



AN IMPROVED ALGORITHM FOR LOGISTIC REGRESSION IN CLASSIFICATION

Tharsika I. ^{1,*}, Kayanan M. ¹ and Yogaraja B. ¹

¹ Department of Physical Science, Faculty of Applied Science, University of Vavuniya, Sri Lanka.

* Corresponding author email: tharsikai1999@gmail.com

Abstract: Logistic regression is an essential tool in the world of machine learning and statistical modeling and is often used for binary classification tasks. The Maximum Likelihood (MLE) estimator is being used to fit the logistic regression model, which is attained by the Iteratively Reweighted Least Squares (IRLS) algorithm. However, when dealing with real-world datasets, the important assumption that the predictor variables are independent often does not hold true, which leads to a multicollinearity problem. Also, having too many predictor variables in the model results in inefficiency in prediction. In these cases, MLE is unstable. In our research, we proposed a new algorithm for logistic regression that handles multicollinearity and variable selection simultaneously using the Least Absolute Shrinkage and Selection Operator (LASSO) concept. For that, we combined IRLS and the Least Angle Regression (LARS) algorithm, which is used to obtain LASSO solutions. Through L1 regularization, our algorithm identified and retained the most relevant predictors while shrinking the coefficients of less important variables toward zero. This not only improved the interpretability of the model but also ultimately led to better predictive performance. Additionally, our algorithm performed well in scenarios with imbalanced datasets, where one class significantly outweighs the other. It achieved highly balanced accuracy and is, therefore, particularly useful for applications where the class distribution is imbalanced. We conducted extensive benchmarking against established algorithms, demonstrating the exceptional performance of our approach in terms of prediction accuracy and feature selection. Our approach not only addressed multicollinearity issues but also showed good performance on imbalanced datasets. These results highlighted how the algorithm can significantly advance the field of logistic regression in binary classification and provide valuable insights for real-world applications.

Keywords: Iteratively reweighted least squares, Least angle regression, Logistic regression least absolute shrinkage and selection operator, Maximum likelihood estimator, Multicollinearity.