21st National Conference on Applied Stochastic Processes
31 August 2019

Minimum Distance Lasso for Robust Estimation in High-Dimensional Sparse Data

Niloofar Ebrahimi, Adel Mohammadpour

Department of Computer Science and Mathematics
Amirkabir University of Technology

Introduction

In recent years, the analysis of high-dimensional data, where the number of parameters is comparable to or exceeds the sample size, has become increasingly important in various fields, such as genomics, neuroimaging, and economics. Traditional likelihood-based estimators often lack robustness to outliers and model misspecification, which can compromise their effectiveness in these contexts. Despite this challenge, there has been limited emphasis on developing robust sparse learning methods for high-dimensional models. In response to this need, the literature has introduced approaches for robust sparse estimation in high-dimensional regression, including the Minimum Distance Lasso Estimation, which is specifically designed for variable selection in sparse linear regression models.

Minimum Distance Lasoo

A prevalent approach in sparse learning involves sparsity-inducing regularization. A well-known example is the Lasso, which utilizes l_1-penalized least squares to identify a parsimonious subset of predictors. Let $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ denote the predictor matrix, where each row represents a p-dimensional vector of variables observed for n training examples. Define $\boldsymbol{X}_i \in \mathbb{R}^p$ as the vector corresponding to the i-th observation across all variables. Similarly, let $\boldsymbol{X}_j \in \mathbb{R}^n$ denote the vector of observations for the j-th variable, and $X_i^j \in \mathbb{R}$ represent the entry in matrix $\boldsymbol{X}$ corresponding to the i-th observation of the j-th variable. Likewise, let $\boldsymbol{Y} \in \mathbb{R}^n$ denote the response vector, with Y_i representing its i-th observation. Consider the general regression model:

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta}^* + \boldsymbol{\eta}$$

We address the sparse estimation of the coefficient vector $\boldsymbol{\beta}^*$ through l_1-penalized loss minimization. Specifically, we consider estimators of the form:

$$\widehat{\beta}_{\lambda_n} = \arg\min_{\beta} \{L(\beta) + \lambda_n \|\beta\|_1\}$$

where the loss function L measures the goodness-of-fit with respect to the response, and λ_n is the regularization parameter for the l_1 penalty. Consequently, this approach remains a likelihood-based estimator and inherits the limitations associated with maximum-likelihood estimators in high-dimensional sparse regression. Specifically, its performance deteriorates substantially in the presence of model misspecification or outliers; a single outlier can render the estimates entirely unreliable.

To overcome the limitations of likelihood-based methods, we propose a penalized minimum distance criterion for robust and consistent estimation in high-dimensional sparse regression. Our approach is inspired by minimum distance estimators, which are well-regarded in nonparametric methods for their robustness and efficiency.

Consider the following optimization problem for the MD-Lasso estimator:

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

Here, the MD-Lasso estimator is formulated as an optimization problem, and we have developed a Gradient Descent algorithm for efficient and scalable optimization. This method is capable of producing local optima that are consistent.

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

In this formulation, the MD-Lasso estimator is framed as an optimization problem. A Gradient Descent algorithm has been developed to efficiently and scalably solve this problem. This method is capable of finding local optima that are consistent.

Experiments

Synthetic Dataset

The performance of our estimator is evaluated using synthetic datasets, compared against Ridge and traditional Lasso estimators. We generated the data from a Gaussian distribution. Each simulation study involves $n = 100$ observations with varying numbers of predictors: $p = 200$, $p = 200$, and $p = 500$, each with standard deviations of 1, 10, and 10, respectively.

Results

The tuning parameters for all methods were selected using five-fold cross-validation. To provide a quantitative assessment, we evaluated the predictive accuracy of each method by randomly partitioning the data into training and test sets, using 80% of the observations for training and the remaining 20% for testing.
To illustrate the effectiveness of the proposed method, we compared the absolute prediction error and squared prediction error of the MD-Lasso estimator with those of Ridge and traditional Lasso estimators, as reported in previous studies.

		Data 1	**Data 2**	**Data 3**
Ridge	**MAE**	90.07	74.73	126.20
	MSE	14514.48	8806.98	19891.34
Lasso	**MAE**	10.98	21.75	41.11
	MSE	3.95	45.28	633.15
MD-Lasso	**MAE**	1.41	11.14	17.81
	MSE	3.02	12.08	283.12

Table 1: Absolute Error (MAE) and Squared Error (MSE) for the models output by MD-Lasso and representative comparison methods on the three simulated datasets. (Smaller values indicate higher predictive accuracy.)

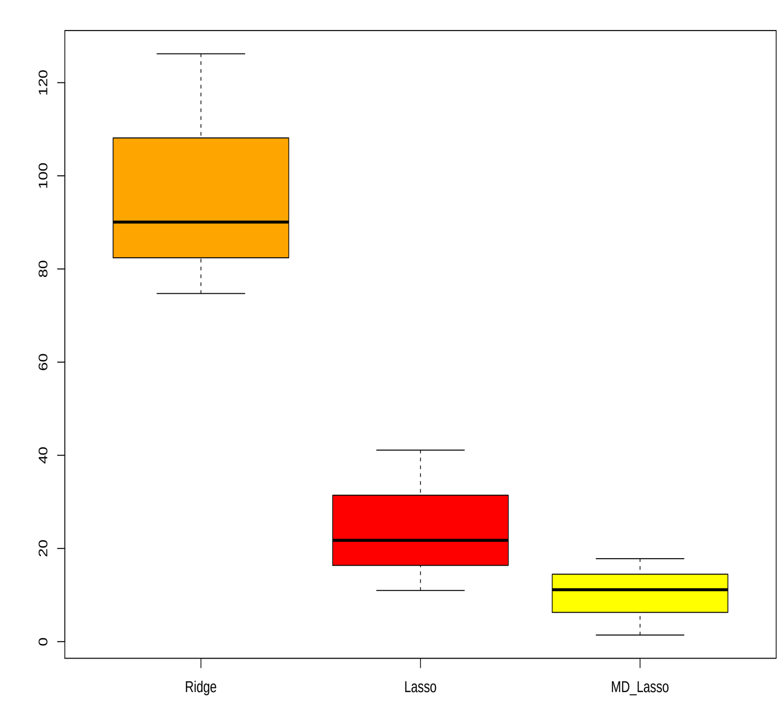

Figure 1: Boxplot of Absolute Error (MAE) for the comparison methods (the lower the better).

We have demonstrated that not only is the predictive performance of MD-Lasso superior to that of other methods, but it also maintains the robustness of the minimum distance function in sparse high-dimensional regression.

References

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

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

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