Package 'IALS'

February 16, 2024

Type Package

Title Iterative Alternating Least Square Estimation for Large-Dimensional Matrix Factor Model

Version 0.1.3

Author Yong He [aut],
Ran Zhao [aut, cre],
Wen-Xin Zhou [aut]

Maintainer Ran Zhao <Zhaoran@mail.sdu.edu.cn>

Description The matrix factor model has drawn growing attention for its advantage in achieving two-directional dimension reduction simultaneously for matrix-structured observations. In contrast to the Principal Component Analysis (PCA)-based methods, we propose a simple Iterative Alternating Least Squares (IALS) algorithm for matrix factor model, see the details in He et al. (2023) <arXiv:2301.00360>.

License GPL-2 | GPL-3

Encoding UTF-8

Depends R (>= 4.0)

Imports RSpectra, pracma, HDMFA

NeedsCompilation no

Repository CRAN

Date/Publication 2024-02-16 11:50:07 UTC

R topics documented:

`Distance` *The distance between the spaces spanned by the column of two matrices.*

Description

Calculate the distance between spaces spanned by the column of two matrices. The distance is between 0 and 1. If the two spaces are the same, the distance is 0. if the two spaces are orthogonal, the distance is 1.

Usage

```
Distance(Z1, Z2)
```

Arguments

`Z1`	Input a matrix with $p \times q_1$.
`Z2`	Input another matrix with $p \times q_2$

Details

Define

$$\mathcal{D}(\boldsymbol{Q}_1, \boldsymbol{Q}_2) = \left(1 - \frac{1}{\max\left(q_1, q_2\right)} \mathrm{Tr}\left(\boldsymbol{Q}_1 \boldsymbol{Q}_1^\top \boldsymbol{Q}_2 \boldsymbol{Q}_2^\top\right)\right)^{1/2}.$$

By the definition of $\mathcal{D}(\boldsymbol{Q}_1, \boldsymbol{Q}_2)$, we can easily see that $0 \leq \mathcal{D}(\boldsymbol{Q}_1, \boldsymbol{Q}_2) \leq 1$, which measures the distance between the column spaces spanned by two orthogonal matrices $\boldsymbol{Q}_1$ and $\boldsymbol{Q}_2$, i.e., $\mathrm{span}(\boldsymbol{Q}_1)$ and $\mathrm{span}(\boldsymbol{Q}_2)$. In particular, $\mathrm{span}(\boldsymbol{Q}_1)$ and $\mathrm{span}(\boldsymbol{Q}_2)$ are the same when $\mathcal{D}(\boldsymbol{Q}_1, \boldsymbol{Q}_2) = 0$, while $\mathrm{span}(\boldsymbol{Q}_1)$ and $\mathrm{span}(\boldsymbol{Q}_2)$ are orthogonal when $\mathcal{D}(\boldsymbol{Q}_1, \boldsymbol{Q}_2) = 1$. The Gram-Schmidt orthogonalization can be used to make $\boldsymbol{Q}_1$ and $\boldsymbol{Q}_2$ column-orthogonal matrices.

Value

Output a number between 0 and 1.

Author(s)

Yong He, Ran Zhao, Wen-Xin Zhou.

References

He, Y., Zhao, R., & Zhou, W. X. (2023). Iterative Least Squares Algorithm for Large-dimensional Matrix Factor Model by Random Projection. <arXiv:2301.00360>.

Examples

```
set.seed(1111)
A=matrix(rnorm(10),5,2)
B=matrix(rnorm(15),5,3)
Distance(A,B)
```

IALS *Iterative Alternating Least Square Estimation for Large-dimensional Matrix Factor Model*

Description

This function is designed to fit the matrix factor model using the Iterative Least Squares (IALS) method, rather than Principal Component Analysis (PCA)-based methods. In detail, in the first step, we propose to estimate the latent factor matrices by projecting the matrix observations with two deterministic weight matrices, chosen to diversify away the idiosyncratic components. In the second step, we update the row/column loading matrices by minimizing the squared loss function under the identifiability condition. The estimators of the loading matrices are then treated as the new weight matrices, and the algorithm iteratively performs these two steps until a convergence criterion is reached.

Usage

```
IALS(X, W1 = NULL, W2 = NULL, m1, m2, max_iter = 100, ep = 1e-06)
```

Arguments

`X`	Input an array with $T \times p_1 \times p_2$, where T is the sample size, p_1 is the the row dimension of each matrix observation and p_2 is the the column dimension of each matrix observation.
`W1`	The initial value for the row factor loading matrix. The default is NULL, with an initial estimate chosen from α-PCA if not provided.
`W2`	The initial value for the column factor loading matrix. The default is NULL, with an initial estimate chosen from α-PCA if not provided.
`m1`	A positive integer indicating the row factor number.
`m2`	A positive integer indicating the column factor number.
`max_iter`	The maximum number of iterations for the algorithm, default is 100.
`ep`	The stopping criterion in the iteration algorithm, default is $10^{-6} \times Tp_1p_2$.

Details

Assume we have two weight matrices $\boldsymbol{W}_i$ of dimension $p_i \times m_i$ for $i = 1, 2$, as substitutes for $\boldsymbol{R}$ and $\boldsymbol{C}$ respectively. Then it is straightforward to estimate $\boldsymbol{F}_t$ simply by

$$\hat{\boldsymbol{F}}_t = \frac{1}{p_1 p_2} \boldsymbol{W}_1^\top \boldsymbol{X}_t \boldsymbol{W}_2.$$

Given $\hat{\boldsymbol{F}}_t$ and $\boldsymbol{W}_1$, we can derive that

$$\hat{\boldsymbol{R}} = \sqrt{p_1}\left(\sum_{t=1}^{T} \boldsymbol{X}_t \boldsymbol{W}_2 \hat{\boldsymbol{F}}_t^\top\right)\left[\left(\sum_{t=1}^{T} \hat{\boldsymbol{F}}_t \boldsymbol{W}_2^\top \boldsymbol{X}_t^\top\right)\left(\sum_{t=1}^{T} \boldsymbol{X}_t \boldsymbol{W}_2 \hat{\boldsymbol{F}}_t^\top\right)\right]^{-1/2}.$$

Similarly, we get the following estimator of the column factor loading matrix

$$\hat{\boldsymbol{C}} = \sqrt{p_2}\left(\sum_{t=1}^{T} \boldsymbol{X}_t^\top \hat{\boldsymbol{R}} \hat{\boldsymbol{F}}_t\right)\left[\left(\sum_{t=1}^{T} \hat{\boldsymbol{F}}_t^\top \hat{\boldsymbol{R}}^\top \boldsymbol{X}_t\right)\left(\sum_{t=1}^{T} \boldsymbol{X}_t^\top \hat{\boldsymbol{R}} \hat{\boldsymbol{F}}_t\right)\right]^{-1/2}.$$

Afterwards, we sequentially update $\boldsymbol{F}$, $\boldsymbol{R}$ and $\boldsymbol{C}$. In simulation, the iterative procedure is terminated either when a pre-specified maximum iteration number (maxiter $= 100$) is reached or when

$$\sum_{t=1}^{T} \|\hat{\boldsymbol{S}}_t^{(s+1)} - \hat{\boldsymbol{S}}_t^{(s)}\|_F \le \epsilon \cdot T p_1 p_2,$$

where $\hat{\boldsymbol{S}}_t^{(s)}$ is the common component estimated at the s-th step, ϵ is a small constant (10^{-6}) given in advance.

Value

The return value is a list. In this list, it contains the following:

`R`	The estimated row loading matrix of dimension $p_1 \times m_1$, satisfying $\boldsymbol{R}^\top \boldsymbol{R} = p_1 \boldsymbol{I}_{m_1}$.
`C`	The estimated column loading matrix of dimension $p_2 \times m_2$, satisfying $\boldsymbol{C}^\top \boldsymbol{C} = p_2 \boldsymbol{I}_{m_2}$.
`F`	The estimated factor matrix of dimension $T \times m_1 \times m_2$.
`iter`	The number of iterations when the stopping criterion is met.

Author(s)

Yong He, Ran Zhao, Wen-Xin Zhou.

References

He, Y., Zhao, R., & Zhou, W. X. (2023). Iterative Alternating Least Square Estimation for Large-dimensional Matrix Factor Model. <arXiv:2301.00360>.

Examples

```
set.seed(11111)
T=20;p1=20;p2=20
k1=3;k2=3

R=matrix(runif(p1*k1,min=-1,max=1),p1,k1)
C=matrix(runif(p2*k2,min=-1,max=1),p2,k2)
```

```
X=E=array(0,c(T,p1,p2))
F=array(0,c(T,k1,k2))

for(t in 1:T){
  F[t,,]=matrix(rnorm(k1*k2),k1,k2)
  E[t,,]=matrix(rnorm(p1*p2),p1,p2)
}
for(t in 1:T){
X[t,,]=R%*%F[t,,]%*%t(C)+E[t,,]
}

#Estimating the matrix factor model using the default initial values
fit1 = IALS(X, W1 = NULL, W2 = NULL,k1, k2, max_iter = 100, ep = 1e-06)
Distance(fit1$R,R);Distance(fit1$C,C)

#Estimating the matrix factor model using one-step iteration
fit2 = IALS(X, W1 = NULL , W2 = NULL, k1, k2, max_iter = 1, ep = 1e-06)
Distance(fit2$R,R);Distance(fit2$C,C)
```

`KIALS` *Estimating the Pair of Factor Numbers via Eigenvalue Ratios Corresponding to IALS*

Description

The function is to estimate the pair of factor numbers via eigenvalue ratios corresponding to IALS method.

Usage

```
KIALS(X, W1 = NULL, W2 = NULL, kmax, max_iter = 100, ep = 1e-06)
```

Arguments

`X`	Input an array with $T \times p_1 \times p_2$, where T is the sample size, p_1 is the the row dimension of each matrix observation and p_2 is the the column dimension of each matrix observation.
`W1`	The initial value for the row factor loading matrix. The default is NULL, with an initial estimate chosen from α-PCA if not provided.
`W2`	The initial value for the column factor loading matrix. The default is NULL, with an initial estimate chosen from α-PCA if not provided.
`kmax`	The user-supplied maximum factor numbers. Here it means the upper bound of the number of row factors and column factors.
`max_iter`	The maximum number of iterations for the algorithm, default is 100. See in `IALS`.
`ep`	The stopping criterion in the iteration algorithm, default is $10^{-6} \times Tp_1p_2$. See in `IALS`.

Details

In detail, we first set $k_{\max}$ is a predetermined upper bound for k_1, k_2 and thus by IALS method, we can obtain the estimate of $\boldsymbol{F}_t$, denote as $\hat{\boldsymbol{F}}_t$, which is of dimension $k_{\max} \times k_{\max}$. Then the dimensions k_1 and k_2 are further determined as follows:

$$\hat{k}_1 = \arg\max_{j \le k_{\max}} \frac{\lambda_j \left(\frac{1}{T} \sum_{t=1}^{T} \hat{\boldsymbol{F}}_t \hat{\boldsymbol{F}}_t^\top \right)}{\lambda_{j+1} \left(\frac{1}{T} \sum_{t=1}^{T} \hat{\boldsymbol{F}}_t \hat{\boldsymbol{F}}_t^\top \right)},$$

$$\hat{k}_2 = \arg\max_{j \le k_{\max}} \frac{\lambda_j \left(\frac{1}{T} \sum_{t=1}^{T} \hat{\boldsymbol{F}}_t^\top \hat{\boldsymbol{F}}_t \right)}{\lambda_{j+1} \left(\frac{1}{T} \sum_{t=1}^{T} \hat{\boldsymbol{F}}_t^\top \hat{\boldsymbol{F}}_t \right)}.$$

Value

k_1	The estimated row factor number.
k_2	The estimated column factor number.

Author(s)

Yong He, Ran Zhao, Wen-Xin Zhou.

References

He, Y., Zhao, R., & Zhou, W. X. (2023). Iterative Alternating Least Square Estimation for Large-dimensional Matrix Factor Model. <arXiv:2301.00360>.

Examples

```
set.seed(11111)
T=20;p1=20;p2=20
k1=3;k2=3

R=matrix(runif(p1*k1,min=-1,max=1),p1,k1)
C=matrix(runif(p2*k2,min=-1,max=1),p2,k2)

X=E=array(0,c(T,p1,p2))
F=array(0,c(T,k1,k2))

for(t in 1:T){
  F[t,,]=matrix(rnorm(k1*k2),k1,k2)
  E[t,,]=matrix(rnorm(p1*p2),p1,p2)
}

for(t in 1:T){
X[t,,]=R%*%F[t,,]%*%t(C)+E[t,,]
}

kmax=8
K=KIALS(X, W1 = NULL, W2 = NULL, kmax, max_iter = 100, ep = 1e-06);K
```

Index