# A COMPARISON BETWEEN MAXIMUM LIKELIHOOD RULE AND LOGISTIC DISCRIMINANT ANALYSIS IN THE CLASSIFICATION OF MIXTURE OF DISCRETE AND CONTINOUS VARIABLES

Iwuagwu Chukwuma E[1]. Onyeagu Sidney I.[2], Chrisogonus K. Onyekwere[3]

[1]Department of Statistics, Abia State Polytechnic Aba, Nigeria

[2,3]Department of Statistics, Nnamdi Azikiwe University Awka, Nigeria

**Corresponding author:** E-mail: chrisogonusjohnson@gmail.com

**ABSTRACT**

An optimal measure of performance is the one that lead to maximization of average error rate or probability of misclassification. This paper aimed to compare between the maximum likelihood rule and logistic discriminant analysis in the classification of mixture of discrete and continuous variables. The efficiency of the methods was tested using simulated and real dataset. The result obtained showed that the maximum likelihood rule performed better than the logistic discriminant analyses, in maximizing the average error rate in both experiment conducted.

**Keyword:** Maximum likelihood rule, Logistic discriminants, error rate, Likelihood ratio, Discriminant analysis.

## 1 Introduction

The dependent variable in regression analysis is assumed to be continuous and normally distributed. Situations arise when the dependent variable is discrete and normal. In this situation, regression analysis becomes invalid. Two methods can be employed, one is discriminant analysis, and the other is logistic method. Discriminant analysis demands that the data be normal and discrete. But when this assumption is violated, the assumption free distribution (logistic regression) can be used. Hamid (2010) describe discriminant analysis as the advancement of a rule for allocating objects into one of some different groups and then the constructed classification rule will be used to determine a group of some future objects. Nocairi, Qannari & Hanafi (2006) developed a simple regularization for discriminant analysis. It is a technique which consists of estimating the covariance matrix within each group by a convex combination of the usual estimate of the covariance matrix and the identity matrix. The coefficient involved in this combination is adjusted to each particular state by means of a cross-validation procedure that targets at minimizing the cross-validated misclassification risk. Leon and Zhu (2007) studied the problem of testing for differences among several groups with correlated mixed data. They adopted the General Location Model (GLM), for the joint distribution of the mixed data and derived a Likelihood Ratio Test (LRT), for comparing the location of mixed-variate population. Kakio, *et al*. (2009) expressed that Logistic Regression (LR) is one of the simplest models for binary classification and can directly estimate the posterior probabilities.

Multinominal Logistic Regression (MLR) is a natural extension of LR to multi-class classification problems. It is known as one of the Generalized Linear Model (GLM) which is a flexible generalization of ordinary least squares regression. By modifying the outputs of the linear predictor by the link function, MLR can naturally estimate the posterior probabilities.

Statisticians can be misled as a result of shortcomings in methods being used to measure performance. An optimal measure of performance that led to maximization of average error rate or probability of misclassification should indicate the degree of effective success in allocating an individual or other subject into a particular category of a dependent variable, as the best classification rule is the one that leads to the smallest probability of misclassification which is called error rates. The classification rule being used may be showing results that appear satisfactory, but they merely camouflage interior performance. The remedy is the adoption of classification rule which avoid the shortcomings of the conventional ones.

The goal of this paper is to compare the efficacy or performance of the maximum likelihood rule and the logistic discriminant analysis in the classification of mixture of discrete and continuous variables in terms of error rate, as an idea discriminant function should not only separate different classes with a minimum misclassification rate for the training set, but possess a good stability such that the predictive variance for unclassified object can be as small as possible.

## 2 Methodology

Data for the research was collected through simulated and real data, a data set of 200 was generated with R-Programming and the average error rates were obtained for $2 \leq q \leq 10$ and $0.1 \leq P_1 P_2 \leq 0.9$ for two situations (a) a case with no interaction between discrete and continuous and situation (b) a case involving interaction between discrete and continuous variables, also q are the components of the discrete variables x and p are the components of the continuous variables, y.

The real data were obtained from primary and secondary data. The primary data was from project implementation while the secondary data came from Human development index.

### 2.2 Maximum Likelihood Rule

Fisher (1920) proposed the maximum likelihood rule and the essential feature of the rule is that, we look at the value of a random sample and choose our estimate of the unknown population parameter, the value for which the probability of obtaining the observed data is a maximum, that is value of $\theta$, that maximizes the probability of error. If the observed sample values are $x_1, x_2, \cdots, x_n$ , we write in discrete case

$p(X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n) = f(x_1, x_2, \cdots, x_n : \theta)$, which is the value of the joint probability distribution of the random variables, $x_1, x_2, \cdots, x_n$ while when the random sample comes from a continuous population, we have $f(x_1, x_2, \cdots, x_n : \theta)$ is the value of the joint probability density at the sample point $(x_1, x_2, \cdots, x_n)$.

#### 2.2.1 Maximum Likelihood Estimation on Gaussian Model

According to Mengsay (2020), a Gaussian model is a d-dimension pattern $x$is given by

$$q(x;\mu,\Sigma) = \frac{1}{(2\pi)^{\frac{d}{2}} \det(\Sigma)^{\frac{1}{2}}} \, exp - \frac{1}{2}(x-\mu)^{\intercal}\Sigma^{-1}(x-\mu) \quad (1)$$

Where, $\mu$ and $\Sigma$ are parameters of the model.

In estimating the model, we assume $n$ sample data $(x,i), i = 1,\cdots,n$ and to estimate the parameters, we obtain the likelihood and log-likelihood of the model as

$$L(\mu,\Sigma) = \prod_{i=1}^{n} q(x;\mu,\Sigma)$$

$$q(x;\mu,\Sigma) = \frac{1}{(2\pi)^{\frac{d}{2}} \det(\Sigma)^{\frac{1}{2}}} \, exp - \frac{1}{2}(x-\mu)^{\intercal}\Sigma^{-1}(x_i-\mu) \quad (2)$$

$$Log(\mu,\Sigma) = \frac{nd}{2}\log(2\pi) = \frac{n}{2}\log\{\det(\Sigma) = -\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^{\intercal}\Sigma^{-1}(x_i-\mu) \quad (3)$$

Then, the likelihood equation can be written as

$$\frac{d}{d\mu} Log\ L(\mu,\Sigma)|_{\mu=\hat{\mu}_{mL}} = 0 \quad (4)$$

$$\frac{d}{d\Sigma} Log\ L(\mu,\Sigma)|_{\Sigma=\hat{\Sigma}_{mL}} = 0 \quad (5)$$

Evaluating equations (4) and (5), we obtain the estimated parameters of the Gaussian model as

$$\hat{\mu}_{mL} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad (6)$$

$$\Sigma_{mL}^{n} = \frac{1}{n}\sum_{i=1}^{n}(x_i-\hat{\mu}_{mL})(x_i-\hat{\mu}_{mL})^{\intercal} \quad (7)$$

### 2.3 Logistic Discriminant Analysis

Logistic Discriminant (Regression) allows predicting an outcome which may be continuous, discrete, and dichotomous or a mix. The goal of logistic is to find the best fitting model that describes the relationship between the outcome (dependent) and a set of independent, explanatory variables.

Cronix and Kristel (2005), suppose there are two p-dimensional sources of population, both normally distributed with different means but the same covariance, then, the variable $x$, can arise from one of the following populations.

$$X \backsim \begin{cases} H_1 = N_p(\mu_1,\Sigma)\ with\ probability\ of\ \pi_1 \\ H_0 = N_p(\mu_0,\Sigma)\ with\ probability\ of\ \pi_0 \end{cases}$$

where, $\pi_1 + \pi_0 = 1$.

Let the variable, $y$, indicate the source population of the corresponding, $x$, then

$$X \backsim \begin{cases} 1 \quad with\ probability\ of\ \pi_1 \\ 0 \quad with\ probability\ of\ \pi_0 = 1 - \pi_1 \end{cases}$$

 

### 2.3.1 Error Rate for Logistic Discriminant Analysis

Let $\pi_{01}$ represent the probability that an observation of population 1, is misclassified, $\pi_{10}$ denote the probability that an observation of population 0, is misclassified and H, the data to estimate logistic discriminant. Then, the error rate (ER) is defined as

$$\mathrm{ER}(H) = \pi_1\pi_{01}(H) + (1-\pi_1)\pi_{10}(H) \quad (8)$$

The probability of misclassifying an observation of population 1 is given by

$$\pi_{01}(H) = \mathrm{P}\left(X^t\beta(H) + A(H) < 0 | X \sim N(\mu_1, \Sigma)\right) \quad (9)$$

$$\pi_{01}(H) = P\left(X^t\beta(H) < -A(H)\right) \Big| X \sim N(\mu_1, \Sigma)$$

$$\pi_{01}(H) = \mathrm{P}\left(\leq Z \frac{-A(H)-\mu_1^t\beta(H)}{\sqrt{\beta^t(H)\Sigma\beta(H)}} \middle| Z \sim N(0,1)\right)$$

$$\pi_{01}(H) = \Phi\left(\frac{-A(H)-\mu_1^t\beta(H)}{\sqrt{\beta^t(H)\Sigma\beta(H)}}\right) \quad (10)$$

where $\Phi$ is the cumulative distribution function of a univariate standard normal. The probability of misclassifying an observation of population 0 is given by:

$$\pi_{10}(H) = \mathrm{P}\left(X^t\beta(H) + A(H) > 0 | X \sim N(\mu_1, \Sigma)\right) \quad (11)$$

$$\pi_{10}(H) = \Phi\left(\frac{A(H)+\mu_1^t\beta(H)}{\sqrt{\beta^t(H)\Sigma\beta(H)}}\right) \quad (12)$$

Therefore, the error from the data from a distribution $H$, is given by

$$ER(H) = \pi_1\Phi\left(\frac{-A(H)-\mu_1^t\beta(H)}{\sqrt{\beta^t(H)\Sigma\beta(H)}}\right) + (1-\pi_1)\Phi\left(\frac{A(H)+\mu_0^t\beta(H)}{\sqrt{\beta^t(H)\Sigma\beta(H)}}\right) \quad (13)$$

where $A(H)$and $\beta(H)$correspond to the intercept and slope estimators.

## 3 Estimations

The result of the average error rate of the simulated was presented in the table 3.1 below. Two situations were considered as stated in methodology. Error rates were computed for both situations, q, taking vales 2, 3, 4, 5, 6, 7, 8, 9 and 10 for a range of values of $P_1$ and $P_2$ between 0.1 and 0.9.

**Table 3.1 Average Error Rate for 2≤ $q$ ≤6 & 0.1 ≤ $P_1$, $P_2$ ≤ 0.9**

| | | | q=2 | | q=3 | | q=4 | | q=5 | | q=6 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| P1 | P2 | Situation | MLR | LDA | MLR | LDA | MLR | LDA | MLR | LDA | MLR | LDA |
| | 0.1 | A | 0.2337 | 0.3070 | 0.3220 | 0.2453 | 0.2400 | 0.3480 | 0.3700 | 0.3877 | 0.2500 | 0.4996 |
| | | B | 0.2430 | 0.3083 | 0.3194 | 0.2563 | 0.3300 | 0.3446 | 0.2480 | 0.3910 | 0.3300 | 0.5000 |
| | 0.3 | A | 0.2500 | 0.2610 | 0.2755 | 0.2463 | 0.3100 | 0.3048 | 0.2600 | 0.3594 | 0.3000 | 0.4998 |
| | | B | 0.3000 | 0.2702 | 0.2770 | 0.2588 | 0.3600 | 0.3056 | 0.3100 | 0.3573 | 0.2900 | 0.4997 |
| 0.1 | 0.5 | A | 0.2337 | 0.2579 | 0.2596 | 0.2465 | 0.3900 | 0.2805 | 0.2450 | 0.3326 | 0.2750 | 0.4996 |
| | | B | 0.2950 | 0.2572 | 0.2632 | 0.2413 | 0.2600 | 0.2854 | 0.2750 | 0.3342 | 0.2600 | 0.4997 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 0.7 | A | 0.2400 | 0.2558 | 0.2613 | 0.2463 | 0.2480 | 0.2764 | 0.3100 | 0.3255 | 0.2400 | 0.4996 |
| | | B | 0.2470 | 0.2556 | 0.2642 | 0.2455 | 0.3800 | 0.2735 | 0.4100 | 0.3210 | 0.2750 | 0.4996 |
| | 0.9 | A | 0.2900 | 0.2559 | 0.2597 | 0.2520 | 0.4500 | 0.2630 | 0.3200 | 0.3163 | 0.2600 | 0.4996 |
| | | B | 0.3300 | 0.2556 | 0.2641 | 0.2515 | 0.2650 | 0.2743 | 0.2430 | 0.3130 | 0.2500 | 0.4997 |
| | 0.3 | A | 0.2400 | 0.3059 | 0.3211 | 0.2490 | 0.3050 | 0.3444 | 0.3100 | 0.3902 | 0.2400 | 0.4997 |
| | | B | 0.3200 | 0.3056 | 0.3223 | 0.2480 | 0.2500 | 0.3458 | 0.3250 | 0.3924 | 0.2450 | 0.4997 |
| 0.3 | 0.5 | A | 0.2400 | 0.2912 | 0.3028 | 0.2380 | 0.2800 | 0.3318 | 0.2700 | 0.3852 | 0.2650 | 0.4997 |
| | | B | 0.3100 | 0.2875 | 0.3049 | 0.2585 | 0.2800 | 0.3352 | 0.2600 | 0.3825 | 0.3500 | 0.4995 |
| | 0.7 | A | 0.2421 | 0.2754 | 0.2905 | 0.2508 | 0.4000 | 0.3194 | 0.3800 | 0.3720 | 0.2400 | 0.4995 |
| | | B | 0.2850 | 0.2684 | 0.2876 | 0.2445 | 0.2100 | 0.3174 | 0.3700 | 0.3702 | 0.2900 | 0.4997 |
| | 0.9 | A | 0.2350 | 0.2603 | 0.2754 | 0.2598 | 0.2900 | 0.3040 | 0.3250 | 0.3619 | 0.2450 | 0.4997 |
| | | B | 0.2500 | 0.2701 | 0.2774 | 0.2345 | 0.2450 | 0.3019 | 0.3150 | 0.3627 | 0.2500 | 0.4990 |
| | 0.5 | A | 0.2600 | 0.3066 | 0.3182 | 0.2510 | 0.2900 | 0.3442 | 0.3900 | 0.3869 | 0.4010 | 0.4990 |
| 0.5 | | B | 0.2450 | 0.3072 | 0.3232 | 0.2488 | 0.2900 | 0.3486 | 0.2400 | 0.3904 | 0.3100 | 0.4990 |
| | 0.7 | A | 0.3100 | 0.3001 | 0.3124 | 0.2510 | 0.2400 | 0.3405 | 0.3600 | 0.3800 | 0.3140 | 0.4990 |
| | | B | 0.2500 | 0.2992 | 0.3161 | 0.2470 | 0.2500 | 0.3404 | 0.3300 | 0.3807 | 0.3200 | 0.4990 |
| | 0.9 | A | 0.2550 | 0.2855 | 0.3034 | 0.2475 | 0.2500 | 0.3295 | 0.3250 | 0.3937 | 0.3330 | 0.4994 |
| | | B | 0.3200 | 0.2832 | 0.3035 | 0.2468 | 0.2490 | 0.3314 | 0.3300 | 0.3924 | 0.3130 | 0.4992 |
| | 0.7 | A | 0.2420 | 0.3081 | 0.3195 | 0.2458 | 0.2450 | 0.3462 | 0.3900 | 0.3865 | 0.2500 | 0.4997 |
| 0.7 | | B | 0.2540 | 0.3036 | 0.3162 | 0.2510 | 0.2540 | 0.3490 | 0.2600 | 0.3859 | 0.2780 | 0.4997 |
| | 0.9 | A | 0.2940 | 0.2982 | 0.3180 | 0.2643 | 0.2520 | 0.3431 | 0.2900 | 0.3877 | 0.3000 | 0.4996 |
| | | B | 0.3140 | 0.2997 | 0.3176 | 0.2535 | 0.3310 | 0.3426 | 0.3900 | 0.3910 | 0.4000 | 0.5000 |

Average Error Rate for $7 \leq q \leq 10$ & $0.1 \leq P_1, P_2 \leq 0.9$

| | | | q=7 | | q=8 | | q=9 | | q=10 | |
|---|---|---|---|---|---|---|---|---|---|---|
| P1 | P2 | Situation | MLR | LDA | MLR | LDA | MLR | LDA | MLR | LDA |
| | 0.1 | A | 0.2500 | 0.5000 | 0.2400 | 0.5000 | 0.2700 | 0.5000 | 0.3100 | 0.5000 |
| | | B | 0.2450 | 0.5000 | 0.2600 | 0.5000 | 0.2900 | 0.5000 | 0.3200 | 0.5000 |
| | 0.3 | A | 0.2200 | 0.5000 | 0.2600 | 0.5000 | 0.2400 | 0.5000 | 0.2850 | 0.5000 |
| | | B | 0.3100 | 0.5000 | 0.2470 | 0.5000 | 0.2200 | 0.5000 | 0.3950 | 0.5000 |
| 0.1 | 0.5 | A | 0.2550 | 0.5000 | 0.3200 | 0.5000 | 0.2500 | 0.5000 | 0.3100 | 0.5000 |
| | | B | 0.3200 | 0.5000 | 0.2600 | 0.5000 | 0.2470 | 0.5000 | 0.3000 | 0.5000 |
| | 0.7 | A | 0.2420 | 0.5000 | 0.2400 | 0.5000 | 0.2450 | 0.5000 | 0.3350 | 0.5000 |
| | | B | 0.2540 | 0.5000 | 0.2490 | 0.5000 | 0.2500 | 0.5000 | 0.3100 | 0.5000 |
| | 0.9 | A | 0.2940 | 0.5000 | 0.2400 | 0.5000 | 0.2420 | 0.5000 | 0.3100 | 0.5000 |
| | | B | 0.3140 | 0.5000 | 0.3400 | 0.5000 | 0.3310 | 0.5000 | 0.2550 | 0.5000 |
| | 0.3 | A | 0.4200 | 0.5000 | 0.3100 | 0.5000 | 0.2900 | 0.5000 | 0.3200 | 0.5000 |
| | | B | 0.4100 | 0.5000 | 0.2450 | 0.5000 | 0.2750 | 0.5000 | 0.3800 | 0.5000 |
| 0.3 | 0.5 | A | 0.2600 | 0.5000 | 0.2750 | 0.5000 | 0.2600 | 0.5000 | 0.2410 | 0.5000 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | B | 0.2600 | 0.5000 | 0.2300 | 0.5000 | 0.2540 | 0.5000 | 0.2650 | 0.5000 |
| | 0.7 | A | 0.3800 | 0.5000 | 0.4100 | 0.5000 | 0.2400 | 0.5000 | 0.3050 | 0.5000 |
| | | B | 0.4300 | 0.5000 | 0.2450 | 0.5000 | 0.2650 | 0.5000 | 0.2500 | 0.5000 |
| | 0.9 | A | 0.2650 | 0.5000 | 0.3400 | 0.5000 | 0.2500 | 0.5000 | 0.2800 | 0.5000 |
| | | B | 0.3050 | 0.5000 | 0.3100 | 0.5000 | 0.2440 | 0.5000 | 0.2800 | 0.5000 |
| | 0.5 | A | 0.2500 | 0.5000 | 0.3250 | 0.5000 | 0.2450 | 0.5000 | 0.2410 | 0.5000 |
| 0.5 | | B | 0.2550 | 0.5000 | 0.3200 | 0.5000 | 0.2500 | 0.5000 | 0.3100 | 0.5000 |
| | 0.7 | A | 0.3200 | 0.5000 | 0.2400 | 0.5000 | 0.2400 | 0.5000 | 0.2900 | 0.5000 |
| | | B | 0.2450 | 0.5000 | 0.2500 | 0.5000 | 0.2450 | 0.5000 | 0.2450 | 0.5000 |
| | 0.9 | A | 0.2540 | 0.5000 | 0.2400 | 0.5000 | 0.2400 | 0.5000 | 0.2900 | 0.5000 |
| | | B | 0.2940 | 0.5000 | 0.2490 | 0.5000 | 0.3020 | 0.5000 | 0.2900 | 0.5000 |
| | 0.7 | A | 0.3140 | 0.5000 | 0.3400 | 0.5000 | 0.3310 | 0.5000 | 0.2500 | 0.5000 |
| 0.7 | | B | 0.4200 | 0.5000 | 0.3100 | 0.5000 | 0.2900 | 0.5000 | 0.3200 | 0.5000 |
| | 0.9 | A | 0.4100 | 0.5000 | 0.2450 | 0.5000 | 0.2750 | 0.5000 | 0.2420 | 0.5000 |
| | | B | 0.3600 | 0.5000 | 0.3750 | 0.5000 | 0.3600 | 0.5000 | 0.2540 | 0.5000 |

**Table 3.2: Summary of the first position in the performance of the two classification rules using simulated data**

| CLASSIFICATION RULE | SITUATION A | NO | SITUATION B | NO |
|---|---|---|---|---|
| MLR | q = 2, q = 3<br>q = 4, q = 5<br>q = 6, q = 7<br>q = 8, q = 9<br>q = 10 | 10 | q = 2, q = 3<br>q = 4, q = 5<br>q = 6, q = 7<br>q = 8, q = 9<br>q = 10 | 10 |
| LDA | Nil | 0 | Nil | 0 |

## 3.2 Application to Real Data

Table 3.3 below shows the result of Average Error Rate for Real Data for Maximum likelihood rule and Logistic Discriminant Analysis

**Table 3.3: MLR and LDA values**

| Data | MLR | LDA |
|---|---|---|
| **Primary** | 0.2510 | 0.5000 |
| **Secondary** | 0.2804 | 0.5000 |

 

### 3.3 Findings

The study compared the performance of the Maximum Likelihood Rule (MLR) and Logistic Discriminant (LD) analysis in the classification of variables that involved mixture of discrete and continuous variables. The findings are as follows:

1. The Maximum Likelihood Rule came first in both situation (a) and situation (b) in terms of lower average error rate in the simulated experiment conducted.
2. The result showed that when there is evidence of interaction between both discrete and continuous variables the Logistic Discriminant Analysis gave poor result than the maximum likelihood rule.
3. The result indicated that when q increases that is the component of the discrete variables the average error rate for the logistic discriminant tends to be higher than the maximum likelihood rule.
4. The result of the application of real data indicated that the maximum likelihood rule achieved a better result in terms of minimized the average error rate.

### 3.4 Conclusion

The aim of this paper was to compare maximum Likelihood Rule and Logistic Discriminant Analysis in the classification of variables that involved mixture of discrete and continuous in order to make better choice between the methods in term of error rate. The best classification rule is the one that leads to a lower error rate than the other. Table 3.3 shows the results obtained for maximum likelihood rule and logistic discriminant analysis. Values obtained for maximum likelihood are smaller than values of logistic discriminant analysis. Hence, maximum likelihood rule performed better than logistic discriminant in terms of minimizing the average error rate. To conclude, we can say that maximum likelihood rule gave better result than the logistic discriminant in the analysis conducted with simulated and real data.

**Author Contributions:** All authors contributed equally to the writing of this paper. All authors read and approved the final manuscript.

**Conflict of interest:** The authors declare no conflict of interest.

### References

Hamid, H. H. (2010). A new a roach for classifying large number of mixed variables. *World Academy of Science Engineering & Technology*. Vol. 46, 10-22.

Nocairi, H.; E. M. Qannari & M. Hanafi (2006). A Simple Regularization Procedure for Discriminant Analysis, *Communications in Statistics - Simulation and Computation*, 35:4, 957-967.

Leon, A. R. & Zhu, Y. (2007). ANOVA extensions for mixed discrete and continuous data. *Computational Statistics & Data Analysis* 52 (4), 2218-2227.

Kakio, M. Silva, D. and Langford, S. (2009). A global geometric framework for nonlinear dimensionality reduction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. Vol. 290, 2319–2323.

Fisher, R. A. (1920). A mathematical examination of the methods of determining the accuracy of observation by the Mean Error, and by the Mean Square Error. *Monthly Notices of the Royal Astronomical Society*, Volume 80, Issue 8, 758–770

Mengsay, L. (2020). Maximum Likelihood Estimation: Gaussian Model and Linear Discriminant Analysis. Obtained on 14 June, 2021, on www.twardsdatascience.com/ml-estimation-gaussian-model-and-linear-discriminant-analysis-

Cronix, G., & Kristel, J. (2005). Logistics discrimination using robust estimator. *Journal of American Statistical Association*. Vol. 66, 1-13