

REDESCENDING M-ESTIMATOR BASED LASSO FOR FEATURE SELECTION

R. MUTHUKRISHNAN¹ AND C. K. JAMES²

•

^{1,2}Department of Statistics

Bharathiar University, Tamil Nadu

Coimbatore-641046, India

¹muthukrishnan1970@gmail.com, ²jamesck74@gmail.com

Abstract

Aim: Regression analysis is one of the statistical methods which helps to model the data and helps in prediction, a large data set with higher number of variables will often create problem due to its dimensionality and hence create difficulties to gather important information from the data, so it is a need of a method which can simultaneously choose important variables which contains most of the information and hence helps to fit the model. Least absolute shrinkage and selection operator (LASSO) is a popular choice for shrinkage estimation and variable selection. But LASSO uses the conventional least squares technique for feature selection which is very sensitive to outliers. As a result, when the data set is contaminated with bad observations (Outliers), the LASSO technique gives unreliable results, so in this paper the focus is to create a method which can resist to outliers in the data and helps in giving a meaningful result. Method: proposed a new procedure, a LASSO method by adding weights which uses the concept of redescending M-estimator, which can resist outliers in both dependent and independent variables. The observation with greater importance receives a higher weight and less weight to the least important observation. Findings: The efficiency of the proposed method has been studied in the real and simulation environment and compared with other existing procedures with measures like Median Absolute Error (MDAE), False Positive Rate (FPR), False Negative Rate (FNR), Mean Absolute Percentage Error (MAPE). The proposed method with the redescending M-estimator shows a higher resistance to outliers compared to conventional LASSO and other robust existing procedures. Conclusion: The study reveals that the proposed method outperforms other existing procedures in terms of MDAE, FPR, FNR and MAPE, indicating its superior performance in variables selection under outlier contaminated datasets.

Keywords: Feature Selection, LASSO, MAPE

I. Introduction

One of the most frequent problems we run into in real-time applications and other scientific fields is data including outliers. The existence of outliers, according to Chatterjee and Hadi [4], may leads to influence the parameter estimation and inaccurate predictions for traditional approaches. Both dependent variables and covariates (predictor variables) may contain outliers. As a result, it's crucial to deal with outliers in regression analysis. For outlier detection problems, numerous robust regression algorithms have been created, such as S- estimator [7], the least median of squares estimator [11], the MM- estimator [18], the τ - estimator [19], and so on. It is well known that there are some M- estimator-based regression methods such as Huber regression which does not delete large residuals, and Tukey regression is not Robust against the outliers in the leverage points. [12], Redescending M-estimators are more resilient than M-estimators since they totally reject extreme outliers. Alamgir et al. [1] proposed an efficient Redescending M-estimator for robust estimation.

In practice, a large number of variables are often incorporated at the beginning of modelling. The

interpretation of models that contain all of the variables is extremely difficult, even when irrelevant factors may increase variance. Therefore, one of the most significant issues in data analysis is the choice of an important variable. Popular methods for variable selection are penalized regression methods such as Least Absolute Shrinkage and Selection Operator (LASSO) [13], Smoothly Clipped Absolute Deviation (SCAD) [6] and adaptive LASSO [20], and so on. Most of the methods mentioned above are closely related to Ordinary Least Squares (OLS) technique. OLS-based methods are not resistant to outliers therefore the outliers can cause problems in a variable selection based on OLS. So, the robust variable selection approach has to be studied. There are numerous effective variable selection techniques in the literature, including the Least Absolute Deviation (LAD)-LASSO [15], which deals with heavy-tailed error, WLAD-LASSO [14], the weighted Wilcoxon-type SCAD method [13], the Huber's criterion and adaptive lasso penalty [9], the quantile regression for analyzing heterogeneity ultra-high dimension [16], the variable selection in the semiparametric varying- coefficient partially linear model via a penalized composite quantile loss [8], the Composite Quantile Regression (CQR) [21], the variable selection with the exponential squared loss [17], Penalized least trimmed square (LTS) [2], Maximum Tangent Likelihood Estimator (MTE) [10], and so on. In this paper, a new robust feature selection method has been introduced. The improved version of the LASSO method uses a weight from a Redescending M-Estimator which can tolerate outliers in the X-Y space. The study based on simulation and real data indicates that the proposed robust feature selection procedure performs better than other existing methods.

The paper is organized as follows. In section II provides brief introduction to LASSO method. A new technique, Alarm Weight LASSO (AW-LASSO) and its corresponding algorithm is described in Section III. In section IV, real data analysis and a simulation study and is carried out to comprehend how well the suggested method works. Finally, section V gives summary and conclusion.

II. LASSO Methods

Regressions models are commonly used in statistical analysis. A popular use is to model the predicted risk of a likely outcome. Unfortunately, using standard regression techniques to create a model from a set of candidate variables often results in overfitting, which increases the number of variables that are eventually included in the model and overestimates how well the included variables explain the observed variability (an effect known as optimism bias). Extreme (extremely low or very high) risk observations are particularly difficult for the model to forecast. A shrinkage and variable selection strategy for regression models is LASSO regression. In order to create a model that minimizes the prediction error, LASSO regression seeks to discover the variables and corresponding regression coefficients. This is accomplished by placing a restriction on the model parameters that causes the regression coefficients to decrease towards zero, or more specifically, by requiring that the total absolute value of the regression coefficients be smaller than a predetermined value (λ). The equation of the LASSO is given below

$$\beta_{LASSO} = \min_{\beta} \left(\sum_{i=1}^n \left(y_i - \sum_j \beta_j X_{ij} \right)^2 + \lambda \sum_j |\beta_j| \right) \quad (1)$$

Since λ controls the amount of regularization the choice of λ is often made by using an automated k-fold cross validation approach. If $\lambda = 0$ the LASSO is same as OLS. As λ increases, the number of non-zero components of β decreases, at $\lambda = \infty$, the LASSO gives the null model. The above LASSO method is based on OLS loss function which is not resistant to outliers, therefore, to address this issue, we modified the LASSO by adding a new weight to form Alarm Weight LASSO which is Elaborately discussed in the section III.

III. Alarm Weight LASSO

Consider the linear regression model

$$y_i = \beta_o + x_i^T \beta + \varepsilon_i, \quad i = 1, 2, 3, \dots, n \quad (2)$$

where Y_i is the response variable $X_i = (x_{i1}, x_{i2}, \dots, x_{ip})^T$ is the p -dimensional covariate vector, $\beta = (\beta_1, \beta_2, \dots, \beta_p)^T$ are the regression parameters, and ε_i are the iid random errors. We assume that $\beta_o = 0$. This can be achieved by centering the covariates and response variable. That is, from now on we will consider the model

$$y_i = x_i^T \beta + \varepsilon_i, \quad i = 1, 2, 3, \dots, n \quad (3)$$

To estimate β is to minimize the ordinary least square (OLS) criterion

$$\sum_{i=1}^n (y_i - x_i^T \beta)^2 \quad (4)$$

The OLS estimates β by minimizing the error sum of squares, i.e.,

$$\hat{\beta}_{OLS} = \min_{\beta} \left\{ (Y - X\beta)^T (Y - X\beta) \right\} \quad (5)$$

The OLS approach to estimate the regression parameter is very sensitive to the outliers. One of the alternatives to OLS is to use weighted OLS. Weighted regression is its robustness against outliers. Weighted regression can assign less weight to outliers and hence reduce their impact on the estimate of the coefficients. Which is obtained by minimizing the OLS criterion

$$\sum_{i=1}^n w_i (y_i - x_i^T \beta)^2 \quad (6)$$

where w_i , for $i = 1, 2, 3, \dots, n$ is the weights which is determined by a redescending M-Estimator which can resist outliers in both X and Y space. The influence function describes the sensitivity of the overall estimate of the outlying data and is defined as

$$\psi(r) = \begin{cases} \frac{16re^{-2(r/c)^2}}{(1 + e^{-(r/c)^2})^4} & , |r| \leq c \\ 0 & , |r| > c \end{cases} \quad (7)$$

The functional relationship between ψ and ρ is given by

$$\psi(r) = \frac{d}{dr} \rho(r) \quad (8)$$

Integrating out the ψ - function under the initial condition, we get the corresponding $\rho(r)$, given by

$$\rho(r) = \begin{cases} \frac{2c^2}{3} \left[1 - \frac{2(1 + 3e^{(r/c)^2})}{(1 + e^{(r/c)^2})^3} \right] & , |r| \leq c \\ \frac{2c^2}{3} \left[1 - \frac{2(1 + 3e)}{(1 + e)^3} \right] & , |r| > c \end{cases} \quad (9)$$

The weight function $w(r) = \psi(r)/r$, is as follows

$$w(r) = \begin{cases} \frac{(4e^{-(r/c)^2})^2}{(1 + e^{-(r/c)^2})^4} & , |r| \leq c \\ 0 & , |r| > c \end{cases} \quad (10)$$

where “ r ” denotes residual and “ c ” is tuning constant.

Efficiency and robustness are two characteristic of a robust procedures that are inversely connected. As a result, choose an estimator with the highest resistance and the lowest efficiency loss. Nobody can afford to select a highly robust estimator that is resistant to outliers at the expense of decreasing efficiency. These two properties should be balanced in some way. The weight function ensures that the residuals with the highest weight (close to 1) correspond to the majority of good observations. Figure 1, represents the weight function $w(r)$ of redescending M-estimator

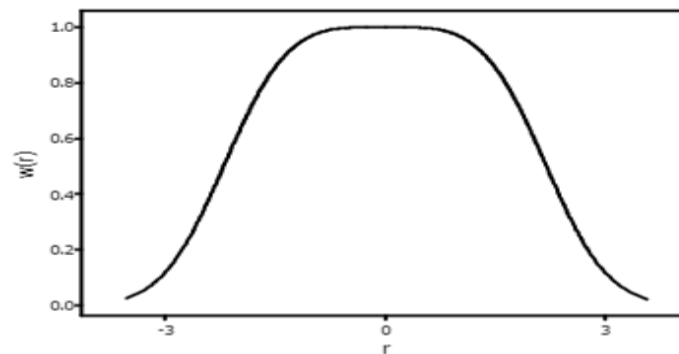


Fig.1: Alarm Weight function of redescending M estimator

From Figure 1, As can be seen, only severely outlying observations are given 0 weights, ensuring that good observations are used to their full potential and that extreme outliers are not overly relied upon. The equation given in (1) thus modified by adding the weight of the redescending M estimator to form Alarm Weight LASSO and the equation is given below.

$$\beta_{AW-LASSO} = \min_{\beta} \left(\sum_{i=1}^n w_i \left(y_i - \sum_j \beta_j X_{ij} \right)^2 + \lambda \sum_j |\beta_j| \right) \quad (11)$$

w_i represents the weight function and λ is the tuning parameter which is chosen by cross validation method.

I. Computational Algorithm

Consider the LASSO model given in (1). We use Iteratively Least Square Method (IRWLS) algorithm for the computation of the AW-LASSO.

Step1: Find the initial estimates of β by using the ridge regression model.

Step2: obtain the corresponding residuals from our initial estimates

Step3: Compute the corresponding weights based on the proposed weight function

Step4: Calculate the new estimate of AW-LASSO coefficients using the IRWLS algorithm

Step5: Repeat step 2 to 4 until convergence.

IV. Experimental Results

In this section various LASSO-type feature selection techniques are compared to the proposed methodology in a real-world setting. Outliers were present in the real data, which were eliminated using the cook distance [5], and the analysis was done using the R programming language. The obtained results such as Median Absolute Error (MDAE), Mean Absolute Percentage Error (MAPE), number of variables, for with and without outliers are also discussed.

I. Real data examples

Here we considered two data sets, namely Boston housing data and diabetes data, detailed descriptions are available in the standard packages. The Boston data set has 506 observations of 15 independent and a dependent variable. The diabetes data set has 442 observations of 9 independent and a dependent variable. The feature selection procedures have been performed after standardizing the variables, with and without outliers and the results are summarised in the table 1.

Table 1: Error values, under with and without outliers

Methods	MDAE	MAPE	No. of Variables Selected	Selected Variables
Boston Housing Data				
LASSO	0.303 (0.123)	4.27 (0.783)	12(12)	Tract, ion, lat, Crim, Zn, nox, rm, dis, tax, ptratio, b, istat (ion, crim, zn, indus, nox, rm, age, dis, tax, ptratio, b, istat)
LAD LASSO	0.267 (0.177)	4.07 (1.20)	12(15)	Tract, ion, crim, zn, nox, rm, age, dis, tax, ptratio, b, istat (Tract, ion, lat, crim, zn, indus, nox, rm, age, dis, rad, tax, ptratio, b, istat)
Huber LASSO	0.291 (0.189)	3.97 (1.17)	11(12)	Tract, ion, crim, nox, rm, age, dis, tax, ptratio, b, istat (Tract, ion, crim, zn, indus, rm, age, dis, tax, ptratio, b, istat)
MTE LASSO	0.306 (0.245)	4.24 (1.32)	10(11)	Tract, ion, crim, nox, rm, dis, tax, ptratio, b, istat (ion, crim, zn, indus, rm, age, dis, tax, ptratio, b, istat)
AW-LASSO	0.301 (0.124)	3.59 (0.783)	6(11)	ion, rm, tax, ptratio, b, istat (ion, crim, zn, indus, rm, age, dis, tax, ptratio, b, istat)
Diabetes Data				
LASSO	0.516 (0.511)	1.15 (1.48)	4(4)	BMI, BP, S3, S5, (BMI, BP, S3, S5)
LAD LASSO	0.490 (0.527)	1.29 (1.57)	4(4)	BMI, BP, S3, S5, (BMI, BP, S3, S5)
Huber LASSO	0.501 (0.510)	1.23 (1.53)	4(4)	BMI, BP, S3, S5, (BMI, BP, S3, S5)
MTE LASSO	0.517 (0.504)	1.15 (1.46)	4(4)	Bmi, bp, S3, S5, (BMI, BP, S3, S5)
AW-LASSO	0.485 (0.510)	1.07 (1.49)	4(4)	bmi, BP, S3, S5, (BMI, BP, S3, S5)

(.) without outlier

From the above table it is observed that under with and without outliers the proposed procedure, AW-LASSO produces the error values are minimum and also select the significant variables when compared with the other procedures.

II. Simulation Study

Simulation studies are carried out to check the efficacy of various methods. In our simulation study, the covariates are generated from a multivariate normal distribution with Mean $\mu = [o]_{p \times 1}$ and variance

$\Sigma = [\sigma_{ij}] = \rho^{|i-j|}$ for various levels of correlation, $\rho = 0.01, 0.5, 0.9$ and number of variables, $p=10, 15, 25$. The model consists of six significant variables and the rest is considered noise variables. The performance of the proposed method is compared with various other robust methods along with the classical LASSO method. Various levels of contamination (0%, 5%, 10%, 20%) are studied for sample size $n = 100, 200, 1000$.

Table 2: False Negative and False Positive rate of each method under various levels of contamination (0% and 5%)

Method	N	Error	$\eta = 0$									$\eta = 5$								
			$p = 10$			$p = 15$			$p = 25$			$p = 10$			$p = 15$			$p = 25$		
			ρ	ρ	ρ	ρ	ρ	ρ	ρ	ρ	ρ	ρ	ρ	ρ	ρ	ρ	ρ	ρ	ρ	ρ
LASSO	100	FPR	0.36	1.02	0.92	0.64	0.75	0.69	0.43	0.55	0.54	0.93	0.99	0.92	0.69	0.73	0.66	0.47	0.56	0.47
		FNR	0.00	0.00	0.03	0.00	0.00	0.05	0.00	0.00	0.00	0.08	0.00	0.00	0.07	0.00	0.00	0.13	0.00	0.01
	200	FPR	0.96	0.99	1.03	0.68	0.72	0.74	0.46	0.57	0.57	0.88	1.00	1.00	0.66	0.76	0.75	0.45	0.57	0.52
		FNR	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.03	0.00	0.00
	1000	FPR	0.90	1.02	0.99	0.67	0.74	0.78	0.44	0.53	0.58	0.96	1.01	1.03	0.62	0.72	0.77	0.42	0.54	0.58
		FNR	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
LAD LASSO	100	FPR	0.59	0.73	0.77	0.27	0.42	0.43	0.10	0.18	0.23	0.49	0.67	0.66	0.22	0.31	0.28	0.06	0.09	0.15
		FNR	0.00	0.01	0.16	0.03	0.06	0.26	0.16	0.18	0.41	0.04	0.05	0.26	0.13	0.14	0.46	0.00	0.40	0.62
	200	FPR	0.67	0.78	0.87	0.31	0.38	0.45	0.09	0.16	0.27	0.49	0.70	0.79	0.20	0.30	0.40	0.06	0.12	0.18
		FNR	0.00	0.00	0.04	0.00	0.00	0.15	0.00	0.04	0.22	0.01	0.00	0.12	0.02	0.04	0.22	0.16	0.14	0.42
	1000	FPR	0.58	0.78	0.81	0.28	0.42	0.52	0.09	0.17	0.29	0.55	0.69	0.78	0.17	0.30	0.43	0.07	0.11	0.22
		FNR	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.08
Huber LASSO	100	FPR	0.56	0.71	0.74	0.24	0.37	0.40	0.09	0.15	0.20	0.47	0.60	0.63	0.21	0.27	0.25	0.05	0.07	0.12
		FNR	0.00	0.01	0.15	0.03	0.06	0.25	0.17	0.19	0.40	0.04	0.05	0.25	0.13	0.15	0.46	0.47	0.42	0.62
	200	FPR	0.64	0.73	0.82	0.28	0.33	0.39	0.08	0.14	0.22	0.48	0.67	0.74	0.18	0.27	0.36	0.06	0.10	0.15
		FNR	0.00	0.00	0.04	0.00	0.00	0.12	0.00	0.04	0.20	0.01	0.00	0.11	0.02	0.04	0.21	0.15	0.14	0.43
	1000	FPR	0.55	0.74	0.77	0.26	0.39	0.46	0.09	0.15	0.23	0.55	0.64	0.75	0.16	0.27	0.39	0.06	0.09	0.18
		FNR	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.04
MTE LASSO	100	FPR	0.51	0.69	0.73	0.15	0.29	0.39	0.05	0.08	0.18	0.33	0.47	0.54	0.11	0.15	0.19	0.02	0.03	0.10
		FNR	0.35	0.16	0.22	0.69	0.40	0.33	0.75	0.66	0.54	0.55	0.38	0.40	0.79	0.66	0.64	0.91	0.87	0.73
	200	FPR	0.94	0.92	0.97	0.60	0.60	0.64	0.15	0.26	0.39	0.61	0.78	0.84	0.25	0.32	0.43	0.03	0.07	0.16
		FNR	0.00	0.00	0.00	0.00	0.00	0.02	0.31	0.04	0.09	0.20	0.09	0.08	0.44	0.24	0.17	0.61	0.47	0.36
	1000	FPR	0.90	0.97	0.93	0.67	0.70	0.74	0.44	0.48	0.53	0.96	0.97	0.96	0.61	0.66	0.71	0.42	0.48	0.53
		FNR	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
AW-LASSO	100	FPR	0.37	0.51	0.51	0.22	0.34	0.35	0.13	0.21	0.22	0.27	0.39	0.37	0.13	0.22	0.17	0.07	0.13	0.12
		FNR	0.00	0.00	0.26	0.00	0.00	0.27	0.00	0.01	0.30	0.00	0.01	0.46	0.01	0.05	0.57	0.00	0.10	0.60
	200	FPR	0.32	0.47	0.54	0.15	0.26	0.32	0.11	0.17	0.23	0.25	0.35	0.46	0.12	0.19	0.27	0.06	0.11	0.17
		FNR	0.00	0.00	0.13	0.00	0.00	0.16	0.00	0.00	0.15	0.00	0.01	0.30	0.00	0.01	0.31	0.00	0.02	0.34
	1000	FPR	0.26	0.33	0.49	0.11	0.18	0.30	0.05	0.09	0.19	0.25	0.27	0.46	0.11	0.13	0.25	0.05	0.06	0.15
		FNR	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.01	0.00	0.00	0.09	0.00	0.00	0.10	0.00	0.00	0.11

The performance of each model is measured by using MDAE, False Negative Rate (FNR), and False Positive Rate (FPR). FNR is defined as the proportion of zero coefficient estimates whose corresponding true coefficients are nonzero and FPR is defined as the proportion of nonzero coefficient estimates whose corresponding true coefficients are zero. The obtained results are summarized in Table 2-5. Also, for effective understanding of the performance of various methods, the pictorial representations of error measures are given in Figure 2 and 3 respectively.

Table 3: False Negative and False Positive Rate of each method under various levels of contamination (10% and 20%)

Method	N	Error	$\eta = 10$									$\eta = 20$								
			$p = 10$			$p = 15$			$p = 25$			$p = 10$			$p = 15$			$p = 25$		
			ρ	ρ	ρ	ρ	ρ	ρ	ρ	ρ	ρ	ρ	ρ	ρ	ρ	ρ	ρ	ρ	ρ	ρ
LASSO	100	FPR	0.93	1.00	0.83	0.65	0.73	0.53	0.48	0.54	0.33	0.90	0.93	0.65	0.63	0.71	0.38	0.45	0.52	0.20
		FNR	0.00	0.00	0.18	0.00	0.01	0.27	0.00	0.02	0.42	0.00	0.04	0.31	0.01	0.04	0.48	0.01	0.08	0.63
	200	FPR	0.92	1.02	0.97	0.65	0.73	0.68	0.45	0.52	0.50	0.93	1.01	0.87	0.65	0.75	0.65	0.41	0.54	0.42
		FNR	0.00	0.00	0.05	0.00	0.00	0.08	0.47	0.00	0.11	0.00	0.00	0.12	0.00	0.00	0.11	0.00	0.01	0.24

LAD LASSO	1000	FPR	0.89	0.97	1.01	0.67	0.74	0.76	0.00	0.57	0.00	0.92	1.02	0.99	0.68	0.75	0.75	0.46	0.53	0.57
		FNR	0.00	0.00	0.00	0.00	0.00	0.00	0.47	0.00	0.20	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00
	100	FPR	0.44	0.65	0.51	0.16	0.23	0.21	0.03	0.06	0.07	0.39	0.49	0.26	0.09	0.18	0.09	0.00	0.04	0.02
		FNR	0.14	0.09	0.44	0.35	0.32	0.64	0.71	0.57	0.82	0.32	0.29	0.66	0.63	0.49	0.83	0.92	0.77	0.94
	200	FPR	0.52	0.64	0.69	0.21	0.24	0.28	0.05	0.08	0.11	0.42	0.55	0.52	0.15	0.21	0.21	0.02	0.05	0.06
		FNR	0.01	0.02	0.19	0.13	0.10	0.38	0.40	0.30	0.58	0.06	0.07	0.39	0.29	0.23	0.51	0.75	0.57	0.76
1000	FPR	0.46	0.63	0.79	0.20	0.27	0.39	0.07	0.10	0.14	0.44	0.60	0.78	0.16	0.23	0.38	0.05	0.07	0.17	
	FNR	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.63	0.00	0.00	0.03	0.00	0.00	0.11	0.00	0.04	0.20	
Huber LASSO	1000	FPR	0.43	0.58	0.50	0.16	0.19	0.19	0.03	0.05	0.06	0.38	0.46	0.26	0.09	0.15	0.08	0.01	0.04	0.01
		FNR	0.14	0.09	0.44	0.35	0.31	0.60	0.69	0.58	0.79	0.31	0.29	0.64	0.61	0.50	0.81	0.86	0.77	0.91
	100	FPR	0.50	0.62	0.63	0.20	0.22	0.24	0.05	0.07	0.08	0.41	0.54	0.48	0.14	0.19	0.18	0.02	0.04	0.05
		FNR	0.01	0.02	0.19	0.14	0.10	0.38	0.39	0.31	0.59	0.05	0.07	0.40	0.28	0.24	0.52	0.72	0.58	0.79
	200	FPR	0.44	0.60	0.76	0.18	0.23	0.34	0.06	0.08	0.11	0.43	0.57	0.71	0.16	0.22	0.33	0.05	0.07	0.13
		FNR	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.60	0.00	0.00	0.02	0.00	0.00	0.08	0.00	0.04	0.20
MTE LASSO	1000	FPR	0.25	0.40	0.41	0.07	0.09	0.15	0.01	0.03	0.04	0.24	0.30	0.21	0.04	0.07	0.06	0.01	0.02	0.01
		FNR	0.65	0.51	0.55	0.85	0.79	0.72	0.90	0.90	0.85	0.72	0.64	0.73	0.89	0.85	0.88	0.92	0.92	0.92
	100	FPR	0.37	0.50	0.59	0.12	0.12	0.21	0.01	0.02	0.08	0.22	0.30	0.34	0.06	0.07	0.10	0.01	0.01	0.03
		FNR	0.46	0.26	0.25	0.71	0.58	0.50	0.86	0.82	0.63	0.67	0.53	0.56	0.84	0.80	0.69	0.92	0.92	0.85
	200	FPR	0.87	0.88	0.94	0.66	0.69	0.69	0.47	0.51	0.00	0.78	0.87	0.90	0.53	0.63	0.65	0.26	0.39	0.47
		FNR	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.18	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.02	0.02
AW-LASSO	1000	FPR	0.29	0.38	0.23	0.12	0.18	0.10	0.06	0.12	0.05	0.27	0.37	0.14	0.17	0.21	0.06	0.08	0.14	0.03
		FNR	0.02	0.11	0.68	0.03	0.16	0.73	0.01	0.16	0.84	0.07	0.25	0.79	0.06	0.27	0.85	0.09	0.31	0.90
	100	FPR	0.25	0.31	0.36	0.11	0.17	0.20	0.06	0.09	0.13	0.26	0.37	0.27	0.12	0.21	0.13	0.06	0.12	0.08
		FNR	0.00	0.04	0.43	0.00	0.04	0.47	0.00	0.07	0.50	0.01	0.10	0.59	0.00	0.11	0.67	0.00	0.15	0.73
	200	FPR	0.25	0.27	0.46	0.11	0.12	0.24	0.05	0.06	0.16	0.25	0.27	0.46	0.11	0.14	0.26	0.05	0.06	0.16
		FNR	0.00	0.00	0.13	0.00	0.00	0.00	0.00	0.00	0.58	0.00	0.00	0.18	0.00	0.00	0.18	0.00	0.00	0.22

False Positive Rate (Figure-2a), the selection of insignificant variables by the method. In these circumstances, the conventional LASSO has a high False Positive Rate relative to other approaches, it is because the LASSO tends to select a greater number of coefficients, while the MTE method's False Positive Rate rises as the sample size increases. In almost all situations, the AW-LASSO approach has a very low False Positive Rate.

False Negative Rate (Figure-2b), or the number of significant variables that the technique failed to choose. The AW-LASSO False Negative Rate is usually always zero at all levels, but when the correlation level rises, the approach produces inconsequential results. The MTE technique shows a high rate of False Negative, however as the sample size grows, the approach tends to converge to zero.

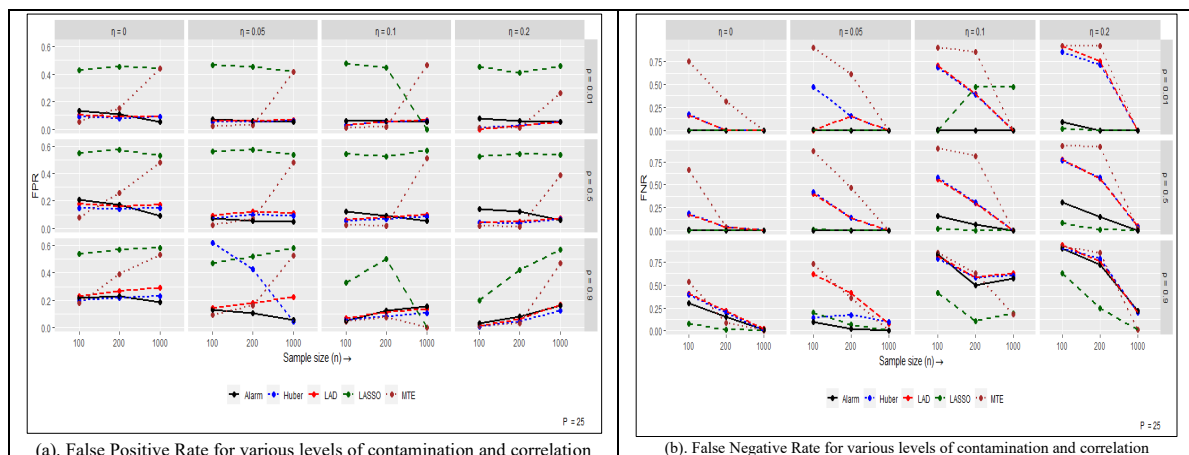


Figure 2: FPR and FNR under various levels of contamination and correlation

Table 4: Median Absolute Error of each method under various levels of contamination (0% and 5%)

Method	N	$\eta = 0$									$\eta = 5$								
		$\rho = 10$			$\rho = 15$			$\rho = 25$			$\rho = 10$			$\rho = 15$			$\rho = 25$		
		ρ			ρ			ρ			ρ			ρ			ρ		
		0.01	0.50	0.90	0.01	0.50	0.90	0.01	0.50	0.90	0.01	0.50	0.90	0.01	0.50	0.90	0.01	0.50	0.90
LASSO	100	1.90	1.91	1.91	1.90	1.87	1.90	1.87	1.84	1.85	2.80	2.49	2.49	2.55	2.70	2.50	2.57	2.56	2.50
	200	1.96	1.96	1.96	1.97	1.96	1.96	1.94	1.93	1.94	2.85	2.37	2.34	2.39	2.75	2.37	2.44	2.41	2.39
	1000	2.02	2.03	2.02	2.02	2.02	2.02	2.01	2.01	2.01	2.80	2.35	2.22	2.30	2.80	2.24	2.35	2.30	2.23
LAD LASSO	100	1.99	2.01	2.02	2.37	2.10	2.05	3.10	2.57	2.09	2.51	2.24	2.23	2.40	2.57	2.36	2.44	2.35	2.43
	200	1.98	1.97	1.99	2.11	2.06	2.02	2.36	2.30	2.11	2.29	2.15	2.13	2.20	2.39	2.25	2.37	2.34	2.35
	1000	2.05	2.06	2.08	2.03	2.03	2.02	2.08	2.06	2.04	2.13	2.13	2.14	2.21	2.17	2.17	2.28	2.25	2.21
Huber LASSO	100	1.97	2.00	1.99	2.37	2.10	2.02	3.10	2.54	2.09	2.48	2.22	2.21	2.41	2.64	2.34	2.43	2.35	2.42
	200	1.98	1.97	2.00	2.11	2.06	2.01	2.37	2.31	2.09	2.28	2.15	2.12	2.22	2.37	2.23	2.39	2.30	2.37
	1000	2.06	2.07	2.08	2.03	2.03	2.02	2.08	2.07	2.04	2.13	2.13	2.14	2.24	2.18	2.16	2.28	2.25	2.20
MTE LASSO	100	3.03	2.15	1.99	4.44	2.75	2.06	4.76	3.44	2.10	2.30	2.20	2.16	2.43	2.63	2.36	2.35	2.26	2.41
	200	1.96	1.96	1.99	2.12	2.01	2.07	3.09	2.10	2.16	2.45	2.30	2.08	2.22	2.57	2.14	2.36	2.28	2.24
	1000	2.06	2.06	2.07	2.02	2.02	2.08	2.20	2.15	2.12	2.14	2.15	2.14	2.21	2.20	2.14	2.16	2.15	2.14
AW-LASSO	100	1.95	1.95	1.96	2.12	2.09	1.97	2.14	2.15	2.07	2.14	2.13	2.10	2.14	2.16	2.14	2.15	2.12	2.12
	200	1.97	1.97	1.98	2.10	2.04	1.98	2.11	2.13	2.09	2.16	2.12	2.12	2.15	2.17	2.14	2.14	2.12	2.12
	1000	2.03	2.02	2.02	2.00	1.99	2.00	2.05	2.00	2.02	2.14	2.15	2.13	2.16	2.26	2.17	2.16	2.15	2.15

Table 5: Median Absolute Error of each method under various levels of contamination (10% and 20%)

Method	N	$\eta = 10$									$\eta = 20$								
		$\rho = 10$			$\rho = 15$			$\rho = 25$			$\rho = 10$			$\rho = 15$			$\rho = 25$		
		ρ			ρ			ρ			ρ			ρ			ρ		
		0.01	0.50	0.90	0.01	0.50	0.90	0.01	0.50	0.90	0.01	0.50	0.90	0.01	0.50	0.90	0.01	0.50	0.90
LASSO	100	2.90	2.81	2.80	2.96	2.97	2.92	2.96	2.91	2.92	3.88	3.82	3.80	4.12	4.06	4.03	4.20	4.21	4.10
	200	2.89	2.64	2.65	2.74	2.91	2.67	2.96	2.90	2.96	3.97	3.86	3.85	4.06	3.95	3.79	4.21	4.09	4.12
	1000	2.92	2.42	2.55	2.45	2.93	2.41	2.98	2.95	2.93	3.91	3.80	3.81	4.93	3.83	3.73	4.18	4.17	4.14
LAD LASSO	100	2.50	2.49	2.47	2.51	2.46	2.63	2.50	2.41	2.85	2.81	2.96	2.96	3.01	2.89	2.70	3.08	3.07	3.39
	200	2.56	2.35	2.42	2.37	2.49	2.57	2.60	2.40	2.71	2.89	2.99	2.99	3.00	2.85	2.81	3.68	3.75	3.33
	1000	2.53	2.28	2.30	2.36	2.35	2.35	2.51	2.47	2.41	2.77	2.63	2.65	2.84	2.78	2.77	3.29	3.11	2.89
Huber LASSO	100	2.45	2.50	2.46	2.35	2.44	2.61	2.53	2.42	2.83	2.60	2.96	2.97	2.99	2.81	2.80	3.04	3.03	3.39
	200	2.50	2.35	2.39	2.39	2.39	2.55	2.62	2.41	2.71	2.88	3.00	3.03	2.98	2.69	2.80	3.64	3.77	3.33
	1000	2.52	2.28	2.30	2.36	2.35	2.34	2.51	2.48	2.41	2.67	2.63	2.65	2.85	2.78	2.76	3.29	3.11	2.89
MTE LASSO	100	2.60	3.46	2.42	5.72	2.34	2.61	2.70	2.64	2.83	2.82	3.44	3.43	2.92	5.16	3.34	3.29	3.47	3.38
	200	2.62	2.90	2.36	5.23	2.81	2.58	2.75	2.48	2.66	2.84	3.47	3.48	2.96	5.01	3.22	3.17	3.49	3.37
	1000	2.58	2.32	2.31	2.36	2.33	2.31	2.72	2.35	2.42	2.94	2.78	2.79	2.93	2.84	2.72	3.12	2.89	2.75
AW-LASSO	100	2.40	2.32	2.36	2.44	2.37	2.32	2.41	2.48	2.52	2.68	2.62	2.63	2.61	2.68	2.65	2.66	2.69	2.65
	200	2.42	2.34	2.35	2.34	2.41	2.37	2.39	2.40	2.36	2.67	2.64	2.66	2.66	2.65	2.66	2.64	2.70	2.70
	1000	2.44	2.32	2.30	2.34	2.33	2.30	2.36	2.33	2.33	2.65	2.76	2.72	2.69	2.77	2.69	2.70	2.77	2.63

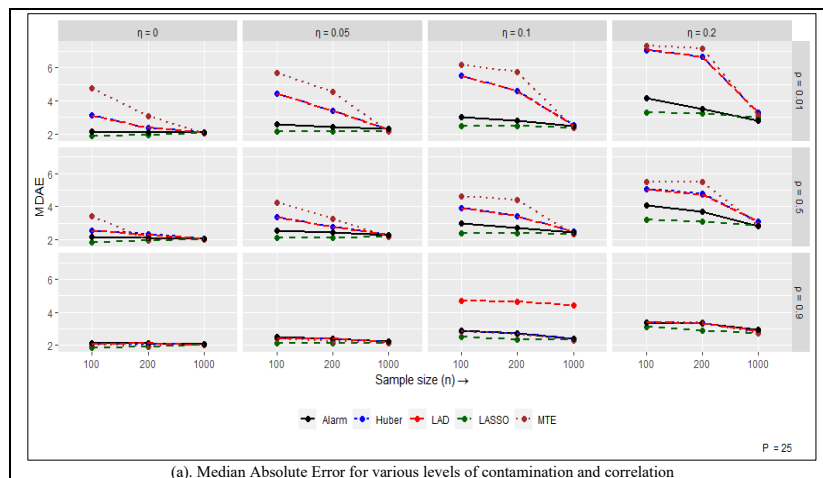


Figure 3: MDAE under various levels of contamination and correlation

The ability of the approaches to predict is seen in Figure 3. In comparison to other approaches AW-LASSO has lower MDAE, but if the data contains no outlier the LASSO exhibits a higher prediction on capacity than other methods and also the MTE method, shows a very high prediction error. However, the AW-LASSO method always has very low prediction error.

IV. Conclusion

Feature selection is a technique that aids in extracting the important variables from a larger range of variables. It is becoming increasingly vital in statistics and is crucial to statistical analysis. In this paper, it is proposed a new feature selection method, which uses a weight function from a redescending M-estimator to modify the ordinary LASSO to form a new feature selection method namely, AW-LASSO. The proposed technique performs well in both with and without outlier condition by examining under the real datasets namely, the Boston Housing and Diabetic data sets. Further, the simulation studies also showed that the superiority of AW-LASSO method over the other methods. The study concluded that the proposed method can be used in the field of statistical learning, specifically in prediction models.

References

- [1] Alamgir, Amjad Ali, Sajjad Ahmad Khan, Dost Muhammad Khan, and Umair Khalil, (2013). A New Efficient Redescending M- Estimator: Alamgir Redescending M- estimator. *Research Journal of Recent Sciences*, 2: 79-91.
- [2] Alfons, C. Croux and Gelper, S. (2013). Sparse least trimmed squares regression for analyzing high-dimensional large data sets. *The Annals of Applied Statistics*, 7: 226-248.
- [3] Arslan, O. (2012). Weighted lad-lasso method for robust parameter estimation and variable selection in regression. *Computational Statistics & Data Analysis*, 56: 1952–1965.
- [4] Chatterjee S. and Hadi A. S. *Sensitivity Analysis in Linear Regression*, John Wiley & Sons, New York, 2009.
- [5] Cook, R. D. (2000). Detection of influential observation in linear regression. *Technometrics*, 42: 65–68.
- [6] Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of American Statistical Association*, 96: 1348–1360.
- [7] Iglewicz, B. and Martinez, J. (1982). Outlier detection using robust measures of scale. *Journal of Statistical Computation and Simulation*, 15: 285–293.
- [8] Kai, B., Li, R., and Zou, H. (2011). New efficient estimation and variable selection methods for semiparametric varying-coefficient partially linear models. *Annals of Statistics*, 39: 305.
- [9] S. Lambert-Lacroix and Zwald, L. (2011). Robust regression through the Huber’s criterion and adaptive lasso penalty. *Electronic Journal of Statistics*, 5: 1015.
- [10] Qin, Y. , Li, S. , Li, Y. and Yu Y. (2017). Penalized maximum tangent likelihood estimation and robust variable selection. *ArXiv Preprint ArXiv: 1708.05439*.
- [11] Wang, H. Li, G. and Jiang, G. (2007). Robust regression shrinkage and consistent variable selection through the lad-lasso. *Journal of Business & Economic Statistics* 25: 347–355.
- [12] Stella Ebele Anekwe, and Sidney Iheanyi Onyeagu, (2021). A Redescending M-Estimator for Detection and Deletion of Outliers in Regression Analysis. *Pakistan Journal of Statistics and Operation Research*, 17: 997-1014.
- [13] Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58: 267-288.
- [14] Wang, L. and Li, R. (2009). Weighted Wilcoxon-type smoothly clipped absolute deviation method. *Biometrics*, 65: 564–571.
- [15] Wang, H. Li, G. and Jiang, G. (2007). Robust regression shrinkage and consistent variable

selection through the lad-lasso. *Journal of Business & Economic Statistics* 25: 347–355.

[16] Wang, L. Wu, Y. and Li, R. (2012). Quantile regression for analysing heterogeneity in ultra-high dimension. *Journal of American Statistical Association*, 107: 214–222.

[17] Wang, X., Jiang, Y., Huang, M. and Zhang, H. (2013). Robust variable selection with exponential squared loss. *Journal of the American Statistical Association*, 108: 632–643.

[18] Yohai, V. J. (1987). High breakdown-point and high efficiency robust estimates for regression. *Annals of Statistics*, 15: 642–656.

[19] Yohai, V. J. and Zamar, R. H. (1988). High breakdown-point estimates of regression by means of the minimization of an efficient scale. *Journal of American Statistical Association*, 83: 406–413.

[20] Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of American Statistical Association*, 101: 1418–1429.

[21] Zou, H and Yuan, M. (2008). Composite quantile regression and the oracle model selection theory. *Annals of Statistics*, 36: 1108–1126.