



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

Takeshi Emura <sup>1,2,\*</sup>, Koutarou Matsumoto <sup>2</sup>, Ryuji Uozumi <sup>3</sup> and Hirofumi Michimae <sup>4</sup>

- <sup>1</sup> Research Center for Medical and Health Data Science, The Institute of Statistical Mathematics, Tokyo 190-8562, Japan
- <sup>2</sup> Biostatistics Center, Kurume University, Kurume 830-0011, Japan; matsumoto\_koutarou@kurume-u.ac.jp
- <sup>3</sup> Department of Industrial Engineering and Economics, Tokyo Institute of Technology, Tokyo 152-8552, Japan; uozumi.r.ab@m.titech.ac.jp
- <sup>4</sup> Department of Clinical Medicine (Biostatistics), School of Pharmacy, Kitasato University, Tokyo 108-8641, Japan; michimaeh@pharm.kitasato-u.ac.jp
- \* 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 that involved generalizing the uniform shrinkage of ridge regression to non-uniform shrinkage; this was shown to perform well in sparse and high-dimensional linear models. In this paper, we introduce our newly developed R package "g.ridge" (first version published on 7 December 2023) that implements both the ridge estimator and generalized ridge estimator. The package is equipped with generalized cross-validation for the 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 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 the R package "glmnet". Hence the generalized ridge estimator may be the most recommended estimator for sparse and high-dimensional models. We demonstrate the package using intracerebral hemorrhage data.

**Keywords:** cross-validation; high-dimensional data; intracerebral hemorrhage; least squares estimator; mean square error; penalized regression; R package; shrinkage estimator; sparse model

# 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 [1,2] is an effective alternative for high-dimensional (p > n) data, and is widely employed in such data encountered in genetics [3,4], medical studies [5], electromagnetic signals [6], grain yields [7], and others.

Hoerl and Kennard [1] originally proposed the ridge estimator to reduce the problem in multicollinearity. Later, the ridge estimator was naturally adopted to high-dimensional (p > n) problems [8,9] as a way to avoid overfitting phenomena.

Ridge regression is a shrinkage estimator that shrinks all regression coefficients uniformly toward zero [1,9–11]. 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. [12] applied the ridge estimator to highdimensional SNP data and performed significance testing for selecting an informative subset of SNPs. Similar applications of ridge regression to high-dimensional data have been performed by Whittaker et al. [13], Cule and De Lorio [14], and Yang and Emura [15].



Citation: Emura, T.; Matsumoto, K.; Uozumi, R.; Michimae, H. g.ridge: An R Package for Generalized Ridge Regression for Sparse and High-Dimensional Linear Models. *Symmetry* 2024, *16*, 223. https:// doi.org/10.3390/sym16020223

Academic Editors: Alexei F. Cheviakov, Toshihiro Abe, Kunio Shimizu and Tomoaki Imoto

Received: 12 January 2024 Revised: 7 February 2024 Accepted: 9 February 2024 Published: 12 February 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). Unlike ordinary ridge regression that shrinks all regression coefficients uniformly, the generalized ridge regression allows different degrees of shrinkage under multiple shrinkage parameters [1,16–18]. Loesgen [18] demonstrated that multiple shrinkage parameters in the generalized ridge estimator arise naturally from prior uncertainty (variance) for regression coefficients. Shen et al. [19] and Hofheinz and Frisch [20] adopted this idea and proposed a generalized ridge estimator for high-dimensional SNPs where multiple shrinkage parameters are estimated preliminarily. Readers are referred to a review paper of Yüzbaşı et al. [21] for more information about generalized ridge regression.

However, multiple shrinkage parameters are difficult to estimate for high-dimensional cases. To overcome this difficulty, Yang and Emura [15] 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 was to assign two different weights (1 or 0.5) for shrinkage parameters via preliminary tests [22–26]. While this approach has sound statistical performance in sparse and high-dimensional models, no 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, accessed on 5 February 2024) that implements both the ridge estimator of Hoerl and Kennard [1] and the generalized ridge estimator of Yang and Emura [15]. The package is equipped with generalized cross-validation for the 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 with sparse and high-dimensional models, and normal and skew-normal distributions for error terms. We conclude that the generalized ridge estimator is superior to the benchmark ridge estimator based on the R package "glmnet". Hence the generalized ridge estimator may be the most recommended estimator for sparse and high-dimensional models. We illustrate the package via the intracerebral hemorrhage dataset.

The remainder of this 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 [1] and the generalized ridge regression proposed by Yang and Emura [15].

### 2.1. Linear Regression

Consider a linear regression model ( $y = X\beta + \varepsilon$ ) where

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

*X* is a design matrix,  $\mathbf{x}_i^{\mathrm{T}} = (x_{i1}, \dots, x_{ip})$  is the transpose of a vector  $\mathbf{x}_i$ ,  $\boldsymbol{\beta}$  is a vector of regression coefficients, and  $\boldsymbol{\varepsilon}$  is a vector of errors with  $E[\boldsymbol{\varepsilon}] = \mathbf{0}$ . In some cases, we assume  $\operatorname{Cov}[\boldsymbol{\varepsilon}] = \sigma^2 I_n$ , where  $I_n$  is the identity matrix of dimension n. The assumption of the covariance is necessary to obtain the standard error (SE) of regression coefficients and test 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, ..., 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 [9].

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

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

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

2.2. Ridge Regression

The ridge estimator is an alternative to LSE, and can be computed even when p > n. Hoerl and Kennard [1] originally defined the ridge regression estimator as

$$\hat{\boldsymbol{\beta}}(\lambda) = \left( \boldsymbol{X}^{\mathrm{T}} \boldsymbol{X} + \lambda \boldsymbol{I}_{n} \right)^{-1} \boldsymbol{X}^{\mathrm{T}} \boldsymbol{y},$$

where  $\lambda > 0$  is called the shrinkage parameter. A diagonal matrix  $\lambda I_n$  added to  $X^T X$  makes all the components of  $\hat{\beta}(\lambda)$  shrink toward zero. Theorem 4.3 of Hoerl and Kennard [1] shows that there exists  $\lambda > 0$ , such that the ridge estimator yields a strictly smaller mean squared error (MSE) compared to that of the LSE.

Golub et al. [8] suggested choosing  $\lambda$  on the basis of the predicted residual error sum of squares, or Allen's PRESS [17] statistics. Golub et al. [8] proposed the rotationinvariant version of the PRESS statistics, and called it the generalized cross-validation (GCV) criterion [8]. The GCV function defined by Golub et al. [8] is

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

where  $A(\lambda) = X(X^TX + \lambda I_n)X^T$  is a "hat matrix". The GCV estimator of  $\lambda$  is defined as

$$\hat{\lambda} = \underset{\lambda \ge 0}{\operatorname{argmin}} V(\lambda).$$

Therefore, the resultant ridge estimator is

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

The GCV theorem [8] justifies the above ridge estimator under the p > n setup. Note that the GCV criterion is different from the cross-validation criterion that is adopted in the R package "glmnet". There are many other criteria for choosing  $\lambda$  [14,27–31], most of which are not applicable or justified for the p > n setup. Therefore, we will adopt the GCV criterion for the following discussions.

## 2.3. Generalized Ridge Regression

Yang and Emura [15] suggest relaxing the uniform shrinkage to the non-uniform shrinkage to yield the generalized ridge regression. For this purpose, Yang and Emura [15] replaced the identity matrix  $I_n$  with the diagonal matrix  $\hat{W}(\Delta)$  (defined later), and proposed

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

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

$$\hat{w}_j(\Delta) = egin{cases} 1/2 & ext{if } z_j \geq \Delta, \ 1 & ext{if } z_j < \Delta \end{cases}$$

where  $z_j = \hat{\beta}_j^0 / \text{SD}(\hat{\beta}^0)$ , and  $\text{SD}(\hat{\beta}^0) = \left\{ \sum_{j=1}^p \left( \hat{\beta}_j^0 - \sum_{j=1}^p \hat{\beta}_j^0 / p \right)^2 / (p-1) \right\}^{1/2}$  for j = 1, ..., p, and  $\hat{\beta}^0 = \left( \hat{\beta}_1^0, ..., \hat{\beta}_p^0 \right)^T$ , defined as  $\hat{\beta}_j^0 = X_j^T y / X_j^T X_j$ , where  $X_j$  is the *j*-th column of X. Note that  $\hat{\beta}^0$  is called the "compound covariate estimator" [32].

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

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

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

$$(\hat{\lambda}, \hat{\Delta}) = \operatorname*{argmin}_{\lambda \ge 0, \, \Delta \ge 0} V(\lambda, \Delta)$$

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

Finally, the generalized ridge estimator is defined as

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

Also, the error variance can be estimated by

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

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

## 2.4. Significance Test

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

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

for j = 1, ..., p. One can perform significance tests, allowing one to access *p*-values of all the *p* regressors [12,14,15]. Since the resultant tests are similar between the ridge and generalized ridge estimators, we will explain the tests based on the generalized ridge estimator below.

Let  $\hat{\beta}_j(\hat{\lambda}, \hat{\Delta})$  be the *j*-th component of  $\hat{\beta}(\hat{\lambda}, \hat{\Delta})$ . As in Cule et al. [12] and Yang and Emura [15], we define a Z-value of  $Z_j = \hat{\beta}_j(\hat{\lambda}, \hat{\Delta}) / \text{SE}\{(\hat{\lambda}, \hat{\Delta})\}$ , where  $\text{SE}\{\hat{\beta}_j(\hat{\lambda}, \hat{\Delta})\}$  is the square root of the *j*-th diagonal of

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

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 two functions: "X.mat" for generating data and "g.ridge(.)" for performing regression.

## 3.1. Generating Data

The design matrix X and response vector y are necessary to perform linear regression (Section 2.1). Therefore, following Section 5 of Yang and Emura ([15], p. 6093), we implemented a convenient R function that generates X with three independent blocks:

$$X = \begin{bmatrix} \mathbf{x}_{1}^{\mathrm{T}} \\ \vdots \\ \mathbf{x}_{n}^{\mathrm{T}} \end{bmatrix} = \begin{bmatrix} x_{11} \cdots x_{1q} & x_{1,q+1} \cdots x_{1,q+r} & x_{1,q+r+1} \cdots x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} \cdots & x_{nq} & x_{n,q+1} \cdots & x_{n,q+r} & \vdots & \ddots & \vdots \\ x_{n,q+r+1} \cdots & x_{np} \end{bmatrix}, \quad (3)$$
$$= [\text{ First block}| \text{ Second block}| \text{ Third block}],$$

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

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

The design matrix mimics the "gene pathway" for two blocks of correlated gene expressions often used in simulation studies [33–36]. 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

$$\boldsymbol{x}_{i}^{\mathrm{T}} = \left(\frac{1}{\sqrt{2}}(z_{i1}+u_{i}), \dots, \frac{1}{\sqrt{2}}(z_{iq}+u_{i})\right| \frac{1}{\sqrt{2}}(z_{i,q+1}+v_{i}), \dots, \frac{1}{\sqrt{2}}(z_{i,q+r}+v_{i}) \mid z_{i,q+r+1}, \dots, z_{ip}),$$

where  $z_{i1}$ , ...,  $z_{ip}$  and  $u_i$ ,  $v_i$  all independently follow N(0, 1) for i = 1, ..., 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)
$ \begin{bmatrix} 1, 1 \\ 0.30286121 \\ -1.4042485 \\ 0.2990863 \\ -0.77945630 \\ 0.16330260 \\ 2, 1 \\ -0.11552023 \\ 1.6003517 \\ -0.5182600 \\ 0.29110555 \\ 1.13324716 \\ 3, 1 \\ -0.39592474 \\ 0.6222520 \\ -0.6380998 \\ 0.06248976 \\ 0.01258271 \\ 4, 1 \\ -0.53638007 \\ -1.4421847 \\ 0.1989783 \\ -0.91435289 \\ -1.10369504 \\ 5, 1 \\ -0.56900668 \\ 0.3999890 \\ 0.6363566 \\ 0.21951083 \\ 1.28595956 \\ 6, 1 \\ 0.06838002 \\ -0.7568795 \\ -0.4469101 \\ -0.07094224 \\ 0.51303380 \\ 7, 1 \\ -1.93387303 \\ -1.5622571 \\ 0.3656942 \\ -0.52163603 \\ 0.88876440 \\ 18, 1 \\ -0.65573210 \\ -1.1243532 \\ 2.0034517 \\ 0.07205347 \\ 0.40843912 \\ 19, 1 \\ -0.4408659 \\ 0.835637 \\ 0.7106600 \\ 0.86011873 \\ -0.42751349 \\ 10, 102837 \\ 0.72431385 \\ 0.0815342 \\ -0.5045873 \\ 0.11688376 \\ -0.41882593 \\ \end{bmatrix} $
X.mat(n=100, p=50, q=10, r=10) # Case I in Section 5 of Yang and Emura (2017)
$ \begin{bmatrix} 1, 1 \\ -1. 628616065 \\ 2, 1 \\ -1. 248901077 \\ 0. 102982061 \\ 0. 350807670 \\ -0. 79449742 \\ -0. 84110652 \\ -0. 94113560 \\ -0. 685971795 \\ -0. 06853613 \\ -1. 511824993 \\ 0. 101449020 \\ -1. 71875097 \\ -2. 63930316 \\ -2. 57129347 \\ -2. 63930316 \\ -2. 57129347 \\ -2. 63930316 \\ -2. 57129347 \\ -2. 63930316 \\ -2. 57129347 \\ -2. 63930316 \\ -2. 57129347 \\ -2. 63930316 \\ -2. 57129347 \\ -2. 63930316 \\ -2. 57129347 \\ -2. 63930316 \\ -2. 57129347 \\ -2. 63930316 \\ -2. 57129347 \\ -2. 63930316 \\ -2. 57129347 \\ -2. 63930316 \\ -2. 57129347 \\ -2. 57129347 \\ -0. 62203494 \\ -7. 1 \\ 0. 577732158 \\ -0. 008340793 \\ 0. 36434856 \\ -0. 17981946 \\ 0. 30900329 \\ -0. 960912240 \\ 0. 737678438 \\ 0. 62334634 \\ -1. 15351705 \\ 1. 55679398 \\ -0. 34899425 \\ -0. 357640114 \\ -0. 414504983 \\ 0. 39977151 \\ 1. 09637574 \\ 0. 08072281 \\ -0. 852747794 \\ 2. 296812377 \\ 0. 64330836 \\ -0. 10850883 \\ 1. 74043088 \\ -2. 572747794 \\ -2. 59274779$

Figure 1. Examples for generating design matrices using "X.mat(.)" [15].

## 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 *y*. Specifically, the command "g.ridge(X, Y, method = "HK", kmax = 500)" can calculate  $\hat{\beta}(\hat{\lambda})$ , where "HK" stands for "Hoerl and Kennard", and "kmax = 500" means the range  $0 \le \lambda \le 500$  for estimating  $\lambda$ . The output displays  $\hat{\lambda} \in [0, 500]$  and  $\hat{\beta}(\hat{\lambda})$ . Similarly, the command "g.ridge(X, Y, method = "YE", kmax = 500)" can calculate  $\hat{\beta}(\hat{\lambda}, \hat{\Delta})$ , 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  $\hat{\beta}$ , 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, one should use the centered responses "Y-mean(Y)" rather than "Y" (this 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  $\hat{\lambda} = 31.66314$ ,  $\hat{\beta}(\hat{\lambda}) = (0.581485036, \dots 0.083260387)$ , 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. Appendix A explains another function, "GCV(.)", that is used for internally computing the GCV function  $V(\lambda)$ .

```
#### ** 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
```

\$heta

Ψ		La				
		estimate	SE	Z	Р	
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	
1	0	0. 494611370	0.09374955	5. 27587995	1.321206e-07	
	$\sim$	$\sim$ $\sim$ $\sim$ $\sim$	$\sim$ $\sim$ $\sim$ $\sim$	$\sim$ $\sim$ $\sim$ $\sim$	$\sim$	
9	5	0.205402132	0.09733490	2.11026194	3.483580e-02	
9	6	-0.002375934	0.09632705	-0. 02466528	9.803219e-01	
9	7	-0.004448922	0.09691848	-0. 04590376	9.633870e-01	
9	8	-0. 056570548	0.09716495	-0. 58221147	5. 604242e-01	
9	9	-0.046582104	0.09396814	-0. 49572230	6. 200904e-01	
1	00	0. 083260387	0.09459359	0.88019056	3.787561e-01	

\$sigma

[1] 1.778618

Figure 2. Cont.



**Figure 2.** The R code and output for calculating the ridge estimator using "g.ridge(.)". The red circle in the graph shows the minimum value at  $\hat{\lambda} = 31.66314$ .

The output of the generalized ridge estimator is similar to that of the ridge estimator. The only 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 have a column for an intercept term (Section 2.1). If *X* is generated by "X.mat", it is already standardized, and hence there is no need to perform 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 a column of ones (for an intercept) since one can always use the reduced model  $y - (\sum_{i=1}^{n} y_i/n)\mathbf{1} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$  derived from the intercept model  $y = \beta_0 \mathbf{1} + X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$  ( $\beta_0$  is estimated by  $\sum_{i=1}^{n} y_i/n$ ), where **1** is a vector of ones. This means that  $\hat{\beta}_0 = \sum_{i=1}^{n} y_i/n$  is applied to estimate  $\beta_0$ .

### 4. Simulations

We conducted simulation studies to test our R package "g.ridge". The goals of the simulations are to show that the generalized ridge estimators in our package exhibit (a) superior performance over the ridge estimator in the R function "glmnet(.)" and (b) sound performance under normally and non-normally distributed errors (a skew-normal distribution will be considered). The 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  $y = X\beta + \varepsilon$ , with  $\beta$  defined by a sparse model of

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

for four cases: (I) b = d = 5; (II) b = d = 10; (III) b = 5, d = -5; (IV) b = 10, d = -10. Errors  $\varepsilon$  were generated independently from the normal distribution or the skew-normal distribution [37]; both distributions had a mean of zero and standard deviation of one, and the skew-normal distribution had a slant parameter of ten (alpha = 10 in the R function "rsn(.)"). Figure 3 shows the remarkable difference between the two distributions. The skew-normal distribution was not previously examined in the simulation setting of Yang and Emura [15].

#### Histogram of errors



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

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

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

where E[.] was performed by a Monte Carlo average over 500 replications.

# 4.2. Simulation Results

Table 1 compares the TMSE 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 this 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 proposed R package seems to be the most recommended estimator for data with sparse and high-dimensional settings.

Error Distribution	Regression Coefficients	p	(i) ridge	(ii) g-ridge	(iii) glmnet
Normal	( <b>I</b> ) $b = d = 5$	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
	<b>(II)</b> $b = d = 10$	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
	( <b>IV</b> ) $b = 10$ 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
	<b>(II)</b> $b = d = 10$	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
	( <b>IV</b> ) $b = 10$ 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

**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 a Monte Carlo average over 500 replications.

Note: We set the sample size n = 100 for 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 data collection was conducted in accordance with the Declaration of Helsinki, and the protocol was approved by the Ethics Committee of Saiseikai Kumamoto Hospital on 29 March 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.

Outcome variables (*y*) are the changes in blood volume (in mL) from the initial value to the follow-up value, which were measured by CT scans. Excluding patients with less than 10 mL blood volume at the initial CT scans and other reasonable 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,  $10^{-3}/\mu$ L; C-reactive protein, mg/dL). The

responses and regressors are centered and standardized before fitting the linear model, as explained in Sections 2.1 and 3.3.

Table 2 shows estimates for regression coefficients to compare their values between the ridge and generalized ridge estimators. Due to the limited space, we only show regression coefficients with p < 0.05 (5% significance). An interesting difference is found in the number of significant regressors: six regressors for the generalized ridge estimator and five regressors for the ridge estimator. That is, C-reactive protein was significant by the generalized ridge estimator and was not by the 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 severe intracerebral hemorrhage [38]. 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. Hence, detecting the significance by the generalized ridge estimator provides an important advantage for this data example. Besides C-reactive protein, other estimated coefficients are convincing for the generalized ridge estimator. For instance, the most significant regressor is lactate dehydrogenase (also the most significant for the ridge estimator). This variable was previously reported as an important predictor of hematoma expansion [39].

Figure 4 displays scatter plots for centered responses  $y - (\sum_{i=1}^{n} y_i/n)\mathbf{1}$  against the predictors  $X\hat{\beta}$  based on the ridge estimator and generalized ridge estimator. While both predictors captured the variability in the changes in blood volume (response variables), this was not fully explained by the predictors since some residual errors remain. Figure 5 depicts the residuals  $y - (\sum_{i=1}^{n} y_i/n)\mathbf{1} - X\hat{\beta}$  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 generalized ridge estimator was shown to work well under asymmetric errors (Section 4).

Therefore, we have demonstrated that the generalized ridge estimator in the proposed R package provides statistically and biologically sound conclusions in real data analysis.



**Figure 4.** Centered responses  $y - (\sum_{i=1}^{n} y_i/n)\mathbf{1}$  against the predictors  $X\hat{\beta}$  based on the ridge estimator and generalized ridge estimator applied to the intracerebral hemorrhage dataset. The red lines were obtained by least squared estimation.

	Ridge			Generalized Ridge		
	$\hat{\hat{m{eta}}}_j$	SE	<i>p</i> -Value	$\hat{\beta}_j$	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

**Residual plot** 

**Table 2.** 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.

## Histogram



**Figure 5.** The residuals  $y - (\sum_{i=1}^{n} y_i/n)\mathbf{1} - X\hat{\beta}$  for the generalized ridge estimator applied to a dataset on patients with intracerebral hemorrhage.

#### 6. Conclusions

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 an abundance of sparse and high-dimensional data [36,40–44] and asymmetrically distributed data [45–48], the proposed package may provide a reliable statistical tool for statisticians and data scientists.

We now compare our R package with some existing R packages for ridge regression. The "bigRR" package (https://cran.r-project.org/package=bigRR, accessed on 5 February 2024) implements the generalized ridge estimator developed by Shen et al. [19]. This package uses a different estimator for shrinkage parameters from the estimator in our package (Section 2.3). However, the bigRR package was already removed from the CRAN. The R package "genridge" is designed for the ridge estimator with a single shrinkage parameter, and hence it actually cannot calculate the generalized ridge estimator of multiple shrinkage parameters. Similarly, the R package "penalize" (https://cran.r-project.org/package=penalized, accessed on 5 February 2024) has the "optL2(.)"

function to perform ridge regression with a single shrinkage, which does not allow multiple shrinkage parameters. In the same manner, an R package "Imridge" [49] was developed to compute the ridge estimator, allowing 25 different ways to estimate the single shrinkage parameter. In summary, these existing R packages are insufficient for using the generalized ridge estimator with multiple shrinkage parameters.

While our simulation study demonstrated the high estimation accuracy of the generalized ridge estimator, the computation time was not fast. For a single simulation run, computation took about 3 s (p = 50), 6 s (p = 100), 12 s (p = 150), and 24 s (p = 200). This means that the time grows steeply as the dimensions increase. In order to analyze very-high-dimensional regressors, such as SNPs, the package needs to be improved for computational efficiency.

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. The extensions may require some efforts for formulating an estimation algorithm and efficient cross-validation criterion [50], performing extensive simulation studies, and developing a software package to fully justify the advantages and usefulness over the usual ridge estimator that is widespread via the "glmnet" package.

**Supplementary Materials:** The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/sym16020223/s1, the R code for simulations

**Author Contributions:** Conceptualization, T.E., and K.M.; software, T.E.; formal analysis, T.E. and K.M.; writing, T.E., K.M., R.U., and H.M.; funding acquisition, T.E., R.U. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by [JSPS KAKENHI] grant number [22K11948] and grant number [20H04147].

**Institutional Review Board Statement:** The data collection (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 29 March 2023).

**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 R code and dataset used in Section 5 are available from the corresponding author on reasonable request.

Acknowledgments: We thank three anonymous reviewers for their helpful comments that improved this manuscript.

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. To compute  $V(\lambda)$  and  $V(\lambda, \Delta)$ , we made an R function "GCV(X,Y,k,W)", where "X" is a design matrix, "Y" is a response vector, and "k" is actually  $\lambda$  (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 "GCV(X,Y,k,W)" is

$$V(k,W) = \frac{1}{n} \left| \left| \{I_n - A(k,W)\} y \right|^2 / \left[ \frac{1}{n} \operatorname{Tr} \{I_n - A(k,W)\} \right]^2,$$
(A1)

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

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

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

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

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

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 using "GCV(X,Y,k,W)".

## References

- Hoerl, A.E.; Kennard, R.W. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* 1970, 12, 55–67. [CrossRef]
- 2. Montgomery, D.C.; Peck, E.A.; Vining, G.G. Introduction to Linear Regression Analysis; John Wiley & Sons: Hoboken, NJ, USA, 2021.
- Arashi, M.; Roozbeh, M.; Hamzah, N.A.; Gasparini, M. Ridge regression and its applications in genetic studies. *PLoS ONE* 2021, 16, e0245376. [CrossRef] [PubMed]
- 4. Veerman, J.R.; Leday, G.G.R.; van de Wiel, M.A. Estimation of variance components, heritability and the ridge penalty in high-dimensional generalized linear models. *Commun. Stat. Simul. Comput.* **2019**, *51*, 116–134. [CrossRef]
- Friedrich, S.; Groll, A.; Ickstadt, K.; Kneib, T.; Pauly, M.; Rahnenführer, J.; Friede, T. Regularization approaches in clinical biostatistics: A review of methods and their applications. *Stat. Methods Med. Res.* 2023, 32, 425–440. [CrossRef]
- Gao, S.; Zhu, G.; Bialkowski, A.; Zhou, X. Stroke Localization Using Multiple Ridge Regression Predictors Based on Electromagnetic Signals. *Mathematics* 2023, 11, 464. [CrossRef]
- Hernandez, J.; Lobos, G.A.; Matus, I.; Del Pozo, A.; Silva, P.; Galleguillos, M. 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.* 2015, 7, 2109–2126. [CrossRef]
- 8. Golub, G.H.; Heath, M.; Wahba, G. Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics* **1979**, *21*, 215–223. [CrossRef]
- 9. Hastie, T.; Tibshirani, R.; Friedman, J. The Elements of Statistical Learning; Springer: New York, NY, USA, 2009.
- 10. Van Wieringen, W.N. Lecture notes on ridge regression. arXiv 2015, arXiv:1509.09169.
- 11. Saleh, A.M.E.; Arashi, M.; Kibria, B.G. *Theory of Ridge Regression Estimation with Applications*; John Wiley & Sons: Hoboken, NJ, USA, 2019.
- 12. Cule, E.; Vineis, P.; De Iorio, M. Significance testing in ridge regression for genetic data. BMC Bioinform. 2011, 12, 372. [CrossRef]
- 13. Whittaker, J.C.; Thompson, R.; Denham, M.C. Marker-assisted selection using ridge regression. *Genet. Res.* 2000, 75, 249–252. [CrossRef]
- 14. Cule, E.; De Iorio, M. Ridge regression in prediction problems: Automatic choice of the ridge parameter. *Genet. Epidemiol.* **2013**, 37, 704–714. [CrossRef]
- 15. Yang, S.-P.; Emura, T. A Bayesian approach with generalized ridge estimation for high-dimensional regression and testing. *Commun. Stat. Simul. Comput.* **2016**, *46*, 6083–6105. [CrossRef]

- 16. Hoerl, A.E.; Kennard, R.W. Ridge regression: Applications to nonorthogonal problems. Technometrics 1970, 12, 69–82. [CrossRef]
- 17. Allen, D.M. The relationship between variable selection and data augmentation and a method for prediction. *Technometrics* **1974**, *16*, 125–127. [CrossRef]
- 18. Loesgen, K.H. A generalization and Bayesian interpretation of ridge-type estimators with good prior means. *Stat. Pap.* **1990**, *31*, 147–154. [CrossRef]
- Shen, X.; Alam, M.; Fikse, F.; Rönnegård, L. A novel generalized ridge regression method for quantitative genetics. *Genetics* 2013, 193, 1255–1268. [CrossRef] [PubMed]
- 20. Hofheinz, N.; Frisch, M. Heteroscedastic ridge regression approaches for genome-wide prediction with a focus on computational efficiency and accurate effect estimation. *G3 Genes Genomes Genet.* **2014**, *4*, 539–546. [CrossRef] [PubMed]
- Yüzbaşı, B.; Arashi, M.; Ahmed, S.E. Shrinkage Estimation Strategies in Generalised Ridge Regression Models: Low/High-Dimension Regime. Int. Stat. Rev. 2020, 88, 229–251. [CrossRef]
- 22. Saleh, E.A.M.; Kibria, G.B.M. Performance of some new preliminary test ridge regression estimators and their properties. *Commun. Stat. Theory Methods* **1993**, *22*, 2747–2764. [CrossRef]
- Norouzirad, M.; Arashi, M. Preliminary test and Stein-type shrinkage ridge estimators in robust regression. *Stat. Pap.* 2017, 60, 1849–1882. [CrossRef]
- 24. Shih, J.-H.; Lin, T.-Y.; Jimichi, M.; Emura, T. Robust ridge M-estimators with pretest and Stein-rule shrinkage for an intercept term. *Jpn. J. Stat. Data Sci.* **2020**, *4*, 107–150. [CrossRef]
- 25. Shih, J.-H.; Konno, Y.; Chang, Y.-T.; Emura, T. A class of general pretest estimators for the univariate normal mean. *Commun. Stat. Theory Methods* **2023**, *52*, 2538–2561. [CrossRef]
- Taketomi, N.; Chang, Y.-T.; Konno, Y.; Mori, M.; Emura, T. Confidence interval for normal means in meta-analysis based on a pretest estimator. *Jpn. J. Stat. Data Sci.* 2023, 1–32. [CrossRef]
- 27. Wong, K.Y.; Chiu, S.N. An iterative approach to minimize the mean squared error in ridge regression. *Comput. Stat.* **2015**, *30*, 625–639. [CrossRef]
- Kibria, B.M.G.; Banik, S. Some ridge regression estimators and their performances. J. Mod. Appl. Stat. Methods 2016, 15, 206–238. [CrossRef]
- 29. Algamal, Z.Y. Shrinkage parameter selection via modified cross-validation approach for ridge regression model. *Commun. Stat. Simul. Comput.* **2020**, *49*, 1922–1930. [CrossRef]
- Assaf, A.G.; Tsionas, M.; Tasiopoulos, A. Diagnosing and correcting the effects of multicollinearity: Bayesian implications of ridge regression. *Tour. Manag.* 2018, 71, 1–8. [CrossRef]
- Michimae, H.; Emura, T. Bayesian ridge estimators based on copula-based joint prior distributions for regression coefficients. Comput. Stat. 2022, 37, 2741–2769. [CrossRef]
- Chen, A.-C.; Emura, T. A modified Liu-type estimator with an intercept term under mixture experiments. *Commun. Stat. Theory Methods* 2016, 46, 6645–6667. [CrossRef]
- Binder, H.; Allignol, A.; Schumacher, M.; Beyersmann, J. Boosting for high-dimensional time-to-event data with competing risks. Bioinformatics 2009, 25, 890–896. [CrossRef]
- 34. Emura, T.; Chen, Y.-H.; Chen, H.-Y. Survival prediction based on compound covariate under cox proportional hazard models. *PLoS ONE* **2012**, *7*, e47627. [CrossRef]
- 35. Emura, T.; Chen, Y.-H. Gene selection for survival data under dependent censoring: A copula-based approach. *Stat. Methods Med. Res.* **2016**, *25*, 2840–2857. [CrossRef] [PubMed]
- Emura, T.; Hsu, W.-C.; Chou, W.-C. A survival tree based on stabilized score tests for high-dimensional covariates. *J. Appl. Stat.* 2023, 50, 264–290. [CrossRef] [PubMed]
- Azzalini, A.; Capitanio, A. The Skew-Normal and Related Families; Cambridge University Press (CUP): Cambridge, UK, 2013; ISBN 9781107029279.
- Wang, D.; Wang, J.; Li, Z.; Gu, H.; Yang, K.; Zhao, X.; Wang, Y. C-reaction protein and the severity of intracerebral hemorrhage: A study from chinese stroke center alliance. *Neurol. Res.* 2021, 44, 285–290. [CrossRef] [PubMed]
- 39. Chu, H.; Huang, C.; Dong, J.; Yang, X.; Xiang, J.; Dong, Q.; Tang, Y. Lactate dehydrogenase predicts early hematoma expansion and poor outcomes in intracerebral hemorrhage patients. *Transl. Stroke Res.* **2019**, *10*, 620–629. [CrossRef] [PubMed]
- 40. Kim, S.Y.; Lee, J.W. Ensemble clustering method based on the resampling similarity measure for gene expression data. *Stat. Methods Med. Res.* **2007**, *16*, 539–564. [CrossRef] [PubMed]
- 41. Zhang, Q.; Ma, S.; Huang, Y. Promote sign consistency in the joint estimation of precision matrices. *Comput. Stat. Data Anal.* 2021, 159, 107210. [CrossRef]
- 42. Bhattacharjee, A. Big Data Analytics in Oncology with R; Taylor & Francis: London, UK, 2022.
- Bhatnagar, S.R.; Lu, T.; Lovato, A.; Olds, D.L.; Kobor, M.S.; Meaney, M.J.; O'Donnell, K.; Yang, A.Y.; Greenwood, C.M. A sparse additive model for high-dimensional interactions with an exposure variable. *Comput. Stat. Data Anal.* 2023, 179, 107624. [CrossRef]
- 44. Vishwakarma, G.K.; Thomas, A.; Bhattacharjee, A. A weight function method for selection of proteins to predict an outcome using protein expression data. *J. Comput. Appl. Math.* **2021**, *391*, 113465. [CrossRef]
- 45. Abe, T.; Shimizu, K.; Kuuluvainen, T.; Aakala, T. Sine-skewed axial distributions with an application for fallen tree data. *Environ. Ecol. Stat.* **2012**, *19*, 295–307. [CrossRef]

- 46. Huynh, U.; Pal, N.; Nguyen, M. Regression model under skew-normal error with applications in predicting groundwater arsenic level in the Mekong Delta Region. *Environ. Ecol. Stat.* **2021**, *28*, 323–353. [CrossRef]
- 47. Yoshiba, T.; Koike, T.; Kato, S. On a Measure of Tail Asymmetry for the Bivariate Skew-Normal Copula. *Symmetry* **2023**, *15*, 1410. [CrossRef]
- Jimichi, M.; Kawasaki, Y.; Miyamoto, D.; Saka, C.; Nagata, S. Statistical Modeling of Financial Data with Skew-Symmetric Error Distributions. Symmetry 2023, 15, 1772. [CrossRef]
- 49. Muhammad, I.U.; Aslam, M.; Altaf, S. Imridge: A Comprehensive R Package for Ridge Regression. R J. 2019, 10, 326. [CrossRef]
- 50. Meijer, R.J.; Goeman, J.J. Efficient approximate k-fold and leave-one-out cross-validation for ridge regression. *Biom. J.* 2013, 55, 141–155. [CrossRef]

**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.