European Journal of Statistics and Probability, 13 (1) 1-10, 2025
b Print ISSN: 2055-0154(Print)
Online ISSN 2055-0162(Online)
Website: https://www.eajournals.org/
Publication of the European Centre for Research Training and Development -UK

Extended Least Absolute Shrinkage with Selection Operator Technique for Sparse Regression Modeling with High Dimensional Date

WB Yahya.[1], M.U Adehi[2], I.M Al-mustapha [3], M.O. Adenomon [4]
[1]Department of Statistics, Faculty of Natural and Applied Sciences, University of Ilorin, Nigeria;
[2]Department of Statistics, Faculty of Natural and Applied Sciences, Nasarawa State University, Keffi-Nigeria;
[3]Department of Statistics, Faculty of Natural and Applied Sciences, Nasarawa State University, Keffi-Nigeria;
[4]Department of Statistics, Faculty of Natural and Applied Sciences, Nasarawa State University, Keffi-Nigeria

doi: https://doi.org/10.37745/ejsp.2013/vol13n1110 Published February 02, 2025

Citation: WB Yahya., M.U Adehi, I.M Al-mustapha and, M.O. Adenomon (2025) Extended Least Absolute Shrinkage with Selection Operator Technique for Sparse Regression Modeling with High Dimensional Date, *European Journal of Statistics and Probability*, 13 (1) 1-10

Abstract: *This study was carried out on extended least absolute shrinkage with selection operator technique for sparse regression modeling with high dimensional date. The objective of this paper is to advance sparse regression modeling techniques for high-dimensional data through the enhancement of the LASSO algorithm and its application, by developing an extended LASSO model to improve variable selection in high-dimensional datasets. The study posits that the extended LASSO algorithm will effectively address key challenges in high-dimensional data analysis, including multicollinearity and over-fitting. The research design focuses on LASSO formulation and sparsity-Inducing properties using least absolute shrinkage and selection operator (LASSO) formulation. Regularization techniques and their impact on bias-variance trade-off. Regularization techniques adjust the model's complexity to achieve an optimal balance between bias and variance, thereby improving the model's performance on unseen data. This paper hypothesized that the extended LASSO algorithm can be successfully applied to real-life high-dimensional datasets, resulting in improved model performance and greater applicability in various fields. Conclusively, this study offers a valuable contribution to both the theoretical framework of sparse regression modeling and its practical use in tackling high-dimensional data challenges, leading to better decision-making across a range of industries.*

Keywords: least absolute shrinkage, selection operator technique, sparse regression modeling, high dimensional date

INTRODUCTION

High-dimensional data has become increasingly prevalent in modern fields such as genomics, finance, marketing, and image processing, where the number of variables often exceeds the number of observations (Papoutsoglou, et al., 2023). This presents unique challenges in statistical modeling, particularly in areas such as variable selection, prediction accuracy, and model interpret-ability. The imbalance between variables and observations can lead to several issues, including over-fitting, multicollinearity, and computational inefficiency, all of which complicate the development of reliable and interpretable models. In traditional regression models like Ordinary Least Squares (OLS), the curse of dimensionality severely limits their ability to perform well in high-dimensional spaces (Bertsimas & Van, 2020). The presence of a large number of predictor variables relative to the number of observations leads to sparse data, resulting in models that are often over-fitted to noise rather than capturing the underlying patterns. Additionally, multicollinearity among predictor variables can further destabilize model estimates, making it difficult to assess the importance of individual variables.

Sparse regression methods, particularly the Least Absolute Shrinkage and Selection Operator (LASSO), have emerged as important tools for handling high-dimensional data. LASSO is capable of performing both variable selection and regularization simultaneously, shrinking some regression coefficients to zero and thus simplifying the model. However, despite its popularity, LASSO has its own limitations, particularly in handling correlated variables and balancing bias and variance trade-offs in high-dimensional settings.

Several extensions of LASSO, such as the Elastic Net and Group LASSO, have been developed to address some of these shortcomings (Mei & Montanari, 2022). However, there remains a need for further advancements to improve model accuracy, variable selection, and computational efficiency. The focus of this study is to develop an extended LASSO technique that enhances the ability to model high-dimensional data, offering improvements in both theoretical and practical applications. This study aims to contribute to the field by proposing an advanced sparse regression modeling technique that provides more robust, interpretable, and computationally efficient solutions for high-dimensional data analysis. High-dimensional data presents significant challenges in statistical modeling, including over-fitting, multicollinearity, and computational inefficiencies, especially when the number of variables exceeds the number of observations. Traditional methods like Ordinary Least Squares (OLS) and even the standard LASSO algorithm struggle with correlated variables, producing biased estimates and often being computationally intensive. While extensions like Elastic Net and Group LASSO offer improvements, there remains a need for a more efficient algorithm that can enhance variable selection, prediction accuracy, and computational performance in high-dimensional settings. This study aims to develop an extended LASSO technique to address these limitations and provide more reliable and interpretable models for real-world applications.

LITERATURE REVIEW

In recent years, advancements in LASSO have been driven by the need to handle high-dimensional datasets more efficiently. High-dimensional data, where the number of variables often exceeds the number of observations, poses challenges such as over-fitting, multicollinearity, and computational inefficiency. Traditional LASSO, introduced by (Tibshirani, 1996), revolutionized the field by performing variable selection and regularization. Group LASSO, introduced by Yuan and Lin (2006), is designed to handle situations where predictors are naturally grouped together, and it is desirable to select entire groups of variables rather than individual variables (Utazirubanda, M., & Ngom, 2021). This is particularly useful in applications where the variables within a group are correlated or represent related features, such as in gene expression data, where genes may be grouped by pathways or biological functions. The exponential growth of data in various fields in recent years such as genomics, finance, and image processing has led to an increasing need for effective regression models that can handle high-dimensional datasets (Chen, Chi, Fan, & Ma, 2021). Traditional regression methods often fail in these scenarios due to the "p > n" problem, where the number of predictors (p) exceeds the number of samples (n), leading to over-fitting and poor generalization (Mei & Montanari, 2022). Sparse regression methods, which aim to select a subset of relevant predictors while simultaneously estimating the associated coefficients, have emerged as a powerful tool for high-dimensional data analysis. Among these methods, (Papoutsoglou, et al., 2023).

High-dimensional data analysis deals with datasets where the number of variables (features) is much larger than the number of observations (samples (Pes, 2020)). This scenario is common in various fields such as genomics, finance, image processing, and social sciences, where data collection technologies have advanced, leading to datasets with hundreds or thousands of variables. Analyzing such datasets presents several challenges and requires specialized techniques to extract meaningful information. Analysis of high-dimensional data poses several challenges that are not typically encountered in low-dimensional settings (Ray, Reddy, & Banerjee, 2021). These challenges arise due to the increased complexity and sparsity of the data, as well as the computational burden of analyzing high-dimensional datasets. Sparse regression is a type of regression analysis that aims to identify a subset of important variables from a larger set of potential predictor variables. In sparse regression, the model includes only a few of the available predictors, with the rest of the coefficients set to zero (Bertsimas & Van, 2020). This sparsity property makes the model more interpretable and can improve its predictive performance, especially in high-dimensional data settings where the number of predictors is much larger than the number of observations. Ridge Regression, also known as Tikhonov regularization (La, Eickenberg, Nunez-elizalde, & Gallant, 2022), is one of the earliest methods developed to handle multicollinearity in regression models, where predictors are highly correlated. It introduces a penalty on the size of the coefficients to prevent them from becoming too large.

METHODS

Least Absolute Shrinkage and Selection Operator (LASSO) Formulation. LASSO is a regression analysis method that performs both variable selection and regularization to enhance the prediction accuracy and interpretability of the statistical model it produces. The LASSO was introduced by Robert Tibshirani in 1996.
The LASSO regression is formulated as (Tibshirani, 1996):

$$\hat{\beta}_{lasso} = \arg min \{\frac{1}{2n}\sum_{i=1}^{n}\left(y_i - \sum_{j=1}^{p} x_{ij}\beta_j\right)^2 + \lambda\sum_{j=1}^{p} |\beta_j|\}$$

where:

y_i is the response variable for the ij^{th} observation.

x_i is the value of the j^{th} predictor variable for the i^{th} observation.

β_j are the coefficients of the regression model.

$\lambda \geq 0$ is the regularization parameter, controlling the strength of the penalty.

The first term in the objective function,
$\frac{1}{2n}\sum_{i=1}^{n}\left(y_i - \sum_{j=1}^{p} x_{ij}\beta_j\right)^2$ the residual sum of squares (RSS), which measures the fit of the model to the data.
The second term,
$\sum_{j=1}^{p} |\beta_j|$ is the ℓ1-norm penalty, which induces sparsity in the coefficient estimates.
The regularization techniques and their impact on bias-variance trade-off, regularization refers to a set of techniques used in statistical modeling and machine learning to prevent over-fitting, especially when dealing with high-dimensional data or complex models. The bias-variance trade-off is a key concept in understanding regularization. It reflects the balance between two sources of error in a model. Regularization techniques adjust the model's complexity to achieve an optimal balance between bias and variance, thereby improving the model's performance on unseen data. Common regularization techniques were implored.
Ridge Regression (L2 Regularization)

$$\beta\text{hat} = \arg min \{\frac{1}{2n}\sum_{i=1}^{n}\left(y_i - \sum_{j=1}^{p} x_{ij}\beta_j\right)^2 + \lambda\sum_{j=1}^{p} {\beta^2}_j |\}$$

Impact on Bias-Variance Trade-off:

Bias: Ridge regression introduces a penalty on the size of the coefficients, which can increase bias by shrinking the coefficients towards zero, leading to a simpler model.

Variance: By shrinking the coefficients, Ridge regression reduces the model's sensitivity to the training data, thus lowering variance.

Trade-off: Ridge regression can reduce variance at the cost of introducing some bias, leading to a model that generalizes better to new data.

LASSO (L1 Regularization)

$$\hat{\beta}_{lasso} = \arg min \{\frac{1}{2n}\sum_{i=1}^{n}\left(y_i - \sum_{j=1}^{p} x_{ij}\beta_j\right)^2 + \lambda\sum_{j=1}^{p} |\beta_j| \}$$

Impact on Bias-Variance Trade-off:

Bias: LASSO introduces bias by penalizing the absolute values of the coefficients, often driving some to zero, which can result in a sparse model.

Variance: By effectively reducing the number of predictors, LASSO reduces variance and improves the model's generalizability.

Trade-off: The LASSO trade-off involves increasing bias (due to the penalty) while significantly reducing variance, especially in high-dimensional datasets where many predictors may be irrelevant.

Elastic Net

$$\beta ha = argmin\left\{\frac{1}{2n}\sum_{i=1}^{n}(y_i - \sum_{j=1}^{p} x_{ij}\,\beta_j)\hat{}2 + \lambda_1\sum_{j=1}^{p} |\beta_j| + \lambda_2\sum_{j=1}^{p} p^2{}_j\right\}$$

Here, λ1 and λ2 are regularization parameters for the ℓ1-norm and ℓ2-norm penalties, respectively.

Impact on Bias-Variance Trade-off:

Bias: Elastic Net introduces bias by shrinking coefficients through both the ℓ1-norm (LASSO) and ℓ2-norm (Ridge) penalties.

Variance: It reduces variance by combining the effects of Ridge (which handles multicollinearity well) and LASSO (which performs variable selection).

Trade-off: Elastic Net balances the strengths of Ridge and LASSO, offering a flexible approach to managing the bias-variance trade-off, especially in cases of highly correlated predictors. In this study, the following assumptions of AWENGL were adopted. The Adaptive Weighted Elastic Net Generalized LASSO (AWENGL) is built upon the following assumptions to ensure its theoretical validity and practical effectiveness in high-dimensional sparse regression. By adhering to these assumptions, AWENGL ensures reliable feature selection, interpretable coefficients, and robust predictive performance, making it a powerful tool for high-dimensional sparse regression tasks

i. High-Dimensional Data Structure

The number of predictors (p) can exceed the number of observations (n), a common scenario in high-dimensional datasets. The design matrix (X) is assumed to have full column rank or to be regularized to handle rank deficiencies.

ii. Sparsity Assumption

The true regression coefficients (β) are sparse, meaning that only a subset of predictors has non-zero values. This assumption ensures that AWENGL can effectively identify and prioritize important predictors.

iii. Feature Correlations

Moderate multicollinearity is allowed among predictors. AWENGL leverages the Elastic Net's capability to group correlated features while penalizing them adaptively through weighted penalties.

iv. Penalty Function

The penalty terms in AWENGL are a combination of LASSO and Ridge penalties:

$$\lambda_1 \sum_{j=1}^{p} w_j \, |\beta_j| + \lambda_2 \sum_{j=1}^{p} \beta^2{}_j$$

v. Model Identifiability

The model is identifiable under the assumed sparsity structure and appropriate tuning of regularization parameters (λ_1 and λ_2).

RESULTS AND DISCUSSION

The objective of this study is to advance the LASSO algorithm with a novel approach called Adaptive Weighted Elastic Net with Group LASSO (AWENGL). Below is a detailed exploration of the proposed improvements and modifications, their rationale, and the expected benefits.

Detailed mathematical derivation for the Adaptive Weighted Elastic-Net Group LASSO (AWENGL) combines adaptive weights, Elastic-Net, and Group LASSO penalties.

Standard LASSO Formulation

The standard LASSO problem is formulated as:

$$\hat{\beta} = \arg min \left\{ \frac{1}{2n} ||y - X\beta||_2^2 + \lambda \sum_{j=1}^{p} |\beta_j| \right.$$

Where $\frac{1}{2n}||y - X\beta||_2^2$ is the mean square error loss function

$$\lambda \sum_{j=1}^{p} |\beta_j| \ is\ the\ penalty\ term (LASSO\ panalty) promoting\ sparsity\ in\ the\ coefficients\ of\ \beta$$

Elastic-Net Penalty

The Elastic-Net penalty is a combination of LASSO ($\ell 1$) and Ridge ($\ell 2$) penalties:

$$Penalty_{Elastic-Net} = \alpha \sum_{j=1}^{p} |\beta_j| + \frac{1-\alpha}{2} \sum_{j-1}^{p} \beta^2{}_j$$

Where $\alpha \epsilon [0.1]$ controls the balance between e1 and e2 penalty

When $\alpha - 0, is\ equivalent\ to\ ridge\ regression.$

Group LASSO Penalty

In Group LASSO, predictors are grouped into GGG groups, and the penalty is applied to the ℓ2 norm of the coefficients within each group:

$$Penalty_{Group} = \lambda_1 \sum_{g=1}^{G} ||\beta_g||_2$$

$where \quad \beta_g\ the$ coefficient for the group g and λ_1 is the regularization parameter for the group penalty

Adaptive Weights

Adaptive LASSO introduces weights to the penalty term to address bias issues. The weights are typically inversely proportional to the magnitude of initial coefficient estimates. Define adaptive weights w_j for each predictor β_j

$$wj = \frac{1}{|\hat{\beta}^{initial}{}_j|\Upsilon}$$

Where $\hat{\beta}^{initial}$ are estimate from initial model e.g ridge regression and ϒis a turning parameters.

Combined improved lasso (AWENGL) formulation

Integrating these components, the improved LASSO problem (AWENGL) is formulated as:

The AWENGL optimization problem is formulated as:

$$\hat{\beta} = \arg min \left\{\frac{1}{2n}||y - X\beta||_2^2 + \lambda_1 \sum_{g=1}^{G} w_g||\beta_g||_2 + \lambda_2\left(\alpha \sum_{j=1}^{p} w_j\ |\beta_j\ | + (1-\alpha)\ ||\beta||_2^2\right)\right.$$

Where

$\left\{\frac{1}{2n}||y - X\beta||_2^2 +\right.$ is the mean squared error loss.

$\lambda_1 \sum_{g=1}^{G} w_g||\beta_g||_2 +$ is the Group LASSO penalty with adaptive weights wg

$\lambda_2\left(\alpha \sum_{j=1}^{p} w_j\ |\beta_j\ | + (1-\alpha)\ ||\beta||_2^2\right)$ is the Elastic-Net penalty with adaptive weights wj

i.e

$$\beta = \arg min \{\frac{1}{2n}\sum_{i=1}^{n}\left(y_i - \sum_{j=1}^{p} x_{ij}\beta_j\right)^2 + \lambda_1 \sum_{j=1}^{p} w_j\ |\beta_j| + \lambda_2 \sum_{j=1}^{p} \beta^2{}_j| + \lambda_3 \sum_{g=1}^{G} |\ |\beta_g|\ |_2\}$$

Where:

Y_i is the response variable for the iiith observation.

X_{ij} is the value of the jth predictor variable for the iii-th observation.

B_j are the coefficients of the regression model.

W_j are adaptive weights, typically set as wj=$1 / |\beta_j^{OLS}$

λ1, λ2, λ3≥0 are regularization parameters controlling the strength of the LASSO, Ridge, and Group LASSO penalties

Rationale for Each Components

1. Elastic Net Component:

   i. Purpose: Mitigates the instability in variable selection caused by LASSO when dealing with highly correlated predictors.

   ii. Impact: By combining ℓ1-norm (LASSO) and ℓ2-norm (Ridge) penalties, the Elastic Net encourages a more stable selection of correlated predictors, reducing the variance of the model without completely discarding relevant variables.

2. Adaptive Weights:

   i. Purpose: Addresses the bias introduced by standard LASSO by allowing different levels of shrinkage for different coefficients.

   ii. Impact: Adaptive weights reduce shrinkage on large coefficients, leading to less biased estimates. This is particularly useful in cases where some predictors have a stronger relationship with the response variable than others.

3. Group LASSO Component:

   i. Purpose: Enables the selection of entire groups of predictors, which is useful when variables are naturally grouped (e.g., genes, time series).

   ii. Impact: Group LASSO allows for group-wise sparsity, ensuring that entire groups of variables can be included or excluded, which is beneficial in settings where predictors are logically or structurally grouped.

Expected Benefits of AWENGL

1. Improved Variable Selection:

   Benefit: AWENGL is expected to select variables more effectively by reducing bias and handling correlated variables more stably. This leads to models that better represent the underlying data structure.

2. Enhanced Predictive Accuracy:

   Benefit: By combining multiple regularization techniques, AWENGL can produce models with lower prediction error, particularly in high-dimensional settings where traditional LASSO might over fit or under-fit.

3. Flexibility in Handling Different Data Structures:

   Benefit: The inclusion of group LASSO allows AWENGL to be applied in a broader range of scenarios, including those where predictors are naturally grouped, leading to more interpretable models.

4. Reduction of Over-fitting:

   Benefit: The Ridge component of Elastic Net, along with adaptive weights, helps in reducing over-fitting, especially in datasets with a high number of predictors relative to observations.

5. Applicability to High-Dimensional Data:

   Benefit: AWENGL is particularly suited for high-dimensional datasets (e.g., genomics, image processing) where the number of predictors far exceeds the number of observations, ensuring that the model remains robust and interpretable.

CONCLUSION

This study advanced the regression modeling techniques for high-dimensional data by developing an extended LASSO model to improve variable selection in high-dimensional datasets. Our model hypothesized that the extended LASSO technique will outperform existing models, such as standard LASSO, Ridge regression, and Elastic Net, in terms of prediction accuracy, variable selection, and computational efficiency. Furthermore, the study posits that the extended LASSO algorithm will effectively address key challenges in high-dimensional data analysis, including multicollinearity and over-fitting.

REFERENCES

Alhamzawi, R., and Ali, H. T. (2018). The Bayesian Adaptive Lasso Regression. *Mathematical Biosciences*, *303*, 75-82.

Bertsimas, D., and Van, P. B. (2020). Sparse High-dimensional Regression. *The Annals of Statistics*, *48*(1), 300-323.

Chen, Y., Chi, Y., Fan, J., and Ma, C. (2021). Spectral Methods for Data Science: a Statistical Perspective. *Foundations and Trends' in Machine Learning*, *14*(5), 566-806.

Davino, C., Romano, R., and Vistocco, D. (2022). Handling Multicollinearity in Quantile Regression through the Use of Principal Component Regression. *Metron*, *80*(2), 153-174.

Desboulets, L. D. (2018). A Review on Variable Selection in Regression Analysis. *Econometrics*, *6*(4), 45.

Gaines, B. R., Kim, J., and Zhou, H. (2018). Algorithms for Fitting the Constrained Lasso. *Journal of Computational and Graphical Statistics*, *27*(4), 861-871.

Garcoa-nieto, P. J., Garcia-gonzalo, E., and Paredes-sonchez, J. P. (2021). Prediction of the Critical Temperature of a Superconductor by Using the Woa/mars, Ridge, Lasso and Elastic-net Machine Learning Techniques. *Neural Computing and Applications*, *33*, 17131-17145.

La, T. T., Eickenberg, M., Nunez-elizalde, A. O., and Gallant, J. L. (2022). Feature-space Selection with Banded Ridge Regression. *Neuroimage*, *264*, 119728.

Mei, S., and Montanari, A. (2022). The Generalization Error of Random Features Regression: Precise Asymptotics and the Double Descent Curve. *Communications on Pure and Applied Mathematics*, *75*(4), 667-766.

Nicholson, W., Matteson, D., and Bien, J. (2017). Bigvar: Tools for Modeling Sparse High-dimensional Multivariate Time Series. *Arxiv Preprint Arxiv:1702.07094*.

Papoutsoglou, G., Tarazona, S., Lopes, M. B., Klammsteiner, T., Ibrahimi, E., Eckenberger, J., Novielli, P., Tonda, A., Simeon, A., and Shigdel, R. (2023). Machine Learning Approaches in Microbiome Research: Challenges and Best Practices. *Frontiers in Microbiology*, *14*, 1261889.

Pes, B. (2020). Ensemble Feature Selection for High-dimensional Data: a Stability Analysis across Multiple Domains. *Neural Computing and Applications*, *32*(10), 5951-5973.

Raj, A. (2024). *Optimization Algorithms for Machine Learning*. Universit't T'bingen.

Ray, P., Reddy, S. S., and Banerjee, T. (2021). Various Dimension Reduction Techniques for High Dimensional Data Analysis: a Review. *Artificial Intelligence Review*, *54*(5), 3473-3515.

Tibshirani, R. (1996). Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *58*(1), 267-288.

Utazirubanda, J. C., M., L. T., and Ngom, P. (2021). Variable Selection with Group Lasso Approach: Application to Cox Regression with Frailty Model. *Communications in Statistics-simulation and Computation*, *50*(3), 881-901.

Wang, W., Liang, J., Liu, R., Song, Y., and Zhang, M. (2022). A Robust Variable Selection Method for Sparse Online Regression Via the Elastic Net Penalty. *Mathematics*, *10*(16), 2985.

Yuan, M., and Lin, Y. (2006). Model Selection and Estimation in Regression with Grouped Variables. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *68*(1), 49-67.

Zou, H. (2006). The Adaptive Lasso and Its Oracle Properties. *Journal of the American Statistical Association*, *101*(476), 1418-1429.