

Package ‘g.ridge’

December 7, 2023

Type Package

Title Generalized Ridge Regression for Linear Models

Version 1.0

Description Ridge regression due to Hoerl and Kennard (1970)<[DOI:10.1080/00401706.1970.10488634](https://doi.org/10.1080/00401706.1970.10488634)> and generalized ridge regression due to Yang and Emura (2017)<[DOI:10.1080/03610918.2016.1193195](https://doi.org/10.1080/03610918.2016.1193195)> with optimized tuning parameters.
These ridge regression estimators (the HK estimator and the YE estimator) are computed by minimizing the cross-validated mean squared errors.
Both the ridge and generalized ridge estimators are applicable for high-dimensional regressors ($p > n$), where p is the number of regressors, and n is the sample size.

License GPL-2

Encoding UTF-8

RoxygenNote 7.2.3

NeedsCompilation no

Author Takeshi Emura [aut, cre],
Szu-Peng Yang [ctb]

Maintainer Takeshi Emura <takeshiemura@gmail.com>

Repository CRAN

Date/Publication 2023-12-07 11:50:05 UTC

R topics documented:

g.ridge	2
GCV	3
X.mat	4

Index	5
--------------	----------

g.ridge

g.ridge (generalized ridge regression)

Description

Generalized ridge regression with the optimal shrinkage parameter. Ridge regression (Hoerl and Kennard, 1970) and generalized ridge regression (Yang and Emura 2017) are implemented. Tuning parameters are optimized by minimizing the CGV function (by the function `CGV(.)`): See Golub et al. (1979), and Sections 2.3 and 3.3 of Yang and Emura (2017).

Usage

```
g.ridge(X, Y, method = "HK", kmax = 500)
```

Arguments

X	design matrix of explanatory variables (regressors)
Y	vector of response variables
method	"HK" or "YE" for Hoerl and Kennard (1970) or Yang and Emura (2017)
kmax	maximum possible value for the shrinkage parameter (the "lambda" parameter), where the parameter is optimized in the interval (0, kmax).

Value

lambda: optimized shrinkage parameter
delta: the optimized thresholding parameter
estimate: regression coefficients (beta)
SE: Standard Error
Z: Z-value for testing beta=0
SE: P-value for testing beta=0
Sigma: variance estimate of the error distribution (the square of the standard deviation)
delta: thresholding parameter

References

Yang SP, Emura T (2017) A Bayesian approach with generalized ridge estimation for high-dimensional regression and testing, *Commun Stat-Simul* 46(8): 6083-105.

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

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)
g.ridge(X,Y-mean(Y),method="YE",kmax=200)

```

GCV

*GCV (generalized cross-validation)***Description**

The CGV function gives the sum of cross-validated squared errors that can be used to optimize tuning parameters in ridge regression and generalized ridge regression. See Golub et al. (1979), and Sections 2.3 and 3.3 of Yang and Emura (2017) for details.

Usage

```
GCV(X, Y, k, W = diag(ncol(X)))
```

Arguments

X	matrix of explanatory variables (design matrix)
Y	vector of response variables
k	shrinkage parameter (>0); it is the "lambda" parameter
W	matrix of weights (default is the identity matrix)

Value

The value of GCV

References

Yang SP, Emura T (2017) A Bayesian approach with generalized ridge estimation for high-dimensional regression and testing, *Commun Stat-Simul* 46(8): 6083-105.

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

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)

```

X.mat

X.mat (generating a design matrix)

Description

A design matrix (X ; $nrow(X)=n$, $ncol(X)=p$) is generated by random numbers as previously used in our simulation studies (Section 5 of Yang and Emura (2017); p.6093). The design matrix has two blocks of correlated regressors (Pearson correlation=0.5): the first q regressors and the second r regressors. Other $p-q-r$ regressors are independent. If regressors are gene expressions, the correlated blocks may be regarded as "gene pathways" (Emura et al. 2012).

Usage

```
X.mat(n, p, q, r)
```

Arguments

n	the number of rows (samples)
p	the number of columns (regressors)
q	the number of correlated regressors in the first block ($1 \leq q < p$, $q+r < p$)
r	the number of correlated regressors in the second block ($1 \leq r < p$, $q+r < p$)

Value

a matrix X ($nrow(X)=n$, $ncol(X)=p$)

References

Yang SP, Emura T (2017) A Bayesian approach with generalized ridge estimation for high-dimensional regression and testing, *Commun Stat-Simul* 46(8): 6083-105

Emura T, Chen YH, Chen HY (2012) Survival prediction based on compound covariate method under Cox proportional hazard models *PLoS ONE* 7(10) doi:10.1371/journal.pone.0047627

Examples

```

X.mat(n=10, p=5, q=2, r=2)
X.mat(n=100, p=50, q=10, r=10) # Case I in Section 5 of Yang and Emura (2017)

```

Index

`g.ridge`, 2

GCV, 3

`X.mat`, 4