

# Package ‘ISR’

April 22, 2022

**Title** The Iterated Score Regression-Based Estimation Algorithm

**Date** 2022-04-22

**Version** 2022.4.22

**Description** Algorithm to handle with PCA-based missing data, where ISR is for PCA-based missing data with high correlation and DISR is for distributed PCA-based missing data. The philosophy of the package is described in Guo G. (2020) <[doi:10.1080/02331888.2020.1823979](https://doi.org/10.1080/02331888.2020.1823979)>.

**License** MIT + file LICENSE

**Encoding** UTF-8

**LazyData** true

**RoxygenNote** 7.1.1

**Imports** MASS, stats

**Suggests** testthat (>= 3.0.0)

**Config/testthat/edition** 3

**Depends** R (>= 3.5.0)

**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>

**Repository** CRAN

**Date/Publication** 2022-04-22 12:00:03 UTC

## R topics documented:

CKD . . . . .	2
DISR . . . . .	3
HCV . . . . .	4
ISR . . . . .	4
Mean . . . . .	5
MMLPCA . . . . .	6

MNIPALS . . . . .	7
MRPCA . . . . .	8
orange . . . . .	9
ozone . . . . .	9
PM2.5 . . . . .	10
review . . . . .	11
SR . . . . .	12
<b>Index</b>	<b>13</b>

---

 CKD

---

*CKD*


---

### Description

chronic kidney disease

### Usage

```
data("CKD")
```

### Format

The format is: num [1:400, 1:18] 48 7 62 48 51 60 68 24 52 53 ... - attr(\*, "dimnames")=List of 2 ..\$ : NULL ..\$ : chr [1:18] "age" "bp" "sg" "al" ...

### Details

There are 1010 missing values in the data set, accounting for 14.03 percent.

### Source

Dr.P.Soundarapandian.M.D.,D.M (Senior Consultant Nephrologist), Apollo Hospitals, Managiri, Madurai Main Road, Karaikudi, Tamilnadu, Indi

### References

Polat, H., Danaei-Mehr, H., and Cetin, A. (2017). Diagnosis of chronic kidney disease based on support vector machine by feature selection methods. Journal of Medical Systems, 41(4), 1-11.

### Examples

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

DISR

*Distributed iterated score regression***Description**

Calculate the estimator on the DISR method

**Usage**

```
DISR(data = 0, data0, real = TRUE, example = FALSE, D)
```

**Arguments**

<code>data</code>	is the original data set
<code>data0</code>	is the missing data set
<code>real</code>	is to judge whether the data set is a real missing data set
<code>example</code>	is to judge whether the data set is a simulation example
<code>D</code>	is the number of nodes

**Value**

<code>XDISR</code>	is the estimator on the DISR method
<code>MSEDISR</code>	is the MSE value of the DISR method
<code>MAEDISR</code>	is the MAE value of the DISR method
<code>REDISR</code>	is the RE value of the DISR method
<code>GCVDISR</code>	is the GCV value of the DISR method
<code>timeDISR</code>	is the time cost of the DISR method

**Examples**

```
library(MASS)
n=100;p=10;per=0.1
X0=data=matrix(mvrnorm(n*p,0,1),n,p)
m=round(per*n*p,digits=0)
mr=sample(1:(n*p),m,replace=FALSE)
X0[mr]=NA;data0=X0
DISR(data=data,data0=data0,real=FALSE,example=FALSE,D=2)
```

---

HCV

*HCV*


---

**Description**

Hepatitis C virus

**Usage**

```
data("HCV")
```

**Format**

The format is: num [1:615, 1:13] 1 1 1 1 1 1 1 1 1 ... - attr(\*, "dimnames")=List of 2 ..\$ : chr [1:615] "1" "2" "3" "4" ... ..\$ : chr [1:13] "Category" "Age" "Sex" "ALB" ...

**Details**

There are 31 missing values in the data set, accounting for 0.39 percent.

**Source**

UCI repository

**References**

Lichtinghagen, R., Pietsch, D., Bantel, H., Manns, M., Brand, K. and Bahr, Matthias. (2013). The Enhanced Liver Fibrosis (ELF) Score: Normal Values, Influence Factors and Proposed Cut-Off Values.. *Journal of hepatology*. 59. 236-242.

**Examples**

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

---

ISR

*Iterated score regression*


---

**Description**

Calculate the estimator on the ISR method

**Usage**

```
ISR(data = 0, data0, real = TRUE, example = FALSE)
```

**Arguments**

data	is the original data set
data0	is the missing data set
real	is to judge whether the data set is a real missing data set
example	is to judge whether the data set is a simulation example

**Value**

XISR	is the estimator on the ISR method
MSEISR	is the MSE value of the ISR method
MAEISR	is the MAE value of the ISR method
REISR	is the RE value of the ISR method
GCVISR	is the GCV value of the ISR method
timeISR	is the time cost of the ISR method

**Examples**

```
library(MASS)
n=100;p=10;per=0.1
X0=data=matrix(mvrnorm(n*p,0,1),n,p)
m=round(per*n*p,digits=0)
mr=sample(1:(n*p),m,replace=FALSE)
X0[mr]=NA;data0=X0
ISR(data=data,data0=data0,real=FALSE,example=FALSE)
```

---

Mean

*Mean method*


---

**Description**

Calculate the estimator on the Mean method

**Usage**

```
Mean(data = 0, data0, real = TRUE, example = FALSE)
```

**Arguments**

data	is the original data set
data0	is the missing data set
real	is to judge whether the data set is a real missing data set
example	is to judge whether the data set is a simulation example

**Value**

XMean	is the estimator on the Mean method
MSEMean	is the MSE value of the Mean method
MAEMean	is the MAE value of the Mean method
REMean	is the RE value of the Mean method
GCVMean	is the GCV value of the Mean method
timeMean	is the time cost of the Mean method

**Examples**

```
library(MASS)
n=100;p=10;per=0.1
X0=data=matrix(mvrnorm(n*p,0,1),n,p)
m=round(per*n*p,digits=0)
mr=sample(1:(n*p),m,replace=FALSE)
X0[mr]=NA;data0=X0
Mean(data=data,data0=data0,real=FALSE,example=FALSE)
```

MMLPCA

*Modified maximum likelihood principal component analysis***Description**

Calculate the estimator on the ISR method

**Usage**

```
MMLPCA(data = 0, data0, real = TRUE, example = FALSE)
```

**Arguments**

data	is the original data set
data0	is the missing data set
real	is to judge whether the data set is a real missing data set
example	is to judge whether the data set is a simulation example

**Value**

XMMLPCA	is the estimator on the MMLPCA method
MSEMMLPCA	is the MSE value of the MMLPCA method
MAEMMLPCA	is the MAE value of the MMLPCA method
REMMLPCA	is the RE value of the MMLPCA method
GCVMMLPCA	is the GCV value of the MMLPCA method
timeMMLPCA	is the time cost of the MMLPCA method

**Examples**

```

library(MASS)
n=100;p=10;per=0.1
X0=data=matrix(mvrnorm(n*p,0,1),n,p)
m=round(per*n*p,digits=0)
mr=sample(1:(n*p),m,replace=FALSE)
X0[mr]=NA;data0=X0
MMLPCA(data=data,data0=data0,real=FALSE,example=FALSE)

```

MNIPALS

*Modified nonlinear iterative partial least squares method***Description**

Calculate the estimator on the MNIPALS method

**Usage**

```
MNIPALS(data = 0, data0, real = TRUE, example = FALSE)
```

**Arguments**

data	is the original data set
data0	is the missing data set
real	is to judge whether the data set is a real missing data set
example	is to judge whether the data set is a simulation example

**Value**

XMNIPALS	is the estimator on the MNIPALS method
MSEMNIPALS	is the MSE value of the MNIPALS method
MAEMNIPALS	is the MAE value of the MNIPALS method
REMNIPALS	is the RE value of the MNIPALS method
GVMNIPALS	is the GCV value of the MNIPALS method
timeMNIPALS	is the time cost of the MNIPALS method

**Examples**

```

library(MASS)
n=100;p=10;per=0.1
X0=data=matrix(mvrnorm(n*p,0,1),n,p)
m=round(per*n*p,digits=0)
mr=sample(1:(n*p),m,replace=FALSE)
X0[mr]=NA;data0=X0
MNIPALS(data=data,data0=data0,real=FALSE,example=FALSE)

```

---

 MRPCA

*Modified regularized PCA*


---

### Description

Calculate the estimator on the MRPCA method

### Usage

```
MRPCA(data = 0, data0, real = TRUE, example = FALSE)
```

### Arguments

data	is the original data set
data0	is the missing data set
real	is to judge whether the data set is a real missing data set
example	is to judge whether the data set is a simulation example

### Value

XMRPCA	is the estimator on the MRPCA method
MSEMRPCA	is the MSE value of the MRPCA method
MAEMRPCA	is the MAE value of the MRPCA method
REMRPCA	is the RE value of the MRPCA method
GCVMRPCA	is the GCV value of the MRPCA method
timeMRPCA	is the time cost of the MRPCA method

### Examples

```
library(MASS)
n=100;p=10;per=0.1
X0=data=matrix(mvrnorm(n*p,0,1),n,p)
m=round(per*n*p,digits=0)
mr=sample(1:(n*p),m,replace=FALSE)
X0[mr]=NA;data0=X0
MRPCA(data=data,data0=data0,real=FALSE,example=FALSE)
```



---

orange	<i>orange</i>
--------	---------------

---

**Description**

orange

**Usage**

```
data("orange")
```

**Format**

The format is: num [1:12, 1:8] 4.79 4.58 4.71 6.58 NA ... - attr(\*, "dimnames")=List of 2 ..\$ : chr [1:12] "1" "2" "3" "4" ... ..\$ : chr [1:8] "Color.intensity" "Odor.intensity" "Attack.intensity" "Sweet" ...

**Details**

There are 19 missing values in the data set, accounting for 19.79 percent.

**Source**

<http://factominer.free.fr/missMDA/index.html>

**References**

Josse J, Husson F (2016). missMDA: A Package for Handling Missing Values in Multivariate Data Analysis. *Journal of Statistical Software*, 70(1), 1–31.

**Examples**

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

---

ozone	<i>ozone</i>
-------	--------------

---

**Description**

ozone

**Usage**

```
data("ozone")
```

**Format**

A data frame with 112 observations on the following 11 variables.

max03 a numeric vector

T9 a numeric vector

T12 a numeric vector

T15 a numeric vector

Ne9 a numeric vector

Ne12 a numeric vector

Ne15 a numeric vector

Vx9 a numeric vector

Vx12 a numeric vector

Vx15 a numeric vector

max03v a numeric vector

**Details**

There are 115 missing values in it, accounting for 9.96 percent.

**Source**

<http://factominer.free.fr/missMDA/index.html>

**References**

Audigier, V., Husson, F., and Josse, J. (2014). A principal components method to impute missing values for mixed data. *Advances in Data Analysis and Classification*, 10(1), 5-26.

**Examples**

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

---

PM2.5

*PM2.5*

---

**Description**

Beijing PM2.5

**Usage**

```
data("PM2.5")
```

**Format**

The format is: num [1:43824, 1:12] 2010 2010 2010 2010 2010 2010 2010 2010 2010 2010 2010 ... - attr(\*, "dimnames")=List of 2 ..\$ : chr [1:43824] "1" "2" "3" "4" ... ..\$ : chr [1:12] "year" "month" "day" "hour" ...

**Details**

It records 43824 daily measurements on 12 variables and there are 2067 missing values on 2067 measurements, accounting for 0.00393.

**Source**

UCI repository

**References**

X. Liang, T. Zou, B. Guo, S. Li, H. Zhang, S. Zhang, H. Huang, and S. Chen. Assessing Beijing's PM2.5 pollution: severity, weather impact, APEC and winter heating. Proceedings of the Royal Society A, 471(2182):1–20, 2015.

**Examples**

```
data(PM2.5)
## maybe str(PM2.5) ; plot(PM2.5) ...
```

---

review

*review*

---

**Description**

Travel reviews

**Usage**

```
data("review")
```

**Format**

The format is: num [1:980, 1:10] 0.93 1.02 1.22 0.45 0.51 0.99 0.9 0.74 1.12 0.7 ... - attr(\*, "dimnames")=List of 2 ..\$ : chr [1:980] "User\_1" "User\_2" "User\_3" "User\_4" ... ..\$ : chr [1:10] "Category\_1" "Category\_2" "Category\_3" "Category\_4" ...

**Details**

980 travelers' reviews of 10 different types of travel facilities in East Asia

**Source**

UCI repository

## References

Renjith, S., Sreekumar, A., and Jathavedan, M. (2018). Evaluation of partitioning clustering algorithms for processing social media data in tourism domain. 2018 IEEE Recent Advances in Intelligent Computational Systems (RAICS), 127-131.

## Examples

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

---

 SR

*Score regression*


---

## Description

Calculate the estimator on the SR method

## Usage

```
SR(data = 0, data0, real = TRUE, example = FALSE)
```

## Arguments

data	is the original data set
data0	is the missing data set
real	is to judge whether the data set is a real missing data set
example	is to judge whether the data set is a simulation example

## Value

XSR	is the estimator on the SR method
MSESR	is the MSE value of the SR method
MAESR	is the MAE value of the SR method
RESR	is the RE value of the SR method
GCVSR	is the GCV value of the SR method
timeSR	is the time cost of the SR method

## Examples

```
library(MASS)
n=100;p=10;per=0.1
X0=data=matrix(mvrnorm(n*p,0,1),n,p)
m=round(per*n*p,digits=0)
mr=sample(1:(n*p),m,replace=FALSE)
X0[mr]=NA;data0=X0
SR(data=data,data0=data0,real=FALSE,example=FALSE)
```

# Index

## \* datasets

- CKD, [2](#)
- HCV, [4](#)
- orange, [9](#)
- ozone, [9](#)
- PM2.5, [10](#)
- review, [11](#)

CKD, [2](#)

DISR, [3](#)

HCV, [4](#)

ISR, [4](#)

Mean, [5](#)

MMLPCA, [6](#)

MNIPALS, [7](#)

MRPCA, [8](#)

orange, [9](#)

ozone, [9](#)

PM2.5, [10](#)

review, [11](#)

SR, [12](#)