

# Package ‘ISR’

September 24, 2021

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

**Date** 2021-09-24

**Version** 2021.9.24

**Description** Algorithms to handle with PCA-based missing data with high correlation. 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

**RoxygenNote** 7.1.1

**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>

**Suggests** magrittr, testthat (>= 3.0.0)

**Config/testthat/edition** 3

**Depends** R (>= 3.5.0)

**Imports** stats, MASS

**Repository** CRAN

**Date/Publication** 2021-09-24 04:40:02 UTC

## R topics documented:

CKD . . . . .	2
HCV . . . . .	3
ISR . . . . .	3
mean . . . . .	4
MMLPCA . . . . .	5
MNIPALS . . . . .	6
MRPCA . . . . .	7
orange . . . . .	8
ozone . . . . .	8

PM2.5 . . . . .	9
review . . . . .	10
SR . . . . .	11

<b>Index</b>	<b>12</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) ...
```

---

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

*Calculate the estimator on the ISR method*


---

**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, MSEISR, MAEISR, REISR, GCVISR,timeISR

**Examples**

```

library(MASS)
etanol=0.9
n=100;p=10;per=0.1
mu=as.matrix(runif(p,0,10))
sigma=as.matrix(runif(p,0,1))
ro=as.matrix(c(runif(p,-1,1)))
R0=ro%*%t(ro);diag(R0)=1
Sigma=sigma%*%t(sigma)*R0
X0=data=mvrnorm(n,mu,Sigma)
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

*Calculate the estimator on the 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, MSEmean, MAEmean, REmean, GCVmean

**Examples**

```

library(MASS)
etanol=0.9
n=100;p=10;per=0.1
mu=as.matrix(runif(p,0,10))
sigma=as.matrix(runif(p,0,1))
ro=as.matrix(c(runif(p,-1,1)))
R0=ro%%t(ro);diag(R0)=1
Sigma=sigma%%t(sigma)*R0
X0=data=mvrnorm(n,mu,Sigma)
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

*Calculate the estimator on the MMLPCA method***Description**

Calculate the estimator on the MMLPCA 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, MSEMMLPCA, MAEMMLPCA, REMMLPCA, GCVMMLPCA,timeMMLPCA

**Examples**

```

library(MASS)
etanol=0.9
n=100;p=10;per=0.1
mu=as.matrix(runif(p,0,10))
sigma=as.matrix(runif(p,0,1))
ro=as.matrix(c(runif(p,-1,1)))
R0=ro%%t(ro);diag(R0)=1
Sigma=sigma%%t(sigma)*R0
X0=data=mvrnorm(n,mu,Sigma)

```

```

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

*Calculate the estimator on the MNIPALS 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, MSEMNIPLALS, MAEMNIPLALS, REMNIPLALS, GCVMNIPALS,timeMNIPALS

### Examples

```

library(MASS)
etanol=0.9
n=100;p=10;per=0.1
mu=as.matrix(runif(p,0,10))
sigma=as.matrix(runif(p,0,1))
ro=as.matrix(c(runif(p,-1,1)))
R0=ro%*%t(ro);diag(R0)=1
Sigma=sigma%*%t(sigma)*R0
X0=data=mvrnorm(n,mu,Sigma)
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***Calculate the estimator on the MRPCA method*

---

**Description**

Calculate the estimator on the MRPCA method

**Usage**

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

**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.

**Value**

XMRPCA, MSEMPCA, MAEMPCA, REMPCA, GCVMRPCA, timeMRPCA

**Examples**

```
library(MASS)
etanol=0.9
n=100;p=10;per=0.1
mu=as.matrix(runif(p,0,10))
sigma=as.matrix(runif(p,0,1))
ro=as.matrix(c(runif(p,-1,1)))
R0=ro%*%t(ro);diag(R0)=1
Sigma=sigma%*%t(sigma)*R0
X0=data=mvrnorm(n,mu,Sigma)
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                      *orange*

---

**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                      *ozone*

---

**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

---

*Calculate the estimator on the SR method*


---

## 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, MSESr, MAESr, RESr, GCVSR

## Examples

```
library(MASS)
etanol=0.9
n=100;p=10;per=0.1
mu=as.matrix(runif(p,0,10))
sigma=as.matrix(runif(p,0,1))
ro=as.matrix(c(runif(p,-1,1)))
R0=ro*%t(ro);diag(R0)=1
Sigma=sigma*%t(sigma)*R0
X0=data=mvrnorm(n,mu,Sigma)
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, [3](#)
  - orange, [8](#)
  - ozone, [8](#)
  - PM2.5, [9](#)
  - review, [10](#)
- 
- CKD, [2](#)
- 
- HCV, [3](#)
- 
- ISR, [3](#)
- 
- mean, [4](#)
  - MMLPCA, [5](#)
  - MNIPALS, [6](#)
  - MRPCA, [7](#)
- 
- orange, [8](#)
  - ozone, [8](#)
- 
- PM2.5, [9](#)
- 
- review, [10](#)
- 
- SR, [11](#)