Package ‘CISE’

April 5, 2018

Type Package

Title Common and Individual Structure Explained for Multiple Graphs

Version 0.1.0

Description Specific dimension reduction methods for replicated graphs (multiple undirected graphs repeatedly measured on a common set of nodes). The package contains efficient procedures for estimating a shared baseline propensity matrix and graph-specific low rank matrices. The algorithm uses block coordinate descent algorithm to solve the model, which alternatively performs L2-penalized logistic regression and multiple partial eigenvalue decompositions, as described in the paper Wang et al. (2017) <arXiv:1707.06360>.

Depends R (>= 3.3.0)

Imports far, gdata, glmnet (>= 2.0-13), MASS, Matrix (>= 1.2-12), rARPACK (>= 0.11-0)


License GPL-2

LazyData true

RoxygenNote 6.0.1

Suggests knitr, rmarkdown, testthat

VignetteBuilder knitr

NeedsCompilation no

Author Lu Wang [aut, cre]

Maintainer Lu Wang <wangronglu22@gmail.com>

Repository CRAN

Date/Publication 2018-04-05 15:19:31 UTC

R topics documented:

A ................................................................. 2
MGRAF1 ...................................................... 2
**Description**

Data from Human Connectome Project (HCP). The brain is segmented into 68 regions. Each subject’s binary network consists of a collection of inter-connections among those brain regions, with ones indicate a connection between a pair of regions and zeros indicate no connection.

**Usage**

data(A)

**Format**

An object of class array of dimension 68 x 68 x 212.

**Source**

Github

**References**

Wang et al. (2017) (Arxiv)

**Description**

MGRAF1 returns the estimated common structure $Z$ and subject-specific low rank components $Q_i$ and $\Lambda_i$ for multiple undirected graphs.

**Usage**

MGRAF1(A, K, tol, maxit)
Arguments

- **A**: Binary array with size VxVxn storing the VxV symmetric adjacency matrices of n graphs.
- **K**: An integer that specifies the latent dimension of the graphs.
- **tol**: A numeric scalar that specifies the convergence threshold of CISE algorithm. CISE iteration continues until the absolute percent change in joint log-likelihood is smaller than this value. Default is tol = 0.01.
- **maxit**: An integer that specifies the maximum number of iterations. Default is maxit = 5.

Details

The subject-specific deviation $D_i$ is decomposed into

$$D_i = Q_i \Lambda_i Q_i^T,$$

where each $Q_i$ is a VxK orthonormal matrix and each $\Lambda_i$ is a KxK diagonal matrix.

Value

A list is returned containing the ingredients below from M-GRAF1 model corresponding to the largest log-likelihood over iterations.

- **Z**: A numeric vector containing the lower triangular entries in the estimated matrix $Z$.
- **Lambda**: Kxn matrix where each column stores the diagonal entries in $\Lambda_i$.
- **Q**: VxKxn array containing the estimated VxK orthonormal matrix $Q_i$, i=1,...,n.
- **D_LT**: Lxn matrix where each column stores the lower triangular entries in $D_i = Q_i \Lambda_i Q_i^T$; $L=V(V-1)/2$.
- **LL_max**: Maximum log-likelihood across iterations.
- **LL**: Joint log-likelihood at each iteration.

Examples

```r
data(A)
n = dim(A)[3]
subs = sample.int(n=n, size=30)
A_sub = A[, , subs]
res = MGRAF1(A=A_sub, K=3, tol=0.01, maxit=5)
```
MGRAF2 returns the estimated common structure $Z$ and $\Lambda$ that are shared by all the subjects as well as the subject-specific low rank matrix $Q_i$ for multiple undirected graphs.

**Usage**

```
MGRAF2(A, K, tol, maxit)
```

**Arguments**

- **A** Binary array with size $V \times V \times n$ storing the $V \times V$ symmetric adjacency matrices of $n$ graphs.
- **K** An integer that specifies the latent dimension of the graphs.
- **tol** A numeric scalar that specifies the convergence threshold of CISE algorithm. CISE iteration continues until the absolute percent change in joint log-likelihood is smaller than this value. Default is $tol = 0.01$.
- **maxit** An integer that specifies the maximum number of iterations. Default is $maxit = 5$.

**Details**

The subject-specific deviation $D_i$ is decomposed into

$$D_i = Q_i \ast \Lambda \ast Q_i^\top,$$

where each $Q_i$ is a $V \times K$ orthonormal matrix and $\Lambda$ is a $K \times K$ diagonal matrix.

**Value**

A list is returned containing the ingredients below from M-GRAF2 model corresponding to the largest log-likelihood over iterations.

- **Z** A numeric vector containing the lower triangular entries in the estimated matrix $Z$.
- **Lambda** $K \times 1$ vector storing the diagonal entries in $\Lambda$.
- **Q** $V \times K \times n$ array containing the estimated $V \times K$ orthonormal matrix $Q_i, i=1,\ldots,n$.
- **D_LT** $L \times n$ matrix where each column stores the lower triangular entries in $D_i = Q_i \ast \Lambda \ast Q_i^\top; L=V(V-1)/2$.
- **LL_max** Maximum log-likelihood across iterations.
- **LL** Joint log-likelihood at each iteration.
Examples

```r
data(A)
n = dim(A)[3]
subs = sample.int(n=n, size=30)
A_sub = A[, , subs]
res = MGRAF2(A=A_sub, K=3, tol=0.01, maxit=5)
```

---

**Description**

MGRAF3 returns the estimated common structure $Z$ and $Q$ that are shared by all the subjects as well as the subject-specific low rank matrix $\Lambda_i$ for multiple undirected graphs.

**Usage**

```r
MGRAF3(A, K, tol, maxit)
```

**Arguments**

- **A** Binary array with size $V \times V \times n$ storing the $V \times V$ symmetric adjacency matrices of $n$ graphs.
- **K** An integer that specifies the latent dimension of the graphs.
- **tol** A numeric scalar that specifies the convergence threshold of CISE algorithm. CISE iteration continues until the absolute percent change in joint log-likelihood is smaller than this value. Default is $\text{tol} = 0.01$.
- **maxit** An integer that specifies the maximum number of iterations. Default is $\text{maxit} = 5$.

**Details**

The subject-specific deviation $D_i$ is decomposed into

$$D_i = Q \ast \Lambda_i \ast Q^\top,$$

where $Q$ is a $V \times K$ orthonormal matrix and each $\Lambda_i$ is a $K \times K$ diagonal matrix.

**Value**

A list is returned containing the ingredients below from M-GRAF3 model corresponding to the largest log-likelihood over iterations.

- **Z** A numeric vector containing the lower triangular entries in the estimated matrix $Z$.
- **Lambda** $K \times n$ matrix where each column stores the diagonal entries in $\Lambda_i$. 
Q    VxK orthonormal matrix
LL_max  Maximum log-likelihood across iterations.
LL  Joint log-likelihood at each iteration.

Examples
```r
data(A)
n = dim(A)[3]
subs = sample.int(n=n, size=30)
A_sub = A[, , subs]
res = MGRAF3(A=A_sub, K=3, tol=0.01, maxit=5)
```

<table>
<thead>
<tr>
<th>VSPL0T</th>
<th>Visuospatial processing ability for 212 subjects</th>
</tr>
</thead>
</table>

Description
Data from Human Connectome Project (HCP). Visuospatial processing is commonly assessed using the Variable Short Penn Line Orientation Test (VSPLOT), where two line segments are presented on the screen and participants are asked to rotate a movable line so that it is parallel to the fixed line. The data contains an indicator of high/low visuospatial processing score for 212 subject. The order of the subjects is in the same order as in the array of adjacency matrices data(A).

Usage
data(VSPLOT)

Format
An object of class character of length 212.

Source
Github

References
Wang et al. (2017) (Arxiv)
Index

*Topic datasets
   A, 2
   VSPLLOT, 6

A, 2
MGRAF1, 2
MGRAF2, 4
MGRAF3, 5
VSPLLOT, 6