Package ‘rbs’

March 10, 2025

Type Package

Title Response Best-subset Selector for Multivariate Regression

Version 1.0.1

Author Jianhua Hu [aut],
Jian Huang [aut],
Xiaoqian Liu [aut],
Xu Liu [aut,cre]

Maintainer Xu Liu <liu.xu@sufe.edu.cn>

Description Provide a procedure to select response variables and estimate regression coefficients simultaneously. It also provides the screening procedure based on the distance correlation, the solutions to the quadratic 0-1 programming problems by transferring to k-flipping optimization problems and to continuous quadratic programming problems, and the multi-test procedure including Benjamini-Hochberg and Bonferroni correction.

License GPL (>= 2)

Depends R (>= 3.2.0)

Imports quadprog

LazyData true

NeedsCompilation yes

Repository CRAN

URL https://github.com/xliusufe/rbs

Encoding UTF-8

Archs x64

Contents

`rbs-package` *Response Best-subset Selector for Multivariate Regression*

Description

Provide a procedure to select response variables and estimate regression coefficients simultaneously. It also provides the screening procedure based on the distance correlation, the solutions to the quadratic 0-1 programming problems by transferring to k-flipping optimization problems and to continuous quadratic programming problems, and the multi-test procedure including Benjamini-Hochberg and Bonferroni correction.

Details

Package:	rbs
Type:	Package
Version:	1.0.1
Date:	2020-08-8
License:	GPL (>= 2)

References

Benjamini, Y. and Hochberg, Y. (1995). Controlling the False Discovery Rate A Practical and Powerful Approach to Multiple testing. Journal of the Royal Statistical Society: Series B (Methodological). 57(1), 289-300.

Chen, W. and L. Zhang (2010). Global Optimality Conditions for Quadratic 0-1 Optimization Problems. Journal of Global Optimization 46(2), 191-206.

Chen, W. (2015). Optimality Conditions for the Minimization of Quadratic 0-1 Problems. SIAM Journal on Optimization, 25(3), 1717-1731.

Fan, J. and Lv, J. (2008). Sure independence screening for ultrahigh dimensional feature space. Journal of the Royal Statistical Society Series B: Statistical Methodology, 70(5), 849-911.

Hu, J., Huang, J., Liu, X. and Qiu F. (2020). Response Best-subset Selector for Multivariate Regression. Manuscript.

Li, R., W. Zhong, and L. Zhu (2012). Feature Screening Via Distance Correlation Learning. Journal of the American Statistical Association, 107 (499), 1129-1139.

Szekely, G.J. and Rizzo, M.L. (2009). Brownian Distance Covariance, Annals of Applied Statistics, 3(4), 1236-1265.

Szekely, G.J. and Rizzo, M.L. (2009). Rejoinder: Brownian Distance Covariance, Annals of Applied Statistics, 3(4), 1303-1308.

Szekely, G.J., Rizzo, M.L., and Bakirov, N.K. (2007). Measuring and Testing Dependence by Correlation of Distances, Annals of Statistics, 35(6), 2769-2794.

Zhu, L. P., Li, L., Li, R. and Zhu, L. X. (2011). Model-free feature screening for ultrahigh-dimensional data. Journal of the American Statistical Association, 106(496), 1464-1475.

`dcorr` *Distance correlation of two multivariates.*

Description

Distance correlation and covariance of two multivariates y and x.

Usage

```
dcorr(y,x)
```

Arguments

`y`	A $n \times q$ numeric matrix.
`x`	A $n \times p$ numeric matrix.

Value

`dcor`	The distance correlation, which is an 4-vactor with the dcorr of both y and x, the dcov of y, the dcov of dcorr x, and the dcov of both y and x. dcov denotes the sample distance covariance, and dcorr denotes the sample distance correlation.

References

Szekely, G.J. and Rizzo, M.L. (2009). Brownian Distance Covariance, Annals of Applied Statistics, 3(4), 1236-1265.

Szekely, G.J. and Rizzo, M.L. (2009). Rejoinder: Brownian Distance Covariance, Annals of Applied Statistics, 3(4), 1303-1308.

Szekely, G.J., Rizzo, M.L., and Bakirov, N.K. (2007). Measuring and Testing Dependence by Correlation of Distances, Annals of Statistics, 35(6), 2769-2794.

Examples

```
n   <- 200
p   <- 5
q   <- 10
q0  <- 5

beta <- matrix(runif(p*q0),p,q0)
eps <- matrix(rnorm(n*q),n,q)

x <- matrix(rnorm(n*p),n,p)
y <- cbind(x%*%beta, matrix(0,n,q-q0)) + eps

dcor <- dcorr(y,x)
```

flip *Optimality conditions for the minimization of quadratic 0-1 problems*

Description

Flip procedure for optimality conditions for the minimization of quadratic 0-1 problems, where one-flip, two-flip and hybrid for both are considered. The hybrid flip applies one-flip and two-flip sequentially.

Usage

```
flip(A,b=NULL,x0=NULL,nflip=1)
```

Arguments

`A`	A p-symmetric matrix.
`b`	A p-vector. Default is zero.
`x0`	The initial value. Default is zero.
`nflip`	An integer $1, 2, 3$ with one-flip if `nflip=1`, two-flip if `nflip=2`, and hybrid if `nflip=3`. Default is `nflip=1` corresponding to one-flip.

Value

`xhat`	The local minimizer.
`obj`	the local minimum.

References

Chen, W. (2015). Optimality Conditions for the Minimization of Quadratic 0-1 Problems. SIAM Journal on Optimization, 25(3), 1717-1731.

Examples

```
data(Qd)
Q <- as.matrix(Qd$Q)
fit <- flip(Q,nflip=1)
fit
```

pcorr *Pearson correlation coefficient*

Description

Pearson correlation coefficient vector between an n-dimensional vector Y and an n-by-p matrix X.

Usage

```
pcorr(Y,X)
```

Arguments

`Y`	An n-dimensional numeric vector.
`X`	An $n \times p$ numeric matrix.

Value

`pcor`	A p-dimensional vector, where the i-th element corresponds to the Pearson correlation coefficient between Y and the i-th column of X.

Examples

```
n   <- 100
p   <- 200

beta <- c(rep(0.5, 20), rep(0, p - 20))
eps <- rnorm(n, mean = 0, sd = 1)

X <- matrix(rnorm(n*p),n,p)
Y <- X %*% beta + eps

pcor <- pcorr(Y,X)
```

pval *P-values for F-test of the separate responses*

Description

P-values for F-test of the separate responses for the multivariate linear regression models.

Usage

```
pval(x,y,criteria=NULL,alpha=0.05,gamma=1.15,family="Fdist",isbic=FALSE)
```

Arguments

x	A $n \times p$ Numeric design matrix for the model.
y	A $n \times q$ Response matrix.
criteria	A criteria to select important variables by a significant level. No corrections if `criteria=NULL`, RBS procedure if `criteria="RBS"`, Benjamini-Hochberg procedure if `criteria="BH"`, and Bonferroni correction if `criteria="Bonf"`.
alpha	A prespecified level.
gamma	A positive separating parameter γ if RBS procedure is used. Default is `gamma=1.15`.
family	A string representing one of the built-in families, by which P-values are calculated. F-test is used if `family="Fdist"` with the first degrees of freedom p and the second degrees of freedom $n-p$, and χ^2-test is used if `family="Chi2"` with degrees of freedom p. Default is `family="Fdist"` (F-test).
isbic	A logical flag. The `BIC` criteria is used (`TRUE`) or not (`default = FALSE`).

Value

Tn	Values of test statistics.
Sigma2	Estimator of the marginal response variance.
pvals	P-values.
pvfdr	The P-values corresponding to selected variables.
signifc	The indices corresponding to selected variables.

References

Benjamini, Y. and Hochberg, Y. (1995). Controlling the False Discovery Rate A Practical and Powerful Approach to Multiple testing. Journal of the Royal Statistical Society: Series B (Methodological). 57(1), 289-300.

Hu, J., Huang, J., Liu, X. and Liu, X. (2020). Response Best-subset Selector for Multivariate Regression. Manuscript.

Examples

```
n   <- 200
p   <- 5
q   <- 10
q0  <- 5

beta <- matrix(runif(p*q0),p,q0)
eps <- matrix(rnorm(n*q),n,q)

x <- matrix(rnorm(n*p),n,p)
y <- cbind(x%*%beta, matrix(0,n,q-q0)) + eps

fit <- pval(x,y)

fit$Tn
fit$pvals
fit$pvfdr
fit$signifc
```

`pvalrp` *P-values for F-test of the separate responses based on the random projection approach*

Description

P-values for F-test of the separate responses for the multivariate linear regression models based on the random projection approach.

Usage

```
pvalrp(x,y,rho=0.4,Pk=NULL,criteria=NULL,alpha=0.05,gamma=1.15,
                                                   family="Fdist",isbic=FALSE)
```

Arguments

`x`	A $n \times p$ Numeric design matrix for the model.
`y`	A $n \times q$ Response matrix.
`rho`	The given ratio $\rho = \lim \frac{k}{n}$, a scale in the unit inteval. Default is `rho = 0.4`.
`Pk`	The preset random matrix. Default is `Pk = NULL`, in the case of which each entry is drawn independently from the standard normal distribution.
`criteria`	A criteria to select important variables by a significant level. No corrections if `criteria=NULL`, RBS procedure if `criteria="RBS"`, Benjamini-Hochberg procedure if `criteria="BH"`, and Bonferroni correction if `criteria="Bonf"`.
`alpha`	A prespecified level.
`gamma`	A positive separating parameter γ if RBS procedure is used. Default is `gamma=1.15`.
`family`	A string representing one of the built-in families, by which P-values are calculated. F-test is used if `family="Fdist"` with the first degrees of freedom p and the second degrees of freedom $n-p$, and χ^2-test is used if `family="Chi2"` with degrees of freedom p. Default is `family="Fdist"` (F-test).
`isbic`	A logical flag. The `BIC` criteria is used (`TRUE`) or not (`default = FALSE`).

Value

`Tn`	Values of test statistics.
`Sigma2`	Estimator of the marginal response variance.
`pvals`	P-values.
`pvfdr`	The P-values corresponding to selected variables.
`signifc`	The indices corresponding to selected variables.
`rho`	The input ratio $\rho = \lim \frac{k}{n}$.

References

Benjamini, Y. and Hochberg, Y. (1995). Controlling the False Discovery Rate A Practical and Powerful Approach to Multiple testing. Journal of the Royal Statistical Society: Series B (Methodological). 57(1), 289-300.

Hu, J., Huang, J., Liu, X. and Liu, X. (2020). Response Best-subset Selector for Multivariate Regression. Manuscript.

Examples

```
n   <- 200
p   <- 5
q   <- 10
q0  <- 5

beta <- matrix(runif(p*q0),p,q0)
eps <- matrix(rnorm(n*q),n,q)

x <- matrix(rnorm(n*p),n,p)
y <- cbind(x%*%beta, matrix(0,n,q-q0)) + eps

fit <- pvalrp(x,y)

fit$Tn
fit$pvals
fit$pvfdr
fit$signifc
```

---

rbs *RBS without covariance of responses*

Description

Select the response variables and estimate regression coefficients simultaneously for multivariate linear regression without covariance of responses.

Usage

```
rbs(x,y,gamma=1.5, lambda=NULL,criteria=2,tau=1)
```

Arguments

x	A $n \times p$ Numeric design matrix for the model.
y	A $n \times q$ Response matrix.
gamma	A positive separating parameter γ. Default is gamma=1.5.
lambda	A user-specified sequence of λ values. By default, a sequence of values of length nlambda is computed, equally spaced on the scale.
criteria	The criteria to be applied to select parameters. Either AIC if criteria=1, BIC (the default) if criteria=2, or GCV if criteria=3. There is no selection if criteria=0, in which case lambda should be a number.
tau	A constant to adjust AIC creteria. Default is tau=1.

Value

delta	The estimation of the δ.
theta	The estimation of the θ.
rss	Residual sum of squares (RSS) without the selection of tuning parameters.
deltapath	The estimation path of the δ with the selection of tuning parameters.

bic	The AIC or BIC or GCV with the selection of tuning parameters.
selected	The index of λ corresponding to lambda_opt with the selection of tuning parameters.

References

Hu, J., Huang, J., Liu, X. and Liu, X (2020). Response Best-subset Selector for Multivariate Regression. Manuscript.

Examples

```
n   <- 200
p   <- 5
q   <- 10
q0  <- 5

beta <- matrix(runif(p*q0),p,q0)
eps <- matrix(rnorm(n*q),n,q)

x <- matrix(rnorm(n*p),n,p)
y <- cbind(x%*%beta, matrix(0,n,q-q0)) + eps

fit <- rbs(x,y,lambda=0.4)
fit$delta

lambda <- seq(0.01, 2, length = 50)
fit <- rbs(x,y,lambda=lambda)
fit$delta
fit$selected
```

---

rbsrp *RBS without covariance of responses based on the random projection approach*

Description

Select the response variables and estimate regression coefficients simultaneously for multivariate linear regression without covariance of responses based on the random projection approach.

Usage

```
rbsrp(x,y,rho=0.4,Pk=NULL,gamma=1.5, lambda=NULL,criteria=2,tau=1)
```

Arguments

x	A $n \times p$ Numeric design matrix for the model.
y	A $n \times q$ Response matrix.
rho	The given ratio $\rho = \lim \frac{k}{n}$, a scale in the unit inteval. Default is rho = 0.4.
Pk	The preset random matrix. Default is Pk = NULL, in the case of which each entry is drawn independently from the standard normal distribution.
gamma	A positive separating parameter γ. Default is gamma=1.5.

lambda	A user-specified sequence of λ values. By default, a sequence of values of length nlambda is computed, equally spaced on the scale.
criteria	The criteria to be applied to select parameters. Either AIC if criteria=1, BIC (the default) if criteria=2, or GCV if criteria=3. There is no selection if criteria=0, in which case lambda should be a number.
tau	A constant to adjust AIC creteria. Default is tau=1.

Value

delta	The estimation of the δ.
theta	The estimation of the θ.
rss	Residual sum of squares (RSS) without the selection of tuning parameters.
deltapath	The estimation path of the δ with the selection of tuning parameters.
bic	The AIC or BIC or GCV with the selection of tuning parameters.
selected	The index of λ corresponding to lambda_opt with the selection of tuning parameters.
rho	The input ratio $\rho = \lim \frac{k}{n}$.

References

Hu, J., Huang, J., Liu, X. and Liu, X (2020). Response Best-subset Selector for Multivariate Regression. Manuscript.

Examples

```
n   <- 200
p   <- 5
q   <- 10
q0  <- 5

beta <- matrix(runif(p*q0),p,q0)
eps <- matrix(rnorm(n*q),n,q)

x <- matrix(rnorm(n*p),n,p)
y <- cbind(x%*%beta, matrix(0,n,q-q0)) + eps

fit <- rbsrp(x,y,lambda=0.4)
fit$delta

lambda <- seq(0.01, 2, length = 50)
fit <- rbsrp(x,y,lambda=lambda)
fit$delta
fit$selected
```

`rbs_qp` *RBSS with considering covariance of responses based on continuous quadratic programming.*

Description

Select the response variables and estimate regression coefficients simultaneously for multivariate linear regression with considering covariance of responses, in which the quadratic 0-1 programming problems are transferred to continuous quadratic programming problems.

Usage

```
rbs_qp(x,y,V=NULL,gamma=1.5,lambda=NULL,criteria=2,tau=1)
```

Arguments

`x`	A $n \times p$ numeric design matrix for the model.
`y`	A $n \times q$ response matrix.
`V`	A user-specified $q \times q$ precision matrix. A estimator is provided if `V=NULL`. Default is `V=NULL`.
`gamma`	A positive separating parameter γ. Default is `gamma=1.5`.
`lambda`	A user-specified sequence of λ values. By default, a sequence of values of length `nlambda` is computed, equally spaced on the scale.
`criteria`	The criteria to be applied to select parameters. Either `AIC` if `criteria=1`, `BIC` (the default) if `criteria=2`, or `GCV` if `criteria=3`. There is no selection if `criteria=0`, in which case `lambda` should be a number.
`tau`	A constant to adjust `AIC` creteria. Default is `tau=1`.

Value

`delta`	The estimation of the δ.
`theta`	The estimation of the θ.
`rss`	Residual sum of squares (RSS) without the selection of tuning parameters.
`deltapath`	The estimation of the δ with the selection of tuning parameters.
`bic`	The `AIC` or `BIC` or `GCV` with the selection of tuning parameters.
`selected`	The index of λ corresponding to `lambda_opt` with the selection of tuning parameters.

References

Chen, W. and L. Zhang (2010). Global Optimality Conditions for Quadratic 0-1 Optimization Problems. Journal of Global Optimization 46(2), 191-206.

Hu, J., Huang, J., Liu, X. and Liu, X (2020). Response Best-subset Selector for Multivariate Regression. Manuscript.

Examples

```
n   <- 200
p   <- 5
q   <- 10
q0  <- 5

Sigma <- matrix(0,q,q)
for(i in 1:q) for(j in 1:q) Sigma[i,j]=0.5^(abs(i-j))
A <- chol(Sigma)
V <- solve(Sigma)

beta <- matrix(runif(p*q0),p,q0)
eps <- matrix(rnorm(n*q),n,q)

x <- matrix(rnorm(n*p),n,p)
y <- cbind(x%*%beta, matrix(0,n,q-q0)) + eps%*%A

fit <- rbs_sig(x,y,lambda=0.4)
fit$delta

fit <- rbs_sig(x,y,V,lambda=0.4)
fit$delta

lambda <- seq(0.01, 2, length = 50)
fit <- rbs_sig(x,y,lambda=lambda)
fit$delta
fit$selected

fit <- rbs_sig(x,y,V,lambda=lambda)
fit$delta
fit$selected
```

---

rbs_sig — *RBS with considering covariance of responses based on k-flipping optimization problems.*

Description

Select the response varibales and estimate regression coefficients simultaneously for multivariate linear regression with considering covariance of responses, in which the quadratic 0-1 programming problems are transferred to k-flipping optimization problems.

Usage

```
rbs_sig(x,y,V=NULL,gamma=1.5, lambda=NULL,criteria=2,nflip=1,tau=1)
```

Arguments

x	A $n \times p$ numeric design matrix for the model.
y	A $n \times q$ response matrix.
V	A user-specified $q \times q$ precision matrix. A estimator is provided if V=NULL. Default is V=NULL.

gamma	A positive separating parameter γ. Default is gamma=1.5.
lambda	A user-specified sequence of λ values. By default, a sequence of values of length nlambda is computed, equally spaced on the scale.
criteria	The criteria to be applied to select parameters. Either AIC if criteria=1, BIC (the default) if criteria=2, or GCV if criteria=3. There is no selection if criteria=0, in which case lambda should be a number.
nflip	An integer $1, 2, 3$ with one-flip if nflip=1, two-flip if nflip=2, and hybrid if nflip=3. Default is nflip=1 corresponding to one-flip.
tau	A constant to adjust AIC creteria. Default is tau=1.

Value

delta	The estimation of the δ.
theta	The estimation of the θ.
rss	Residual sum of squares (RSS) without the selection of tuning parameters.
deltapath	The estimation of the δ with the selection of tuning parameters.
bic	The AIC or BIC or GCV with the selection of tuning parameters.
selected	The index of λ corresponding to lambda_opt with the selection of tuning parameters.

References

Chen, W. (2015). Optimality Conditions for the Minimization of Quadratic 0-1 Problems. SIAM Journal on Optimization, 25(3), 1717-1731.

Hu, J., Huang, J., Liu, X. and Liu, X (2020). Response Best-subset Selector for Multivariate Regression. Manuscript.

Examples

```
n   <- 200
p   <- 5
q   <- 10
q0  <- 5

Sigma <- matrix(0,q,q)
for(i in 1:q) for(j in 1:q) Sigma[i,j]=0.5^(abs(i-j))
A <- chol(Sigma)
V <- solve(Sigma)

beta <- matrix(runif(p*q0),p,q0)
eps <- matrix(rnorm(n*q),n,q)

x <- matrix(rnorm(n*p),n,p)
y <- cbind(x%*%beta, matrix(0,n,q-q0)) + eps%*%A

fit <- rbs_sig(x,y,lambda=0.4)
fit$delta

fit <- rbs_sig(x,y,V,lambda=0.4)
fit$delta

lambda <- seq(0.01, 2, length = 50)
```

```
fit <- rbs_sig(x,y,lambda=lambda)
fit$delta
fit$selected

fit <- rbs_sig(x,y,V,lambda=lambda)
fit$delta
fit$selected
```

---

sirs *Sure independence ranking and screening (SIRS)*

Description

A model-free screening method, proposed by Zhu et al. (2011), which combines a soft cutoff value and a hard cutoff.

Usage

```
sirs(Y, X, standardize_X = TRUE, N = NULL, d = 10, ntop = 10)
```

Arguments

Y	An n-dimensional numeric vector.
X	An $n \times p$ numeric matrix.
standardize_X	Logical flag for X standardization, prior to performing screening. Default is standardize_X=TRUE.
N	An integer specified by the user for the hard threshold rule, representing the number of top N ranked variables to retain after sorting. The default value is NULL, which corresponds to $\lfloor n/\log(n) \rfloor$ (floor function).
d	An integer specified by the user for the soft threshold rule, representing the dimension of the auxiliary variables. The default value is 10.
ntop	An integer, which is integer that the indices of the top ntop most correlated variables will be output.

Value

indn	The indices of the top ntop most correlated variables. If ntop exceeds the number of variables selected by the combined soft and hard threshold rules, the indices of all variables selected by the threshold rules are returned. Otherwise, the indices of the top ntop variables are returned.

References

Zhu, L. P., Li, L., Li, R. and Zhu, L. X. (2011). Model-free feature screening for ultrahigh-dimensional data. Journal of the American Statistical Association, 106(496), 1464-1475.

Examples

```
n   <- 100
p   <- 200

beta <- c(rep(0.5, 20), rep(0, p - 20))
eps <- rnorm(n, mean = 0, sd = 1)

X <- matrix(rnorm(n*p),n,p)
Y <- X %*% beta + eps

fit <- sirs(Y,X)
```

---

sis *Sure independence screening (SIS)*

Description

Screening procedure based on Pearson correlation coefficient of Y and X.

Usage

```
sis(Y, X, ntop=10)
```

Arguments

`Y`	An n-dimensional numeric vector.
`X`	An $n \times p$ numeric matrix.
`ntop`	An integer specifying the number of top correlated variables (`ntop`) whose indices will be output.

Value

`pcor`	The whole Pearson correlation.
`indn`	The indices of the top `ntop` most correlated variables.

References

Fan, J. and Lv, J. (2008). Sure independence screening for ultrahigh dimensional feature space. Journal of the Royal Statistical Society Series B: Statistical Methodology, 70(5), 849-911.

Examples

```
n   <- 100
p   <- 200

beta <- c(rep(0.5, 20), rep(0, p - 20))
eps <- rnorm(n, mean = 0, sd = 1)

X <- matrix(rnorm(n*p),n,p)
Y <- X %*% beta + eps

fit <- sis(Y,X)
```

`sisdc` *Screening procedure based on the distance correlation.*

Description

Screening procedure based on the distance correlation of two multivariates y and x.

Usage

```
sisdc(y, x, d=1, ntop=10)
```

Arguments

`y`	A $n \times q$ numeric matrix.
`x`	A $n \times p$ numeric matrix.
`d`	An integer. Screening variable y if `d=1`, and Screening variable x if `d=2`.
`ntop`	An integer, which is integer that the indices of the top `ntop` most correlated variables will be output.

Value

`dcor`	The whole distance correlation.
`indn`	The indices of the top `ntop` most correlated variables.

References

Li, R., W. Zhong, and L. Zhu (2012). Feature Screening Via Distance Correlation Learning. Journal of the American Statistical Association, 107 (499), 1129-1139.

Szekely, G.J. and Rizzo, M.L. (2009). Brownian Distance Covariance, Annals of Applied Statistics, 3(4), 1236-1265.

Szekely, G.J. and Rizzo, M.L. (2009). Rejoinder: Brownian Distance Covariance, Annals of Applied Statistics, 3(4), 1303-1308.

Szekely, G.J., Rizzo, M.L., and Bakirov, N.K. (2007). Measuring and Testing Dependence by Correlation of Distances, Annals of Statistics, 35(6), 2769-2794.

Examples

```
n   <- 200
p   <- 5
q   <- 10
q0  <- 5

beta <- matrix(runif(p*q0),p,q0)
eps <- matrix(rnorm(n*q),n,q)

x <- matrix(rnorm(n*p),n,p)
y <- cbind(x%*%beta, matrix(0,n,q-q0)) + eps

fit <- sisdc(y,x)
fit
```

Index