bootstrap algorithm for linear discriminant analysis. The apparent error rate in the Linear discriminant method and the proposed bootstrap method were compared. From the result, it is evident that the proposed Bootstrap method compares favorably with the Linear discriminant method with a reduced error rate.
Keywords: bootstrap, linear discriminant, apparent error, multivariate, algorithm

1.0 Introduction

Linear Discriminant Analysis is a multivariate method of finding a linear combination of variables which best separates two or more classes. It is used when dealing with continuous independent variables and a categorical dependent variable. The performance of a discriminant criterion could be evaluated by estimating the probabilities of misclassification of observation.

According to Efron (1979), the Bootstrap method is a non-parametric technique that resamples the original data. The idea behind Bootstrap is to use the data of a sample study for the purpose of approximating the sampling distribution of a statistic.

Linear discriminant analysis is used to discriminate and classify set of data as long as the data involved do not have linear dependencies and are drawn from a multivariate normal distribution and the groups have the same covariance matrix. Therefore, this work intends to compare the Bootstrap method and the Linear Discriminant method to identify which of the method performs better based on their error rate.

2.0 Materials and Method

2.1 Method

The classification rule is to assign an object to the group with highest conditional probability (that is the Bayes rule). Let C and D represent two groups, the Bayes rule is to assign the object to group C if $P(C) > P(D)$ for all $C \neq D$. According to Teknomo (2006) the probability $P(C/X)$ that an observation belongs to group C, given a set of the independent variables X, and, the probability $P(D/X)$ that an observation belongs to group D given a set of independent variables X are

$$P(C/X) = \frac{P(X/C).P(C)}{P(X/C).P(C) + P(X/D).P(D)}$$

$$P(D/X) = \frac{P(X/D).P(D)}{P(X/C).P(C) + P(X/D).P(D)} \qquad 1$$

Where,
X is a matrix of the set of independent variables.
$P(C/X)$ the probability that an observation belongs to group C given a set of independent variables X
$P(X/C)$ the probability of getting a particular set of independent variables X given the observation comes from group C
$P(C)$ is the prior probability about group 'C', therefore, $P(C) = \frac{n_C}{N}$
$P(D/X)$ the probability that an observation belongs to group D given a set of independent variables X
$P(D)$ is the prior probability about group 'D', therefore, $P(D) = \frac{n_D}{N}$
$P(X/D)$ the probability of getting a particular set of independent variables X given that the observation comes from group D.
By Bayes rule, assign observations to group C if

$$\frac{P(X/C).P(C)}{P(X/C).P(C) + P(X/D).P(D)} > \frac{P(X/D).P(D)}{P(X/C).P(C) + P(X/D).P(D)} \text{ for all } C \neq D \qquad 2$$

simplifying

$$P(X/C).P(C) > P(X/D).P(D) \ for\ all\ C \neq D \qquad 3$$

Using the assumption that in LDA the data comes from multivariate normal distribution whose probability density function (PDF) is given by

$$\left.\begin{aligned} P(X/C) &= \left(\frac{1}{(2\pi)^{p/2}|C_C|^{1/2}}\right)\exp\left(-\frac{1}{2}(X-\mu_C)'C_C^{-1}(X-\mu_C)\right) \\ P(X/D) &= \left(\frac{1}{(2\pi)^{p/2}|C_D|^{1/2}}\right)\exp\left(-\frac{1}{2}(X-\mu_D)'C_D^{-1}(X-\mu_D)\right) \end{aligned}\right\} \quad 4$$

Where μ_C is the vector mean, C_C is the covariance matrix of group C, μ_D is the vector mean, C_D is the covariance matrix of group D and p is the number of independent variables

Substituting $P(X/C)$ and $P(X/D)$ into the inequality 3 we have

Assign observations k to group C if

$$\left(\frac{P(C)}{(2\pi)^{p/2}|C_C|^{1/2}}\right)\exp\left(-\frac{1}{2}(X-\mu_C)'C_C^{-1}(X-\mu_C)\right) > \left(\frac{P(D)}{(2\pi)^{p/2}|C_D|^{1/2}}\right)\exp\left(-\frac{1}{2}(X-\mu_D)'C_D^{-1}(X-\mu_D)\right) \text{ for all } C \neq D \quad 5$$

Simplifying both sides of (5), we obtain

$$\left(\frac{P(C)}{|C_C|^{1/2}}\right)\text{Exp}\left(-\frac{1}{2}(X-\mu_C)'C_C^{-1}(X-\mu_C)\right) > \left(\frac{P(D)}{|C_D|^{1/2}}\right)\exp\left(-\frac{1}{2}(X-\mu_D)'C_D^{-1}(X-\mu_D)\right) \text{ for all } C \neq D \quad 6$$

Taking log of both sides of (6)

$$-\frac{1}{2}\ln(|C_C|) + \ln(P(C)) - \frac{1}{2}(X-\mu_C)'C_C^{-1}(X-\mu_C) > -\frac{1}{2}\ln(|C_D|) + \ln(P(D)) - \frac{1}{2}(X-\mu_D)'C_D^{-1}(X-\mu_D) \text{ for all } C \neq D \quad 7$$

Since all covariance matrices are the same in LDA $C = C_i = C_j$,

$(X-\mu_C)'C_C^{-1}(X-\mu_C) = XC^{-1}X - 2\mu_C C^{-1}X + \mu_C C^{-1}\mu_C$ and

$(X-\mu_D)'C_D^{-1}(X-\mu_D) = XC^{-1}X - 2\mu_D C^{-1}X + \mu_D C^{-1}\mu_D$

inequality 7 becomes

$$-\frac{1}{2}\ln(|C|) + \ln(P(C)) - \frac{1}{2}XC^{-1}X + \mu_C C^{-1}X - \frac{1}{2}\mu_C C^{-1}\mu_C > -\frac{1}{2}\ln(|C|) + \ln(P(D)) - \frac{1}{2}XC^{-1}X + \mu_D C^{-1}X - \frac{1}{2}\mu_D C^{-1}\mu_D \quad \forall\, C \neq D \quad 8$$

hence from 8 we have

$$\ln(P(C)) + \mu_C C^{-1}X - \frac{1}{2}\mu_C C^{-1}\mu_C > \ln(P(D)) + \mu_D C^{-1}X - \frac{1}{2}\mu_D C^{-1}\mu_D \quad \forall\, C \neq D \quad 9$$

Let $\boldsymbol{f_C = \ln(P(C)) + \mu_C C^{-1}X - \frac{1}{2}\mu_C C^{-1}\mu_C}$ and

$\boldsymbol{f_D = \ln(P(D)) + \mu_D C^{-1}X - \frac{1}{2}\mu_D C^{-1}\mu_D}$ (Teknomo 2006)

We assign an individual to group "C" if $f_C > f_D$ and to group "D" if $f_C < f_D$

2.2 Proposed Bootstrap Algorithms for Estimating Error Rate in Discriminant Analysis

Let n sized sample $Z_k = (Y \quad X_{ji})$ where Y (n x 1) is a column vector containing the groups and $X_{ij} = (x_{1j}, x_{2j}, \dots, x_{nj})'$ is a matrix of dimension n× p where i= 1, 2 ,…, n and j= 1,2,…, p,

1. Draw a sample $(z_1^{(b)} z_2^{(b)}, \dots, z_n^{(b)})$ with replacement from the original sample, with $\frac{1}{n}$ probability of sampling Z_i , label the element of each vector $z_k^{(b)} = (y \quad x_{ij}^{(b)})'$ and obtain the matrix $X_{ij}{}^{(b)} = (x_{1j}{}^{(b)}, x_{2j}{}^{(b)}, \dots, x_{nj}{}^{(b)})'$ where i= 1,2,…, n and j=1,2,…, p
2. Obtain the matrix for the independent variables say X matrix of dimension n × p and Y (n×1) column vector for groups of observation, and partition the matrix X into X_C and X_D the number of groups available, p is the number of independent variables and n is the total number of observations for the groups combined.
3. Compute the various means for each predictor variable for both groups and then obtain the Bootstrap group means which is given as $\frac{\sum_{i=1}^{B}\mu}{B}$ (Obiora-ilouno and Mbegbu(2012)).
4. Computing the discriminant function using the Bayes criterion for classification given as $f_C = \mu_C C^{-1}X_k^{'} - \frac{1}{2}\mu_C C^{-1}\mu_C{}^{'} + \ln(P(C))$ and $f_D = \mu_D C^{-1}X_k^{'} - \frac{1}{2}\mu_D C^{-1}\mu_D{}^{'} + \ln(P(D))$

5. Classifying the observation in the sample with the Bayes classification rule we obtain probability of misclassification $\frac{m_C}{n_{C.}}$ and $\frac{m_D}{n_{D.}}$ and Error Rate (APER) $\frac{m_C+m_D}{n}$
6. Repeating steps 1 - 5 r times (r = 1, 2, …, B) where B is the number of repetition.
7. Compute the mean of all Bootstrap Error Rate obtained from the Bootstrap samples,
8. $\frac{\sum_{r=1}^{B}\left(\frac{m_C+m_D}{n}\right)}{B} = \hat{\beta}$ which is the Bootstrap Error Rate.

2.3 Data Collection

The data used to implement this algorithm is a Secondary data collected from the pre-science unit of (Nnamdi Azikiwe University, Awka). The number of students admitted and not admitted into the University through the Pre-science programme were used. The student's individual scores in the University Tertiary Matriculation Examination (UTME) and their corresponding scores in Pre-Science examinations were considered as independent variable in order to determine true classification of students admitted into the university. R code was used for the statistical analysis of these data

Using the Linear Discriminant Analysis to classify students into their various groups of 'admitted' or 'Not admitted'

Using the Linear Discriminant Analysis to classify students into their various groups of 'admitted' or 'Not admitted'
Let Y_i be groups which assumes the values either 1 (admitted) or 2 (Not admitted) for i= 1,2, …, n

$$\text{Let } Y_i = \begin{cases} 1, \text{admitted} \\ 2, \text{Not admitted} \end{cases}$$

Let X_1, X_2, X_3, X_4, X_5 be independent variables namely UTME score and the four highest scores from five different subjects. Data obtained are shown in appendix 1

3.0 Results and Discussion.

3.1. Results

The tables below shows the results gotten using the R code using the data in Appendix I which is the student scores in UTME and the four highest scores from five different subject taken by 50 students

Table 2: Table showing Group means for each of the independent variables for the analytical method

	UTME	**Subject 1**	**Subject 2**	**Subject 3**	**Subject 4**
1	207.421	58.474	64.947	84.263	66.737
2	180.161	26.000	23.516	19.677	36.774

Table 3: Table showing the Coefficients of linear discriminant function

UTME	0.0017522282
Subject 1	-0.0302208034
Subject 2	-0.0007046323
Subject 3	-0.0890596209
Subject 4	0.0137982325

Table 4: Table showing the true classification of students

Confusion Matrix

	PREDICTED	
ACTUAL	ADMITTED	NOT ADMITTED
ADMITTED	19	1
NOT ADMITTED	0	30
TOTAL	19	31

From the Confusion Matrix in Table 4 the Apparent Error Rate for The LDA is 0.02 and the Percentage of correctly classified (PCC) for LDA 98%

Estimating the Bootstrap error rate for 100 Bootstrap samples each of size n = 50 was used as shown in Table 5 below.

Table 5: Group means for the 100 Bootstrap Samples

	UTME	**SUBJECT 1**	**SUBJECT 2**	**SUBJECT 3**	**SUBJECT 4**
1	207.9474	56.57895	66.05263	82.57895	60.94737
2	178.4839	27.93548	24.80645	18.22581	40.83871

Table 6: Coefficients of linear discriminant function

UTME	0.003145706
SUBJECT 1	-0.041049879
SUBJECT 2	-0.024705002
SUBJECT 3	-0.139888570
SUBJECT 4	0.047445213

Table 7: The probabilities of Correctly Classified (PCC) observation for the 100 Bootstrap sample**.**

Bootstrap samples	Probability of correctly classified observation						
1 - 7	0.930	0.930	0.943	0.917	0.903	0.927	0.910
8 - 14	0.910	0.920	0.927	0.920	0.923	0.923	0.940
15 - 91	. . .	. . .	. . .	. . .	. . .	. . .	. . .
92 - 98	0.937	0.890	0.913	0.910	0.920	0.940	0.930
99 - 100	0.933	0.900					

The mean of the 100 Bootstrap samples which is the percentage of correctly classified observation is 0.983 or 98.3%

Bootstrap Error Rate = $1 - \frac{\sum_{i=1}^{B} X_i}{B} \Rightarrow 1 - 0.983$

Therefore, Bootstrap error rate is 0.017

Table 11: Summary of the coefficient of linear discriminant function for Normal Linear discriminant function and Bootstrap Linear discriminant function (n=50, B=100 &1000)

Variables	**Normal Linear discriminant function**	**Bootstrap values for B=100**	**Bootstrap values for B=1000**
UTME	-0.0151	0.0031	0.0028
SUBJECT 1	-0.0089	-0.0410	0.0068
SUBJECT 2	-0.0254	-0.0247	-0.0332
SUBJECT 3	-0.0486	-0.1399	-0.1934
SUBJECT 4	-0.0112	0.0474	0.0225

Table 12 shows the summary of error rates obtained from the Linear discriminant method and proposed bootstrap method with bootstrap samples of 100 and 1000 respectively.

Table 12: Comparison of Error Rates

Linear discriminant method APER	**Bootstrap Error Rate**	
Linear discriminant	**B=100**	**B=1000**
0.020	0.017	0.019

3.2 Discussion

From the result of the analysis, the analytical method has an error rate of 0.02, the bootstrap error rate for B=100 and 1000 yielded an error rate of 0.017 and 0.019 respectively, indicating that the Bootstrap error compared favorably with the analytical method with a reduced error than the analytical method.

4.0 Conclusions

Bootstrap method for estimating the error rate using the linear discriminant analysis has been proposed in the paper. The results obtained as shown in Tables 11 and 12 indicates that the Bootstrap methods produced smaller error rate indicating that the Bootstrap algorithm proposed yielded a better reduced error rate.

References

Chernick, M.R (2008), Bootstrap Methods; A Guide For Practitioners And Researchers, Second Edition, Wiley, Hoboken.

Efron B. (1979), Bootstrap Methods, Another Look at the Jackknife, the Annals of Statistics, Vol. 7, pp 1-26

Obiora-Ilouno H.O & Mbegbu J. I (2012) Bootstrap algorithm for the estimation of logistic regression parameters,Journal of Nigerian Statistical Association, Vol.24, pp10-19

Sahinler S. And Topuz D (2007), Bootstrap And Jackknife Resampling Algorithm For Estimation Of Regression Parameters, Journal Of Applied Quantitative Method, Vol. 2, No. 2, pp 188-199.

Teknomo, K. (2006), Discriminant Analysis Tutorials. http://people.revoledu.com/kardi/tutorial/LDA/.

APPENDIX I

S/N	Group	Jamb	A	B	C	D
1	1	211	48	60	90	54
2	1	193	55	48	87	54
3	1	196	73	52	85	74
4	1	218	48	58	88	57
5	1	222	63	82	83	74
6	1	186	51	56	80	56
7	1	251	59	75	81	60
8	1	209	50	80	89	68
9	1	199	55	64	63	48
10	1	214	69	84	94	77
11	1	196	71	70	90	92
12	1	195	60	80	88	72
13	1	209	57	46	87	67
14	1	182	69	84	86	70
15	1	204	48	54	84	61
16	1	231	58	54	83	69
17	1	205	67	73	81	76
18	1	237	63	40	80	60
19	1	183	47	74	82	79
20	2	168	43	60	77	53
21	2	262	37	35	42	57
22	2	200	19	28	23	35
23	2	155	28	18	23	39
24	2	200	53	45	22	61
25	2	183	23	15	22	63
26	2	219	25	19	20	37
27	2	180	13	16	19	22
28	2	128	33	33	19	22
29	2	161	26	20	19	31
30	2	199	33	20	19	37
31	2	191	14	14	18	26
32	2	231	17	14	18	28
33	2	175	25	19	18	30
34	2	169	20	06	18	36
35	2	201	31	21	18	37
36	2	129	28	39	18	44
37	2	176	36	45	18	45
38	2	170	24	38	18	49
39	2	183	10	20	17	22
40	2	167	17	22	17	26
41	2	173	36	26	17	28
42	2	218	20	16	17	33
43	2	171	25	04	17	35
44	2	164	10	08	12	24
45	2	153	24	14	12	31
46	2	172	41	18	12	33
47	2	154	39	37	12	50
48	2	160	13	11	12	27
49	2	156	11	22	08	31
50	2	217	32	26	08	48

APPENDIX II

The Computer Program in R For the Bootstrap

```
DATAb=read.table("data.txt", header=TRUE)
NewData <-matrix(0,50,6)
for(j in 1:6)
  {
   NewData[,j] <- DATAb[,j]
}
nsampl<-50
B = 100
boot.samples<-array(rep(0), dim=c(nsampl,6,B) )
for( i in 1:B){
 SG1<-c(sample(1:19, 19 , replace=TRUE))
 SG2<-c(sample(20:50,31, replace=TRUE))
 boot.samples[,,i] = (rbind( NewData[SG1,],NewData[SG2,] ))
}
NewData = as.data.frame(NewData)
cf = rep(0,B)
for(i in 1:B)
  {  Cla <-lda(boot.samples[,1,i]~ boot.samples[,2,i] + boot.samples[,3,i] + boot.samples[,4,i] + boot.samples[,5,i]
        + boot.samples[,6,i] )
DATALL<-predict(Cla,newdata=NewData[, 2:6])$class
tab=table(DATALL,NewData[,1])
cf[i] = (tab[1,1] + tab[2,2])/50 }
cf
BER=1-mean(cf)
BER
```

The computer program in R for the Linear Discriminant Analysis

```
rm(list=ls(all=TRUE))
library(MASS)
DATA<-read.table("r.txt", header=TRUE)
head(DATA)
plot(DATA[,c(2,3,4,5,6)],col=DATA[,1])

DATAL<-lda(Group~Jamb+A+B+C+D,data=DATA)
DATAL
DATALL<-predict(DATAL, newdata=DATA[,c(2,3,4,5,6)])$class
tab=table(DATALL,DATA[,1])
dimnames(tab)<-list(Actual=c("Admitted","Notadmitted"), Predicted=c("Admitted","Notadmitted"))
#con<-rbind(tab[1,]/sum(tab[1,]),tab[2,]/sum(tab[2,]))
#dimnames(con)<-list(Actual=c("Admitted","Notadmitted"), Predicted=c("Admitted","Notadmitted"))
print(round(tab,3))
N=sum(tab[1,])+sum(tab[2,])
APER=(tab[1,2]+tab[2,1])/N
APER
PCC=((tab[1,1]+tab[2,2])/N)*100
PCC
```