Package ‘PSIM’

December 5, 2023

Type Package

Title Penalized Subgroup Identification Model

Version 0.1.0

Author Fuxin Wang, Dong Liu, Yong He, Mingjuan Zhang

Maintainer Fuxin Wang <wangfuxin2001@163.com>

Description In various domains, many datasets exhibit both high variable dependency and group structures, necessitating their simultaneous estimation. To analyze such data, this software package primarily offers two core estimation methods. The first is the Center Augmented Factor Adjustment Regression Model (CAFARM) we developed. Utilizing center-augmented penalty functions and factor structures, CAFARM eliminates data dependencies while identifying subgroups within datasets. We provide CAFARM methods optimized using Coordinate Descent and the Difference of Convex-Alternating Direction Method of Multipliers (DC-ADMM), applicable for both L1 and L2 distance cases. The second method is the Factor-Adjusted Pairwise Fusion Penalty (FA-PFP) approach, which integrates factor augmentation into the pairwise fusion penalty framework to estimate subgroup structures and address high inherent data dependency. Additionally, functions based on the Bayesian Information Criterion (BIC) are included for selecting tuning parameters in CAFARM and FA-PFP. Furthermore, the package includes the Standard CAR (S-CAR) method without factor augmentation, enabling comparative analysis with our approaches.

License GPL

Encoding UTF-8

LazyData true

RoxygenNote 7.2.3

Imports MASS,
glmnet,
prclust,
stats,
Ckmeans.1d.dp

R topics documented:

`BIC_CAFARM` *Selecting Tuning Parameter for CAFARM via corresponding BIC*

Description

This function is to select tuning parameters simultaneously for CAFARM via minimizing the BIC.

Usage

```
BIC_CAFARM(Y, Fhat, Uhat, epsilon)
```

Arguments

`Y`	The response vector.
`Fhat`	The estimated common factors matrix.
`Uhat`	The estimated idiosyncratic factors matrix.
`epsilon`	User-supplied stopping error.

Examples

```
n <- 50
p <- 50
r <- 3
alpha <- sample(c(-3,3),n,replace=TRUE,prob=c(1/2,1/2))
beta <- c(rep(1,2),rep(0,48))
B <- matrix((rnorm(p*r,1,1)),p,r)
F_1 <- matrix((rnorm(n*r,0,1)),n,r)
U <- matrix(rnorm(p*n,0,0.1),n,p)
X <- F_1%*%t(B)+U
Y <- alpha + X%*%beta + rnorm(n,0,0.5)
BIC_CAFARM(Y,F_1,U,0.3)
```

`BIC_PFP` *Selecting Tuning Parameter for FA-PFP via corresponding BIC*

Description

This function is to select tuning parameters simultaneously for FA-PFP via minimizing the BIC.

Usage

```
BIC_PFP(Y, Fhat, Uhat, epsilon)
```

Arguments

`Y`	The response vector.
`Fhat`	The estimated common factors matrix.
`Uhat`	The estimated idiosyncratic factors matrix.
`epsilon`	User-supplied stopping error.

Examples

```
n <- 50
p <- 50
r <- 3
alpha <- sample(c(-3,3),n,replace=TRUE,prob=c(1/2,1/2))
beta <- c(rep(1,2),rep(0,48))
B <- matrix((rnorm(p*r,1,1)),p,r)
F_1 <- matrix((rnorm(n*r,0,1)),n,r)
U <- matrix(rnorm(p*n,0,0.1),n,p)
X <- F_1%*%t(B)+U
Y <- alpha + X%*%beta + rnorm(n,0,0.5)
BIC_PFP(Y,F_1,U,0.3)
```

CAFARM — *CAFARM for Subgroup Identification and Variable Selection Based on Coordinate Descent Algorithm under l2 Distance*

Description

This function employs CAFARM in conjunction with the Coordinate Descent Algorithm to effectively identify subgroup structures and perform variable selection.

Usage

```
CAFARM(Y, X_aug, r, lam_CAR, lam_lasso, tng_init, K, epsilon)
```

Arguments

Y	Response vector.
X_aug	Augmented matrix of the estimations of common factors and idiosyncratic factors.
r	User supplied number of common factors.
lam_CAR	The tuning parameter for CAR regularization.
lam_lasso	The tuning parameter for LASSO.
tng_init	The tuning parameter of initial value.
K	User-supplied group number.
epsilon	User-supplied stopping error.

Examples

```
n <- 50
p <- 50
r <- 3
alpha <- sample(c(-3,3),n,replace=TRUE,prob=c(1/2,1/2))
beta <- c(rep(1,2),rep(0,48))
B <- matrix((rnorm(p*r,1,1)),p,r)
F_1 <- matrix((rnorm(n*r,0,1)),n,r)
U <- matrix(rnorm(p*n,0,0.1),n,p)
X <- F_1%*%t(B)+U
Y <- alpha + X%*%beta + rnorm(n,0,0.5)
CAFARM(Y,cbind(F_1,U),3,0.01,0.05,0.1,2,0.3)
```

`DCADMM_iter_l1` *CAFARM-Based Subgroup Identification and Variable Selection Using DC-ADMM under the L1 Distance*

Description

This function employs CAFARM and the corresponding DC-ADMM to identify subgroup structures and conduct variable selection under the L1 Distance.

Usage

```
DCADMM_iter_l1(
  Y,
  F_hat,
  U_hat,
  r_1,
  r_2,
  r_3,
  lambda_1,
  lambda_2,
  n,
  K,
  p,
  init_para,
  epsilon_1,
  epsilon_2
)
```

Arguments

`Y`	Response vector.
`F_hat`	The estimated factor matrix.
`U_hat`	The estimated idiosyncratic factors matrix.
`r_1`	The Lagrangian augmentation parameter for constraints of intercepts.
`r_2`	The Lagrangian augmentation parameter for constraints of group centers.
`r_3`	The Lagrangian augmentation parameter for constraints of coeffcient.
`lambda_1`	The tuning parameter for CAR.
`lambda_2`	The tuning parameter for LASSO.
`n`	sample size.
`K`	estimated group number.
`p`	Dimension of covariates.
`init_para`	parameter for initialization
`epsilon_1`	User-supplied stopping error for outer loop.
`epsilon_2`	parameter for initialization for inner loop.

Examples

```
n <- 50
p <- 50
r <- 3
alpha <- sample(c(-3,3),n,replace=TRUE,prob=c(1/2,1/2))
beta <- c(rep(1,2),rep(0,48))
B <- matrix((rnorm(p*r,1,1)),p,r)
F_1 <- matrix((rnorm(n*r,0,1)),n,r)
U <- matrix(rnorm(p*n,0,0.1),n,p)
X <- F_1%*%t(B)+U
Y <- alpha + X%*%beta + rnorm(n,0,0.5)
DCADMM_iter_l1(Y,F_1,U,0.5,0.5,0.5,0.01,0.05,50,2,50,0.1,1,0.3)
```

`DCADMM_iter_l2` *CAFARM-Based Subgroup Identification and Variable Selection Using DC-ADMM under the L2 Distance*

Description

This function employs CAFARM and the corresponding DC-ADMM to identify subgroup structures and conduct variable selection under the L2 Distance.

Usage

```
DCADMM_iter_l2(
  Y,
  F_hat,
  U_hat,
  r_1,
  r_2,
  r_3,
  lambda_1,
  lambda_2,
  n,
  K,
  p,
  init_para,
  epsilon_1,
  epsilon_2
)
```

Arguments

`Y`	The response vector.
`F_hat`	The estimated factor matrix.
`U_hat`	The estimated idiosyncratic factors matrix.
`r_1`	The Lagrangian augmentation parameter for constraints of intercepts.
`r_2`	The Lagrangian augmentation parameter for constraints of group centers.
`r_3`	The Lagrangian augmentation parameter for constraints of coeffcient.
`lambda_1`	The tuning parameter for CAR

lambda_2	The tuning parameter for LASSO
n	sample size
K	estimated group number
p	dimension
init_para	parameter for initialization
epsilon_1	User-supplied stopping error for outer loop.
epsilon_2	parameter for initialization for inner loop.

Examples

```
n <- 50
p <- 50
r <- 3
alpha <- sample(c(-3,3),n,replace=TRUE,prob=c(1/2,1/2))
beta <- c(rep(1,2),rep(0,48))
B <- matrix((rnorm(p*r,1,1)),p,r)
F_1 <- matrix((rnorm(n*r,0,1)),n,r)
U <- matrix(rnorm(p*n,0,0.1),n,p)
X <- F_1%*%t(B)+U
Y <- alpha + X%*%beta + rnorm(n,0,0.5)
DCADMM_iter_l2(Y,F_1,U,0.5,0.5,0.5,0.01,0.05,50,2,50,0.1,1,0.3)
```

FA_PFP *Factor Adjusted-Pairwise Fusion Penalty (FA-PFP) Method for Subgroup Identification and Variable Selection*

Description

This function utilizes the FA-PFP approach implemented via the ADMM algorithm to identify subgroup structures and conduct variable selection.

Usage

```
FA_PFP(Y, Fhat, Uhat, vartheta, lam, gam, tng_init, lam_lasso, epsilon)
```

Arguments

Y	Response vector.
Fhat	The estimated common factors matrix.
Uhat	The estimated idiosyncratic factors matrix.
vartheta	The Lagrangian augmentation parameter for intercepts.
lam	The tuning parameter for PFP.
gam	User-supplied parameter for ADMM.
tng_init	User-supplied parameter for initial value.
lam_lasso	The tuning parameter for LASSO.
epsilon	User-supplied stopping error.

Examples

```
n <- 50
p <- 50
r <- 3
alpha <- sample(c(-3,3),n,replace=TRUE,prob=c(1/2,1/2))
beta <- c(rep(1,2),rep(0,48))
B <- matrix((rnorm(p*r,1,1)),p,r)
F_1 <- matrix((rnorm(n*r,0,1)),n,r)
U <- matrix(rnorm(p*n,0,0.1),n,p)
X <- F_1%*%t(B)+U
Y <- alpha + X%*%beta + rnorm(n,0,0.5)
FA_PFP(Y,F_1,U,1,0.67,3,0.1,0.05,0.3)
```

SCAR *Standard Center Augmented Regularization (S-CAR) Method for Subgroup Identification and Variable Selection*

Description

This function employs the S-CAR approach combined with the Coordinate Descent Algorithm to identify subgroup structures and execute variable selection.

Usage

```
SCAR(Y, X, lam_CAR, lam_lasso, tng_init, K, epsilon)
```

Arguments

Y	The response vector.
X	The covariate matrix.
lam_CAR	The tuning paramter for CAR.
lam_lasso	The tuning paramter for lasso.
tng_init	User-supplied parameter for initial value.
K	estimated group number.
epsilon	User-supplied stopping error.

Examples

```
n <- 50
p <- 50
r <- 3
alpha <- sample(c(-3,3),n,replace=TRUE,prob=c(1/2,1/2))
beta <- c(rep(1,2),rep(0,48))
B <- matrix((rnorm(p*r,1,1)),p,r)
F_1 <- matrix((rnorm(n*r,0,1)),n,r)
U <- matrix(rnorm(p*n,0,0.1),n,p)
X <- F_1%*%t(B)+U
Y <- alpha + X%*%beta + rnorm(n,0,0.5)
SCAR(Y,X,0.01,0.05,0.1,2,0.3)
```

Index