Iraqi Journal for Computer Science and Mathematics

Journal Homepage: http://journal.esj.edu.iq/index.php/IJCM
e-ISSN: 2788-7421 p-ISSN: 2958-0544

The Early Warning of Financial Failure for Iraqi Banks Based on Robust Adaptive Lasso Logistic Regression

Aqeel Jaafar Abbas[1], Hassan S. Uraibi[2]*

[1]Dept. of Statistics, Faculty of Administration and Economics,
University of Al-Qadisiyah, Diwaniyah, IRAQ

[2]Dept. of Statistics, Faculty of Administration and Economics,
University of Al-Qadisiyah, Diwaniyah, IRAQ

*Corresponding Author: Hassan S. Uraibi

DOI: https://doi.org/10.52866/ijcsm.2024.05.01.008
Received June 2023 ; Accepted September 2023 ; Available online January 2024

ABSTRACT: It is well known that there are many mathematical financial failure models that have been proposed in the financial literature for specific stock markets. Some researchers are not aware these mathematical models were constructed to be fitted for that market data, not for other markets. Iraq stock market exchange is one of these markets in which the researchers used imported models such as Kida, Sherrod, Altman, and others to predict financial failure. Therefore, the development of a financial failure warning model for banks has become very crucial for the Iraqi bank sector in the stock market exchange. Unfortunately, there is no clear information about the financial failure of Iraqi banks as a response variable, and the financial indicators contain outliers. The objective of this paper is to propose an algorithm to know the performance of efficient and inefficient banks based on their indicators during specific time periods. The output of this algorithm will be considered as response variables. Then, a weighted adaptive lasso logistic regression algorithm that has a high breakdown point is used to tackle outliers' problem. Thirteen banks have been chosen as the most traded during the period (2010-2017), and for each bank (27) financial indicators were collected. Our proposed model is compared with adaptive lasso logistic regression by using Deviance, Misclassification, Area Under Curve, Mean Square Errors, and Mean Absolute Errors. Consequently, the results showed that the weighted Adaptive Lasso Logistic Regression model is more robust and relevant than others to be a financial model to warn of the failure of the banks in Iraq's stock market.

Keywords: Lasso, Outliers, Leverage Points, Financial Failure, and Logistic Regression

1. INTRODUCTION

The prediction model of financial failure is a very important statistical model and sends early warning signals of financial risk for business firms. Its management has to revise its strategic plans and decision-making to improve its financial activities and avoid these failure risks. Models such as Altman [1] and Sherrod [2] were put forward to specific markets, which are considered strong forms of efficiency markets. It is well-known that the international stock market exchanges are classified into three forms of efficiency: strong, semi-strong, and weak [3]. Unfortunately, many authors used these models even for weak efficiency markets. Many authors pointed out that Iraq's Stock market exchange is considered a weak form of efficiency, but many of them used these models with Iraq's market data. Statistically, this procedure is incorrect because the models were built from samples randomly selected from its markets, which are known as strong efficiency and cannot be generalized for weak markets.

So, building a particular model for predicting the financial failure of the Iraq stock market exchange has become necessary. Due to the binary response of the financial failure model and the existence of many financial indicators, the logistic regression variable selection methods are recommended. Recently, many papers have addressed building models for predicting the financial failure of companies and banks in the financial literature. The Lasso variable selection method of Logistic regression has been recently applied to select the most important financial indicators to predict financial failure [4].

Indeed, the assumptions of logistic regression are not easy to meet in the real world of data. It is probable that some violations of these assumptions occur; therefore, we need to consider it [5]. Silverstone [6] pointed out that the Maximum Likelihood is the best method for estimating the logistic curve when the response is binary and is binomially distributed. It is considered one unbiased estimate method when the sample size is large enough. Moreover, it has a unique solution when its assumptions are met [7].

Unfortunately, the presence of outliers leads to violation of these assumptions and the parameters estimation of logistic regression will be misleading. Outliers in the logistic regression can appear on both sides of the model, the response variable and independent variables. Let us focus on the good and bad leverage points. They appear in the space of the independent variables. Croux et al. [8] found that the bad leverage points are the most dangerous types of outliers because they are observations of classification error and, at the same time, an unusual observation that moves away from their counterparts in the variables space Illustration.

Some of the robust approaches that have been presented in the statistical literature to tackle this problem replaced the robust objective function instead of the Maximum likelihood function, such as ([9]; [10]). Pregibon [11], Bianco and Yohai [12], Carroll and Pederson [13], and Croux & Haesbroeck [14] added weights to this estimator to reduce the effect of leverage points in the space of the independent variables. In 1985, [15] suggested using the robust Mahalanobis distance instead of the traditional one to address the problem of leverage points. Most of these weights depend on the robust Mahalanobis distance, which depends on the Minimum Covariance Determinant (MCD) estimates proposed by [16].

Rousseuw [17] mentioned that High Leverage Points (HLP) do not allow the model to fit due to it having a significant impact on the method of Maximum likelihood. Imon (2006) [18] showed that this type of leverage point is not only responsible for giving wrong estimates but also has the ability to create the masking problem ([19], [20],[21]). In 1986, [22] mentioned that logistic regression analysis might not be accurate when the number of independent variables is large. Sometimes, the number of independent variables exceeds the sample size. In this case, Lasso [23] and its siblings are the best choices.

However, this paper concentrated on building a prediction model of financial failure for the bank sector of the Iraq stock market exchange. Two challenges have faced us. First, there is not enough information about the financial failure of banks except in two cases of bankruptcy banks. The second challenge is that the financial indicators collected for this paper are contaminated by the high leverage points with masking and swamping cases. The standard algorithm is suggested to overcome the first challenge and then obtain the binary response for failure and success cases as a dependent variable. A weighted adaptive Lasso algorithm that has a high breakdown point is proposed to tackle the second challenge simultaneously as a robust estimation and variable selection. The rest of this paper is organized to present the Penalized Logistic Regression in Section 2. Section 3 describes the Weighted Adaptive Lasso Logistic Regression Algorithm. Section 4 explains Iraq banks' financial indicators data, and Section 5 is the conclusion.

2. PENALIZED LOGISTIC REGRESSION

Suppose that the dichotomous response y_i (financial failure and success) may be connected to a set of linear combinations of financial indicators $\beta_j X_{ij}$, where β_j is the coefficients vector, X_{ij} is the financial indicators, $i = 1,2, \dots, n$ and, $j = 1,2, \dots, p$. These types of data usually lend themselves to logistic regression since the fitted values will be within the acceptable range of binary responses, which are only restricted by the values 0 and 1. The distribution of i^{th} case is Bernoulli, $y_i \sim Ber(\pi_i)$ where π_i is the probability of success case when $(y_i = 1)$,

$$\pi_i = p(y_i|x_i) = \frac{e^{\beta_0+\beta_j X_{ij}}}{1+e^{\beta_0+\beta_j X_{ij}}} \; ; \; 0 < \pi_i < 1. \tag{1}$$

When $(y_i = 0)$ is the probability (1-π_i) of failure case is equivalent to, we then have

$$1 - \pi_i = 1 - \frac{e^{\beta_0+\beta_j X_{ij}}}{1+e^{\beta_0+\beta_j X_{ij}}} = \frac{1}{1+e^{\beta_0+\beta_j X_{ij}}}, \tag{2}$$

where β_0 is the intercept, $(\beta_j)_{n\times 1}$ is the vector of unknown financial ratio coefficients, $i = 1,2, \dots, n, \;\; j = 1,2, \dots, p$ if all cases have been considered together, $y_i \sim Bin(n, \pi_i)$ is binomial distribution and the relationship between π_i and x_i is nonlinear. In 1944, [24] suggested logistic transformation of probability estimated vector $p(y_i = 1|x_i)$ to be a linear function. The logit transformation can be written as follows:

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \log\left(\frac{\frac{e^{\beta_0+\beta_j X_{ij}}}{1+e^{\beta_0+\beta_j X_{ij}}}}{\frac{1}{1+e^{\beta_0+\beta_j X_{ij}}}}\right) = \log\left(e^{\beta_0+\beta_j X_{ij}}\right) = \beta_0 + \beta_j X_{ij} \tag{3}$$

The likelihood function for the logistic regression model is then given by

$$L(\beta_0,\beta_j) = \prod_{i=1}^{n} \pi_i(\beta_0+\beta_j X_{ij})^{y_i}\left(1-\pi_i(\beta_0+\beta_j X_{ij})\right)^{1-y_i} \quad (4)$$

Then, the coefficients β_0 and β_j can be estimated by minimizing the negative log-likelihood function, which is defined as:

$$L(\beta_0,\beta_j) = \sum_{i=1}^{n}\left[y_i \log\left(\pi_i(\beta_0+\beta_j X_{ij})\right) + (1-y_i)\log(1-\pi_i(\beta_0+\beta_j X_{ij}))\right] \quad (5)$$

The weakness of Eq. (5) is that it cannot control the overfitting problem. In 1996, Tibshirani [23] added non-negative L_1 Lasso penalty term to negative log in Eq. (5) to control the size of financial indicators coefficients as follows:

$$Q_{n,\lambda}(\beta) = -\log L(\beta) + \lambda \sum_{j=1}^{p}|\beta_j| \quad (6)$$

where $\beta = \binom{\beta_0}{\beta_j}$ vector, $\sum_{j=1}^{p}|\beta_j| \le t$ and t is a constant such that the smaller values of $0 < t < \sum_{j=1}^{p}|\beta_j|$ produce a shrunken estimate of the Lasso coefficient. Choosing appropriate t, Lasso tends to force some coefficients to be exactly zero so that Lasso can select the variable with non-zero coefficients like subset selection methods.

Suppose that with the initial value of λ, $\hat{\beta}_j^{lasso} = \hat{\beta}_j^{ols} - \frac{\lambda}{2} sign(\hat{\beta}_j^{lasso})$. Since both $\hat{\beta}_j^{lasso}$ and $\hat{\beta}_j^{ols}$ share the same sign, we can rewrite the previous formula as follows:
$\hat{\beta}_j^{lasso} = \hat{\beta}_j^{ols} - \frac{\lambda}{2} sign(\hat{\beta}_j^{lasso}) \Leftrightarrow \hat{\beta}_j^{ols} = \hat{\beta}_j^{lasso} + \frac{\lambda}{2} sign(\hat{\beta}_j^{lasso})$. It is clear that the part of $\hat{\beta}_j^{lasso} + \frac{\lambda}{2} sign(\hat{\beta}_j^{lasso})$ must be positive or negative if the sign of $\hat{\beta}_j^{lasso}$ is positive or negative, respectively.
Consequentially, $sign(\hat{\beta}_j^{lasso}) = sign(\hat{\beta}_j^{ols})$. Then

$$\hat{\beta}_j^{lasso} = sign(\hat{\beta}_j^{ols})\left(\hat{\beta}_j^{ols} - \frac{\lambda}{2}\right)^{+}, j = 1,2,\dots,p. \quad (7)$$

where $(+)$ denotes the positive part of the expression inside the parenthesis. Hence, $\frac{\lambda}{2}$ is a crucial threshold point for non-zero or zero estimate estimates, where the corresponding $|\hat{\beta}_j^{ols}| < \frac{\lambda}{2}$ and the algorithm set $\hat{\beta}_j^{lasso} = 0$. The parameter λ can be found by solving the equation (last one) subject to $\sum_j \hat{\beta}_j^{lasso} = t$. Therefore, each value of λ corresponds to a unique value of t. For instance, if we have two predictors, without loss of generality, we may assume that the OLS estimates are positive. Now, λ can be computed as follows:

$$\begin{Bmatrix} \hat{\beta}_1^{lasso} = \left(\hat{\beta}_1^{ols} - \frac{\lambda}{2}\right)^{+} \\ \hat{\beta}_2^{lasso} = \left(\hat{\beta}_2^{ols} - \frac{\lambda}{2}\right)^{+} \end{Bmatrix} \Longrightarrow \hat{\beta}_1^{lasso} + \hat{\beta}_2^{lasso} = t \quad (8)$$

$$\left(\hat{\beta}_1^{ols} - \frac{\lambda}{2}\right)^{+} + \left(\hat{\beta}_2^{ols} - \frac{\lambda}{2}\right)^{+} = t$$
$$\Longrightarrow \left(\hat{\beta}_1^{ols} - \frac{\lambda}{2}\right)^{+} + \left(\hat{\beta}_2^{ols} - \frac{\lambda}{2}\right)^{+} - t = 0$$

$$\hat{\beta}_1^{ols} + \hat{\beta}_2^{ols} - \lambda - t = 0 \Longrightarrow \lambda = \hat{\beta}_1^{ols} + \hat{\beta}_2^{ols} - t \quad (9)$$

By substituting Eq. (9) into Eq. (8), then

$$\begin{Bmatrix} \hat{\beta}_1^{lasso} = \left(\hat{\beta}_1^{ols} - \frac{\hat{\beta}_1^{ols}+\hat{\beta}_2^{ols}-t}{2}\right)^{+} = \left(\frac{t}{2} + \frac{\hat{\beta}_1^{ols}-\hat{\beta}_2^{ols}}{2}\right) \\ \hat{\beta}_2^{lasso} = \left(\hat{\beta}_2^{ols} - \frac{\hat{\beta}_1^{ols}+\hat{\beta}_2^{ols}-t}{2}\right)^{+} = \left(\frac{t}{2} - \frac{\hat{\beta}_1^{ols}-\hat{\beta}_2^{ols}}{2}\right) \end{Bmatrix} \quad (10)$$

Eq. (10) may be generalized for multivariate cases:

$$t = \sum_{j=1}^{p}|\hat{\beta}_j^{lasso}| = \sum_{j=1}^{p}\left[sign(\hat{\beta}_j^{ols})\left(\hat{\beta}_j^{ols} - \frac{\lambda}{2}\right)^{+}\right] \quad (11)$$

$$t = \sum_{j=1}^{p} sign(\hat{\beta}_j^{ols})\hat{\beta}_j^{ols} - p\frac{\lambda}{2} \quad (12)$$

Mathematically, the relationship between tuning parameters t and λ is one-to-one. This tuning parameter is considered the absolute bound.
When $\lambda = 0$, the logistic parameters can be estimated by maximum likelihood, $\lambda = \infty$ when all parameters are removed. Consequently, as λ increases step by step, more coefficients are either set to zero or removed, yielding sparse subset non-zero coefficients. The regularized coefficient vector β can be estimated by minimizing $Q_{n,\lambda}(\beta)$.

In 2003, [25] introduced sparse logistic regression with L_1 penalty. Meanwhile, Zhu and Hastie [26], Park and Hastie [27], Sun and Wang [28], Plan and Vershynin [29], and Yang and Qian [30] found intriguing findings and applications of the Lasso approach in logistic regression. However, the Lasso penalty has come under fire for being biased because it frequently chooses numerous noisy variables [31]. Fan and Li [32] stated that a good selection procedure is one that has the following oracle properties.

- Diagnostic the correct submodel, $\Gamma = \{j: \hat{\beta}_j \neq 0\}$
- Has the optimal estimation $\sqrt{n}(\hat{\beta}_\Gamma - \beta_\Gamma) \xrightarrow{d} N(0, \ddot{\Sigma})$, where $\ddot{\Sigma}$ is the variance-covariance matrix of the true model.

In general, the Lasso path is inconsistent where the probability mass is greater than zero. Even if the whole path contains the true parameter value, it cannot be achieved using prediction accuracy as the selection criterion [33]. In the context of linear regression models, Zou [34] developed the adaptive Lasso penalty to address the bias drawback. The adaptive Lasso criterion for logistic regression specifically takes into account the objective function [35], given by

$$Q_{n,\lambda}(\beta) = -\log L(\beta) + \lambda \sum_{j=1}^{p} \frac{|\beta_j|}{|\tilde{\beta}_j|}, \qquad (13)$$

where $\tilde{\beta}_j$ is the initial estimator for β_j.

The first part of Eq. (5) and Eq. (13) is a negative log-likelihood loss function, which is sensitive to the presence of leverage points that lead to erroneous conclusions due to its influence. Some interesting applications of the penalized method of logistic regression for high-dimensional data can be seen in [36], [37], and [38]. All of these methods are focused on the selection of robust loss functions instead of negative log-likelihood functions.

3. WEIGHTED ADAPTIVE LASSO LOGISTIC REGRESSION ALGORITHM

The breakdown point of Adaptive Lasso Logistic Regression (ALLR) can be zero in the presence of at least a single leverage point. Our proposed algorithm-weighted ALLR (WALLR) can be summarized to reduce the influence of leverage points in ALLR as follows:

1. Computing Minimum Volume Ellipsoid (MVE) of multivariate location $(\bar{X}_{Rob})$ and scatter $(c(X)_{Rob})$ of financial indicators. The estimator of MVE is a concentration algorithm that covers at least half plus one of the data set using Order statistics.
2. Finding Robust Mahalanobis Distance (RMD) relies on step (1)

$$RMD_i = \sqrt{(X_i - \bar{X}_{Rob})' c(X)^{-1}_{Rob} (X_i - \bar{X}_{Rob})} \qquad (14)$$

in which any value of $\left(RMD_i > \sqrt{\chi^2_{(p,0.95)}}\right)$ is considered a leverage point.

The weights' function is calculated as follows:

$$W_i = \begin{cases} 1 & \text{if } RMD_i \leq \sqrt{\chi^2_{(p,0.95)}} \\ \dfrac{\sqrt{\chi^2_{(p,0.95)}}}{RMD_i} & \text{if } RMD_i > \sqrt{\chi^2_{(p,0.95)}} \end{cases} \qquad (15)$$

Finally, WALLR can be written as follows:

$$Q_{n,\lambda}(\beta^W) = -\log L(\beta^W) + \lambda_W \sum_{j=1}^{p} \frac{|\beta^W_j|}{|\tilde{\beta}_j|}, \qquad (16)$$

in which vector β^W can be estimated by minimizing $Q_{n,\lambda}(\beta^W)$.

4. IRAQ BANKS FINANCIAL INDICATORS DATA

The data set is collected from the Iraq stock market exchange. We have taken a random sample of (13) banks from (26) banks registered in the market and were the most traded without interruption for the period (2010-2017), which are (The National Bank of Iraq (NBI), Ashur International Bank (AIB), Babylon Bank (BabyB), Bank of Baghdad (BaghB, Elaf Islamic Bank (EIB), Credit Bank of Iraq (CrBI), Investment Bank of Iraq (IBI), Trade Bank of Iraq (TBI), Gulf Commercial Bank (GCB), Al-Mansour Bank for Investment (MBI), Sumer Bank(SB), Commercial Bank of Iraq (CoBI) and finally, the National Islamic bank (NIB)), as we collected starting (27) financial indicators. Unfortunately, we did not find any information about the failure and success of banks on the website of the Iraq Stock Market Exchange. Therefore, we propose the standard algorithm to determine the performance of efficient (1) and inefficient banks (0) based on their indicators for the period (2010-2017). The output of this algorithm will be considered as binary response variables.

Let $X_{(104\times27)}$ be the design matrix of indicator (27) for the period (2010-2017), which includes (104) observations. The National Bank of Iraq has (8) observations for each indicator placed at the first eight rows of X, in which the ninth row to the sixteenth row have been assigned to Ashur International Bank, and so on for other banks according to their sequences in the collected data.

Each bank should be subjected to the following steps,

1. Compute the column median of X, $Med_{27} = median(X)$, which represents the sector median of (27) indicator for the full period, see Table 1.
2. Let $b = \{1,9,17,\dots,96,104\}$ # as a vector
3. Let Y be an empty vector
4. For $K = 1:13$ Do { # is accouter
5. Z is zeros vector # the length of this vector being 27
6. $Bank^{(K)} = X[b[K]:(b[K+1]-1),]$ # Determining the specific data of bank K
7. Let R be an empty vector
8. For $j = 1:[b[K+1]-1]$ Do { # is a counter,

$$Z = \begin{cases} 1 & if\ X[j,] > Med_{27} \\ 0 & \text{Otherwise} \end{cases}$$

$$R[j] = \begin{cases} 1 & if\ \sum Z_{27}^{(K)} \geq 14 \\ 0 & \text{Otherwise} \end{cases}$$

}

$Y = (Y, R)$

}

9. END

The length of $Z_{27}^{(K)}$ vector is (27), which is the indicator of K^{th} bank for a certain year. We consider the K^{th} bank such that when the number of successful indicators exceeds (14), it is efficient because it has (14) cases of success, which represent more than half of the indicators. For this reason, we indicate putting the number (1) for that year in the response variable and vice versa. The vector Y is the binary response, which represents the failure and annual success of banks.

Table 1: The medians of the banking sector of (27) indicators for the period (2010-2017)

X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9
0.056	0.47	15.555	1.85	1.134	0.9	0.9	1.03	0.76
X_{10}	**X_{11}**	**X_{12}**	**X_{13}**	**X_{14}**	**X_{15}**	**X_{16}**	**X_{17}**	**X_{18}**
0.91	1.147	0.103	0.056	0.048	0.06	0.062	0.158	0.522
X_{19}	**X_{20}**	**X_{21}**	**X_{22}**	**X_{23}**	**X_{24}**	**X_{25}**	**X_{26}**	**X_{27}**
0.073	0.923	0.488	15	0.432	0.951	0.489	0.022	1.909

4.1 DETECTION OUTLIERS AND LEVERAGE POINTS

For the detection of outliers in the response variable, we used the Standardized Pearson residuals, $spr_i = \frac{r_i}{\sqrt{(1-h_{ii})}}$, where $r_i = \frac{y_i - m_i\hat{\pi}_i}{\sqrt{m_i\hat{\pi}_i(1-\hat{\pi}_i)}}$, h_{ii} is the diagonal elements of $H = \sqrt{\widehat{W}}X(X'\widehat{W}X)^{-1}X'\sqrt{\widehat{W}}$ (for more details, see [39]). The preliminary results showed that there are no outliers in these residuals.

Robust Mahalanobis Distance (RMD) is identified as leverage points in the indicators. The results showed that there is a large number of Leverage Points (LP) in the data set, similar to the RMD method after comparing it with the cut-off point ($\chi^2_{(0.975,27)} = 6.762$). It turns out that about (34%) of the data are LP and that three of them are high LPs, while the traditional method MD detected only (19%) of the LP without HLP. Fig.1 explains there are three (HLP) detected by the (RMD) method that have masked about (16) observations by the MD method, and this phenomenon is called (masking), which was referred to [40].

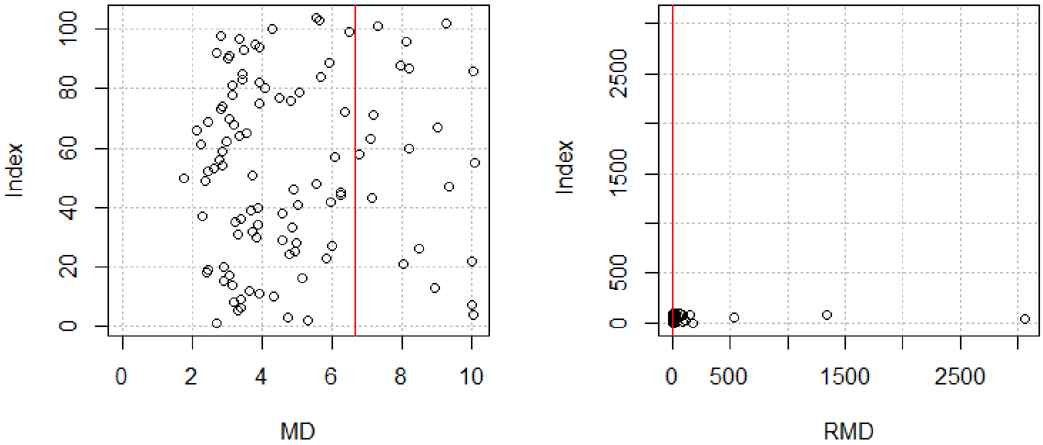

Fig.1: Robust Mahalanobis Distance vs Classical Mahalanobis Distance for detecting leverage points

4.2 ROBUST DIAGNOSTIC of MULTICOLLINEARITY PROBLEM

From the aforementioned, it is clear that the influence of the LP should be reduced to the least possible before proceeding with testing the presence of the multicollinearity problem because it may be a reason for the emergence of this problem. Therefore, the researcher used the data weighting method using a weight function based on the (RMD). It is extracted by using the location and measurement matrix (MCD) and then multiplying the explanatory variables matrix (X) by the function:

$$w_i = \min\left(1, \frac{\chi^2_{(0.975,27)}}{RMD_i}\right) \quad (17)$$

$$X^*_{(i,j)} = X_{(i,j)} \times w_i$$

Testing the existence of a multicollinearity problem with the presence of LPs is not feasible. Hence, we weighted the indicators (X_((W,i))) to get rid of the effect of LPs followed by the Pearson Correlation Coefficient. Fig. 2 clearly shows that there are high correlations between some of the indicators, as the color gradient on the left of the figure indicates the degree of correlation starting from (-1) and ending with (1). Meanwhile, the dark blue color indicates strength of a strong to perfect correlation. We notice that the points of the main diagonal have completely taken the dark blue color because they represent the complete correlation (1) between the variable and itself, while other colors close to this color indicate a strong or very strong correlation as follows:

$\mathrm{cor}(X^*_2, X^*_{14}) > 0.90,\ \mathrm{cor}(X^*_7, X^*_8) = \mathrm{cor}(X^*_7, X^*_9) = \mathrm{cor}(X^*_7, X^*_{10}) > 0.90,$

$\mathrm{cor}(X^*_{16}, X^*_{17}) > 0.90\ \mathrm{cor}(X^*_5, X^*_{22}) = \mathrm{cor}(X^*_5, X^*_{24}) = \mathrm{cor}(X^*_5, X^*_{26}) > 0.90, \mathrm{cor}(X^*_6, X^*_{25}) > 0.90,$

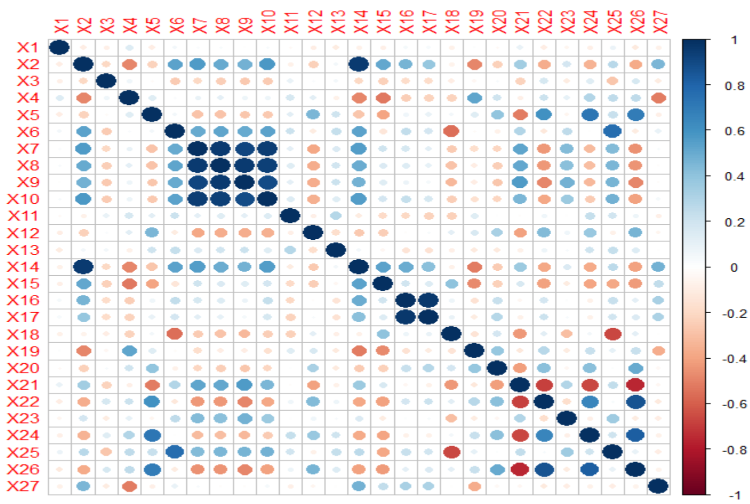

Fig.2: The correlation coefficient between unweighted indicators

4.3 BUILDING the EARLY WARNING MODE of FINANCIAL FAILURE

According to the initial description of the data, there is a violation of the assumptions of logistic regression, such as the presence of LP and multicollinearity problems with high dimensional data, so selecting the suitable estimation method is crucial. We employed an adaptive Lasso with weighted and unweighted data for dimensional reduction and to eliminate the multicollinearity problem. The adaptive Lasso is a robust method when it is used with weighted data, but it is not that case with unweighted data. However, the performance of both methods results in two financial failure models. The assessment of each model relies on some criteria to select the best ones as follows,

- $\text{Deviance}(\pi|Y) = -2\sum_{i=1}^{n}(Y_i \log(\pi_i)) + (1-Y_i)\ \log(1-\pi_i)$
- $\text{Specificity} = \frac{FP}{FP+TN}$ which is equivalent to False Positive Rate (Misclassification)
- The ROC from Area Under the ROC Curve (AUC) stands for the Receiver Operating Characteristic Curve, which can show the powerful performance of the model for classification at all classification thresholds. AUC assesses how well they are rated Instead of assessing predictions' absolute values. Moreover, it evaluates the accuracy of the model's predictions regardless of the categorization threshold that is used. The plotting of AUC refers to the tradeoff relationship between Sensitivity and Specificity. $\text{Sensitivity} = \frac{TP}{TP+FN}$, $\text{Specificity} = \frac{FP}{FP+TN}$, where TP, FP, FN and TN are the True Positive, False Positive, False Negative and True Negative, respectively. Mean of Squares Errors (MSE) and mean absolute errors (MAE), in which $MSE = \frac{\sum_{i=1}^{n}(Y_i-\hat{Y}_i)^2}{n}$ and $MAE = \frac{|Y_i-\hat{Y}_i|}{n}$, respectively.

4.3.1 ADAPTIVE LASSO LOGISTIC REGRESSION (ALLR)

Based on the foregoing, we think Adaptive Lasso of Logistic Regression (ALLR) is an important penalized method that can be used with our data before and after weighting the indicators. The r-code of ALLR is available in the glmnet package of the R programming statistical packages. For this purpose, the original full dataset is divided into two subsets: a training data set that includes (80) samples specific to (NBI, AIB, BabyB, BaghB, EIB, CrBI, IBI, TBI, GCB, MBI), and another one testing data set that includes (24) samples and specific to (SB, CoBI and finally, the NIB). The ALLR has been applied with an unweighted training dataset to get the best indicators with non-zero coefficients. Table (2) displays the results of regression coefficients of the ALLR method when the tuning parameter $\lambda = 0.017$, which is computed using the cross-validation method.

The ALLR method selects (14) indicators with non-zero coefficients. Consequently, the best indicators that have to be in the Early Warning of Financial Failure Model are (Dividend return ratio X_1, Current ratio X_4, Liquidity ratio X_{11}, Operating ratio X_{12}, Profit ratio X_{13}, asset turnover ratio X_{16}, The ratio of revenue to total deposits X_{17}, The ratio of expenses to total deposits X_{19}, Deposit-capital ratio X_{20}, Market value indicator X_{22}, Liquidity / Total assets ratio X_{24}, Shareholders' equity/ assets ratio X_{25}, and Total assets / Total liquidity ratio X_{27}). It is notable that all coefficients presented in Table (2) are statistically significant. Thus, the ALLR model suggests that the indicators with non-zero coefficients do influence the financial failure of banks. The positive values of coefficients result in an increased probability of financial failure of the banks, while the negative values reduce the probability of financial failure of banks. For instance, $e^{\beta_4} = e^{0.629205} = 1.876$ will be the odds ratio that is associated with the Current ratio, which has approximately 88% more odds of having a financial failure of banks than successful ones. Meanwhile, $e^{\beta_{11}} = e^{-0.0134} \approx 0.99$ refers to the Annual closing price associated with a (1%) reduction in the relative risk of financial failure.

It is evidence that ALLR chooses some significantly correlated indicators $(16,17)$ and $(22,24)$ with the non-zero indicators. Certainly, ALLR will not be the best choice for building the warning early financial failure model because it is an inefficient method in the presence of multicollinearity. The LPs may be the source of this problem. If the WALLR method avoids this problem and selects uncorrelated indicators, it surely will be the best.

The model of best indicators of ALLR has been used with the testing data set to predict the probability of success of the banks during the period of study. Table (3) presents the expected probability π_i, and annual prediction class, either failure or success for Sumer Bank, Commercial Bank of Iraq, and the National Islamic Bank, which are considered in the testing dataset.

It is clear that there is a significant difference in the performance of banks with this model. This showed that the performance of the CoBI bank was distinguished throughout (7) years except in 2014, while the NIB did not suffer from any failure case in its performance. The SB started the year (2010) in failure and then improved its performance until (2015) and (2016), which were failure years, and finally, in 2017, it began gaining success again.

Table 2: The ALLR coefficients of Indicators for unweighted data

Indicators	$\hat{\beta}$	Indicators	$\hat{\beta}$	Indicators	$\hat{\beta}$
X1	47.65771	X10	0	X19	2.691213
X2	0	X11	-0.0134	X20	6.821134
X3	0	X12	-0.68404	X21	0
X4	0.629205	X13	-0.65431	X22	-8.05046
X5	0	X14	0	X23	0
X6	0	X15	0	X24	-4.33356
X7	0	X16	52.76402	X25	29.072
X8	0	X17	15.39094	X26	0
X9	0	X18	0	X27	-26.9476

Based on the π_i presented in Table (3), the values of all criteria are shown in Table (4). It is notable that the Deviance is 0.86, and the FPR equals (0.17), which means from all banks that were recognized as "success" in the original data, what percentage did we miss-classify?

Table 3: The π_i and Prediction of Classification for the banks in an unweighted testing dataset

Bank	π_i	Prob.	Prediction of Classification	Bank	π_i	Prob.	Prediction of Classification
Sumer Bank	π_1	0.46	F	Commercial Bank of IRAQ	π_{13}	0.03	F
	π_2	0.72	S		π_{14}	0.97	S
	π_3	0.99	S		π_{15}	0.96	S
	π_4	1.00	S		π_{16}	0.96	S
	π_5	0.99	S		π_{17}	0.97	S
	π_6	0.04	F	National Islamic Bank	π_{18}	0.99	S
	π_7	0.12	F		π_{19}	0.80	S
	π_8	0.85	S		π_{20}	0.99	S
al Bank of	π_9	0.99	S		π_{21}	0.99	S
	π_{10}	0.99	S		π_{22}	0.99	S

π_{11}	0.81	S	π_{23}	0.51	S
π_{12}	0.81	S	π_{24}	0.99	S

AUC is approximately 89%, meaning that the model has the ability to distinguish the data of the response variable by (89%) (see Fig. 3). This is confirmed by the fact that the classification error amounted to about (17%), the MSE of about (0.24), and the MAE about (0.38).

Table 4: The values of model assessment criteria of ALLR

Method	**Deviance**	**FPR**	**AUC**	**MSE**	**MAE**
ALLR	0.86	0.17	0.89	0.24	0.38

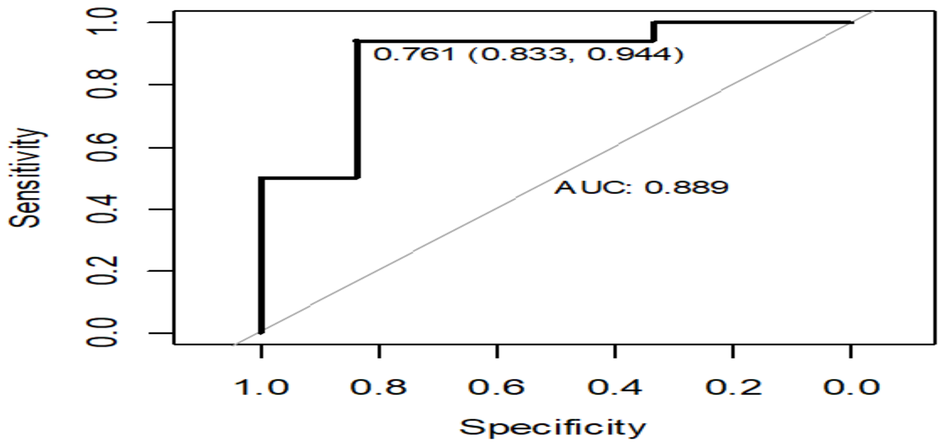

Fig.3: AUC of ALLR

4.3.2 WEIGHTED ADAPTIVE LASSO LOGISTIC REGRESSION (WALLR)

We conducted the process of estimating the robust parameters using WALLR when the tuning parameter equals ($\lambda = 0.86$). It is evidenced from Table (5) that WALLR selected only (7) indicators to be in the best model. The indicators selected are Equity ratio X_2 Profit ratio X_{13}, Deposit-capital ratio X_{20}, Shareholders' equity/ assets ratioX_{25}, Net profit after tax/ Total assets ratio X_{26}, and Total assets / Total liquidity ratio X_{27}. However, the X_2, X_{20}, and X_{25} indicators possess the greatest relative impact of early warning of a financial failure of banks while the X_{26}, and X_{27} indicators have reduced this impact.

Table 5: The WALLR coefficients of Indicators for unweighted data

Indicators	$\hat{\beta}$	Indicators	$\hat{\beta}$	Indicators	$\hat{\beta}$
X_1	0	X_{10}	0	X_{19}	0
X_2	6.238317	X_{11}	0	X_{20}	3.708139
X_3	0	X_{12}	0	X_{21}	0
X_4	0	X_{13}	-0.55391	X_{22}	0
X_5	0	X_{14}	0	X_{23}	0
X_6	0	X_{15}	0	X_{24}	0

X_7	0	X_{16}	0	X_{25}	13.25335
X_8	0	X_{17}	0	X_{26}	-7.23084
X_9	0	X_{18}	0	X_{27}	-10.0431

We have plotted the correlation matrix of weighted indicators to be sure that there is a multicollinearity problem. Fig. 4 confirms no multicollinearity problem between the selected indicators.

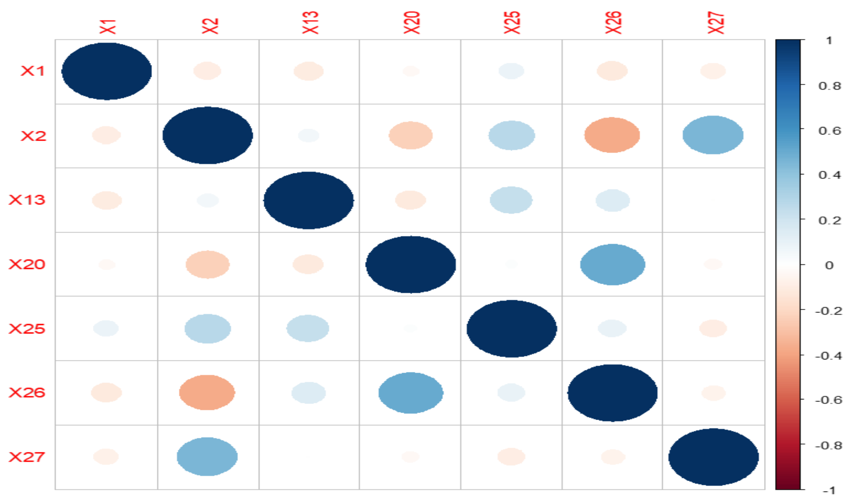

Fig.4: The correlation coefficient between weighted indicators

After obtaining the robust estimators, it is also possible to obtain the robust probabilities $\boldsymbol{\pi_i}$, which led to the result presented in Table (6).

Table 6: The $\boldsymbol{\pi_i}$ and Prediction of Classification for the banks in a weighted testing dataset

Bank	$\boldsymbol{\pi_i}$	Prob.	Prediction of Classification	Bank	$\boldsymbol{\pi_i}$	Prob.	Prediction of Classification
Sumer Bank	π_1	0.02	F	Commercial Bank of IRAQ	π_{13}	0.01	F
	π_2	0.00	F		π_{14}	0.99	S
	π_3	0.11	F		π_{15}	1.00	S
	π_4	0.50	S	National Islamic Bank	π_{16}	0.99	S
	π_5	0.74	S		π_{17}	0.99	S
	π_6	0.75	S		π_{18}	1.00	S
	π_7	0.98	S		π_{19}	0.74	S
	π_8	0.98	S		π_{20}	0.77	S
cial Ban k of	π_9	0.19	F		π_{21}	0.65	S

π_{10}	0.08	F	π_{22}	0.98	S
π_{11}	0.73	S	π_{23}	0.37	F
π_{12}	0.48	F	π_{24}	0.15	F

The predictions of this model are considered robust. Therefore, the results presented in Table (6) showed the first three years of the SB were a failure. Still, in the years (2013-2017), it is back again to success, while the CBI started oscillating between failure and success in the first five years dominated by cases of failure. Still, in the last three years, it showed very good efficiency. Unlike the previous cases, the NBI started a strong beginning of success listed in the first six years in Table (6), but in the last two years, it started to fail.

Table 7: The values of model assessment criteria of WALLR

Method	Deviance	FPR	AUC	MSE	MAE
WALLR	0.47	0.13	0.98	0.15	0.35

The results obtained from the WALLR method are presented in Table (7). Moreover, the value of (AUC) amounted to approximately (98%), which is a very high percentage that confirms the ability of the chosen model to predict a new observation that is either (0) or (1) (see Fig. 5).

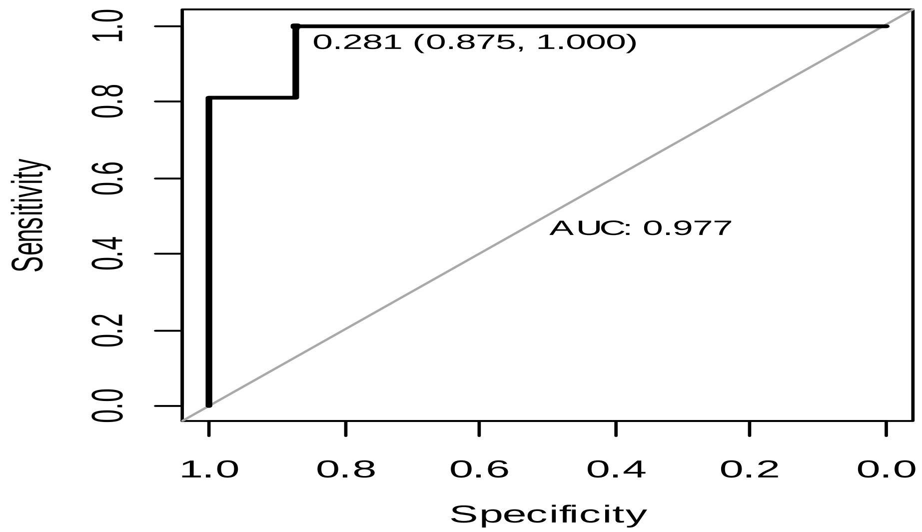

Fig.5 AUC of WALLR

5. CONCLUSION

This paper has suggested a robust prediction model for the financial failure of the banks in the Iraq stock market exchange. The procedures to get such a model require assumptions of logistic regression to be met. One of the important assumptions is the response variable should be binary. Unfortunately, this variable is not available in the Iraq stock market exchange. Therefore, we have proposed a simple algorithm to overcome this problem. The data of indicators has LP and collinearity problems. A weighted function has been used to tackle the LP problem and Adaptive Lasso logistic regression is used to process the problem of collinearity and reduce the dimensionality. The proposed method (WALLR) produced a prediction model of financial failure, which is compared with the model of traditional adaptive Lasso logistic regression (ALLR) by using some criteria. The results show the model size of WALLR is smaller than the ALLR model, and it is more efficient than it is due to very low values for (Deviance, misclassification, MSE, and MAE) and, moreover, very high AUC. Consequently, we recommend that interested researchers and bank managers use this algorithm to evaluate their bank performances as warning messages and avoid financial failure.

Funding

None

ACKNOWLEDGEMENT

We are grateful to the reviewers for reviewing this manuscript.

CONFLICTS OF INTEREST

The author declares no conflict of interest.

REFERENCES

[1] E. I. Altman, “Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy” *The Journal of Finance*, vol. 23, no. 4, pp. 589–609, Sep. 1968, doi: https://doi.org/10.1111/j.1540-6261.1968.tb00843.x.

[2] Tomas A.“Detecting Financial Distress with the b-Sherrod Model : a Case Study”, Finanse Rynki Finansowe, Szczecin, Ubezpieczenia nr 74 t 2, 2015,doi:10.18276/frfu.2015.74/2-21

[3] E. F. Fama, “Efficient Capital Markets: A Review of Theory and Empirical Work,” The Journal of Finance, vol. 25, no. 2, pp. 383–417, May 1970, doi: https://doi.org/10.1111/j.1540-6261.1970.tb00518.x.

[4] S. Tian, Y. Yu, and H. Guo, “Variable selection and corporate bankruptcy forecasts,” Journal of Banking & Finance, vol. 52, pp. 89–100, Mar. 2015, doi: https://doi.org/10.1016/j.jbankfin.2014.12.003.

[5] J. W. Fraas and I. Newman, “Ordinary Least Squares Regression, Discriminant Analysis, and Logistic Regression: Questions Researchers and Practitioners Should Address When Selecting an Analytic Technique.,” Feb. 2003.

[6] H. Silverstone, “Estimating the Logistic Curve,” Journal of the American Statistical Association, vol. 52, no. 280, pp. 567–577, Dec. 1957, doi: https://doi.org/10.1080/01621459.1957.10501414.

[7]C. Antle, L. Klimko, and W. Harkness, “Confidence intervals for the parameters of the logistic distribution,” Biometrika, vol. 57, no. 2, pp. 397–402, 1970, doi: https://doi.org/10.1093/biomet/57.2.397.

[8] C. Croux, C. Flandre, and G. Haesbroeck, “The breakdown behavior of the maximum likelihood estimator in the logistic regression model,” Statistics & Probability Letters, vol. 60, no. 4, pp. 377–386, Dec. 2002, doi: https://doi.org/10.1016/s0167-7152(02)00292-4.

[9] V. J. Yohai, “Local and global robustness of regression estimators,” Journal of Statistical Planning and Inference, vol. 57, no. 1, pp. 73–92, Jan. 1997, doi: https://doi.org/10.1016/s0378-3758(96)00037-7.

[10] F. R. Hampel, “Contributions to the theory of robust estimation,” Thesis, University of California., 1968.

[11] D. Pregibon, “Logistic Regression Diagnostics,” The Annals of Statistics, vol. 9, no. 4, pp. 705–724, Jul. 1981, doi: https://doi.org/10.1214/aos/1176345513.

[12] A. M. Bianco, and V. J. Yohai, “ Robust Estimation in the Logistic Regression Model. In: Rieder, H. (eds) Robust Statistics, Data Analysis, and Computer Intensive Methods. Lecture Notes in Statistics, vol 109. Springer, New York, NY. 1996, https://doi.org/10.1007/978-1-4612-2380-1_2

[13] R. J. Carroll and S. Pederson, “On Robustness in the Logistic Regression Model,” Journal of the Royal Statistical Society: Series B (Methodological), vol. 55, no. 3, pp. 693–706, Jul. 1993, doi: https://doi.org/10.1111/j.2517-6161.1993.tb01934.x.

[14] Christophe Croux and Gentiane Haesbroeck, “Implementing the Bianco and Yohai estimator for logistic regression,” vol. 44, no. 1–2, pp. 273–295, Oct. 2003, doi: https://doi.org/10.1016/s0167-9473(03)00042-2.

[15] L. A. Stefanski, “The effects of measurement error on parameter estimation,” Biometrika, vol. 72, no. 3, pp. 583–592, Jan. 1985, doi: https://doi.org/10.1093/biomet/72.3.583.

[16] P. J. Rousseeuw and A. M. Leroy, Robust regression and outlier detection. Hoboken, Nj: Wiley-Interscience, 2003.

[17] P. J. Rousseeuw, “Tutorial to robust statistics,” Journal of Chemometrics, vol. 5, no. 1, pp. 1–20, Jan. 1991, doi: https://doi.org/10.1002/cem.1180050103.

[18] A. Fitrianto and T. Wendy, “Identification of high leverage points in binary logistic regression,” AIP Conference Proceedings, 2016, doi: https://doi.org/10.1063/1.4966096.

[19] A. H. M. R. Imon and A. S. Hadi, “Identification of multiple high leverage points in logistic regression,” Journal of Applied Statistics, vol. 40, no. 12, pp. 2601–2616, Jul. 2013, doi: https://doi.org/10.1080/02664763.2013.822057.

[20]B. A. Syaiba and M. Habshah, “Robust Logistic Diagnostic for the Identification of High Leverage Points in Logistic Regression Model,” Journal of Applied Sciences, vol. 10, no. 23, pp. 3042–3050, Dec. 2010, doi: https://doi.org/10.3923/jas.2010.3042.3050.

[21] S. K. Sarkar, H. Midi, and S. Rana, “Detection of Outliers and Influential Observations in Binary Logistic Regression: An Empirical Study,” Journal of Applied Sciences, vol. 11, no. 1, pp. 26–35, Dec. 2010, doi: https://doi.org/10.3923/jas.2011.26.35.

[22] T. J. Santner, and D. E. Duffy (1986), “A note on A. Albert and JA Anderson’s conditions for the existence of maximum likelihood estimates in logistic regression models”, Biometrika, vol.73, no. 3, pp.755-758, Dec1986, https://doi.org/10.1093/biomet/73.3.755

[23]R. Tibshirani, “Regression Shrinkage and Selection Via the Lasso,” Journal of the Royal Statistical Society: Series B (Methodological), vol. 58, no. 1, pp. 267–288, Jan. 1996, doi: https://doi.org/10.1111/j.2517-6161.1996.tb02080.x.
[24]J. Berkson, “Application to the Logistic Function to Bio-Assay,” Journal of the American Statistical Association, vol. 39, no. 227, p. 357, Sep. 1944, doi: https://doi.org/10.2307/2280041.
[25] S. K. Shevade and S. S. Keerthi, “A simple and efficient algorithm for gene selection using sparse logistic regression,” Bioinformatics, vol. 19, no. 17, pp. 2246–2253, Nov. 2003, doi: https://doi.org/10.1093/bioinformatics/btg308.
[26] J. Zhu, “Classification of gene microarrays by penalized logistic regression,” Biostatistics, vol. 5, no. 3, pp. 427–443, Jul. 2004, doi: https://doi.org/10.1093/biostatistics/kxg046.
[27] M. Y. Park and T. Hastie, “Penalized logistic regression for detecting gene interactions,” Biostatistics, vol. 9, no. 1, pp. 30–50, Apr. 2007, doi: https://doi.org/10.1093/biostatistics/kxm010.
[28] H. Sun and S. Wang, “Penalized logistic regression for high-dimensional DNA methylation data with case-control studies,” Bioinformatics, vol. 28, no. 10, pp. 1368–1375, Mar. 2012, doi: https://doi.org/10.1093/bioinformatics/bts145.
[29] Y. Plan and R. Vershynin, “Robust 1-bit Compressed Sensing and Sparse Logistic Regression: A Convex Programming Approach,” vol. 59, no. 1, pp. 482–494, Jan. 2013, doi: https://doi.org/10.1109/tit.2012.2207945.
[30] L. Yang and Y. Qian, “A sparse logistic regression framework by difference of convex functions programming,” Applied Intelligence, vol. 45, no. 2, pp. 241–254, Feb. 2016, doi: https://doi.org/10.1007/s10489-016-0758-2.
[31] C.-H. Zhang and J. Huang, “The sparsity and bias of the Lasso selection in high-dimensional linear regression,” The Annals of Statistics, vol. 36, no. 4, Aug. 2008, doi: https://doi.org/10.1214/07-aos520.
[32] J. Fan and R. Li, “Variable Selection via Nonconcave Penalized Likelihood and its Oracle Properties,” Journal of the American Statistical Association, vol. 96, no. 456, pp. 1348–1360, Dec. 2001, doi: https://doi.org/10.1198/016214501753382273.
[33] C. Leng, Y. Lin, and G. Wahba, “A Note on the Lasso and Related Procedures in Model Selection” Jan. 2006.
[34] H. Zou, “The Adaptive Lasso and Its Oracle Properties,” Journal of the American Statistical Association, vol. 101, no. 476, pp. 1418–1429, Dec. 2006, doi: https://doi.org/10.1198/016214506000000735.
[35] A. Basu, A. Ghosh, M. Jaenada, and L. Pardo, “Robust adaptive Lasso in high-dimensional logistic regression,” arXiv.org, Apr. 07, 2023. https://arxiv.org/abs/2109.03028 (accessed Aug. 11, 2023).
[36] A. Ghosh and S. Majumdar, “Ultrahigh-Dimensional Robust and Efficient Sparse Regression Using Non-Concave Penalized Density Power Divergence,” IEEE Transactions on Information Theory, vol. 66, no. 12, pp. 7812–7827, Dec. 2020, doi: https://doi.org/10.1109/tit.2020.3013015.
[37] A. Ghosh, M. Jaenada, and L. Pardo, “Robust adaptive variable selection in ultra-high dimensional linear regression models,” arXiv.org, Sep. 24, 2020. https://arxiv.org/abs/2004.05470 (accessed Aug. 12, 2023).
[38] A. M. Bianco, G. Boente, and G. Chebi, “Penalized robust estimators in sparse logistic regression,” Test, vol. 31, no. 3, pp. 563–594, Nov. 2021, doi: https://doi.org/10.1007/s11749-021-00792-w.
[39] Hosmer, D.W., Lemeshow, S., (2000). Applied Logistic Regression. 2nd ed., Wiley, New York.
[40] Rousseeuw, P. J., & Driessen, K. V. (1999). A fast algorithm for the minimum covariance determinant estimator. Technometrics, 41(3), 212-223.