Minimum Distance Lasso for Robust Estimation in High-Dimensional Data

Adel Mohammadpour[1], **Seyedeh Niloofar Ebrahimi**[2]
August 18, 2020

Abstract

This paper introduces the Minimum Distance Lasso [1] (MD-Lasso) estimation method for variable selection and parameter estimation in high-dimensional sparse linear regression models. High-dimensional data analysis involves models where the number of parameters, p, is comparable to or larger than the sample size, n. Traditional regression methods often face challenges, such as overfitting and sensitivity to outliers, when applied to such data. The proposed MD-Lasso method integrates minimum distance functionals, com- monly used in non-parametric estimation for robustness, with L_1-regularization to achieve variable selection and parameter estimation simultaneously. The MD-Lasso method is governed by a scaling parameter that limits the influence of outliers, maintaining robustness in high-dimensional regression problems. Compared to conventional optimization methods like those previously used for MD estimators, this paper employs a gradient descent optimization approach, demonstrating its superiority through numerical simulations. The results show that MD-Lasso outperforms traditional methods, such as Ridge and standard Lasso re- gression, in terms of Mean Absolute Error (MAE) and Mean Squared Error (MSE). The MD-Lasso estimator provides robust and accurate estimates in the presence of noise and outliers, making it an effective approach for analyzing high-dimensional data.

1 Introduction

In recent years, advancements in technology have enabled the simultaneous recording and analysis of numerous variables in datasets, where the number of variables is often larger than the sample size. This is commonly referred to as high-dimensional data. Such datasets require new approaches for analysis, as traditional regression methods, which often face challenges with outliers, are not reliable. Methods that are robust against outliers, such as those utilizing statistical robustness techniques, are more suitable. Sparse models in high-dimensional settings often need efficient estimators that can handle outliers effectively.

One of the efficient estimation methods for high-dimensional regression is the Least Absolute Shrinkage and Selection Operator (Lasso), introduced by Tibshirani [3]. Lasso estimates regression coefficients by minimizing the sum of squared errors while applying an L_1 penalty to reduce the number of non-zero coefficients. This method performs well in variable selection and model interpretation.

[1]**Dr. Adel Mohammadpour**, *Associate Professor, Department of Mathematics and Computer Science, Amirkabir University of Technology.*

[2]nilufarebrahimi@aut.ac.ir

The Minimum Distance Lasso method proposed by Yang and Lozano (2016) is recognized for its robustness and efficiency in handling outliers. It uses minimum distance functionals in a non-parametric context with an L_1 penalty, combining robustness with sparsity.

2 Problem Overview

In statistical data analysis, one of the influential factors is the presence of outliers observations that deviate significantly from the rest of the data. Robust methods, which are not overly influenced by outliers, are necessary for reliable inference. In high-dimensional models, where the number of predictors is large, the issue of variable selection becomes crucial. Regularization and shrinkage methods, such as Lasso, are used to achieve accurate estimates by incorporating a penalty term.

2.1 Least Absolute Shrinkage and Selection Operator (Lasso)

Lasso is a regularization method proposed by Tibshirani (1996). It performs variable selection and shrinkage simultaneously by imposing an L_1 penalty on the absolute values of the regression coefficients. The Lasso regression model is formulated as:

$$\min_{\beta} \left\{ \sum_{i=1}^{N} \left(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j \right)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \right\}$$

In this formula, λ is the regularization parameter that controls the sparsity of the model. A larger λ leads to more coefficients being shrunk to zero, simplifying the model.

The penalty function used in Lasso is:

$$\sum_{j=1}^{p} |\beta_j|$$

which causes some coefficients to be exactly zero when λ is large, effectively performing variable selection.

Model Introduction

2.2 Problem Formulation and Notation

Consider a matrix $X \in \mathbb{R}^{n\times p}$, where each row represents an observation and each column represents a predictor variable. Let $Y \in \mathbb{R}^n$ be the response vector, and $\beta \in \mathbb{R}^p$ be the vector of coefficients. The goal is to estimate β using a model:

$$Y = X\beta + \eta$$

where $\eta \in \mathbb{R}^n$ represents the error term. For simplicity, we assume that $\beta_0 = 0$. To estimate β, the following regularized loss function is minimized:

$$\beta_{\lambda_n} = \arg\min_{\beta} \left(L(\beta) + \lambda_n \|\beta\|_1 \right)$$

where $L(\beta)$ represents the loss function, and λ_n is the regularization parameter for the L_1 penalty.

2.3 Minimum Distance Estimator for High-Dimensional Regression

To evaluate the effectiveness of the Minimum Distance Lasso (MD-Lasso) estimator, we start with the Minimum Distance (MD) estimator for high-dimensional regression. Suppose we have a random vector $X \in \mathbb{R}^p$ and a response variable $Y \in \mathbb{R}$. The goal is to estimate the conditional distribution of Y given X. The MD estimator aims to minimize the L_2 distance between the estimated and true conditional distributions.

The Minimum Distance (MD) function is defined as:

$$d(\beta) = \int [f(Y|X;\beta) - f(Y|X)]^2 \, dY$$

Expanding this function gives:

$$d(\beta) = \int f^2(Y|X;\beta) \, dY - 2\mathbb{E}[f(Y|X;\beta)] + \text{constant}$$

Here, $f(Y|X;\beta)$ is the estimated conditional density function, and $f(Y|X)$ is the true conditional density. For simplicity, assume that the error term follows a multivariate normal distribution $N(X'\beta, \sigma^2 I)$. Thus, the conditional density function can be written as:

$$f(Y|X;\beta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(Y - X'\beta)^2\right)$$

The empirical expectation $\mathbb{E}[f(Y|X;\beta)]$ is approximated by:

$$\frac{1}{n}\sum_{i=1}^{n} f(Y_i|X_i;\beta)$$

Therefore, the empirical distance function becomes:

$$d_n(\beta) = -\frac{2}{n}\sum_{i=1}^{n} f(Y_i|X_i;\beta)$$

Using this, the objective function for the MD-Lasso method is defined as:

$$L(\beta) = -c\log\left(\sum_{i=1}^{n} \exp\left(-\frac{1}{2c}(Y_i - X_i'\beta)^2\right)\right)$$

where c is a scaling parameter. The MD-Lasso estimator is given by:

$$\hat{\beta}_{\lambda_n} = \arg\min_{\beta}\left(-c\log\left[\sum_{i=1}^{n} \exp\left(-\frac{1}{2c}(Y_i - X_i'\beta)^2\right)\right] + \lambda_n\|\beta\|_1\right)$$

This formulation shows that the MD-Lasso method remains robust in the presence of outliers and performs well for high-dimensional sparse data.

3 Numerical Calculations and Results

In this section, we analyze the performance of the MD-Lasso method for parameter estimation in high-dimensional models. The MD-Lasso, Ridge, and Lasso methods are compared based on their Mean Absolute Error (MAE) and Mean Squared Error (MSE) across different levels of noise and model complexity.

3.1 Synthetic Data

A simulation study is conducted in three stages using Gaussian noise to assess the effectiveness of the MD-Lasso method:

- **Stage 1:** 100 samples with 200 variables and a noise level of 10.
- **Stage 2:** 100 samples with 500 variables and a noise level of 10.
- **Stage 3:** 100 samples with 200 variables and a noise level of 10.

In all stages, the MD-Lasso method demonstrates superior results in terms of robustness and efficiency for high-dimensional data with noise.

Table 1: Summary of Results from Three-Stage Simulation Study

Method	Stage 1 (MAE/MSE)	Stage 2 (MAE/MSE)	Stage 3 (MAE/MSE)
Ridge	90.07/14514.48	74.73/8806.98	126.20/19891.34
Lasso	1.98/3.95	14.75/45.28	21.11/633.15
MD-Lasso	1.41/3.02	11.14/12.08	17.81/283.12

3.2 Data Analysis

The performance of different methods is assessed using a 5-fold cross-validation method for tuning hyperparameters. In each iteration, 80% of the data is randomly selected for training, and 20% is used for testing. For MD-Lasso, the gradient descent algorithm is applied to obtain optimal estimates, aiming to minimize Mean Absolute Error (MAE) and Mean Squared Error (MSE).

As demonstrated in Table 1, MD-Lasso outperforms Ridge and traditional Lasso in predictive performance, with lower MAE and MSE. The robustness of MD-Lasso against outliers makes it a more practical and effective approach for high-dimensional noisy data.

4 Discussion

The MD-Lasso method shows great potential for robust parameter estimation and variable selection in high-dimensional regression models, particularly when dealing with data that contains noise and outliers. The gradient descent optimization approach further enhances the method's applicability, providing more reliable results. Future research could explore additional optimization techniques or consider different types of penalty functions to further improve the performance of the MD-Lasso estimator.

5 References

References

[1] Lozano, A.C., Meinshausen, N., and Yang, E. (2016). Minimum Distance Lasso for robust high-dimensional regression. *Electronic Journal of Statistics*, 10, 1296–1340.

[2] Bouveyron, C., and Brunet, C. (2014). Model-based clustering of high-dimensional data: A review. *Computational Statistics and Data Analysis*, 71, 52–78.

[3] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society*, Series B (Methodological), 58, 267–288.