The Conditional Breakdown Properties of LAD-LASSO Regression

Bernice Feng

bf1318@nyu.edu

Advised by Prof. Avi Giloni & Prof. Jeffrey Simonoff

LAD-LASSO Regression

The least squares estimator minimizes

$$\sum_{i=1}^{n}(y_i - [\beta_0 + \beta_1 x_{1i} + \cdots + \beta_p x_{pi}])^2 \tag{1}$$

The LAD estimator minimizes

$$\sum_{i=1}^{n}|y_i - [\beta_0 + \beta_1 x_{1i} + \cdots + \beta_p x_{pi}]| \tag{2}$$

The LASSO estimator minimizes

$$\sum_{i=1}^{n}(y_i - [\beta_0 + \beta_1 x_{1i} + \cdots + \beta_p x_{pi}])^2 + \lambda \sum_{j=1}^{p}|\beta_j| \tag{3}$$

The LAD-LASSO estimator minimizes

$$\sum_{i=1}^{n}|y_i - [\beta_0 + \beta_1 x_{1i} + \cdots + \beta_p x_{pi}]| + \lambda \sum_{j=1}^{p}|\beta_j| \tag{4}$$

LAD-LASSO as LP

LAD as LP:

$$\begin{array}{ll} \min & \mathbf{e}_n^T \mathbf{r}^+ + \mathbf{e}_n^T \mathbf{r}^- \\ \text{s.t.} & \mathbf{X}\beta + \mathbf{r}^+ - \mathbf{r}^- = \mathbf{y} \\ & \beta \text{ free}, \ \mathbf{r}^+ \geq \mathbf{0}, \ \mathbf{r}^- \geq \mathbf{0} \end{array} \tag{5}$$

$$\mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} 1 & x_{11} & \dots & \dots & x_{p1} \\ \vdots & & & & \vdots \\ 1 & x_{1n} & \dots & \dots & x_{pn} \end{pmatrix} = \begin{pmatrix} \mathbf{x}^1 \\ \vdots \\ \mathbf{x}^n \end{pmatrix} \tag{6}$$

LAD-LASSO as LP with design matrix $\mathbf{X}^*$ and response vector $\mathbf{y}^*$

$$\mathbf{y}^* = \begin{pmatrix} y_1 \\ \vdots \\ y_n \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \mathbf{X}^* = \begin{pmatrix} 1 & x_{11} & \dots & \dots & x_{p1} \\ \vdots & & & & \vdots \\ 1 & x_{1n} & \dots & \dots & x_{pn} \\ 0 & -\lambda & 0 & \dots & 0 \\ 0 & 0 & -\lambda & & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & 0 & -\lambda \end{pmatrix} = \begin{pmatrix} \mathbf{x}^{*1} \\ \vdots \\ \mathbf{x}^{*n+p} \end{pmatrix} \tag{7}$$

Conditional Breakdown

The conditional breakdown value of LAD regression is

$$a(\mathbf{y}|\mathbf{X}) = \min_{1 \leq m < n} \left\{ m : b(m, \mathbf{y}|\mathbf{X}) \text{ is infinite} \right\} \tag{8}$$

where

$$b(m, \mathbf{y}|\mathbf{X}) = \sup_{\tilde{y}} \|\boldsymbol{\beta}^{LAD}(\mathbf{X}, \tilde{\mathbf{y}}) - \boldsymbol{\beta}^{LAD}\| \tag{9}$$

Here we contaminate m $(1 \leq m < n)$ values of the response vector $\mathbf{y}$ so that row i is replaced by $(\mathbf{x}^i, \tilde{\mathbf{y}}_i)$, and we obtain new data $(\mathbf{X}, \tilde{\mathbf{y}})$

MIP Approach

$$
\begin{aligned}
\min \quad & \sum_{i=1}^{n} u_i + l_i = a(\mathbf{y}|\mathbf{X}) && \text{(10a)}\\
\text{s.t.} \quad & \mathbf{x}^{i'}\boldsymbol{\xi} + \eta^{+} - \eta^{-} + s_i - t_i = 0 \qquad \text{for } i = 1, \ldots, n, && \text{(10b)}\\
& s_i - Mu_i \le 0, \quad t_i - Ml_i \le 0 \qquad \text{for } i = 1, \ldots, n, && \text{(10c)}\\
& \eta_i^{+} + \eta_i^{-} + Mu_i + Ml_i \le M \qquad \text{for } i = 1, \ldots, n, && \text{(10d)}\\
& u_i + l_i \le 1 \qquad \text{for } i = 1, \ldots, n, && \text{(10e)}\\
& \sum_{i=1}^{n} \eta_i^{+} + \eta_i^{-} - s_i - t_i \le 0, \quad \sum_{i=1}^{n} s_i + t_i \ge \varepsilon, && \text{(10f)}\\
& \boldsymbol{\xi} \text{ free}, \ \boldsymbol{\eta}^{+} \ge \mathbf{0}, \ \boldsymbol{\eta}^{-} \ge \mathbf{0}, \ \boldsymbol{s} \ge \mathbf{0}, \ \boldsymbol{t} \ge \mathbf{0}, && \text{(10g)}\\
& u_i, l_i \in \{0, 1\} \text{ for } i = 1, \ldots, n && \text{(10h)}
\end{aligned}
$$

Enumerative Algorithm

$a(\mathbf{y}|\mathbf{X}) = |E|$ where $|E|$ is the smallest integer such that

$$\max \frac{\sum_{i \in E} |\mathbf{x}^i \boldsymbol{\xi}|}{\sum_{i \in \mathbb{N}} |\mathbf{x}^i \boldsymbol{\xi}|} \geq \frac{1}{2}, \tag{11}$$

where

$$\boldsymbol{\xi} = \begin{pmatrix} 1 & \mathbf{0} \\ & \mathbf{X}_B \end{pmatrix}^{-1} \begin{pmatrix} \gamma \\ \mathbf{0} \end{pmatrix}, \tag{12}$$

Note that the algorithm enumerates through all $(p) \times (p+1)$ submatrices $\mathbf{X}_B$ of $\mathbf{X}$.

Data Processing

1. direct LAD-LASSO with no scaling
2. direct LAD-LASSO with predictors scaled by SD's: $\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x_j^i - \mu(\mathbf{x}_j)\right)^2}$, where $\mu(\mathbf{x}_j) = \frac{1}{n}\sum_{i=1}^{n} x_j^i$
3. direct LAD-LASSO with predictors scaled by MAD's: $\frac{1}{n}\sum_{i=1}^{n}\left|x_j^i - \text{Median}(\mathbf{x}_j)\right|$
4. center the variables with mean, no scaling of predictors: $\frac{1}{n}\sum_{i=1}^{n} y_i$, $\frac{1}{n}\sum_{i=1}^{n} x_j^i$
5. center the variables with median, no scaling of predictors: $\text{Median}(\mathbf{y}), \text{Median}(\mathbf{x}_j)$
6. center the variables with mean, scale predictors with SD's
7. center the variables with median, scale predictors with MAD's

Table

lambda	beta	active variable	LP objective	CB for LAD	CB for LAD-LASSO	CB for relaxed LAD-LASSO
0	[14.212097, 0.739684, -0.111747, -0.457656]	3	24.416163	4	4	4
3	[12.887341, 0.721444, -0.018468, -0.404982]	3	28.073166	4	4	4
6	[12.955865, 0.720236, 0.0, -0.408271]	2	31.486949	4	5	4
9	[12.955865, 0.720236, 0.0, -0.408271]	2	34.872471	4	4	4
12	[12.929997, 0.71308, 0.0, -0.405993]	2	38.24543	4	4	4
15	[12.929997, 0.71308, 0.0, -0.405993]	2	41.602652	4	4	4
18	[11.094048, 0.741692, 0.0, -0.337113]	2	44.901918	4	3	4
21	[11.094048, 0.741692, 0.0, -0.337113]	2	48.138332	4	3	4
24	[7.805414, 0.729168, 0.0, -0.202704]	2	51.084108	4	2	4
27	[4.801362, 0.635543, 0.0, -0.027241]	2	53.402418	4	2	4
30	[4.801362, 0.635543, 0.0, -0.027241]	2	55.390768	4	2	4
33	[4.76056, 0.618195, 0.0, -0.020626]	2	57.331264	4	2	4
36	[4.76056, 0.618195, 0.0, -0.020626]	2	59.247727	4	2	4
39	[4.76056, 0.618195, 0.0, -0.020626]	2	61.16419	4	2	4
42	[5.638474, 0.555051, 0.0, -0.027241]	2	63.060168	4	2	4
45	[5.022222, 0.555556, 0.0, 0.0]	1	64.733333	4	1	6
48	[5.792593, 0.481481, 0.0, 0.0]	1	66.333333	4	1	6
51	[5.792593, 0.481481, 0.0, 0.0]	1	67.777778	4	1	6
54	[8.451613, 0.225806, 0.0, 0.0]	1	68.83871	4	1	6
57	[8.451613, 0.225806, 0.0, 0.0]	1	69.516129	4	1	6
60	[10.8, 0.0, 0.0, 0.0]	0	69.9	4	1	14

Figure 1: Breakdown values for aircraft data

Graph

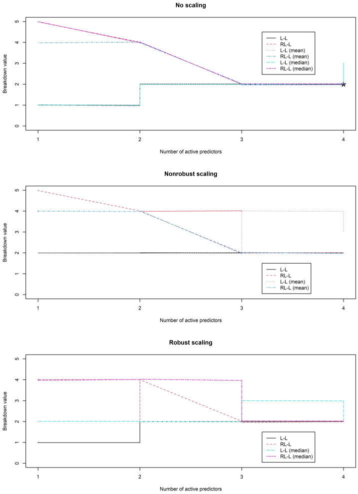

Figure 2: Breakdown values for aircraft data

References I

S.M. Ajeel and H.A. Hashem.
Comparison some robust regularization methods in linear regression via simulation study.
Academic Journal of Nawroz University, 9(2):244–252, 2020.
https://doi.org/10.25007/ajnu.v9n2a818.

O. Arslan.
Weighted LAD-LASSO method for robust parameter estimation and variable selection in regression.
Computational Statistics and Data Analysis, 56:1952–1965, 2012.

S.D. Cahya, B. Sartono, Indahwati, and E. Purnaningrum.
Performance of LAD-LASSO and WLAD-LASSO on high dimensional regression in handling data containing outliers.
Jurnal Teori dan Aplikasi Matematika, 6:844–856, 2022.
https://journal.ummat.ac.id/index.php/jtam/article/view/8968.

References II

X. Chen, Z. Wang, and M. McKeown.
Asymptotic analysis of robust LASSOs in the presence of noise with large variance.
IEEE Transactions on Information Theory, 56:5131–5149, 2010.

X. Gao and Y. Feng.
Penalized weighted least absolute deviation regression.
Statistics and Its Interface, 11:79–89, 2018.

X. Gao and J. Huang.
Asymptotic analysis of high-dimensional LAD regression with LASSO.
Statistica Sinica, 20:1485–1506, 2010.

A. Giloni and M. Padberg.
Alternative methods of linear regression.
Mathematical and Computer Modelling, 35:361–374, 2002.

A. Giloni and M. Padberg.
The finite sample breakdown point of ℓ_1-regression.
SIAM Journal of Optimization, 14:1028–1042, 2004.

References III

A. Giloni, B. Sengupta, and J.S. Simonoff.
A mathematical programming approach for improving the robustness of least sum of absolute deviations regression.
Naval Research Logistics, 53:261–271, 2006.

A. Giloni, J.S. Simonoff, and B. Sengupta.
Robust weighted LAD regression.
Computational Statistics and Data Analysis, 50:3124–3140, 2006.

Y. Jiang, Y. Wang, J. Zhang, B. Xie, J. Liao, and W. Liao.
Outlier detection and robust variable selection via the penalized weighted LAD-LASSO method.
Journal of Applied Statistics, 48:234–246, 2021.

H. Li, X. Xu, Y. Lu, X. Yu, T. Zhao, and R. Zhang.
Robust variable selection based on relaxed lad lasso.
Symmetry, 14:2161, 2022.
https://doi.org/10.3390/sym14102161.

References IV

I. Mizera and C. H. Müller.
Breakdown points and variation exponents of robust m-estimators in linear models.
Annals of Statistics, 27:1164–1177, 1999.

I. Mizera and C. H. Müller.
The influence of the design of ℓ_1-type m-estimators.
In P. Hackl A. Atkinson and C.H. Müller, editors, *MODA6 – Advances in Model-Oriented Design and Analysis*, pages 193–200. Physica-Verlag, Heidelberg, 2001.

S. Rahardiantoro and A. Kurnia.
LAD-LASSO: Simulation study of robust regression in high dimensional data.
Indonesian Journal of Statistics, 18:105–107, 2015.
`https://journal.ipb.ac.id/index.php/statistika/article/view/16775`.

References V

R. Tibshirani.
Regression shrinkage and selection via the lasso.
Journal of the Royal Statistical Society, Series B, 58:267–288, 1996.

H. Wang, G. Li, and G. Jiang.
Robust regression shrinkage and consistent variable selection through the LAD-Lasso.
Journal of Business and Economic Statistics, 25:347–355, 2007.

L. Wang.
The L_1 penalized LAD estimator for high dimensional linear regression.
Journal of Multivariate Analysis, 120:135–151, 2013.

J. Xu and Z. Ying.
Simultaneous estimation and variable selection in median regression using Lasso-type penalty.
Annals of the Institute of Statistical Mathematics, 62:487–514, 2010.