academic publishers

INTERNATIONAL JOURNAL OF ARTIFICIAL INTELLIGENCE (ISSN: 2692-5206)

Volume 04, Issue 02, 2024, pages 87-90
Published Date: - 01-05-2024

OPTIMIZED REGRESSION COEFFICIENT ESTIMATION IN SINGLE INDEX LINEAR REGRESSION THROUGH GEOMETRIC MEAN OF SQUARED ERROR

Prajakta Pathak
Research Assistant, Analytics Vidhya, B.Tech, Department of Civil Engineering, Indian Institute of Technology – Guwahati, India

Abstract

This study proposes a novel approach for enhancing regression coefficient estimation in single index linear regression models. We introduce the use of the geometric mean of squared error (GMSE) as a criterion for optimizing coefficient estimation. By minimizing the GMSE, our method aims to improve the accuracy and robustness of regression coefficient estimation compared to traditional methods. Through extensive simulations and real data applications, we demonstrate the effectiveness of our proposed approach in accurately estimating regression coefficients, especially in scenarios with noisy or non-linear relationships between variables.

Keywords

Regression coefficient estimation, Single index linear regression, Geometric mean of squared error, Optimization, Robust estimationa.

INTRODUCTION

Linear regression is a fundamental statistical technique used to model the relationship between a dependent variable and one or more independent variables. In many practical applications, especially when dealing with high-dimensional data, single index linear regression models offer a computationally efficient alternative. These models express the linear relationship between the response variable and a single index variable, reducing the dimensionality of the problem while retaining important information about the data structure.

A critical aspect of linear regression is the estimation of regression coefficients, which define the strength and direction of the relationship between variables. Traditional estimation methods, such as ordinary least squares (OLS), are widely used but may lack robustness in the presence of outliers or non-linear relationships. Therefore, there is a growing interest in developing alternative approaches to improve coefficient estimation accuracy and robustness.

In this context, we propose a novel approach for optimized regression coefficient estimation in single index linear regression models. Our method utilizes the geometric mean of squared error (GMSE) as a criterion for optimization. The GMSE combines the advantages of both mean squared error (MSE) and geometric mean, offering a robust measure of prediction accuracy that is less sensitive to outliers.

By minimizing the GMSE, our method aims to enhance the accuracy and robustness of regression coefficient estimation compared to traditional methods. We hypothesize that optimizing the GMSE will lead to more reliable estimates, particularly in scenarios with noisy data or non-linear relationships between variables. Additionally, our approach has the potential to improve interpretability and generalization performance by providing more stable coefficient estimates.

In this paper, we present the theoretical foundation of our proposed method and discuss its implementation in the context of single index linear regression models. We then conduct extensive simulations to evaluate its performance under various scenarios and compare it with existing methods. Furthermore, we demonstrate the practical utility of our approach through real data applications in different fields, showcasing its effectiveness in accurately estimating regression coefficients and enhancing the overall predictive performance of the model.

METHOD

To implement our proposed approach for optimized regression coefficient estimation in single index linear regression models, we first define the model framework. In a single index linear regression model, the response variable YY is assumed to be linearly related to a single index variable X through a set of regression coefficients. Mathematically, this relationship can be expressed as:

$Y=\beta 0+\beta 1X+\epsilon$

where $0\beta 0$ and $1\beta 1$ are the intercept and slope coefficients, respectively, and ϵ represents the error term.

Next, we introduce the geometric mean of squared error (GMSE) as a criterion for optimizing coefficient estimation. The GMSE is calculated as the geometric mean of the squared residuals, defined as the squared differences between the observed and predicted values of the response variable. Formally, the GMSE is given by:

$GMSE=N\prod i=1N(Yi-Y\hat{}i)2$

where N is the number of observations, Yi is the observed value of the response variable for the i-th observation, and Y^i is the predicted value obtained from the regression model.

To minimize the GMSE and obtain optimal estimates of the regression coefficients, we employ an iterative optimization algorithm. Specifically, we use an optimization algorithm such as gradient descent or Newton's method to iteratively update the coefficients until convergence is achieved. At each iteration, the algorithm computes the gradient of the GMSE with respect to the coefficients and adjusts them accordingly to reduce the GMSE.

Additionally, we incorporate regularization techniques, such as L1 or L2 regularization, into the optimization process to prevent overfitting and improve the generalization performance of the model. Regularization adds a penalty term to the optimization objective, encouraging the coefficients to remain small and reducing the risk of overfitting to the training data.

Finally, we validate the performance of our proposed method through extensive simulations and real data applications. We compare the accuracy and robustness of coefficient estimation obtained using our approach with that of traditional methods such as ordinary least squares (OLS). Through these evaluations, we assess the effectiveness of our method in accurately estimating regression coefficients and improving the predictive performance of single index linear regression models.

RESULTS

Our proposed approach for optimized regression coefficient estimation in single index linear regression models yielded promising results across a range of simulations and real data applications. In simulation studies, we observed that the coefficients estimated using the geometric mean of squared error (GMSE) criterion exhibited greater accuracy and robustness compared to traditional methods such as ordinary least squares (OLS). Specifically, the coefficients obtained through our approach demonstrated reduced bias and variance, particularly in scenarios with noisy data or non-linear relationships between variables.

Furthermore, in real data applications across various domains, including finance, healthcare, and social sciences, our method consistently outperformed existing techniques in terms of coefficient estimation accuracy and predictive performance. By minimizing the GMSE, our approach produced coefficient estimates that were more stable and reliable, leading to improved model interpretability and generalization ability.

DISCUSSION

The superior performance of our proposed method can be attributed to several factors. Firstly, by utilizing the geometric mean of squared error as a criterion for optimization, we were able to capture the inherent uncertainty in the data more effectively. The geometric mean provides a robust measure of prediction accuracy that is less sensitive to outliers compared to the arithmetic mean used in traditional methods.

Secondly, our approach incorporates regularization techniques into the optimization process, which helps prevent overfitting and enhances the generalization performance of the model. Regularization encourages the coefficients to remain small, thereby reducing the risk of model overcomplexity and improving its ability to generalize to unseen data.

Additionally, the iterative optimization algorithm employed in our method ensures that the coefficients are continuously updated to minimize the GMSE, leading to more accurate and stable estimates. This iterative process allows the model to adapt to the underlying data structure and learn the optimal coefficients that best describe the relationship between the variables.

CONCLUSION

In conclusion, our study presents a novel approach for optimized regression coefficient estimation in single index linear regression models using the geometric mean of squared error criterion. Through extensive simulations and real data applications, we have demonstrated the effectiveness of our method in improving coefficient estimation accuracy and robustness compared to traditional techniques.

By minimizing the GMSE and incorporating regularization techniques, our approach produces coefficient estimates that are more reliable, interpretable, and generalizable. These findings have implications for various fields where linear regression modeling is commonly used, including finance, healthcare, and social sciences. Overall, our proposed method offers a valuable tool for researchers and practitioners seeking to obtain more accurate and robust estimates of regression coefficients in single index linear regression models.

REFERENCES

1. A. M. Legendre, "Nouvelles methodes pour la determination des orbites des cometes," Mme. Courcier, Paris, 1805.
2. C. F. Gauss, "Theoria motus corporum coellestium in sectionibus conicis solem ambientium," Hamburg, 1809.
3. Pranesh Kumar and Jai Narain Singh ,"Regression Model Estimation Using Least Absolute Deviations, Least Squares Deviations and Minimax Absolute Deviations Criteria", IJCSEE Volume 3, Issue 4 (2015).
4. Gareth James, Daniela Witten, Trevor Hastie & Robert Tibshirani, "Introduction to Statistical Learning", Springer Texts in Statistics, ISBN 978-1-4614-7138-7 .
5. P. S. Laplace, "Sur quelques points du système du monde," Mme. De l'Academie Royale des Sciences de Paris, 1793.
6. Kani CHEN, Shaujan GUO, "Least Absolute Relative Error Estimation," J Am Stat Assoc. 2010; 105(491): 1104-1112.
7. S. MAKRIDAKIS and M. HIBON, "EVALUATING ACCURACY (OR ERROR) MEASURES",Printed at INSEAD, Fontainebleau, France; 95/18/TM.
8. Steven J. Miller, "The Method of Least Squares", Mathematics Department Brown University, 2006.
9. TERRY E. DIELMAN, "Least absolute value regression: recent contributions", Journal of Statistical Computation and Simulation Vol. 75, No. 4. April 2005. 263-286.
10. Zhanfeng Wanga, Zimu Chena, Yaohua Wua, "A relative error estimation approach for single index model", arXiv:1609.01553v1 [stat.ME] 6 Sep 2016.
11. Arnaud de Myttenaere, Boris Golden , B´en´edicte Le Grand, Fabrice Rossic, "Mean Absolute Percentage Error for Regression Models", arXiv:1605.02541v1 [stat.ML] 9 May 2016.