Effect of outliers on the variable selection by the regularized regression

Junho Jeong[a], Choongrak Kim[1, a]

[a]Department of Statistics, Pusan National University, Korea

Abstract

Many studies exist on the influence of one or few observations on estimators in a variety of statistical models under the "large n, small p" setup; however, diagnostic issues in the regression models have been rarely studied in a high dimensional setup. In the high dimensional data, the influence of observations is more serious because the sample size n is significantly less than the number variables p. Here, we investigate the influence of observations on the least absolute shrinkage and selection operator (LASSO) estimates, suggested by Tibshirani (*Journal of the Royal Statistical Society, Series B*, **73**, 273–282, 1996), and the influence of observations on selected variables by the LASSO in the high dimensional setup. We also derived an analytic expression for the influence of the k observation on LASSO estimates in simple linear regression. Numerical studies based on artificial data and real data are done for illustration. Numerical results showed that the influence of observations on the LASSO estimates and the selected variables by the LASSO in the high dimensional setup is more severe than that in the usual "large n, small p" setup.

Keywords: high-dimension, influential observation, LASSO, outlier, regularization

1. Introduction

Much work has been done in regression diagnostics since Cook's distance (Cook, 1977) was introduced forty years ago. The concept of regression diagnostics, considered in the classical linear model, has been extended to the Box-Cox transformation model (Box and Cox, 1964), ridge regression model (Hoerl and Kennard, 1970), and nonparametric regression models such as spline smoothing model, local polynomial regression, semiparametric model, and varying coefficient model. All regression diagnostic results in various regression models are done under the assumption of "large n, small p", i.e., the number of unknown parameters which are estimated is less than the number of samples in the data.

High-dimensional data (small n, large p) are very popular in the areas of information technology, bioinformatics, astronomy, and finance. Classical statistical inferences such as the least squares estimation in the linear model cannot be used in high-dimensional data. Recently, many methodological and computational advances have allowed high-dimensional data to be efficiently analyzed; in addition, least absolute shrinkage and selection operator (LASSO), as introduced by Tibshirani (1996), remains an important statistical tools for high-dimensional data.

In this paper, we study diagnostic issues in the LASSO regression model for high-dimensional data. Kim *et al.* (2015) recently derived an approximate version of Cook's distance in the LASSO

[1] Corresponding author: Department of Statistics, Pusan National University, 2, Busandaehak-ro 63 beon-gil, Geumjeong-gu, Busan 46241, Korea. E-mail: crkim@pusan.ac.kr

regression; however, it is based on the "large n, small p" assumption. Cook's distance in the LASSO model suggested by Kim *et al.* (2015) cannot be directly used in high-dimensional data because the covariance estimator of the LASSO estimator, which is necessary in defining a version of Cook's distance, is not easily derived in the high-dimensional data. Further, under high dimensional setup, we are more interested in influential observations on the variable selection rather than the influence on estimators. Studies on the diagnostic measures for the LASSO model in the high-dimensional data are relatively few. Among them, Zhao *et al.* (2013) proposed an influence measure for marginal correlations between the response and all the predictors, and Jang and Anderson-Cook (2017) suggested influence plots for LASSO. In this paper, we focus the influence of one or few observations on the variable selection by the LASSO using the deletion method. The influence on the variable selection in the classical model via the least squares was studied by Bae *et al.* (2017), and the selection of a smoothing parameter in the robust LASSO was done by Kim and Lee (2017).

In high-dimensional data, the influence of one or few observations on some estimators could be more serious and important because the number of observations is small compared to the "large n, small p" setup. We investigate that a variable selection result based on the LASSO regression can be significantly different if one or few observations are deleted. LASSO estimates often do not have a analytic form; therefore, we assume the design matrix is orthogonal.

This paper is organized as follows. In Section 2, the difference between LASSO estimates based on the full samples and the partial samples after deleting some observations, respectively, is derived under simple setup of the design. Numerical studies based on artificial data sets are done in Section 3, and an illustrative example based on a real data set is given in Section 4. Finally, concluding remarks are given in Section 5.

2. Case influence diagnostics in LASSO

2.1. LASSO estimator based on partial samples

Consider a simple linear regression model with no intercept, i.e.,

$$y_i = \beta x_i + \varepsilon_i, \quad i = 1, \ldots, n, \tag{2.1}$$

where $\sum x_i = 0$, $\sum x_i^2 = 1$, and $\sum y_i = 0$. Then, it can be shown in Tibshirani (1996), for example, that the LASSO estimator of β under model (2.1) is

$$\hat{\beta}_L(\lambda) = \text{sgn}\left(\hat{\beta}\right)\left(\left|\hat{\beta}\right| - \lambda\right)_+,$$

where $\text{sgn}(x)$ denotes the sign of x and $\hat{\beta} = \sum x_i y_i$ is the least squares estimator (LSE) of β. Now, let $K = \{i_1, i_2, \ldots, i_k\}$ be an index set of size k, and let $\hat{\beta}_{L(K)}(\lambda)$ be a LASSO estimator of β based on $(n - k)$ observations after deleting k observations in K. Then, we have the following result.

Proposition 1. *Under model (2.1),*

$$\hat{\beta}_{L(K)}(\lambda) = sgn\left(\hat{\beta} - \sum_{i \in K} x_i y_i\right)\left(1 - \sum_{i \in K} x_i^2\right)^{-1}\left(\left|\hat{\beta} - \sum_{i \in K} x_i y_i\right| - \lambda\right)_+.$$

Proof: Note that $\hat\beta_{L(K)}(\lambda) = \arg\min_\beta f(\beta)$, where $f(\beta) = (1/2)\sum_{j\notin K}(y_j - \beta x_j)^2 + \lambda|\beta|$. Since

$$f(\beta) = \frac{1}{2}\sum_{j=1}^{n}\left(y_j - \beta x_j\right)^2 - \frac{1}{2}\sum_{i\in K}(y_i - \beta x_i)^2 + \lambda|\beta|$$

$$= \frac{1}{2}\sum_{j=1}^{n}\left(y_j - \hat\beta x_j\right)^2 + \frac{1}{2}\left(\hat\beta - \beta\right)^2 - \frac{1}{2}\sum_{i\in K}(y_i - \beta x_i)^2 + \lambda|\beta|.$$

Now, by the first derivative of $f(\beta)$ with respect to β, we have

$$f'(\beta) = -\hat\beta + \beta + \sum_{i\in K}x_i(y_i - \beta x_i) + \lambda\cdot\mathrm{sgn}(\beta).$$

Therefore, by noting $x = \mathrm{sgn}(x)|x|$, we have

$$\hat\beta = \left(1 - \sum_{i\in K}x_i^2\right)\beta + \sum_{i\in K}x_iy_i + \lambda\cdot\mathrm{sgn}(\beta)$$

$$= \mathrm{sgn}(\beta)|\beta|\left(1 - \sum_{i\in K}x_i^2\right) + \sum_{i\in K}x_iy_i + \lambda\cdot\mathrm{sgn}(\beta)$$

$$= \mathrm{sgn}(\beta)\left\{\left(1 - \sum_{i\in K}x_i^2\right)|\beta| + \lambda\right\} + \sum_{i\in K}x_iy_i,$$

i.e.,

$$\hat\beta - \sum_{i\in K}x_iy_i = \mathrm{sgn}(\beta)\left\{\left(1 - \sum_{i\in K}x_i^2\right)|\beta| + \lambda\right\}. \tag{2.2}$$

Since the second term of Equation (2.2) is positive, we must have $\mathrm{sgn}(\hat\beta - \sum_{i\in K}x_iy_i) = \mathrm{sgn}(\beta)$. Now,

$$\left(1 - \sum_{i\in K}x_i^2\right)\beta = \hat\beta - \sum_{i\in K}x_iy_i - \lambda\cdot\mathrm{sgn}(\beta)$$

$$= \mathrm{sgn}\left(\hat\beta - \sum_{i\in K}x_iy_i\right)\left|\hat\beta - \sum_{i\in K}x_iy_i\right| - \lambda\cdot\mathrm{sgn}(\beta)$$

$$= \mathrm{sgn}\left(\hat\beta - \sum_{i\in K}x_iy_i\right)\left(\left|\hat\beta - \sum_{i\in K}x_iy_i\right| - \lambda\right)$$

$$= \mathrm{sgn}\left(\hat\beta - \sum_{i\in K}x_iy_i\right)\left(\left|\hat\beta - \sum_{i\in K}x_iy_i\right| - \lambda\right)_+.$$

Therefore,

$$\hat\beta_{L(K)}(\lambda) = \mathrm{sgn}\left(\hat\beta - \sum_{i\in K}x_iy_i\right)\left(1 - \sum_{i\in K}x_i^2\right)^{-1}\left(\left|\hat\beta - \sum_{i\in K}x_iy_i\right| - \lambda\right)_+$$

which completes the proof.   □

Remark 1. As a simple consequence of Proposition 1, the LASSO estimator based on $(n-1)$ observations after deleting the ith observation is

$$\hat{\beta}_{L(i)}(\lambda) = \text{sgn}\left(\hat{\beta} - x_i y_i\right)\left(1 - x_i^2\right)^{-1}\left(\left|\hat{\beta} - x_i y_i\right| - \lambda\right)_+.$$

2.2. Case influence in LASSO

Proposition 2. *Without loss of generality, we assume that $\hat{\beta} > 0$. Then, $\hat{\beta}_L(\lambda) = (\hat{\beta} - \lambda)_+$. Now, we further assume that $\hat{\beta} \geq \sum_{i \in K} x_i y_i \geq 0$. Then, it is easy to show that*

$$\hat{\beta}_{L(K)}(\lambda) = \left(\sum_{i \notin K} x_i^2\right)^{-1}\left(\hat{\beta}_L(\lambda) - \sum_{i \in K} x_i y_i\right).$$

To see the relationship between $\hat{\beta}_L(\lambda)$ and $\hat{\beta}_{L(K)}(\lambda)$ in terms of λ (see Figure 1), we consider for the single case deletion i.e., $K = \{i\}$. First, if $\lambda \geq \hat{\beta}$, then $\hat{\beta}_L(\lambda) = \hat{\beta}_{L(i)}(\lambda) = 0$. Second, if $\lambda \leq \hat{\beta} - x_i y_i$, then $\hat{\beta}_{L(i)}(\lambda) = (\hat{\beta}_L(\lambda) - x_i y_i)/\sum_{j \neq i} x_j^2$. Finally, if $\hat{\beta} - x_i y_i < \lambda < \hat{\beta}$, then $\hat{\beta}_L(\lambda) = \hat{\beta} - \lambda$ while $\hat{\beta}_{L(i)}(\lambda) = 0$, i.e., if $\hat{\beta} - x_i y_i < \lambda < \hat{\beta}$, then the deletion of the ith observation results in not selecting the covariate. Therefore, if λ is chosen in this range, the ith observation is said to be influential on the feature selection.

3. Numerical studies

3.1. Influence on coefficient estimates

Consider a simple linear regression model

$$y_i = \beta x_i + \varepsilon_i, \quad i = 1, \ldots, n,$$

where $\sum x_i = 0$, $\sum x_i^2 = 1$, and $\sum y_i = 0$. We generate $n = 20$ random numbers by the following steps.

Step 1: Generate x_i from $N(0, 1)$

Step 2: Generate ε_i from $N(0, 0.1^2)$

Step 3: Let $Y_i = 0.1 x_i + \varepsilon_i$

Step 4: Do centering and scaling to meet $\sum x_i = 0$, $\sum x_i^2 = 1$, and $\sum y_i = 0$.

The LSE based on the artificial data is $\hat{\beta} = 0.346$. When the 20th observation is deleted, $\hat{\beta} - x_{20} y_{20} = 0.179$. Therefore, if $0.179 < \lambda < 0.346$, $\hat{\beta}_L(\lambda) = 0.346 - \lambda$; however, $\hat{\beta}_{L(20)}(\lambda) = 0$.

3.2. Influence on variable selections

In this numerical study, we investigate three important aspects of the influence of the ith observation on the variable selection in the "small n, large p" setup. First, we want to see that the number of selected variables via the LASSO is very sensitive to the deletion of one observation. The number of selected variables has very important implication in the sense of the prediction and the estimation of the degrees of freedom. Second, we want to see the variable selection performance of the LASSO
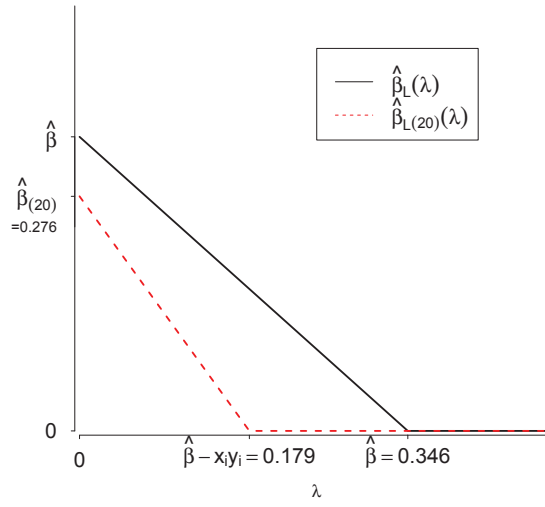
Figure 1: $\hat{\beta}_L(\lambda)$ and $\hat{\beta}_{L(20)}(\lambda)$ in terms of λ.

in the outlier model (a model with one outlying observation) when the tuning parameter is given so that the LASSO selects the true number of variables. Third, we want to see how correctly the LASSO selects variables in the outlier model when the tuning parameter is estimated by the cross-validation. To do this, we generate random numbers from the model given as follows.

Consider a linear regression model

$$y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i, \quad i = 1, \ldots, n,$$

where $\boldsymbol{\beta} = (5, 5, 5, 5, 0, \ldots, 0)^t$ is a 100-dimensional vector. We generate $n = 20$ random numbers by the following steps.

Step 1: Generate each X_j, $j = 1, \ldots, 100$ from $N(5, 1)$.

Step 2: Generate ε_i from $N(0, 0.1^2)$.

Step 3: Let $Y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i$.

- Simulation (I) - Sensitivity of LASSO to a single observation

  In this simulation study, we want to see that the number of selected variables via the LASSO is very sensitive to the deletion of one observation. Table 1 shows the LASSO estimators based on the full samples (i.e., $n = 20$) selected variables 1, 2, 3, and 4, which are true nonzero variables. However, the LASSO estimators based on 19 observations after deleting the ith ($i = 1, 2, \ldots, 20$) observation show different selection results. For example, if we delete the 2nd observation, the LASSO selected 10 variables; in addition, the LASSO selected just one variable if we deleted the 16th observation.

- Simulation (II) - Sensitivity of LASSO to a single outlier (λ : given)

  Here, we want to see the variable selection performance of the LASSO in the outlier model (a model with one outlying observation) when the tuning parameter is given so that the LASSO selects the 4 variables, i.e., true number of variables. We considered three outlier models, where each model

Table 1: Selected variables by the LASSO with the full data set ($n = 20$) and 19 observations after deleting one observation, respectively, based on the artificial data

Deleted observation	Selected variables	Number of selected variables
Original data	1 2 3 4	4
1	1 2 3 4 52 96	6
2	1 2 3 4 10 22 42 48 68 99	10
3	1 2 3 4 48 62	6
4	1 2 3 4	4
5	1 2 3 4	4
6	1 2 3 4 76	5
7	1 2 3 4	4
8	1 2 3 4 10	5
9	1 2 3 4	4
10	1 2 3 4 10 42	6
11	1 2 3 4 48	5
12	1 2 3 4 48	5
13	1 2 3 4 36 48	6
14	1 2 3 4 10 46 48 66 96	9
15	1 2 3 4 36	5
16	4	1
17	1 2 3 4	4
18	1 2 3 4 36 48 62 76	8
19	1 2 3 4 48	5
20	1 2 3 4	4

Table 2: The average proportion of the number of correctly selected variables among 4 variables selected by the LASSO out of 100 replications

Deleted observation	$y_1 + 10$	$y_1 + 30$	$y_1 + 50$
None	0.294	0.186	0.120
1	0.512	0.492	0.441
2	0.296	0.168	0.112
3	0.284	0.155	0.109
4	0.272	0.173	0.108
5	0.283	0.170	0.104
6	0.304	0.174	0.097
7	0.274	0.171	0.104
8	0.283	0.177	0.109
9	0.292	0.157	0.109
10	0.295	0.171	0.102
11	0.280	0.148	0.102
12	0.299	0.170	0.095
13	0.274	0.172	0.104
14	0.283	0.165	0.104
15	0.290	0.181	0.093
16	0.296	0.189	0.100
17	0.281	0.160	0.096
18	0.287	0.172	0.104
19	0.290	0.171	0.093
20	0.291	0.185	0.110

Three outlier models with $y_1 + c$, $c = 10, 30, 50$ are considered.

contains outlier with $y_1 + c$, $c = 10, 30, 50$. Table 2 shows the average proportion of the number of correctly selected variables among 4 variables selected by the LASSO out of 100 replications. Table 2 also shows the shows that the proportion of selecting a true model is about 0.5 when we delete the 1st observation (outlier); however, the proportion is less than 0.3 either when we use the

Table 3: The average proportion of the number of correctly selected variables by the LASSO out of 100 replications when the tuning parameter is estimated by the cross-validation

Deleted observation	$y_1 + 10$	$y_1 + 30$	$y_1 + 50$
None	0.490	0.350	0.220
1	0.550	0.505	0.530
2	0.480	0.280	0.210
3	0.475	0.330	0.200
4	0.465	0.330	0.225
5	0.450	0.310	0.230
6	0.480	0.295	0.215
7	0.480	0.325	0.220
8	0.465	0.320	0.200
9	0.455	0.325	0.190
10	0.485	0.300	0.225
11	0.470	0.295	0.215
12	0.465	0.300	0.235
13	0.450	0.300	0.210
14	0.475	0.310	0.195
15	0.450	0.275	0.240
16	0.490	0.315	0.190
17	0.480	0.305	0.235
18	0.460	0.310	0.210
19	0.445	0.305	0.235
20	0.480	0.305	0.215

If the number of selected variables less than four, we assumed that 4 variables are selected by the LASSO.
Three outlier models with $y_1 + c$, $c = 10, 30, 50$ are considered.

full data or when we delete the other observation than the outlier.

- Simulation (III) - Sensitivity of LASSO to a single outlier (λ : estimated)

  Finally, we want to see the variable selection performance of the LASSO in the outlier model when the tuning parameter is estimated by the cross-validation, for example. To do this we compute the average proportion of the number of the selected variables by LASSO out of 100 replications. If the number of selected variables less than four, we assumed that 4 variables are selected by the LASSO. Table 3 shows the average proportion of the number of correctly selected variables. We considered three outlier models, where each model contains an outlier with $y_1 + c$, $c = 10, 30, 50$. Table 3 also indicates the proportion of selecting true model is above 0.5 when we delete the 1st observation (outlier), but the proportion is less than 0.5 either when we use the full data or when we delete other observation than the outlier.

Remark 2. Even though the formula for the multiple cases deletion was given in Proposition 1, the simulation study was done for a single case deletion only. Due to different aspects of multiple cases deletion (swamping phenomenon and masking effect), further simulation studies for multiple cases deletion would be helpful.

4. Example

As an illustrative example for the influence of observations on the variable selection in LASSO, we use the brain aging data of Lu *et al.* (2004). This data set contains measurements of $p = 403$ genes and $n = 30$ human brain samples, and the response is the age of each human. We fit this data set by the LASSO based on the original sample ($n = 30$), and fit based on $n = 29$ observations after deleting

Table 4: Selected variables by the LASSO fit in the brain aging data based the original sample ($n = 30$) and fit based on $n = 29$ observations after deleting the ith observation

Deleted observation	Selected variables	# of selected variables
Original data	36 60 73 103 123 127 140 160 180 182 195 217 238 243 244 262 268 297 318 336 346 361 362 363 389	25
1	60 71 73 83 103 123 127 134 140 148 160 180 182 207 217 243 262 268 297 308 318 322 325 336 339 361 389	27
2	36 60 73 103 123 127 140 160 180 182 195 216 217 243 244 262 268 297 318 322 336 346 362 363 376 389 400	27
3	36 60 73 103 123 127 140 160 180 182 195 227 243 244 262 268 283 297 315 318 336 346 361 362 363 389	26
4	36 73 82 83 103 123 124 127 140 160 180 182 195 207 239 244 262 268 297 336 361 363 369 374 389	25
5	57 59 71 140 141 148 239 298 305 371	10
6	17 36 73 103 123 124 127 140 160 182 238 243 244 262 268 297 318 336 346 362 363 364 389	23
7	36 60 73 103 123 127 140 160 180 182 195 243 244 262 268 297 318 336 346 361 362 363 376 389	24
8	57 59 140 141 148 239 298 305 371	9
9	57 59 63 141 239 298 301 348 355 361	10
10	36 60 73 103 123 127 140 160 180 182 195 238 243 244 262 268 297 318 336 346 361 362 363 389	24
11	36 43 73 83 103 126 130 138 139 140 160 182 216 243 245 262 268 291 297 315 318 361 362 389	25
12	36 60 73 75 103 123 127 140 160 180 182 195 217 238 243 244 262 268 297 318 325 336 346 361 363 376 389	27
13	5 73 83 103 123 127 140 148 160 180 182 207 217 244 262 268 274 297 336 346 361 362 363 389	24
14	36 54 60 71 73 103 123 124 127 140 160 182 195 217 243 244 262 268 297 315 318 336 362 363 364 389 400	28
15	36 54 60 73 103 123 127 140 160 174 180 182 195 217 243 244 262 268 297 318 336 346 362 363 389	25
16	36 73 83 103 123 127 160 166 182 238 244 262 297 336 346 362 363 389	18
17	36 60 73 103 123 127 140 160 180 182 195 217 243 244 262 268 297 318 336 346 361 362 363 389	24
18	17 36 73 75 103 123 124 127 140 148 160 174 182 217 238 244 262 268 336 347 361 362 363 389 400	25
19	49 59 105 123 140 148 166 225 239 301 305 364 377	13
20	17 36 60 73 82 103 123 127 140 160 180 182 195 212 217 239 243 244 262 268 281 283 322 336 346 362 363 369 389	29
21	36 60 73 103 123 127 140 160 180 182 195 217 238 243 244 262 268 297 318 336 346 361 362 363 376 389	26
22	36 60 73 103 123 127 140 160 180 182 195 238 243 244 262 268 297 318 336 346 361 362 363 389	23
23	5 73 123 124 127 140 148 180 182 238 243 244 262 274 297 308 318 336 346 361 362 363 389	23
24	36 60 73 103 123 127 140 160 180 182 195 216 217 243 244 262 268 297 318 336 346 361 362 363 389	25
25	5 36 54 73 75 82 103 123 127 140 160 167 182 185 217 238 243 244 262 268 297 332 334 336 339 346 355 362 363 389	28
26	36 73 103 123 124 127 140 160 167 182 185 217 238 243 244 262 268 270 297 308 336 346 364 374 389 396	26
27	36 73 103 123 127 140 151 160 166 182 238 239 244 262 268 286 297 315 336 346 361 362 363 364 389	25
28	17 36 73 111 123 127 148 160 213 238 244 262 268 297 318 326 334 336 362 363 389	21
29	36 60 73 103 123 127 140 160 180 182 207 243 244 262 268 297 315 318 336 361 362 363 369 389	24
30	36 73 103 123 127 140 160 180 182 217 243 244 262 268 297 318 325 336 363 389	20

the ith ($i = 1, \ldots, 30$) observation to see the influence of the ith observation on the variable selection by the LASSO.

Table 4 shows that the LASSO selects 25 variables among 403 variables. If we delete one observation, then most cases result in selecting around 25 variables except the cases of deleting the 8th, 9th, and 19th observation, respectively. We may conclude that those three observations are quite influential as far as the number of variable selection is concerned.

5. Concluding remarks

One or few observations could be very influential on estimators in "small p, large n" case, and this phenomenon becomes more serious in "small n, large p" case. In this paper, we investigate the influence of observations on the LASSO estimates and the selected variables by the LASSO. Also, we derived analytic expression for the influence of the ith observation on the LASSO estimates in the simple linear regression. Simulation results show that the influence of an outlier is more serious in the high dimensional case than in the low dimensional case.

For further studies, it will be worth studying the basic building blocks which affect variable selection results. Despite difficulties, it is also necessary to modify the LASSO model to a robust LASSO that is not sensitive to outliers.

Acknowledgments

References

Bae W, Noh S, and Kim C (2017). Case influence diagnostics for the significance of the linear regression model, *Communications for Statistical Applications and Methods*, **24**, 155–162.

Box GEP and Cox DR (1964). An analysis of transformations, *Journal of the Royal Statistical Society, Series B*, **26**, 211–252.

Cook RD (1977). Detection of influential observation in linear regression, *Technometrics*, **19**, 15–18.

Hoerl AE and Kennard RW (1970). Ridge regression: biased estimation for nonorthogonal problems, *Technometrics*, **12**, 55–67.

Jang DH and Anderson-Cook CM (2017). Influence plots for LASSO, *Quality and Reliability in Engineering International*, **33**, 1317–1326.

Kim C, Lee J, Yang H, and Bae W (2015). Case influence diagnostics in the lasso regression, *Journal of the Korean Statistical Society*, **44**, 271–279.

Kim J and Lee S (2017). A convenient approach for penalty parameter selection in robust lasso regression, *Communications for Statistical Applications and Methods*, **24**, 651–662.

Lu T, Pan Y, Kao SY, Kohane I, and Chan J (2004). Gene regulation and DNA damage in the ageing human brain, *Nature*, **429**, 883–891.

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

Zhao J, Leng C, Li L, and Wang H (2013). High-dimensional influence measure, *The Annals of Statistics*, **41**, 2639–2667.