Package ‘COR’

October 12, 2022

Title The COR for Optimal Subset Selection in Distributed Estimation

Date 2021-11-30

Version 0.0.1

Description An algorithm of optimal subset selection, related to Covariance matrices, Observation matrices and Response vectors (COR) to select the optimal subsets in distributed estimation. The philosophy of the package is described in Guo G. (2020) <doi:10.1080/02331888.2020.1823979>.

License MIT + file LICENSE

Encoding UTF-8

RoxygenNote 7.1.1

Imports stats

NeedsCompilation no

Author Guangbao Guo [aut, cre] (<https://orcid.org/0000-0002-4115-6218>),
Haoyue Song [aut],
Lixing Zhu [aut]

Maintainer Guangbao Guo <ggb11111111@163.com>

Depends R (>= 3.5.0)

Repository CRAN

Date/Publication 2021-12-06 17:40:02 UTC

R topics documented:

beta_AD *Caculate the estimators of beta on the A-opt and D-opt*

Description

Caculate the estimators of beta on the A-opt and D-opt

Usage

```
beta_AD(K = K, nk = nk, alpha = alpha, X = X, y = y)
```

Arguments

K	is the number of subsets
nk	is the length of subsets
alpha	is the significance level
X	is the observation matrix
y	is the response vector

Value

betaA, betaD

Examples

```
p=6;n=1000;K=2;nk=200;alpha=0.05;sigma=1
e=rnorm(n,0,sigma); beta=c(sort(c(runif(p,0,1))));
data=c(rnorm(n*p,5,10));X=matrix(data, ncol=p);
y=X%*%beta+e;
beta_AD(K=K,nk=nk,alpha=alpha,X=X,y=y)
```

beta_cor *Caculate the estimator of beta on the COR*

Description

Caculate the estimator of beta on the COR

Usage

```
beta_cor(K = K, nk = nk, alpha = alpha, X = X, y = y)
```

Arguments

`K`	is the number of subsets
`nk`	is the length of subsets
`alpha`	is the significance level
`X`	is the observation matrix
`y`	is the response vector

Value

betaC

Examples

```
p=6;n=1000;K=2;nk=200;alpha=0.05;sigma=1
e=rnorm(n,0,sigma); beta=c(sort(c(runif(p,0,1))));
data=c(rnorm(n*p,5,10));X=matrix(data, ncol=p);
y=X%*%beta+e;
beta_cor(K=K,nk=nk,alpha=alpha,X=X,y=y)
```

`communities` *The communities and crime data set*

Description

A data set about the communities and crime

Usage

```
data("communities")
```

Format

A data frame with 1994 observations on the following 128 variables.

- `V1` a numeric vector
- `V2` a numeric vector
- `V3` a numeric vector
- `V4` a character vector
- `V5` a numeric vector
- `V6` a numeric vector
- `V7` a numeric vector
- `V8` a numeric vector
- `V9` a numeric vector
- `V10` a numeric vector

V11 a numeric vector
V12 a numeric vector
V13 a numeric vector
V14 a numeric vector
V15 a numeric vector
V16 a numeric vector
V17 a numeric vector
V18 a numeric vector
V19 a numeric vector
V20 a numeric vector
V21 a numeric vector
V22 a numeric vector
V23 a numeric vector
V24 a numeric vector
V25 a numeric vector
V26 a numeric vector
V27 a numeric vector
V28 a numeric vector
V29 a numeric vector
V30 a numeric vector
V31 a numeric vector
V32 a numeric vector
V33 a numeric vector
V34 a numeric vector
V35 a numeric vector
V36 a numeric vector
V37 a numeric vector
V38 a numeric vector
V39 a numeric vector
V40 a numeric vector
V41 a numeric vector
V42 a numeric vector
V43 a numeric vector
V44 a numeric vector
V45 a numeric vector
V46 a numeric vector
V47 a numeric vector

V48 a numeric vector
V49 a numeric vector
V50 a numeric vector
V51 a numeric vector
V52 a numeric vector
V53 a numeric vector
V54 a numeric vector
V55 a numeric vector
V56 a numeric vector
V57 a numeric vector
V58 a numeric vector
V59 a numeric vector
V60 a numeric vector
V61 a numeric vector
V62 a numeric vector
V63 a numeric vector
V64 a numeric vector
V65 a numeric vector
V66 a numeric vector
V67 a numeric vector
V68 a numeric vector
V69 a numeric vector
V70 a numeric vector
V71 a numeric vector
V72 a numeric vector
V73 a numeric vector
V74 a numeric vector
V75 a numeric vector
V76 a numeric vector
V77 a numeric vector
V78 a numeric vector
V79 a numeric vector
V80 a numeric vector
V81 a numeric vector
V82 a numeric vector
V83 a numeric vector
V84 a numeric vector

V85 a numeric vector
V86 a numeric vector
V87 a numeric vector
V88 a numeric vector
V89 a numeric vector
V90 a numeric vector
V91 a numeric vector
V92 a numeric vector
V93 a numeric vector
V94 a numeric vector
V95 a numeric vector
V96 a numeric vector
V97 a numeric vector
V98 a numeric vector
V99 a numeric vector
V100 a numeric vector
V101 a numeric vector
V102 a numeric vector
V103 a numeric vector
V104 a numeric vector
V105 a numeric vector
V106 a numeric vector
V107 a numeric vector
V108 a numeric vector
V109 a numeric vector
V110 a numeric vector
V111 a numeric vector
V112 a numeric vector
V113 a numeric vector
V114 a numeric vector
V115 a numeric vector
V116 a numeric vector
V117 a numeric vector
V118 a numeric vector
V119 a numeric vector
V120 a numeric vector
V121 a numeric vector

V122 a numeric vector

V123 a numeric vector

V124 a numeric vector

V125 a numeric vector

V126 a numeric vector

V127 a numeric vector

V128 a numeric vector

Source

UCI repository

References

Redmond, M. A. and A. Baveja: A Data-Driven Software Tool for Enabling Cooperative Information Sharing Among Police Departments. European Journal of Operational Research 141 (2002) 660-678.

Examples

```
data(communities)
## maybe str(communities) ; plot(communities) ...
```

COR *Caculate the optimal subset lengths on the COR*

Description

Caculate the optimal subset lengths on the COR

Usage

```
COR(K = K, nk = nk, alpha = alpha, X = X, y = y)
```

Arguments

K	is the number of subsets
nk	is the length of subsets
alpha	is the significance level
X	is the observation matrix
y	is the response vector

Value

seqL, seqN,lWMN

Examples

```
p=6;n=1000;K=2;nk=200;alpha=0.05;sigma=1
e=rnorm(n,0,sigma); beta=c(sort(c(runif(p,0,1))));
data=c(rnorm(n*p,5,10));X=matrix(data, ncol=p);
y=X%*%beta+e;
COR(K=K,nk=nk,alpha=alpha,X=X,y=y)
```

`ethylene_CO` *The chemical sensor data set*

Description

A data set about chemical sensor

Usage

```
data("ethylene_CO")
```

Format

A data frame with 4001 observations on the following 19 variables.

`V1` a character vector

`V2` a character vector

`V3` a character vector

`V4` a character vector

`V5` a character vector

`V6` a character vector

`V7` a character vector

`V8` a character vector

`V9` a character vector

`V10` a character vector

`V11` a character vector

`V12` a character vector

`V13` a character vector

`V14` a character vector

`V15` a character vector

`V16` a character vector

`V17` a character vector

`V18` a character vector

`V19` a character vector

Details

We selected the first 4001 rows on the original data set about 1048576 observations on 19 variables.

Source

UCI Repository

References

Wang, H. Y., Zhu, R., and Ma, P. (2018). Optimal subsampling for large sample logistic regression. Journal of the American Statistical Association, 113(522), 829-844.

Examples

```
data(ethylene_CO)
## maybe str(ethylene_CO) ; plot(ethylene_CO) ...
```

MSEcom *Caculate the MSE values of the COR criterion in simulation*

Description

Caculate the MSE values of the COR criterion in simulation

Usage

```
MSEcom(K = K, nk = nk, alpha = alpha, X = X, y = y)
```

Arguments

K	is the number of subsets
nk	is the length of subsets
alpha	is the significance level
X	is the observation matrix
y	is the response vector

Value

MSEx,MSEA,MSEc,MSEm,MSEa

Examples

```
p=6;n=1000;K=2;nk=500;alpha=0.05;sigma=1
e=rnorm(n,0,sigma); beta=c(sort(c(runif(p,0,1))));
data=c(rnorm(n*p,5,10));X=matrix(data, ncol=p);
y=X%*%beta+e;
MSEcom(K=K,nk=nk,alpha=alpha,X=X,y=y)
```

`MSEver` *Caculate the MSE values of the COR criterion for redundant data in simulation*

Description

Caculate the MSE values of the COR criterion for redundant data in simulation

Usage

```
MSEver(K = K, nk = nk, alpha = alpha, X = X, y = y)
```

Arguments

`K`	is the number of subsets
`nk`	is the length of subsets
`alpha`	is the significance level
`X`	is the observation matrix
`y`	is the response vector

Value

minE,Mcor,Mx,MA

Examples

```
p=6;n=1000;K=2;nk=200;alpha=0.05;sigma=1
e=rnorm(n,0,sigma); beta=c(sort(c(runif(p,0,1))));
data=c(rnorm(n*p,5,10));X=matrix(data, ncol=p);
y=X%*%beta+e;
MSEver(K=K,nk=nk,alpha=alpha,X=X,y=y)
```

Index