Package 'DLPCA'

October 12, 2022

Type Package

Title The Distributed Local PCA Algorithm

Version 0.0.5

Date 2022-08-07

Maintainer Guangbao Guo <ggb11111111@163.com>

Description Algorithm to handle with optimal subset selection for distributed local principal component analysis. The philosophy of the package is described in Guo G. (2020) <doi:10.1080/02331888.2020.1823979>.

License MIT + file LICENSE

NeedsCompilation no

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

Depends R (>= 3.5.0)

RoxygenNote 7.2.0

Suggests testthat (>= 3.0.0)

Config/testthat/edition 3

Repository CRAN

Date/Publication 2022-08-07 02:20:02 UTC

R topics documented:

`Application` *Application*

Description

Application data set

Usage

```
data("Application")
```

Format

The format is: int [1:48, 1:15] 6 9 7 5 6 7 9 9 9 4 ... - attr(*, "dimnames")=List of 2 ..$: NULL ..$: chr [1:15] "FL" "APP" "AA" "LA" ...

Details

It is the scoring of 15 indicators on 48 interviewees

Examples

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

`DLPCA` *Distributed local PCA*

Description

Calculate the estimator on the DLPCA method

Usage

```
DLPCA(X = X, n = n, p = p, m = m, K = K, L = L)
```

Arguments

`X`	is the original data matrix
`n`	is the sample size
`p`	is the number of variables
`m`	is the number of eigenvalues
`K`	is the number of nodes
`L`	is the number of subgroups

Value

time	is the time cost
V	is the right singular matrix
Vm	is the m-right singular matrix
Smean	is the mean covariance matrix
MMSER	is the mean MSE values of the robust covariance matrix sub-estimators
MMSES	is the mean MSE values of the covariance matrix sub-estimators
MMSEX	is the mean MSE values of the sub-estimators of the matrix X
MSER	is the min MSE values of the robust covariance matrix sub-estimators
MSES	is the min MSE values of the covariance matrix sub-estimators
MSEX	is the min MSE values of the sub-estimators of the matrix X
wMSER	is the location of the min MSE values of the robust covariance matrix sub-estimators
wMSES	is the location of the min MSE values of the covariance matrix sub-estimators
wMSEX	is the location of the min MSE values of the sub-estimators of the matrix X
sigm	is the estimator of the covariance matrix of the matrix X

Examples

```
data(Application)
X=Application
n=nrow(Application);p=ncol(Application)
m=5;L=4;K=4
DLPCA_result=DLPCA(X=X,n=n,p=p,m=m,K=K,L=L)
```

---

gt2011 *Gas-Turbine CO and NOx Emission Data*

Description

Gas-Turbine CO and NOx Emission Data in 2011

Usage

```
data("gt2011")
```

Format

A data frame with 7411 observations on the following 11 variables.

`AT` a numeric vector

`AP` a numeric vector

`AH` a numeric vector

`AFDP` a numeric vector

`GTEP` a numeric vector

`TIT` a numeric vector

`TAT` a numeric vector

`TEY` a numeric vector

`CDP` a numeric vector

`CO` a numeric vector

`NOX` a numeric vector

Details

The dataset contains 36733 instances of 11 sensor measures aggregated over one hour, from a gas turbine located in Turkey for the purpose of studying flue gas emissions, namely CO and NOx.

Source

Heysem Kaya, Department of Information and Computing Sciences, Utrecht University, 3584 CC, Utrecht, The Netherlands

Examples

```
data(gt2011)
```

---

gt2012 *Gas-Turbine CO and NOx Emission Data*

Description

Gas-Turbine CO and NOx Emission Data in 2012

Usage

```
data("gt2012")
```

Format

A data frame with 7628 observations on the following 11 variables.

`AT` a numeric vector

`AP` a numeric vector

`AH` a numeric vector

`AFDP` a numeric vector

`GTEP` a numeric vector

`TIT` a numeric vector

`TAT` a numeric vector

`TEY` a numeric vector

`CDP` a numeric vector

`CO` a numeric vector

`NOX` a numeric vector

Details

The dataset contains 36733 instances of 11 sensor measures aggregated over one hour, from a gas turbine located in Turkey for the purpose of studying flue gas emissions, namely CO and NOx.

Source

Heysem Kaya, Department of Information and Computing Sciences, Utrecht University, 3584 CC, Utrecht, The Netherlands

Examples

```
data(gt2012)
```

---

gt2013 *Gas-Turbine CO and NOx Emission Data*

Description

Gas-Turbine CO and NOx Emission Data in 2013

Usage

```
data("gt2013")
```

Format

A data frame with 7152 observations on the following 11 variables.

`AT` a numeric vector

`AP` a numeric vector

`AH` a numeric vector

`AFDP` a numeric vector

`GTEP` a numeric vector

`TIT` a numeric vector

`TAT` a numeric vector

`TEY` a numeric vector

`CDP` a numeric vector

`CO` a numeric vector

`NOX` a numeric vector

Details

The dataset contains 36733 instances of 11 sensor measures aggregated over one hour, from a gas turbine located in Turkey for the purpose of studying flue gas emissions, namely CO and NOx.

Source

Heysem Kaya, Department of Information and Computing Sciences, Utrecht University, 3584 CC, Utrecht, The Netherlands

Examples

```
data(gt2013)
```

---

gt2014 *Gas-Turbine CO and NOx Emission Data*

Description

Gas-Turbine CO and NOx Emission Data in 2014

Usage

```
data("gt2014")
```

Format

A data frame with 7158 observations on the following 11 variables.

`AT` a numeric vector

`AP` a numeric vector

`AH` a numeric vector

`AFDP` a numeric vector

`GTEP` a numeric vector

`TIT` a numeric vector

`TAT` a numeric vector

`TEY` a numeric vector

`CDP` a numeric vector

`CO` a numeric vector

`NOX` a numeric vector

Details

The dataset contains 36733 instances of 11 sensor measures aggregated over one hour, from a gas turbine located in Turkey for the purpose of studying flue gas emissions, namely CO and NOx.

Source

Heysem Kaya, Department of Information and Computing Sciences, Utrecht University, 3584 CC, Utrecht, The Netherlands

Examples

```
data(gt2014)
```

---

gt2015 *Gas-Turbine CO and NOx Emission Data*

Description

Gas-Turbine CO and NOx Emission Data in 2015

Usage

```
data("gt2015")
```

Format

A data frame with 7384 observations on the following 11 variables.

`AT` a numeric vector

`AP` a numeric vector

`AH` a numeric vector

`AFDP` a numeric vector

`GTEP` a numeric vector

`TIT` a numeric vector

`TAT` a numeric vector

`TEY` a numeric vector

`CDP` a numeric vector

`CO` a numeric vector

`NOX` a numeric vector

Details

The dataset contains 36733 instances of 11 sensor measures aggregated over one hour, from a gas turbine located in Turkey for the purpose of studying flue gas emissions, namely CO and NOx.

Source

Heysem Kaya, Department of Information and Computing Sciences, Utrecht University, 3584 CC, Utrecht, The Netherlands

Examples

```
data(gt2015)
```

Iris *Iris*

Description

Iris data set

Usage

```
data("Iris")
```

Format

A data frame with 150 observations on the following 5 variables.

`Sepal.length` a numeric vector

`Sepal.width` a numeric vector

`Petal.length` a numeric vector

`Petal.width` a numeric vector

`Species` a character vector

Details

It contains 150 samples with 5 variables

Source

Gaspar peninsula in Canada

Examples

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

`MSEpca` *MSE on PCA*

Description

Caculate the MSE value on PCA

Usage

```
MSEpca(V = V, X = X, n = n, p = p, m = m, K = K, L = L)
```

Arguments

`V`	is the right singular matrix
`X`	is the orignal data set
`n`	is the sample size
`p`	is the number of variables
`m`	is the number of eigenvalues
`K`	is the number of nodes
`L`	is the number of subgroups

Value

`MSEpca`	the MSE value on PCA

Examples

```
data(Application)
X=Application
n=nrow(Application);p=ncol(Application)
m=5;L=4;K=4
DLPCA_result=DLPCA(X=X,n=n,p=p,m=m,K=K,L=L)
V=DLPCA_result$V
MSEpca_result=MSEpca(V=V,X=X,n=n,p=p,m=m,K=K,L=L)
MSE_PCA=MSEpca_result$MSEpca
```

Index