HRLR regression

Lasanthi C. R. Pelawa Watagoda, Alan T. Arnholt and Hasthika S. Rupasinghe Arachchige Don

Department of Mathematical Sciences, Appalachian State University, Boone, NC, USA

ABSTRACT

Selecting the most important predictor variables and achieving high prediction accuracy are the two main goals in Statistical Learning. In this work, we propose a new regularization method, HRLR (a Hybrid of Relaxed Lasso and Ridge Regression), where we use properties of both relaxed lasso and ridge regression. We also demonstrate the effectiveness of our method and the accuracy of the algorithm on simulated as well as real-world data. Simulation results suggest that the HRLR regression outperforms the existing well-known methods including lasso and Elastic net in many scenarios described, especially when the error distribution is non-normal. The proposed algorithm is implemented in R and links are provided in this paper along with the R functions used for the simulation.

1. Introduction

Suppose that the response variable $Y_i$ and at least one predictor variable $x_{ij}$ are quantitative with $x_{i1} \equiv 1$. Let $x_i^T = (x_{i1}, \ldots, x_{ip}) = (1, u_i^T)$ and $\beta = (\beta_1, \ldots, \beta_p)^T$ where $\beta_1$ corresponds to the intercept, then the multiple linear regression (MLR) model is

$$Y_i = \beta_1 + x_{i2}\beta_2 + \cdots + x_{ip}\beta_p + \epsilon_i = x_i^T \beta + \epsilon_i \quad (1)$$

for $i = 1, \ldots, n$. This model is also called the full model. Here $n$ is the sample size, and we assume that the zero mean random variables $\epsilon_i$ are independent and identically distributed (iid) with variance $V(\epsilon_i) = \sigma^2$. In matrix notation, these $n$ equations become

$$Y = X\beta + \epsilon \quad (2)$$

where $Y$ is an $n \times 1$ vector of responses, $X$ is an $n \times p$ matrix of predictors, $\beta$ is a $p \times 1$ vector of unknown coefficients, and $\epsilon$ is an $n \times 1$ vector of unknown errors. The $i$th fitted value $\hat{Y}_i = X_i^T \hat{\beta}$ and the $i$th residual $r_i = Y_i - \hat{Y}_i$ where $\hat{\beta}$ is an estimator of $\beta$.

There are many methods for estimating $\beta$, including, lasso by Tibshirani (1996), Eelastic net by Zou and Hastie (2005), relaxed lasso by Meinshausen (2007) and ridge regression by Hoerl and Kennard (1970). The lasso by Tibshirani (1996) is an attractive regularization method for high dimensional regression. The lasso coefficients, $\hat{\beta}^l$, minimize the quantity

$$\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^{p}|\beta_j| \quad (3)$$

There are some benefits and limitations of using lasso. In terms of benefits, first, it does variable selection by setting some of the coefficients equal to zero. Second, it uses an efficient computational algorithm. According to Meinshausen (2007), one of the limitations of lasso is when $p > n$ the rate of convergence of the lasso is very slow. Also, if the data set contains highly correlated independent variables, lasso tends to produce poor coefficient estimates.

Relaxed lasso is a generalization of the lasso shrinkage technique for linear regression. Relaxed lasso does variable selection and parameter estimation by using two penalty parameters in contrast to lasso which uses only one. According to Meinshausen (2007), relaxed lasso produces all standard lasso solutions with the same computational effort as regular lasso and at the same time, generates models that are similar or have even slightly better predictive performance.

Elastic net by Zou and Hastie (2005) is another regularization and variable selection method. It is a hybrid of ridge regression and lasso regularization. Elastic net also generates zero-valued coefficients. Zou and Hastie (2005) has shown that the elastic net outperforms lasso on data with highly correlated predictors.

The ridge regression coefficient estimates denoted by $\hat{\beta}^R$ are values that minimize
\[ \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^{p} \beta_j^2 = \text{RSS} + \lambda \sum_{j=1}^{p} \beta_j^2, \quad (4) \]

where \( \lambda \geq 0 \) is a tuning parameter. The quantity \( \lambda \sum_{j=1}^{p} \beta_j^2 \) is the shrinkage penalty in this technique. The final model with ridge regression will include all \( p \) predictors.

Apart from the methods described above, there are numerous techniques for variable selection section have been discovered more recently. See the studies by Y. Wang et al. (2019), S. Wang et al. (2011) and Xie and Zeng (2010) for examples.

This work proposes a new regularization and variable selection method, which is a Hybrid of Relaxed Lasso regularization and Ridge regression (HRLR). Theoretical and numerical results demonstrate that the HRLR produces parsimonious models with equal or lower prediction loss than the regular lasso and relaxed lasso estimators for high dimensional data where the number of predictor variables \( p \) is very large, possibly much larger than the number of observations \( n \). In Section 2, we define the estimator and describe the theory and method. We present a simulation in Section 3 as well as two real data examples in Section 4.

2. Theory and method

**Definition 1.** The HRLR regression estimator is defined for \( \lambda_1 \in [0, \infty), \lambda_2 \in [0, \infty) \) and \( \phi \in (0, 1) \) as

\[ \hat{\beta}^{\lambda_1, \lambda_2, \phi} = \arg \min_{\beta} n^{-1} \sum_{i=1}^{n} (Y_i - X_i^T \{ \beta \cdot 1_{\mathcal{M}_i} \})^2 + \phi \lambda_1 \| \beta \|_1 + \lambda_2 \| \beta \cdot 1_{\mathcal{M}_i} \|_2 \]

\[ (5) \]

Where \( 1_{\mathcal{M}_i} \) is the indicator function on the set of variables \( \mathcal{M} \subseteq \{1, \ldots, p \} \) so that for all \( k \in \{1, \ldots, p\} \)

\[ \beta \cdot 1_{\mathcal{M}_i} = \begin{cases} 0 & k \notin \mathcal{M}_i \\ \beta_k & k \in \mathcal{M}_i \end{cases} \]

The HRLR estimator is a two-stage procedure: for each fixed \( \lambda_2 \) we first find the ridge regression coefficients, and then compute the relaxed lasso-type shrinkage along the relaxed lasso coefficient solution paths (along \( \phi \) and \( \lambda_1 \)). The occurrence of double shrinkage in the process takes care of using a scaling transformation.

For ease of implementation, we introduce Theorem 1, where we consider a data set with \( n \) observations and \( p \) predictors. Let \( y = (y_1, \ldots, y_n)^T \) be the response and \( X = (x_1, \ldots, x_n) \) be the model matrix, where \( x_j = (x_{1j}, \ldots, x_{nj})^T, \ j = 1, \ldots, p \) are the predictors. Also, assume that the response is centered and the predictors are standardized.

**Theorem 1.** Given dataset \((y, X)\) and \((\lambda_1, \lambda_2)\), define an artificial dataset \((y^*, X^*)\) by

\[ X^*_{(n+p) \times p} = (1 + \lambda_2)^{-1/2} \left( \begin{array}{c} X \\ \sqrt{\lambda_2} I \end{array} \right), \]

\[ y^*_{(n+p)} = \left( \begin{array}{c} y \\ 0 \end{array} \right) \]

Let \( y = \frac{\lambda_1}{\sqrt{1 + \lambda_2}} \) and \( \beta^* = \sqrt{1 + \lambda_2} \beta \), then the HRLR criterion can be written as

\[ L(y, \phi, \beta) = L(y, \phi, \beta^*) = \| y^* - X^* \{ \beta^* \cdot 1_{\mathcal{M}_i} \} \|^2 + \phi y^* \| \beta^* \|_1 \]

Then \( \hat{\beta} = \frac{1}{\sqrt{1 + \lambda_2}} \hat{\beta}^* \).

**Proof.** From Equation (5), we have

\[ y^T y - 2y^T X \{ \beta \cdot 1_{\mathcal{M}_i} \} + \{ \beta \cdot 1_{\mathcal{M}_i} \}^T X^T X \{ \beta \cdot 1_{\mathcal{M}_i} \} + \lambda_2 \{ \beta \cdot 1_{\mathcal{M}_i} \}^T \{ \beta \cdot 1_{\mathcal{M}_i} \} + \phi \lambda_1 \| \beta \|_1 \]

Substituting the identities

\[ y^* X^* (\sqrt{1 + \lambda_2}) = y^T X, \]

\[ \beta^* = \sqrt{1 + \lambda_2} \beta. \]

\[ X^* X^* = \frac{X^T X + \lambda_2}{1 + \lambda_2} \]

we have

\[ y^{*T} y^* - 2y^{*T} X^* \sqrt{1 + \lambda_2} \{ \beta \cdot 1_{\mathcal{M}_i} \} + \{ \beta \cdot 1_{\mathcal{M}_i} \}^T \]

\[ \left[ (1 + \lambda_2) X^T X^* - \lambda_2 \{ \beta \cdot 1_{\mathcal{M}_i} \}^T \right] + \lambda_2 \{ \beta \cdot 1_{\mathcal{M}_i} \}^T \{ \beta \cdot 1_{\mathcal{M}_i} \} + \phi \lambda_1 \| \beta \|_1 \]

which is equivalent to

\[ y^{*T} y^* - 2 \sqrt{1 + \lambda_2} y^{*T} X^* \{ \beta^* \cdot 1_{\mathcal{M}_i} \} + (1 + \lambda_2) \]

\[ \{ \beta^* \cdot 1_{\mathcal{M}_i} \}^T X^{*T} X^* \{ \beta^* \cdot 1_{\mathcal{M}_i} \} - \lambda_2 \{ \beta^* \cdot 1_{\mathcal{M}_i} \}^T \]

\[ \{ \beta^* \cdot 1_{\mathcal{M}_i} \} + \lambda_2 \{ \beta^* \cdot 1_{\mathcal{M}_i} \}^T \{ \beta^* \cdot 1_{\mathcal{M}_i} \} + \phi \lambda_1 \| \beta \|_1 \]

by simplifying we get

\[ y^{*T} y^* - 2y^{*T} X^* \{ \beta^* \cdot 1_{\mathcal{M}_i} \}

\[ + (1 + \lambda_2) \{ \beta^* \cdot 1_{\mathcal{M}_i} \}^T X^{*T} X^* \{ \beta^* \cdot 1_{\mathcal{M}_i} \}

\[ + \phi \lambda_1 \| \beta \|_1 \]

which results in

\[ \| y^* - X^* \{ \beta^* \cdot 1_{\mathcal{M}_i} \} \|^2 + \phi y^* \| \beta^* \|_1 \]
Theorem 1 says that we can transform the HRLR problem into an equivalent relaxed lasso problem on augmented data. Therefore the set of solutions of the HRLR problem is given by

\[ \hat{\beta}^* = \frac{1}{\sqrt{1 + \lambda_2}} \hat{\beta}^* \]

2.1. Asymptotic results

As shown in Theorem 1, HRLR problem can be transformed into an equivalent relaxed lasso problem on augmented data. Therefore, many of the asymptotic results of HRLR regression directly follow the asymptotic results of the relaxed lasso procedure on augmented data. In the study by Meinshausen (2007), it is conjectured and later supported by numerical examples that the relaxed lasso is a consistent variable selection procedure when the penalty parameters are obtained by cross validation. Therefore, according to Theorem 1, this conjecture is true for HRLR estimates on augmented data as it uses cross validated penalty parameters. It has been further shown by Meinshausen (2007) that the conjecture is not applicable for lasso.

3. Simulations

In this section, we use simulations to demonstrate the proposed HRLR method, and compare HRLR to four standard algorithms (lasso, relaxed lasso, ridge and elastic net). This simulation closely follows the simulation done by Tibshirani (1996). We simulate data from the true model \( y = X \beta + \sigma \varepsilon \) with four error distributions:

1. \( \varepsilon \sim N(0, 1) \)
2. \( \varepsilon \sim t_3 \)
3. \( \varepsilon \sim \text{exp}(1) - 1 \)
4. \( \varepsilon \sim \text{unif}(-1, 1) \)

The same examples used by Tibshirani (1996) were used in this paper to compare the prediction performance of the HRLR, lasso, ridge regression and elastic net systematically. Cross validation was used to select the final model. Examples are described below. The statistical software R (R Core Team, 2017) was used to conduct the simulation.

(1) Example 1: We simulated 50 data sets consisting of 20 observations in the training set and 200 observations in the testing set with eight predictors in both the training and testing sets. We let \( \beta = (3, 1.5, 2, 0, 0, 0, 0, 0) \) and \( \sigma = 3 \). The pairwise correlation between \( x_i \) and \( x_j \) was set to be \( \text{cor}(i, j) = 0.5^{i-j} \).

(2) Example 2: Same as Example 1, except that \( \beta_j = 0.85 \) for all \( j \).

(3) Example 3: We simulated 50 data sets consisting of 100 observations in the training set and 200 observations in the testing set with 40 predictors in both the training and testing sets. We set

\[ \beta = (0, \ldots, 0, 2, \ldots, 2, 0, \ldots, 0, 2, \ldots, 2) \]

and \( \sigma = 15 \) with \( \text{cor}(i, j) = 0.5 \) for all \( i \) and \( j \).

3.1. Simulation results

In this section, we use simulations to compare the proposed HRLR method to a collection of commonly used variable selection and/or shrinkage methods. As displayed in Table 1 and Figure 1, from the Example 1 simulations, we see that HRLR has the lowest median test root mean square errors (denoted as TRMSEs) compared to all other methods for each error type. The standard error of the TRMSE (denoted as SE) is slightly higher for HRLR for normal errors compared to other methods, but was significantly lower for all the other error distributions.

According to Table 2 and Figure 2, HRLR has slightly higher TRMSEs compared to other methods for all the error distributions in Example 2 except when compared to relaxed lasso, but the standard error was lower in most cases, especially when the errors are uniformly distributed. As discussed by S. Wang et al. (2011) the lower TRMSE for elastic net in Example 2 might be due to the setup of the Example. It is clear from Table 3 and Figure 3 that HRLR has consistently smaller TRMSEs than all other methods in Example 3 for all error types except for normal errors. It is

| Error Dist. | HRLR | relaxed | lasso | ridge | enet |
|-------------|------|---------|-------|-------|------|
| 1           | 3.381257 | 4.015348 | 3.969064 | 3.393395 | 3.383454 |
| SE          | 0.537501 | 0.722440 | 0.503587 | 0.459631 | 0.503179 |
| 2           | 1.927973 | 2.451771 | 2.126671 | 2.231816 | 2.116753 |
| SE          | 0.440730 | 0.849136 | 0.541952 | 0.577668 | 0.491403 |
| 3           | 1.253082 | 2.142927 | 1.327364 | 1.486484 | 1.352891 |
| SE          | 0.357699 | 1.045325 | 0.577658 | 0.595865 | 0.588552 |
| 4           | 0.792765 | 1.707721 | 1.046893 | 1.187845 | 1.037665 |
| SE          | 0.389845 | 0.925814 | 0.595209 | 0.996269 | 0.936685 |
Table 2. Median TRMSE and the SE values for Example 2 for the five methods considered with four error distributions

| Error Dist. | TRMSE | HRLR | relaxed | lasso | ridge | enet |
|-------------|-------|------|---------|-------|-------|------|
| 1           | 3.625042 | 3.775470 | 3.507422 | 3.262799 | 3.414795 |
| 2           | 2.270207 | 2.311954 | 2.161612 | 1.972279 | 2.118149 |
| 3           | 1.540670 | 1.786979 | 1.529741 | 1.263564 | 1.458482 |
| 4           | 0.286257 | 0.374682 | 0.254703 | 0.286257 | 0.374682 |

interesting to compare the HRLR method to other methods when the error distribution is uniform.

4. Real data example 1

We considered a diabetes data set originally used in (Efron et al., 2004) shown in Table 1. This data set contains 10 baseline variables, age, sex, body mass index (BMI), average blood pressure (BP) and six blood serum measurements. The response of interest is a quantitative measure of disease progression 1 year after baseline.

We split the data into training and testing sets, having roughly 50% of the observations in each set. Table 5 shows the coefficient estimates for diabetes data based on the training set. The testing set was used to calculate the Test Root Mean Square Error (TRMSE) for each model built.

HRLR and lasso dropped the same three variables (age, ldl and tch) while relaxed lasso dropped four variables (age, sex, ldl and tch). Elastic net dropped three variables (age, sex and tch). All models dropped the age variable. Sex is typically an important predictor of diabetes. Consequently, one might question the decision to drop the sex variable made in the relaxed lasso and elastic net models.

The last row of Table 5 shows the TRMSE values. Among these TRMSE values, HRLR had the smallest TRMSE while elastic net had the highest.

5. Real data example 2

In the article, Fitting Percentage of Body Fat to Simple Body Measurements, Johnson (1996) uses the data at http://jse.amstat.org/datasets/fat.dat.txt provided to him by Dr. A. Garth Fischer in a personal communication on 5 October 1994, as a multiple linear regression activity with his students. A subset of the variables at http://jse.amstat.org/datasets/fat.dat.txt is available in the R package mfp by Ambler and Benner (2015) and the data set is frequently used in the text Statistical
Regression and Classification by Matloff (2017). We used a clean version of the data found at https://raw.githubusercontent.com/alanarnholt/MISCD/master/bodyfatClean.csv. The cleaned data set contains 17 potential predictor variables and a response variable named brozek_C.

The data were partitioned into a training and a testing set where roughly 70% of the observations were assigned to the training set and 30% to the testing set. Table 6 shows the coefficient estimates for “BodyFat” data. In this example, HRLR selected the least number of predictors followed by relaxed lasso, lasso, and elastic net, respectively. All the TRMSE values were very close. Among these values, relaxed lasso had the smallest TRMSE while ridge regression had the highest. The predictor variable abdomen_wrist was found by subtracting wrist measurements from the abdomen measurements (i.e. abdomen_wrist = abdomen-wrist). Therefore, the variables abdomen_cm, wrist_cm and abdomen_wrist are highly correlated and one of these three variables must be dropped from the model to reduce multicollinearity. HRLR, relaxed lasso and lasso agreed on dropping abdomen_cm while elastic net did not.

The collection of R functions hrlrpack, available from https://hasthika.github.io/hrlrpack.txt was used to create tables and figures. Both Tables 1 and 2, as well as Figures 1, 2, and 3 were made with hrlrsim. An implementation of the algorithm can be found in the R function hrlr and was used to produce Tables 4 and 5 for real data examples presented in Section 4.

Table 3. Median TRMSE and the SE values for Example 3 for the five methods considered with four error distributions

| Error Dist. | TRMSE | HRLR | relaxed | lasso | ridge | enet |
|-------------|-------|------|---------|-------|-------|------|
| 1           | 16.63881 | 17.73929 | 16.45876 | 16.05251 | 16.35268 |
| SE          | 1.049380 | 1.694161 | 0.750628 | 0.75907 | 0.823232 |
| 2           | 3.230074 | 7.631660 | 3.771314 | 3.589136 | 3.650914 |
| SE          | 1.353897 | 1.319291 | 1.626837 | 1.58390 | 1.581940 |
| 3           | 1.584614 | 6.978842 | 2.144613 | 2.168660 | 2.093443 |
| SE          | 0.278420 | 0.627340 | 1.390488 | 1.224328 | 1.302390 |
| 4           | 1.217255 | 6.957786 | 1.785902 | 1.820204 | 1.723091 |
| SE          | 0.291042 | 0.781399 | 1.427069 | 1.251697 | 1.335950 |
6. Conclusion

We have proposed a computationally efficient method, a hybrid of relaxed lasso and ridge regression (HRLR), for variable selection in linear models. The idea of HRLR is similar to the elastic net method where we used relaxed lasso instead of lasso. HRLR is likely to eliminate highly correlated variables more efficiently. Simulation and examples suggest that the number of variables selected by HRLR is usually smaller than the number of variables selected by

![Figure 3. TRMSE distributions of the five methods considered with four error distributions for Example 3.](image)

| Table 4. Diabetes data |
|------------------------|
| Patient | Age | Sex | BMI | BP | Serum measurements | y |
| 1 | 59 | 2 | 32.1 | 101 | 157 | 93.2 | 38 | 4 | 4.9 | 87 | 151 |
| 2 | 48 | 1 | 21.6 | 87 | 183 | 103.2 | 70 | 3 | 3.9 | 69 | 75 |
| 3 | 72 | 2 | 30.5 | 93 | 156 | 93.6 | 41 | 4 | 4.7 | 85 | 141 |
| 4 | 24 | 1 | 25.3 | 84 | 198 | 131.4 | 40 | 5 | 4.9 | 89 | 206 |
| 5 | 50 | 1 | 23.0 | 101 | 192 | 125.4 | 52 | 4 | 4.3 | 80 | 135 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 441 | 36 | 1 | 30.0 | 95 | 201 | 125.2 | 42 | 5 | 5.1 | 85 | 220 |
| 442 | 36 | 1 | 19.6 | 71 | 250 | 133.2 | 97 | 3 | 4.6 | 92 | 57 |

| Table 5. Coefficient estimates for diabetes data for the five methods considered |
|------------------------|
| Variable | HRLR | relaxed lasso | ridge | enet |
| age | 0.00000000 | 0.00000000 | 0.00000000 | 0.01742117 | 0.00000000 |
| sex | -0.02908175 | 0.00000000 | -0.01539929 | -0.03656159 | 0.00000000 |
| bmi | 0.33552927 | 0.35071797 | 0.34203828 | 0.30647396 | 0.33093482 |
| map | 0.13958292 | 0.13658879 | 0.13781028 | 0.14539792 | 0.13229055 |
| tc | -0.13690801 | -0.14827839 | -0.12538283 | -0.03965807 | -0.01895922 |
| ldl | 0.00000000 | 0.00000000 | 0.00000000 | -0.07896509 | -0.06693646 |
| hdl | -0.04423102 | -0.03158335 | -0.04174263 | -0.08563881 | -0.07253518 |
| tch | 0.00000000 | 0.00000000 | 0.00000000 | 0.02245709 | 0.00000000 |
| tlg | 0.33620499 | 0.35037166 | 0.33560931 | 0.26525703 | 0.28081150 |
| glu | 0.07397086 | 0.07484722 | 0.06815829 | 0.06763129 | 0.06180458 |
| TRMSE | 0.6782023 | 0.6853457 | 0.6799639 | 0.6796829 | 0.6860057 |
the other methods studied. The simulation results show that the HRLR compares favorably with other well-known modeling techniques, especially when the error distribution is non-normal. The analysis of both diabetes and body fat data set demonstrates the usefulness of the proposed method with real-world data. Finally, the HRLR method can be used for model fitting as well as a reliable variable selection technique with many error distributions.

**PUBLIC INTEREST STATEMENT**

When working with “Big Data”, variable selection has become a very important and a necessary step in model building across all disciplines. Among the numerous methods available for variable selection, lasso, relaxed lasso, and elastic net are widely used. In this article, we have proposed a novel method, a Hybrid of Relaxed Lasso and Ridge Regression (HRLR), where we integrate properties from both relaxed lasso and ridge regression to develop a computationally efficient variable selection method for linear models. The practitioners can produce linear models using HRLR method, by simply using the R functions provided in this article.

**Disclosure statement**

No potential conflict of interest was reported by the authors.

**Funding**

The authors received no direct funding for this research.

**References**

Ambler, O. B. G., & Benner, M. B. A. (2015). *mfp: Multivariable fractional polynomials*. CRAN (The Comprehensive R Archive Network). https://CRAN.R-project.org/package=mfp

R Core Team. (2017). *R: A language and environment for statistical computing*. R Foundation for statistical computing. https://www.R-project.org

Efron, B., Hastie, T., Johnstone, I., & Tibshirani, R. (2004). Least angle regression. *The Annals of Statistics*, 32(2), 407–499. https://doi.org/10.1214/009053604000000067

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1), 55–67. https://doi.org/10.1080/00401706.1970.1048634

Johnson, R. W. (1996). Fitting percentage of body fat to simple body measurements. *Journal of Statistics Education*, 4(1). https://doi.org/10.1080/10691989

Meinshausen, N. (2007). Relaxed lasso. *Computational Statistics & Data Analysis*, 52(1), 374–393. https://doi.org/10.1016/j.csda.2006.12.019

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

Wang, S., Nan, B., Rosset, S., & Zhu, J. (2011). Random lasso. *Annals of Applied Statistics*, 5(1), 468-485. https://arxiv.org/abs/1104.3398v1

Wang, Y., Jiang, Y., Zhang, J., Chen, Z., Xie, B., & Zhao, C. (2019). Robust variable selection based on the random quantile LASSO. *Communications in Statistics - Simulation and Computation*, 1–11. Taylor & Francis eprint. https://doi.org/10.1080/03610918.2019.1643886

Xie, J., & Zeng, L. (2010). Group variable selection methods and their applications in analysis of genomic data. In J. Feng, W. Fu, & F. Sun (Eds.), *Frontiers in computational and systems biology* (Vol. 15, pp. 231–248). Springer London. series Title: Computational Biology http://link.springer.com/10.1007/978-1-84996-196-7_12

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2), 301–320. https://doi.org/10.1111/j.1467-9868.2005.00503.x

**Table 6. Coefficient estimates for body fat data for the five methods considered**

| Variable         | HRLR         | relaxed     | lasso        | ridge        | enet         |
|------------------|--------------|-------------|--------------|--------------|--------------|
| age              | 0.0791059    | 0.04604482  | 0.04391435   | 0.090061076  | 0.05667671  |
| weight_lbs       | 0.000000000  | 0.000000000 | 0.000000000 | 0.000000000 | 0.000000000 |
| height_in        | −0.07738279  | −0.05938509 | −0.04079588  | −0.030997615 | −0.03472945 |
| neck_cm          | 0.000000000  | −0.057102588| −0.11440429  | −0.111206362 | −0.11676469 |
| chest_cm         | 0.000000000  | 0.000000000 | 0.000000000 | 0.000000000 | 0.000000000 |
| abdomen_cm       | 0.000000000  | 0.000000000 | 0.000000000 | 0.000000000 | 0.000000000 |
| hip_cm           | 0.000000000  | 0.000000000 | 0.000000000 | 0.000000000 | 0.000000000 |
| thigh_cm         | 0.000000000  | 0.022723755 | 0.05315623   | 0.075930628  | 0.05893404 |
| knee_cm          | 0.000000000  | 0.000000000 | −0.03966939  | −0.034524515 | −0.04357893 |
| ankle_cm         | 0.000000000  | 0.000000000 | 0.05916103   | 0.042586820  | 0.06092822 |
| biceps_cm        | 0.000000000  | 0.000000000 | 0.03821674   | 0.031027182  | 0.03830752 |
| forearm_cm       | 0.000000000  | 0.002414099 | 0.04565175   | 0.0352540075 | 0.04583286 |
| wrist_cm         | −0.12437328  | −0.10851797 | −0.16645279  | −0.197619699 | −0.20312188 |
| bmi_C            | 0.000000000  | 0.000000000 | 0.000000000 | 0.079378228  | 0.000000000 |
| age_sq           | 0.000000000  | 0.045628778 | 0.10885672   | 0.093390394  | 0.10213594 |
| abdomen_wrist    | 0.83846021   | 0.846912806 | 0.91846915   | 0.384097332  | 0.53899245 |
| am               | 0.000000000  | 0.034282390 | 0.07859840   | 0.080403867  | 0.09020931 |
| TRMSE            | 0.5210758    | 0.5182996   | 0.5236091    | 0.534444     | 0.5253579   |