Article

g.ridge: An R Package for the Generalized Ridge Regression for Sparse and High-Dimensional Linear Models

Takeshi Emura *, Koutarou Matsumoto , Ryuji Uozumi , Hirofumi Michimae

Posted Date: 15 January 2024

doi: 10.20944/preprints202401.1119.v1

Keywords: Cross-validation; High-dimensional data; Least squared estimator; Mean square error; Penalized regression; R package; Shrinkage estimator; Sparse model

Article

g.ridge: An R Package for the Generalized Ridge Regression for Sparse and High-Dimensional Linear Models

Takeshi Emura [1,2,*], Kotaro Matsumoto [3], Ryuji Uozumi [4] and Hirofumi Michimae [5]

[1] Research Center for Medical and Health Data Science, the Institute of Statistical Mathematics, 10-3 Midori-cho, Tachikawa, Tokyo 190-8562, Japan
[2] Biostatistics Center, Kurume University, Kurume, Japan
[3] Biostatistics Center, Kurume University, Kurume, Japan
[4] Department of Industrial Engineering and Economics, Tokyo Institute of Technology, Tokyo, Japan
[5] School of Pharmacy, Department of Clinical Medicine (Biostatistics), Kitasato University, Tokyo, Japan
* Correspondence: takeshiemura@gmail.com

Abstract: Ridge regression is one of the most popular shrinkage estimation methods for linear models. Ridge regression effectively estimates regression coefficients in the presence of high-dimensional regressors. Recently, a generalized ridge estimator was suggested by generalizing the uniform shrinkage of ridge regression to the non-uniform shrinkage, which was shown to perform well under sparse and high-dimensional linear models. In this paper, we introduce our newly developed R package "g.ridge" (the first version published on 2023-12-07 at https://cran.r-project.org/package=g.ridge) that implements both the ridge estimator and generalized ridge estimators. The package equips with the generalized cross-validation for automatic estimation of shrinkage parameters. The package also includes a convenient tool for generating a design matrix. By simulations, we test the performance of the R package under sparse and high-dimensional settings with the normal and skew-normal error distributions. From the simulation results, we conclude that the generalized ridge estimator is superior to the benchmark ridge estimator based on "glmnet", and hence, it can be the most recommended estimator under sparse and high-dimensional models. We demonstrate the package using the intracerebral hemorrhage data.

Keywords: Cross-validation; High-dimensional data; Least squared estimator; Mean square error; Penalized regression; R package; Shrinkage estimator; Sparse model

MSC: 62J05; 62J07;62P10; 62-04

1. Introduction

In linear regression, the least squares estimator (LSE) may not be suitable to estimate regression coefficients if the number of regressors p is higher than the sample size n (i.e., $p > n$). Ridge regression (Hoerl and Kennard 1970; Montgomery et al. 2021) is an effective alternative for the high-dimensional ($p > n$) data, and is widely employed in such data encountered in genetics (Arashi et al. 2021; Vishwakarma et al. 2021), medical studies (Friedrich et al. 2023), electromagnetic signals (Gao et al. 2023), grain yields (Hernandez et al., 2015), and others.

Hoerl and Kennard (1970) originally proposed the ridge estimator to reduce the problem in multicollinearity. Later, the ridge estimator is naturally adopted to the high-dimensional ($p > n$) problem (Golub *et al.*, 1979; Hastie et al., 2009) as a way to avoid overfitting phenomenon.

Ridge regression is a shrinkage estimator that shrinks all regression coefficients uniformly toward zero (Hoerl and Kennard 1970; Hastie *et al.*, 2009). This approach is particularly suitable for modeling microarrays or single nucleotide polymorphism (SNP) data, where many coefficients are close to zero (sparse models). For instance, Cule *et al.* (2011) applied the ridge estimator to the high-dimensional SNP data and performed significance testing for selecting an informative subset of SNPs. Similar applications of ridge regression to high-dimensional data include Whittaker *et al*. (2000), Cule and De Lorio (2013), and Yang and Emura (2017).

Unlike the ordinary ridge regression that shrinks all regression coefficients uniformly, the generalized ridge regression allows different degrees of shrinkage under multiple shrinkage parameters. Allen (1974) and Loesgen (1990) demonstrated that the multiple shrinkage parameters in the generalized ridge estimator arise naturally by utilizing prior information about regression coefficients; see also an extensive review paper of Yüzbaşı et al. (2020) for generalized ridge regression methods.

However, multiple shrinkage parameters are difficult to be estimated for high-dimensional cases. To overcome this difficulty, Yang and Emura (2017) suggested reducing the number of shrinkage parameters for the case of $p > n$ in their formulation of a generalized ridge estimator. The idea behind their approach is to assign two different weights (1 or 0.5) for shrinkage parameters by preliminary tests (Saleh and Kibria 1993; 2019; Norouzirad and Arashi 2019; Shih et al., 2021;2023; Taketomi et al. 2023). While this approach is shown to be promising due to its sound statistical performance under sparse and high-dimensional models, none of the software packages were implemented for the generalized ridge estimator.

This paper introduces our R package "g.ridge" (https://cran.r-project.org/package=g.ridge) that implements both the ridge estimator (Hoerl and Kennard 1970) and generalized ridge estimator (Yang and Emura 2017). The package equips with the generalized cross-validation for automatic estimation of shrinkage parameters. The package also includes a convenient tool for generating a design matrix. By simulations, we test the performance of the R package under the sparse and high-dimensional models, and the normal and skew-normal distributions for error terms. We conclude that the generalized ridge estimator is superior to the benchmark ridge estimator based on "glmnet", and hence, it can be the most recommended estimator under sparse and high-dimensional models. We illustrate the package via the intracerebral hemorrhage data.

The remainder of the paper is structured as follows. Section 2 reviews the ridge and generalized ridge estimators. Section 3 introduces the proposed R package. Section 4 tests the package via simulations. Section 5 gives a real data example. Section 6 concludes the paper.

2. Ridge regression and generalized ridge regression

This section introduces the ridge regression method proposed by Hoerl and Kennard (1970) and the generalized ridge regression proposed by Yang and Emura (2017).

2.1. Linear regression

Consider the linear regression model $\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where

$$\boldsymbol{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \qquad X = \begin{bmatrix} \boldsymbol{x}_1^{\mathrm{T}} \\ \vdots \\ \boldsymbol{x}_n^{\mathrm{T}} \end{bmatrix} = \begin{bmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{bmatrix}, \qquad \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}, \qquad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix};$$

X is a design matrix, $\boldsymbol{x}_i^{\mathrm{T}} = (x_{i1}, \ldots, x_{ip})$ is the transpose of the vector $\boldsymbol{x}_i$, $\boldsymbol{\beta}$ is a vector of regression coefficients, and $\boldsymbol{\varepsilon}$ is a vector of errors with $E[\boldsymbol{\varepsilon}] = \boldsymbol{0}$. In some case, we assume $\mathrm{Cov}[\boldsymbol{\varepsilon}] = \sigma^2 I_n$, where I_n is the identify matrix of dimension n. The assumption of the covariance is necessary to obtain the standard error (SE) of regression coefficients and testing their significance. Assume that X is standardized such that $\sum_{i=1}^{n} x_{ij} = 0$ and $\sum_{i=1}^{n} x_{ij}^2 = c$ for $j = 1, \ldots, p$, where c is n or $n-1$. Also, we assume that X does not include the intercept term (see Section 3.3 for details). These settings for X are usually imposed in ridge regression (Hastie et al. 2009).

If $X^{\mathrm{T}}X$ is invertible (non-singular), the LSE is defined as

$$\hat{\boldsymbol{\beta}}^{\mathrm{LSE}} = (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}\boldsymbol{y}.$$

Clearly, the LSE is not available when $X^{\mathrm{T}}X$ is singular, especially when $p > n$.

2.2. Ridge regression

Ridge estimator is an alternative to the LSE, which can be computed even when $p > n$. Hoerl and Kennard (1970) originally defined the ridge regression estimator as

$$\widehat{\boldsymbol{\beta}}(\lambda) = (X^{\mathrm{T}}X + \lambda I_n)^{-1}X^{\mathrm{T}}\boldsymbol{y},$$

where $\lambda > 0$ is called shrinkage parameter. A diagonal matrix λI_n added to $X^{\mathrm{T}}X$ makes all the components of $\widehat{\boldsymbol{\beta}}(\lambda)$ shrunk toward zero. Theorem 4.3 of Hoerl and Kennard (1970) shows that there exists $\lambda > 0$ such that the ridge estimator yields strictly smaller mean squared error (MSE) than that of the LSE.

Golub *et al*. (1979) suggested choosing λ on the basis of the predicted residual error sum of squares, or Allen's PRESS (Allen, 1974) statistics. Golub *et al*. (1979) proposed the rotation-invariant version of the PRESS statistics, and called it as the generalized cross-validation (GCV) criterion (Golub *et al*.,1979). The GCV function defined by Golub *et al*. (1979) is

$$V(\lambda) = \frac{1}{n}||\{I_n - A(\lambda)\}\boldsymbol{y}||^2 \Big/ \left[\frac{1}{n}\mathrm{Tr}\{I_n - A(\lambda)\}\right]^2. \tag{1}$$

where $A(\lambda) = X(X^{\mathrm{T}}X + \lambda I_n)X^{\mathrm{T}}$ is the "hat matrix". The GCV estimator of λ is defined as

$$\hat{\lambda} = \underset{\lambda \geq 0}{\mathrm{argmin}} V(\lambda).$$

Therefore, the resultant ridge estimator is

$$\widehat{\boldsymbol{\beta}}(\hat{\lambda}) = (X^{\mathrm{T}}X + \hat{\lambda} I_n)^{-1}X^{\mathrm{T}}\boldsymbol{y}.$$

The GCV theorem (Golub *et al*. 1979) guarantees the use of the above ridge estimator under the $p > n$ setup. Noted that the GCV criterion is different from the cross-validation criterion that is available in the widely used R function "cv.glmnet(.)" in the R package "glmnet". There are many other available criteria for choosing λ (Cule et al. 2013; Wong and Chiu 2015; Kibria and Banik 2016; Assaf et al. 2019; Michimae and Emura 2020), most of which are not applicable for the $p > n$ setup. Therefore, we will adopt the GCV criterion for the following discussions, which works well for both $p < n$ and $p > n$ setups.

2.3. Generalized ridge regression

Yang and Emura (2017) suggests relaxing the uniform shrinkage to the non-uniform shrinkage to yield the generalized ridge regression. For this purpose, Yang and Emura (2017) replaced the identity matrix I_n with the diagonal matrix $\widehat{W}(\varDelta)$ (defined later), and proposed

$$\widehat{\boldsymbol{\beta}}(\lambda, \varDelta) = \{X^{\mathrm{T}}X + \lambda \widehat{W}(\varDelta)\}^{-1}X^{\mathrm{T}}\boldsymbol{y},$$

where $\lambda > 0$ is a shrinkage parameter and $\varDelta \geq 0$ is called the "threshold" parameter. The diagonal components of $\widehat{W}(\varDelta)$ were suggested to be larger values (stronger shrinkage) for the components $\boldsymbol{\beta}$ closer to zero, yielding $\widehat{W}(\varDelta) = \mathrm{diag}\{\widehat{w}_1(\varDelta), \ldots, \widehat{w}_p(\varDelta)\}$, where

$$\widehat{w}_j(\varDelta) = \begin{cases} 1/2 & \text{if} \quad z_j \geq \varDelta, \\ 1 & \text{if} \quad z_j < \varDelta \end{cases}$$

where $z_j = \hat{\beta}_j^0/\mathrm{SD}(\widehat{\boldsymbol{\beta}}^0)$, and $\mathrm{SD}(\widehat{\boldsymbol{\beta}}^0) = \{\sum_{j=1}^{p}(\hat{\beta}_j^0 - \sum_{j=1}^{p}\hat{\beta}_j^0/p)^2/(p-1)\}^{1/2}$ for $j = 1, \ldots, p$, and $\widehat{\boldsymbol{\beta}}^0 = (\hat{\beta}_1^0, \ldots, \hat{\beta}_p^0)^{\mathrm{T}}$, defined as $\hat{\beta}_j^0 = X_j^{\mathrm{T}}\boldsymbol{y}/X_j^{\mathrm{T}}X_j$, where X_j is the j-th row of X. Note that $\widehat{\boldsymbol{\beta}}^0$ is called "compound covariate estimator" (Chen and Emura 2017).

The optimal value of $(\lambda, \varDelta)$ is estimated by the modified GCV function defined as

$$V(\lambda, \varDelta) = \frac{1}{n}||\{I_n - A(\lambda, \varDelta)\}\boldsymbol{y}||^2 \Big/ \left[\frac{1}{n}\mathrm{Tr}\{I_n - A(\lambda, \Delta)\}\right]^2, \tag{2}$$

where $A(\lambda, \varDelta) = X\{X^{\mathrm{T}}X + \lambda \widehat{W}(\varDelta)\}^{-1}X^{\mathrm{T}}$. Then the estimators $(\hat{\lambda}, \hat{\varDelta})$ are defined as

$$(\hat{\lambda}, \hat{\varDelta}) = \underset{\lambda \geq 0,\, \varDelta \geq 0}{\mathrm{argmin}}\, V(\lambda, \varDelta).$$

Computation of $(\hat{\lambda}, \hat{\Delta})$ is not difficult. Given Δ, the function $V(\lambda, \Delta)$ is continuous in λ, and hence it is easily minimized using any optimization scheme, such as the R function "optim(.)" to get $\hat{\lambda}(\Delta)$. Under the sparse model ($\boldsymbol{\beta} \approx \mathbf{0}$), the histogram of $\hat{\beta}_j^0/\mathrm{SD}(\widehat{\boldsymbol{\beta}}^0)$, $j = 1, \ldots, p$, is well-approximated by $N(0{,}1)$. This implies that $|\hat{\beta}_j^0|/\mathrm{SD}(\widehat{\boldsymbol{\beta}}^0)$ falls in the range $[0{,}3]$ with nearly 99.73%, and hence, a search range $\Delta \in [0{,}3]$ suffices. Since $V(\hat{\lambda}(\Delta), \Delta)$ is discontinuous in Δ, a grid search was suggested on the grid $D = \{0, 3/100, \ldots, 300/100\}$.

Finally, the generalized ridge estimator is defined as

$$\widehat{\boldsymbol{\beta}}(\hat{\lambda}, \hat{\Delta}) = \{X^{\mathrm{T}}X + \hat{\lambda}\widehat{W}(\hat{\Delta})\}^{-1}X^{\mathrm{T}}\boldsymbol{y}.$$

Also, the error variance can be estimated by

$$\hat{\sigma}^2 = \left\|\boldsymbol{y} - X\widehat{\boldsymbol{\beta}}(\hat{\lambda}, \hat{\Delta})\right\|^2 / \nu,$$

where $\nu = \mathrm{Tr}\{I_n - A(\hat{\lambda}, \hat{\Delta})\}^2$ and $A(\hat{\lambda}, \hat{\Delta}) = X\{X^{\mathrm{T}}X + \hat{\lambda}\widehat{W}(\hat{\Delta})\}^{-1}X^{\mathrm{T}}$.

2.4. Significance test

Ridge and generalized ridge estimators provide methods to test the null hypothesis

$$H_{0j}: \ \beta_j = 0 \quad \text{vs.} \quad H_{1j}: \ \beta_j \neq 0,$$

for $j = 1, \ldots, p$. One can perform significance testing, allowing one to access P-values of all the p regressors (Cule *et al.* 2011; Cule and De Lorio 2013; Yang and Emura 2017). Since the significance tests are similar between the ridge and generalized ridge estimators, we will explain the significance tests based on the generalized ridge estimator below.

Let $\hat{\beta}_j(\hat{\lambda}, \hat{\Delta})$ be the j-th component of $\widehat{\boldsymbol{\beta}}(\hat{\lambda}, \hat{\Delta})$. As in Cule *et al*. (2011) and Yang and Emura (2017), we define a Z-value $Z_j = \hat{\beta}_j(\hat{\lambda}, \hat{\Delta})/\mathrm{SE}\{(\hat{\lambda}, \hat{\Delta})\}$, where $\mathrm{SE}\{\hat{\beta}_j(\hat{\lambda}, \hat{\Delta})\}$ is the square root of the j-th diagonal of

$$\mathrm{Cov}\{\widehat{\boldsymbol{\beta}}(\hat{\lambda}, \hat{\Delta})\} = \hat{\sigma}^2\{X^{\mathrm{T}}X + \hat{\lambda}\widehat{W}(\hat{\Delta})\ \}^{-1}X^{\mathrm{T}}X\{X^{\mathrm{T}}X + \hat{\lambda}\widehat{W}(\hat{\Delta})\}^{-1},$$

We define the P-value as $P_j = 2\{1 - \Phi(|Z_j|)\}$, where $\Phi(.)$ is the cumulative distribution function of $N(0, 1)$. One can then select a subset of regressors by specifying a significance level.

3. R package: g.ridge

We implemented the methods of Section 2 in the R package "g.ridge". This section explains the main R function "g.ridge(.)", and the other function "X.mat(.)". Appendix A explains another function "GCV(.)" that may not be directly used, but is useful for internal computing.

3.1. Generating data

The design matrix X and response vector $\boldsymbol{y}$ are necessary to perform linear regression (Section 2.1). Therefore, following Section 5 of Yang and Emura (2017, p.6093), we implemented a convenient R function that generates X having three independent blocks:

$$X = \begin{bmatrix} \boldsymbol{x}_1^{\mathrm{T}} \\ \vdots \\ \boldsymbol{x}_n^{\mathrm{T}} \end{bmatrix} = \left[\begin{array}{ccc|ccc|ccc} x_{11} & \cdots & x_{1q} & x_{1,q+1} & \cdots & x_{1,q+r} & x_{1,q+r+1} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nq} & x_{n,q+1} & \cdots & x_{n,q+r} & x_{n,q+r+1} & \cdots & x_{np} \end{array}\right], \tag{3}$$

$$= [\quad \text{First block} \quad | \quad \text{Second block} \quad | \quad \text{Third block} \quad],$$

where the first block consists of $q \geq 0$ correlated regressors (correlation = 0.5), and the second block consists of $r \geq 0$ correlated regressors (correlation = 0.5), and the third block consists of $p - q - r > 0$ independent regressors. That is,

$$\mathrm{Corr}(x_{ij}, x_{ik}) = \begin{cases} 0.5 & \text{if} \quad j, k \in \{1, \ldots, q\}, \\ 0.5 & \text{if} \quad j, k \in \{q+1, \ldots, q+r\}, \\ 0 & \text{otherwise}. \end{cases}$$

doi:10.20944/preprints202401.1119.v1

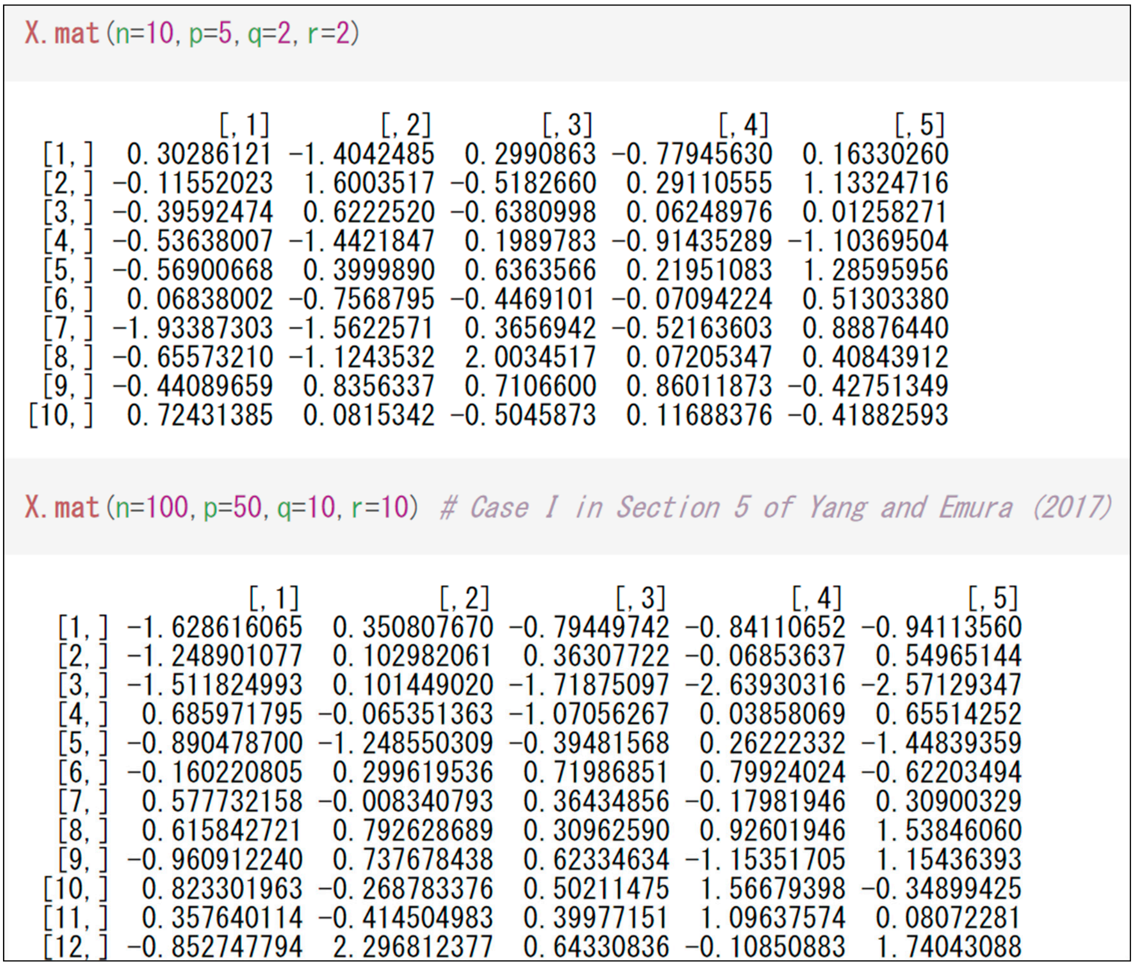
The design matrix mimics the "gene pathway" for two blocks of correlated gene expressions often used in simulation studies (Binder et al. 2009; Emura et al. 2012; 2016; 2023). The values $q = 0$ and $r = 0$ give a design matrix with all independent columns.

The marginal distributions of p regressors follow $N(0,1)$, which is achieved by

$$\boldsymbol{x}_i^{\mathrm{T}} = \left(\frac{1}{\sqrt{2}}(z_{i1} + u_i), \dots, \frac{1}{\sqrt{2}}(z_{iq} + u_i) \,\middle|\, \frac{1}{\sqrt{2}}(z_{i,q+1} + v_i), \dots, \quad \frac{1}{\sqrt{2}}(z_{i,q+r} + v_i) \,\middle|\, z_{i,q+r+1}, \dots, \quad z_{ip} \right),$$

where $z_{i1}, \dots, z_{ip}$ and u_i, v_i all independently follow $N(0,1)$ for $i = 1, \dots, n$.

Figure 1 shows an example for generating design matrices using "X.mat(.)". One can simply input n, p, q, and r into "X.mat(.)" by noting the constraint $q + r < p$.



```
X.mat(n=10, p=5, q=2, r=2)
```

```
             [,1]       [,2]       [,3]        [,4]        [,5]
 [1,]  0.30286121 -1.4042485  0.2990863 -0.77945630  0.16330260
 [2,] -0.11552023  1.6003517 -0.5182660  0.29110555  1.13324716
 [3,] -0.39592474  0.6222520 -0.6380998  0.06248976  0.01258271
 [4,] -0.53638007 -1.4421847  0.1989783 -0.91435289 -1.10369504
 [5,] -0.56900668  0.3999890  0.6363566  0.21951083  1.28595956
 [6,]  0.06838002 -0.7568795 -0.4469101 -0.07094224  0.51303380
 [7,] -1.93387303 -1.5622571  0.3656942 -0.52163603  0.88876440
 [8,] -0.65573210 -1.1243532  2.0034517  0.07205347  0.40843912
 [9,] -0.44089659  0.8356337  0.7106600  0.86011873 -0.42751349
[10,]  0.72431385  0.0815342 -0.5045873  0.11688376 -0.41882593
```

```
X.mat(n=100, p=50, q=10, r=10) # Case I in Section 5 of Yang and Emura (2017)
```

```
              [,1]         [,2]        [,3]        [,4]        [,5]
 [1,] -1.628616065  0.350807670 -0.79449742 -0.84110652 -0.94113560
 [2,] -1.248901077  0.102982061  0.36307722 -0.06853637  0.54965144
 [3,] -1.511824993  0.101449020 -1.71875097 -2.63930316 -2.57129347
 [4,]  0.685971795 -0.065351363 -1.07056267  0.03858069  0.65514252
 [5,] -0.890478700 -1.248550309 -0.39481568  0.26222332 -1.44839359
 [6,] -0.160220805  0.299619536  0.71986851  0.79924024 -0.62203494
 [7,]  0.577732158 -0.008340793  0.36434856 -0.17981946  0.30900329
 [8,]  0.615842721  0.792628689  0.30962590  0.92601946  1.53846060
 [9,] -0.960912240  0.737678438  0.62334634 -1.15351705  1.15436393
[10,]  0.823301963 -0.268783376  0.50211475  1.56679398 -0.34899425
[11,]  0.357640114 -0.414504983  0.39977151  1.09637574  0.08072281
[12,] -0.852747794  2.296812377  0.64330836 -0.10850883  1.74043088
```

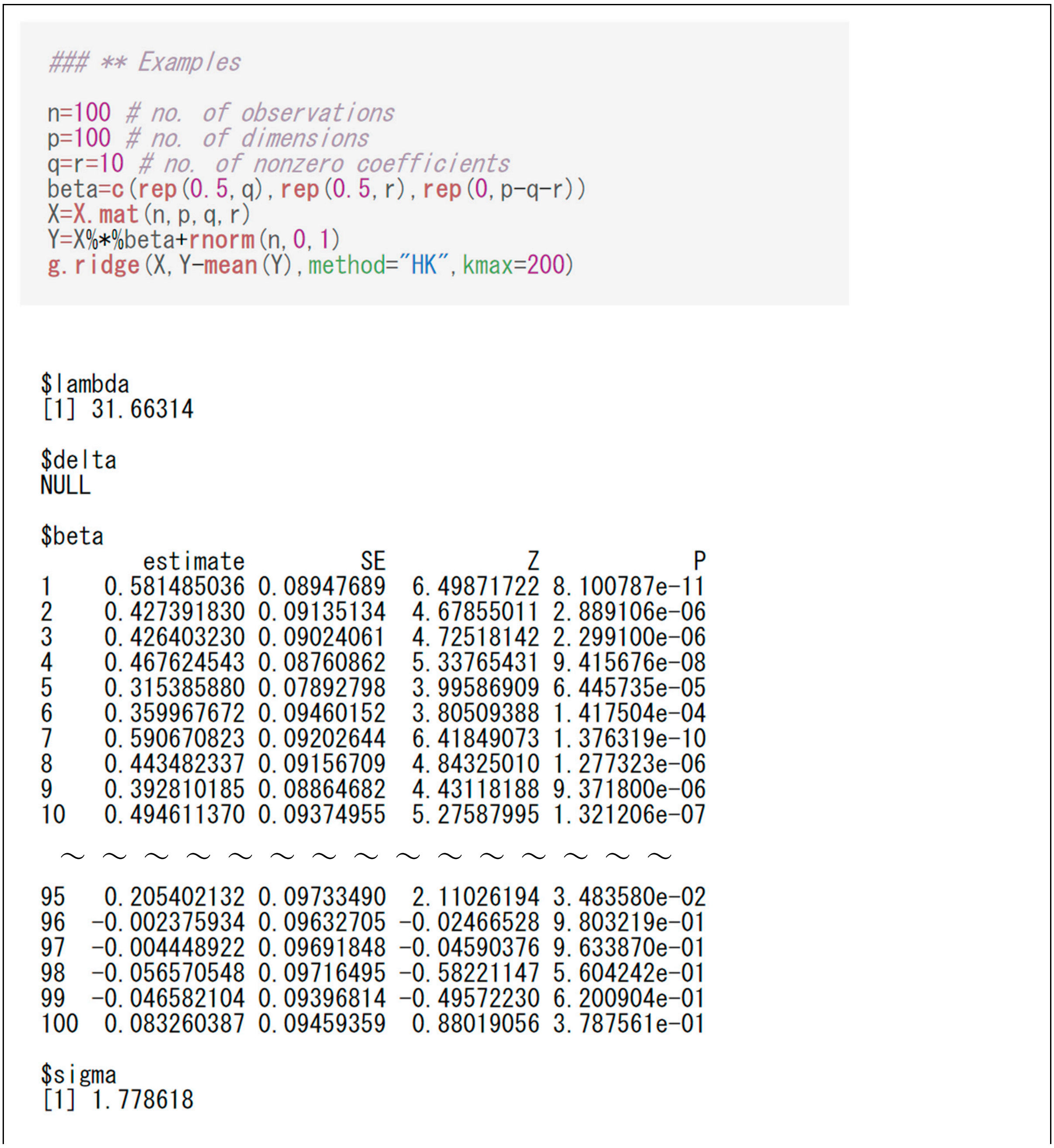
Figure 1. Examples for generating design matrices using "X.mat(.)".

3.2. Performing regression

The ridge and generalized ridge estimators can be computed by the R function "g.ridge(.)" whose input requires a design matrix X and a response vector $\boldsymbol{y}$. Specifically, the command "g.ridge(X, Y, method = "HK", kmax = 500)" can calculate the ridge estimator $\widehat{\boldsymbol{\beta}}(\hat{\lambda})$, where "HK" stands for "Hoerl and Kennard", and "kmax=500" means the range $0 \le \lambda \le 500$ for estimating λ. The output of the command includes $\hat{\lambda} \in [0, 500]$ and $\widehat{\boldsymbol{\beta}}(\hat{\lambda})$. Similarly, the command "g.ridge(X, Y, method = "YE", kmax = 500)" can calculate $\widehat{\boldsymbol{\beta}}(\hat{\lambda}, \hat{\Delta})$, where "YE" stands for "Yang and Emura". The output also displays the plot of $V(\lambda)$ against $\lambda \in [0, 500]$, and its minimizer $\hat{\lambda} \in [0, 500]$ (the plot of $V(\lambda, \hat{\Delta})$ for the generalized ridge).

The R function "g.ridge(.)" can also compute the SE of $\widehat{\boldsymbol{\beta}}$, Z-value and P-value for significance tests (Section 2.4), and the estimate of the error variance $\hat{\sigma}^2$ (Section 2.3). As in the typical practice of ridge regression, we used the centered responses "Y-mean(Y)" rather than "Y" (will be explained in Section 3.3). Also, if "X" were not generated by "X.mat", the scaled design matrix "scale(X)" would be recommended rather than "X" itself.

Figure 2 shows the code and output for the ridge estimator. The output shows $\hat{\lambda} = 31.66314$, $\widehat{\boldsymbol{\beta}}(\hat{\lambda}) = (0.581485036, \ldots 0.083260387)$, SE, Z-value, P-value, and $\hat{\sigma}^2 = 1.778618$. The output also displays the GCV function $V(\lambda)$ against $\lambda \in [0, 200]$, showing its minimum at $\hat{\lambda} = 31.66314$. In the code, we changed "kmax" from the default value (500) to 200 for a better visualization of the plot. In many cases, users will need to try different values of "kmax" to reach a desirable plot for the GCV function.



```
### ** Examples

n=100 # no. of observations
p=100 # no. of dimensions
q=r=10 # no. of nonzero coefficients
beta=c(rep(0.5,q),rep(0.5,r),rep(0,p-q-r))
X=X.mat(n,p,q,r)
Y=X%*%beta+rnorm(n,0,1)
g.ridge(X,Y-mean(Y),method="HK",kmax=200)
```

```
$lambda
[1] 31.66314

$delta
NULL

$beta
       estimate         SE           Z            P
1    0.581485036 0.08947689  6.49871722 8.100787e-11
2    0.427391830 0.09135134  4.67855011 2.889106e-06
3    0.426403230 0.09024061  4.72518142 2.299100e-06
4    0.467624543 0.08760862  5.33765431 9.415676e-08
5    0.315385880 0.07892798  3.99586909 6.445735e-05
6    0.359967672 0.09460152  3.80509388 1.417504e-04
7    0.590670823 0.09202644  6.41849073 1.376319e-10
8    0.443482337 0.09156709  4.84325010 1.277323e-06
9    0.392810185 0.08864682  4.43118188 9.371800e-06
10   0.494611370 0.09374955  5.27587995 1.321206e-07

 ～ ～ ～ ～ ～ ～ ～ ～ ～ ～ ～ ～ ～ ～ ～

95   0.205402132 0.09733490  2.11026194 3.483580e-02
96  -0.002375934 0.09632705 -0.02466528 9.803219e-01
97  -0.004448922 0.09691848 -0.04590376 9.633870e-01
98  -0.056570548 0.09716495 -0.58221147 5.604242e-01
99  -0.046582104 0.09396814 -0.49572230 6.200904e-01
100  0.083260387 0.09459359  0.88019056 3.787561e-01

$sigma
[1] 1.778618
```

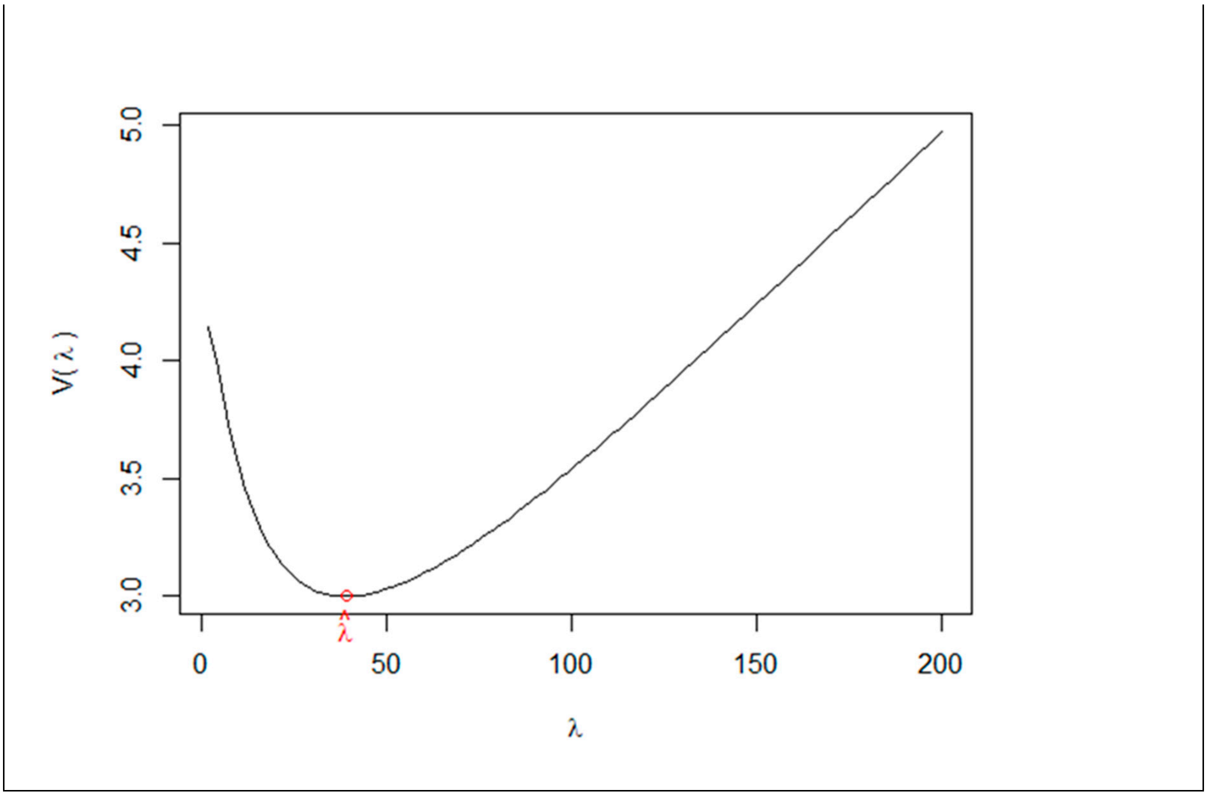


Figure 2. The R code and output for calculating the ridge estimator using "g.ridge(.)".

The output of the generalized ridge estimator is similar to that of the ridge estimator. Only the difference is an additional parameter estimate $\hat{\Delta} \in [0, 3]$ shown at "delta$".

3.3. Technical remarks on centering and standardization

We assume that X is standardized and does not include the intercept term (Section 2.1). If X is generated by "X.mat", it is already standardized, and hence there is no need to do the standardization. However, in other cases, X has to be standardized, e.g. by the R command "scale(X)". By assumption, the design matrix X does not include the intercept term (a column of ones) since one can always use the reduced model $\boldsymbol{y} - (\sum_{i=1}^{n} y_i / n)\mathbf{1} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ for the intercept model $\boldsymbol{y} = \beta_0 \mathbf{1} + X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ (β_0 is estimated by $\sum_{i=1}^{n} y_i / n$), where $\mathbf{1}$ is a vector of ones. This means that the usual unbiased estimator is applied to the intercept.

4. Simulations

We conducted simulation studies to test our R package "g.ridge". The goals of the simulations are to show: (a) the generalized ridge estimators in our package exhibits superior performance over the ridge estimator in the package "glmnet(.)", and (b) the sound performance under the normally and non-normally distributed errors (the skew-normal distribution will be considered). Supplementary Materials include the R code to reproduce the results of the simulation studies.

4.1. Simulation settings

We generated X by "X.mat" with $n = 100$, $p \in \{50,\ 100,\ 150,\ 200\}$, and $q = r = 10$ based on Equation (3). Given X, we set $\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, with $\boldsymbol{\beta}$ defined by the sparse model:

$$\boldsymbol{\beta} = (\overbrace{b/q,\dots,b/q}^{q},\ \overbrace{d/r,\dots,d/r}^{r},\ \overbrace{0,\dots,0}^{p-q-r})^{\mathrm{T}},$$

for four cases: (**I**) $b = d = 5$; (**II**) $b = d = 10$; (**III**) $b = 5,\ d = -5$; (**IV**) $b = 10,\ d = -10$. Errors $\boldsymbol{\varepsilon}$ were generated independently from the normal distribution, or the skew-normal distribution (Azzalini 2014); both distributions have mean zero and standard deviation one, and the skew-normal distribution has the slant parameter ten (alpha=10 in the R function "rsn(.)"). Figure 3 shows the remarkable difference of the two distributions. The skew-normal distribution was not previously examined in the simulation setting of Yang and Emura (2017).

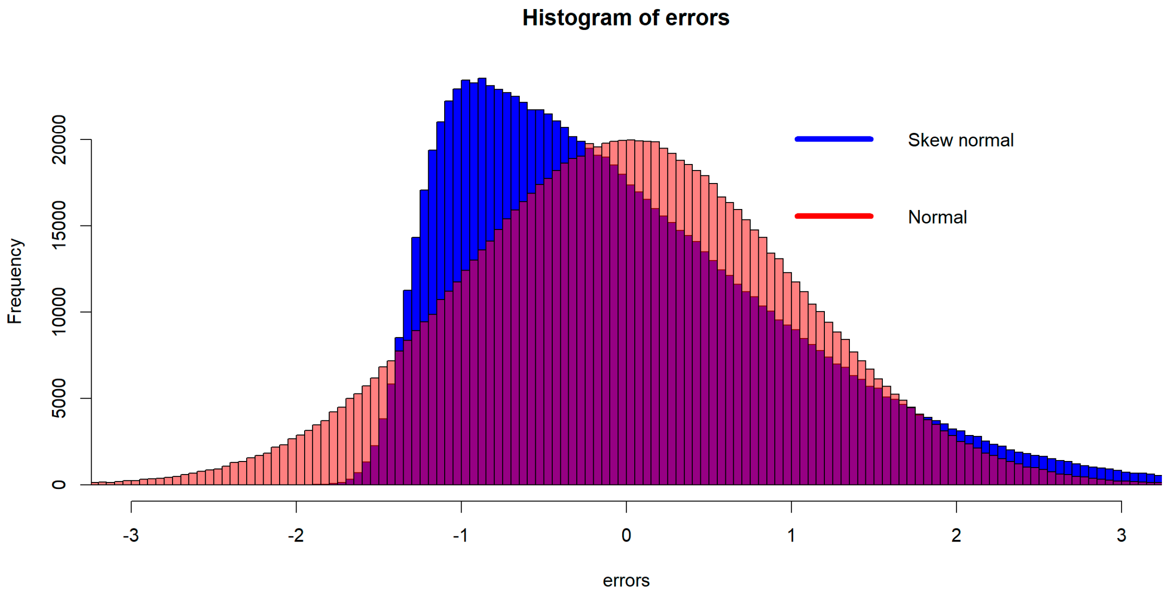


Figure 3. The histogram of the normal and skew-normal distributions (the slant parameter ten; alpha=10 in the R function "rsn(.)"). Both distributions have mean 0 and standard deviation 1.

For a given dataset $(X, \boldsymbol{y})$, we computed $\widehat{\boldsymbol{\beta}}$ that can be (i) the ridge estimator by "g.ridge(.)", (ii) the generalized ridge estimator by "g.ridge(.)", or (iii) the ridge estimator by "glmnet(.)". Based on 500 replications (on random errors $\boldsymbol{\varepsilon}$), the performances of the three proposed estimators were compared by the total mean squared error (TMSE) criterion defined as

$$\mathrm{TMSE}(\widehat{\boldsymbol{\beta}}) = \mathrm{E}\left[\left\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right\|^2\right] = \mathrm{E}\left[\sum\nolimits_{j=1}^{p} \left(\hat{\beta}_j - \beta_j\right)^2\right],$$

where $\mathrm{E}[.]$ was performed by the Monte Carlo average over 500 replications.

4.2. Simulation results

Table 1 compares the MSE among the three estimators: (i) the ridge by "g.ridge(.)", (ii) the generalized ridge by "g.ridge(.)", and (iii) the ridge by "glmnet(.)". We see that the generalized ridge estimator is superior to the ridge estimator since the former has smaller TMSE values for all cases. Also, the generalized ridge estimator is superior to the "glmnet(.)" estimator for cases of $p \in \{100, 150, 200\}$, while it is not the case for $p = 50$. This conclusion holds consistently across the four parameter settings (**I**)-(**IV**) and two error distributions (normal and skew-normal). In conclusion, the generalized ridge estimator in the R proposed package is the most recommended estimator for data with sparse and high-dimensional settings.

Table 1. The total mean squared error (TMSE) of the three estimators: (i) the ridge by "g.ridge(.)", (ii) the generalized (g-) ridge by "g.ridge(.)", and (iii) the ridge by "glmnet(.)". The TMSE is computed by the Monte Carlo average over 500 replications.

Error distribution	Regression coefficients	p	(i) ridge	(ii) g-ridge	(iii) glmnet
Normal	(**I**) $b = d = 5$	50	0.463	0.385	0.306
		100	0.950	0.682	2.182
		150	1.146	0.658	1.996
		200	1.520	0.920	2.199

	(**II**) $b = d = 10$	50	0.855	0.681	0.545
		100	2.151	1.562	8.688
		150	3.008	1.482	7.904
		200	4.929	2.687	8.691
	(**III**) $b = 5$ and $d = -5$	50	0.602	0.539	0.388
		100	0.990	0.628	2.025
		150	1.219	0.703	2.132
		200	1.589	0.953	2.226
	(**IV**) $b = 10$ and $d = -10$	50	1.541	1.290	0.737
		100	2.398	1.580	8.046
		150	3.231	1.614	8.434
		200	4.651	2.770	8.804
Skew-normal	(**I**) $b = d = 5$	50	0.440	0.361	0.294
		100	0.957	0.670	2.182
		150	1.162	0.678	2.000
		200	1.500	0.910	2.197
	(**II**) $b = d = 10$	50	0.821	0.655	0.527
		100	2.285	1.705	8.691
		150	3.021	1.509	7.905
		200	4.883	2.673	8.686
	(**III**) $b = 5$ and $d = -5$	50	0.576	0.519	0.376
		100	0.974	0.622	2.029
		150	1.233	0.721	2.137
		200	1.582	0.949	2.243
	(**IV**) $b = 10$ and $d = -10$	50	1.504	1.273	0.720
		100	2.449	1.508	8.054
		150	3.224	1.616	8.453
		200	4.618	2.731	8.860

NOTE: We set the sample size $n = 100$ all cases.

5. Data analysis

This section analyzes a real dataset to illustrate the generalized ridge estimator in the proposed package. We retrospectively analyzed a dataset from patients with intracerebral hemorrhage who were hospitalized at Saiseikai Kumamoto Hospital, Kumamoto city, Japan. The study was conducted in accordance with the Declaration of Helsinki, and the protocol was approved by the Ethics Committee of Saiseikai Kumamoto Hospital on March 29, 2023. Saiseikai Kumamoto Hospital is a regional tertiary hospital that serves patients with stroke in southern Japan and provides acute care in a comprehensive stroke care unit.

The outcome variables ($\boldsymbol{y}$) are the changes of the blood volume (in mL) from the initial value and to the follow-up value, which were measured by the CT scan. Excluding patients with less than 10 mL blood volume at the initial CT scan and other inclusion/exclusion criteria, we arrive at $n = 172$ patients. Regressor variables (X) consist of $p = 35$ continuous measurements, including histological variables (e.g., age), vital signs (e.g., blood pressure, mmHG; respiratory rate, times/minute), and blood measurements (e.g. albumin, g/dL; Gamma-glutamyl transpeptidase (Gamma-GT), IU/L; lactate dehydrogenase, IU/L; Prothrombin time, second; Blood platelet count, 10^{3}/μL; C-reactive protein, mg/dL). The responses and regressors are centered and standardized before fitting a linear model as explained in Sections 2.1 and 3.3.

Figure 3 displays scatter plots for the centered responses $\boldsymbol{y} - (\sum_{i=1}^{n} y_i / n)\mathbf{1}$ against the predictors $X\widehat{\boldsymbol{\beta}}$ based on the ridge estimator and generalized ridge estimator. We observe that the predictors reasonably captured the variability of the changes in the blood volume (response variables). However, the figure also shows that the changes in the blood volumes were not fully explained by the predictors since some residual errors remain. Figure 4 depicts the residuals constructed by the generalized ridge estimator. We observe that the residuals are asymmetric around zero. The asymmetry in the errors does not yield any problem as the proposed ridge estimators was verified for the asymmetric errors (Section 4).

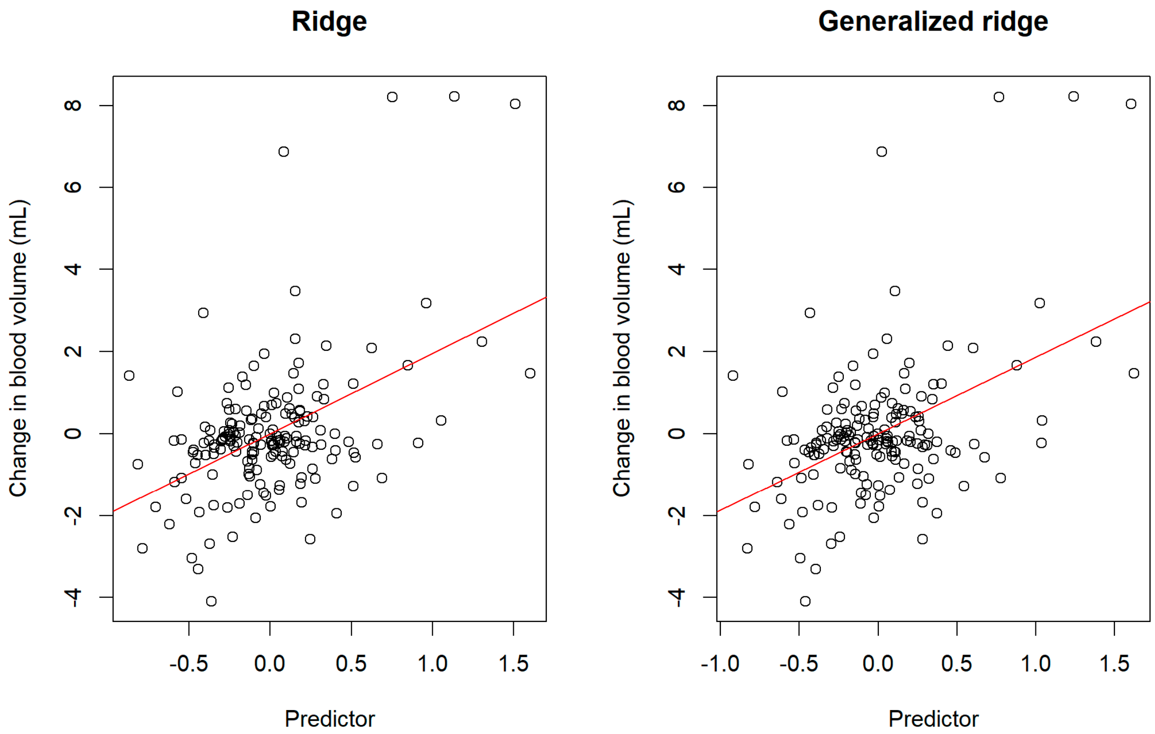


Figure 3. The responses $\boldsymbol{y} - (\sum_{i=1}^{n} y_i \,/n)\mathbf{1}$ against the predictors $X\widehat{\boldsymbol{\beta}}$ based on the ridge estimator and generalized ridge estimator applied to the intracerebral hemorrhage data.

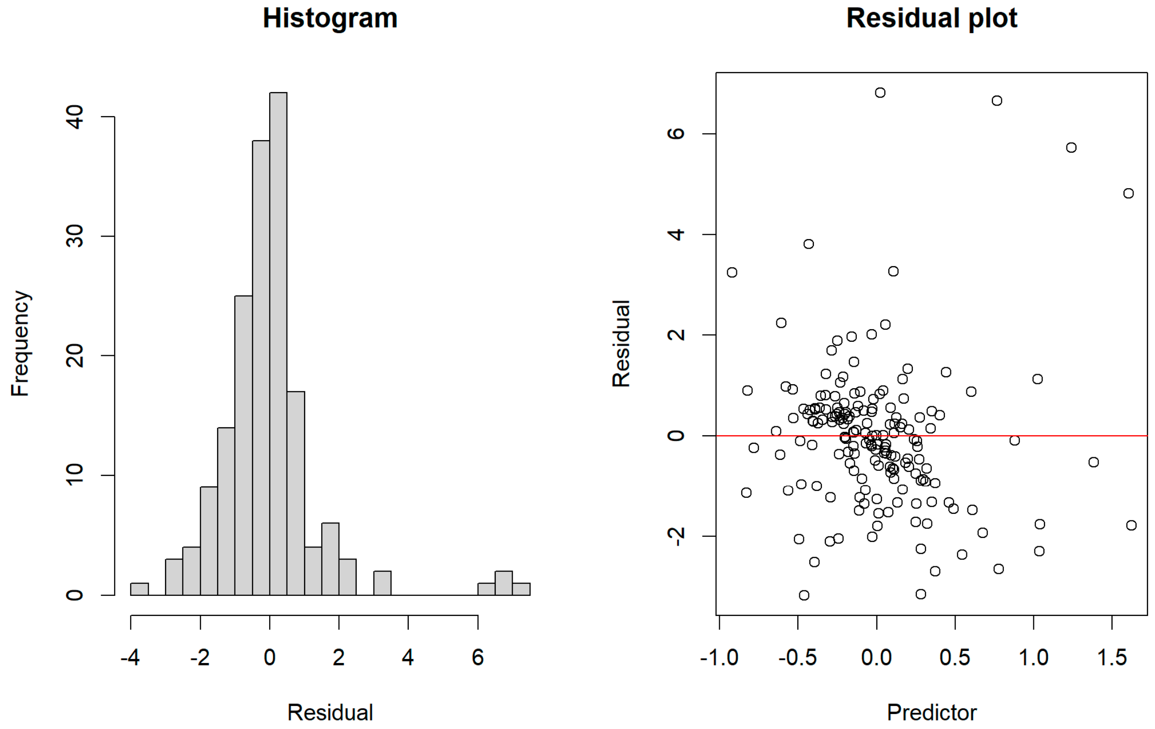


Figure 4. The fitted residuals for the generalized ridge estimator applied to a dataset on patients with intracerebral hemorrhage.

Table 2 shows the fitted results for estimated regression coefficients sorted by *P*-values (only with *P* < 0.05). The fitted results are quite similar between the ridge and generalized ridge estimators. The most significant regressor is the lactate dehydrogenase. This variable was previously reported as an important predictor of hematoma expansion (Chu et al. 2019). An interesting difference is found in the number of significant regressors: 6 regressors for the generalized ridge estimator, and 5 regressors for the ridge estimator. That is, the C-reactive protein as a regressor selected solely by the generalized ridge estimator. A study conducted in a large population in China reported that a high C-reactive protein level was an independent risk factor for the severe intracerebral hemorrhage (Wang et al. 2021). The selection of C-reactive protein as a variable associated with increased hematoma volume may reflect the tendency of patients with severe intracerebral hemorrhage to have easier hematoma expansion.

Table 2. The fitted results for estimated regression coefficients (only with *P*-value < 0.05) sorted by P-values applied to a dataset on patients with intracerebral hemorrhage.

	Ridge			**Generalized ridge**		
	$\widehat{\boldsymbol{\beta}}_j$	**SE**	***P*-value**	$\widehat{\boldsymbol{\beta}}_j$	**SE**	***P*-value**
Lactate dehydrogenase	0.122	0.047	0.008	0.145	0.055	0.008
Gamma-GT	0.116	0.048	0.016	0.143	0.056	0.010
Respiratory rate	-0.120	0.052	0.020	-0.140	0.059	0.018
Prothrombin time	0.077	0.036	0.031	0.083	0.040	0.038
Blood platelet count	-0.100	0.049	0.040	-0.114	0.056	0.044
C-reactive protein	None	None	> 0.05	0.112	0.057	0.049

Finally, we compared the performance of the ridge and generalized ridge estimators by means of the residual mean squared error (RMSE) defined as

$$\text{RMSE} = \left\| \boldsymbol{y} - \left(\frac{\sum_{i=1}^{n} y_i}{n} \right) \mathbf{1} - X\widehat{\boldsymbol{\beta}} \right\|^2 .$$

The RMSE were 2.288 (the ridge estimator) and 2.284 (the generalize ridge estimator). Therefore, the predictor constructed from the generalized ridge estimator had slightly better performance over the one from the ridge estimator.

Therefore, we have demonstrated that the ridge and generalize ridge estimators in the proposed R package provide statistically and biologically sound conclusions on the real data analysis.

6. Conclusion

This paper introduced the new R package “g.ridge” that can perform the ridge and generalized ridge regression methods. We showed that the generalized ridge estimator in the proposed package performs better than the widely used ridge estimator in the “glmnet” package. Furthermore, the ridge and generalized ridge estimators in the proposed package can deal with asymmetric error distributions. With the abundance of sparse and high-dimensional data (Kim et al. 2007; Zhang et al. 2021; Vishwakarma et al. 2021; Bhattacharjee 2022; Bhatnagar et al. 2023; Emura et al. 2023) and asymmetrically distributed data (Abe et al. 2012; Huynh et al. 2021; Yoshiba et al. 2023; Jimichi et al. 2023), the proposed package may provide a reliable statistical tool for statisticians and data scientists.

The generalized ridge estimator considered in this article may be extended to logistic regression, Poisson regression, and Cox regression. Such extensions have not been explored yet. While the extensions might not be technically difficult, well-designed simulation studies and implementations in some software packages will be necessary to fully justify the advantage and usefulness over the usual ridge estimator that is widespread via the “glmnet” package.

Institutional Review Board Statement: The study (of Section 5) was conducted in accordance with the Declaration of Helsinki, and the protocol was approved by the Ethics Committee of Saiseikai Kumamoto Hospital (Approval No. 1156 and March 29, 2023).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 January 2024 doi:10.20944/preprints202401.1119.v1

Informed Consent Statement: Given the retrospective nature of the study and the use of anonymized patient data in Section 5, written informed consent was not required. Patients preferred not to be included in this study were given the opportunity to opt out.

Data Availability Statement:Supplementary Materials include the R code to reproduce the results of the simulation studies. The dataset used in Section 5 is available from the corresponding author on reasonable request.

Acknowledgements: Emura T is supported financially by the research grant from the JSPS KAKENHI (grant no. 22K11948, and grant no. 20H04147).

Conflicts of Interest: The authors declare that they have no competing interests.

Appendix A. GCV function

The GCV functions are defined as $V(\lambda)$ and $V(\lambda, \Delta)$ in Equations (1) and (2), respectively. The ridge estimator uses $V(\lambda)$ while the generalized ridge estimator uses $V(\lambda, \Delta)$ for estimating shrinkage parameters. Therefore, we made an R function "CGV(X,Y,k,W)" to help computing $V(\lambda)$ and $V(\lambda, \Delta)$, where "X" is a design matrix, "Y" is a response vector, "k" is actually λ (because "lambda" is a long name). Note that "W" can be any square matrix of dimension p to allow for general use. Thus, what we actually compute in "CGV(X,Y,k,W)" is

$$V(k,W) = \frac{1}{n}||\{I_n - A(k,W)\}\boldsymbol{y}||^2 \Big/ \left[\frac{1}{n}\mathrm{Tr}\{I_n - A(k,W)\}\right]^2, \tag{3}$$

where $A(k,W) = X\{X^{\mathrm{T}}X + kW\}^{-1}X^{\mathrm{T}}$ for any symmetric matrix W. However, we normally use $W = I_n$ for the ridge or $W = \widehat{W}(\Delta)$ for the generalized ridge as defined in Section 2.3.

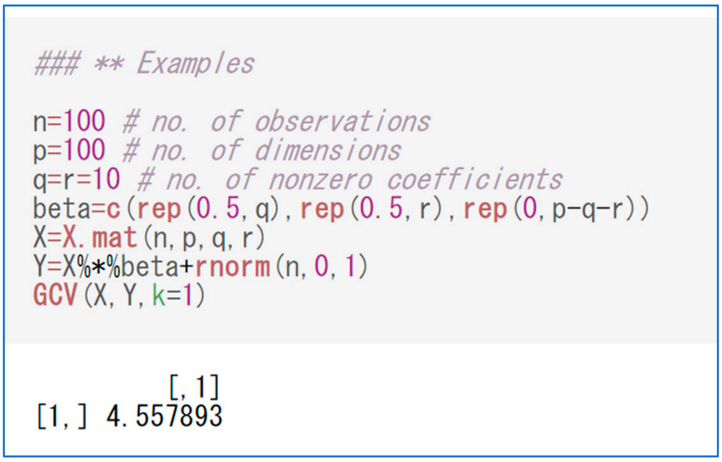
Figure A1 shows the R code for using "CGV(X,Y,k,W)". The default for "W" is $W = I_n$ and therefore "GCV(X,Y,k=1)" can compute

$$V(1,I_n) = \frac{1}{n}||\{I_n - A(1,I_n)\}\boldsymbol{y}||^2 \Big/ \left[\frac{1}{n}\mathrm{Tr}\{I_n - A(1,I_n)\}\right]^2$$

for $A(1,I_n) = X\{X^{\mathrm{T}}X + I_n\}^{-1}X^{\mathrm{T}}$, or equivalently,

$$V(1) = \frac{1}{n}||\{I_n - A(1)\}\boldsymbol{y}||^2 \Big/ \left[\frac{1}{n}\mathrm{Tr}\{I_n - A(1)\}\right]^2$$

for $A(1) = X\{X^{\mathrm{T}}X + I_n\}^{-1}X^{\mathrm{T}}$.



```
### ** Examples

n=100 # no. of observations
p=100 # no. of dimensions
q=r=10 # no. of nonzero coefficients
beta=c(rep(0.5,q),rep(0.5,r),rep(0,p-q-r))
X=X.mat(n,p,q,r)
Y=X%*%beta+rnorm(n,0,1)
GCV(X,Y,k=1)
```

```
         [,1]
[1,] 4.557893
```

Figure A1. An example for using "CGV(X,Y,k,W)".

References

Abe, T., Shimizu, K., Kuuluvainen, T., & Aakala, T. (2012). Sine-skewed axial distributions with an application for fallen tree data. *Environmental and Ecological Statistics*, **19**, 295-307.

Allen, D. M. (1974). The relationship between variable selection and data augmentation and a method for prediction. *Technometrics* **16**, 125-127.

Arashi, M., Roozbeh, M., Hamzah, N. A., & Gasparini, M. (2021). Ridge regression and its applications in genetic studies. *Plos one*, **16**(4), e0245376.

Assaf, A. G., Tsionas, M., & Tasiopoulos, A. (2019). Diagnosing and correcting the effects of multicollinearity: Bayesian implications of ridge regression. *Tourism Management*, **71**, 1-8.

Azzalini, A. with the collaboration of Capitanio, A. (2014). *The Skew-Normal and Related Families*. Cambridge University Press, IMS Monographs series

Binder, H., Allignol, A., Schumacher, M., Bayersmmann, J. (2009). Boosting for high-dimensional time-to-event data with competing risks, *Bioinformatics*, **25**, 890-896.

Bhattacharjee, A. (2022). *Big Data Analytics in Oncology with R*. CRC Press.

Bhatnagar, S. R., Lu, T., Lovato, A., Olds, D. L., Kobor, M. S., Meaney, M. J., et al. (2023). A sparse additive model for high-dimensional interactions with an exposure variable. *Computational Statistics & Data Analysis*, **179**, 107624.

Chen, A. C., Emura, T. (2017). A modified Liu-type estimator with an intercept term under mixture experiments, *Communications in Statistics-Theory and Method*, **46**(13): 6645-6667.

Chu, H., Huang, C., Dong, J., Yang, X., Xiang, J., Dong, Q., & Tang, Y. (2019). Lactate dehydrogenase predicts early hematoma expansion and poor outcomes in intracerebral hemorrhage patients. *Translational Stroke Research*, **10**, 620-629.

Cule, E., and De Iorio, M. (2013). Ridge regression in prediction problems: automatic choice of the ridge parameter. *Genetic Epidemiology* **37**, 704-714.

Cule, E., Vineis, P. and De Iorio, M. (2011). Significance testing in ridge regression for genetic data. *BMC Bioinformatrics* **12**, 372.

Emura, T., Chen, Y. H., and Chen, H. Y. (2012). Survival prediction based on compound covariate under Cox proportional hazard models. *PLoS ONE* **7**, e47627.

Emura, T., Chen, Y. H. (2016). Gene selection for survival data under dependent censoring: a copula- based approach. *Statistical Methods in Medical Research*, **25**(6), 2840-2857.

Emura, T., Hsu, W. C., & Chou, W. C. (2023). A survival tree based on stabilized score tests for high-dimensional covariates. *Journal of Applied Statistics*, **50**(2), 264-290.

Friedrich, S., Groll, A., Ickstadt, K., Kneib, T., Pauly, M., Rahnenführer, J., & Friede, T. (2023). Regularization approaches in clinical biostatistics: A review of methods and their applications. *Statistical Methods in Medical Research*, **32**(2), 425-440.

Gao S, Zhu G, Bialkowski A, Zhou X (2023). Stroke Localization Using Multiple Ridge Regression Predictors Based on Electromagnetic Signals. *Mathematics*. **11**(2):464.

Golub, G. H., Heath, M. and Wahba, G. (1979). Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics* **21**, 215-223.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning*. Springer-Verlag, New York.

Hernandez, J.; Lobos, G.A.; Matus, I.; Del Pozo, A.; Silva, P.; Galleguillos, M (2015). Using Ridge Regression Models to Estimate Grain Yield from Field Spectral Data in Bread Wheat (Triticum Aestivum L.) Grown under Three Water Regimes. *Remote Sens*. **7**, 2109-2126

Hoerl, A. E. and Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* **12**, 55-67.

Huynh, U., Pal, N., & Nguyen, M. (2021). Regression model under skew-normal error with applications in predicting groundwater arsenic level in the Mekong Delta Region. *Environmental and Ecological Statistics*, **28**, 323-353.

Jimichi, M., Kawasaki, Y., Miyamoto, D., Saka, C., & Nagata, S. (2023). Statistical Modeling of Financial Data with Skew-Symmetric Error Distributions. *Symmetry*, **15**(9), 1772.

Kibria, B. M. G. and Banik, S. (2016). Some ridge regression estimators and their performances. *Journal of Modern Applied Statistical Methods* **15** (1), 206-238.

Kim, S.-Y. and Lee, J.-W. (2007). Ensemble clustering method based on the resampling similarity measure for gene expression data. *Statistical Methods in Medical Research* **16**, 539-564.

Loesgen, K.-H. (1990). A generalization and Bayesian interpretation of ridge-type estimators with good prior means. *Statistical Papers* **31**, 147-154.

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2021). *Introduction to Linear Regression Analysis*. John Wiley & Sons.

Michimae, H., & Emura, T. (2022). Bayesian ridge estimators based on copula-based joint prior distributions for regression coefficients. *Computational Statistics*, **37**(5), 2741-2769.

Norouzirad, M., & Arashi, M. (2019). Preliminary test and Stein-type shrinkage ridge estimators in robust regression. *Statistical Papers*, **60**, 1849-1882.

Saleh, E. A. M., & Kibria, G. B. M. (1993). Performance of some new preliminary test ridge regression estimators and their properties. *Communications in Statistics-Theory and Methods*, **22**(10), 2747-2764.

Saleh, A. M. E., Arashi, M., & Kibria, B. G. (2019). *Theory of Ridge Regression Estimation with Applications*. John Wiley & Sons.

Shih, J. H., Lin, T. Y., Jimichi, M., & Emura, T. (2021). Robust ridge M-estimators with pretest and Stein-rule shrinkage for an intercept term. *Japanese Journal of Statistics and Data Science*, **4**, 107-150.

Shih, J. H., Konno, Y., Chang, Y. T., & Emura, T. (2023). A class of general pretest estimators for the univariate normal mean. *Communications in Statistics-Theory and Methods*, **52**(8), 2538-2561.

Taketomi, N., Chang, Y. T., Konno, Y., Mori, M., & Emura, T. (2023). Confidence interval for normal means in meta-analysis based on a pretest estimator. *Japanese Journal of Statistics and Data Science*, https://doi.org/10.1007/s42081-023-00221-2.

Vishwakarma, G. K., Thomas, A., & Bhattacharjee, A. (2021). A weight function method for selection of proteins to predict an outcome using protein expression data. Journal of Computational and Applied *Mathematics*, **391**, 113465.

Wang, D., Wang, J., Li, Z., Gu, H., Yang, K., Zhao, X., & Wang, Y. (2022). C-reaction protein and the severity of intracerebral hemorrhage: a study from chinese stroke center alliance. *Neurological Research*, **44**(4), 285-290.

Whittaker, J. C., Thompson, R., and Denham, M. C., (2000). Marker-assisted selection using ridge regression. *Genetical Research* **75**, 249-252.

Wong, K. Y., Chiu, S. N. (2015). An iterative approach to minimize the mean squared error in ridge regression. *Computational Statistics*, **30**(2):625-639.

Yang, S. P., & Emura, T. (2017). A Bayesian approach with generalized ridge estimation for high-dimensional regression and testing. *Communications in Statistics-Simulation and Computation*, **46**(8), 6083-6105.

Yoshiba, T.; Koike, T.; Kato, S. (2023). On a Measure of Tail Asymmetry for the Bivariate Skew-Normal Copula. *Symmetry*, **15**, 1410.

Yüzbaşı, B., Arashi, M., & Ejaz Ahmed, S. (2020). Shrinkage Estimation Strategies in Generalised Ridge Regression Models: Low/High‐Dimension Regime. *International Statistical Review*, **88**(1), 229-251.

Zhang, Q., Ma, S., & Huang, Y. (2021). Promote sign consistency in the joint estimation of precision matrices. *Computational Statistics & Data Analysis*, **159**, 107210.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.