

# Package ‘SOPC’

August 29, 2022

**Title** The Sparse Online Principal Component Estimation Algorithm

**Date** 2022-08-29

**Version** 0.0.4

**Description** The sparse online principal component can not only process the on-line data set, but also obtain a sparse solution of the online data set. The philosophy of the package is described in Guo G. (2022) <[doi:10.1007/s00180-022-01270-z](https://doi.org/10.1007/s00180-022-01270-z)>.

**License** MIT + file LICENSE

**Encoding** UTF-8

**LazyData** true

**RoxygenNote** 7.2.0

**Imports** elasticnet, stats

**Suggests** testthat (>= 3.0.0)

**Config/testthat/edition** 3

**Maintainer** Guangbao Guo <[ggb11111111@163.com](mailto:ggb11111111@163.com)>

**NeedsCompilation** no

**Author** Guangbao Guo [aut, cre],  
Chunjie Wei [aut],  
Guoqi Qian [aut]

**Depends** R (>= 3.5.0)

**Repository** CRAN

**Date/Publication** 2022-08-29 06:40:05 UTC

## R topics documented:

Heart	2
Hugging	3
IPC	4
OPC	4
PC	5
PPC	6

PSA . . . . .	6
SAPC . . . . .	7
SOPC . . . . .	8
SPC . . . . .	9

<b>Index</b>	<b>10</b>
--------------	-----------

---

Heart	<i>Heart failure</i>
-------	----------------------

---

### Description

Heart failure

### Usage

```
data("Heart")
```

### Format

A data frame with 299 observations on the following 13 variables.

age a numeric vector  
 anaemia a numeric vector  
 creatinine\_phosphokinase a numeric vector  
 diabetes a numeric vector  
 ejection\_fraction a numeric vector  
 high\_blood\_pressure a numeric vector  
 platelets a numeric vector  
 serum\_creatinine a numeric vector  
 serum\_sodium a numeric vector  
 sex a numeric vector  
 smoking a numeric vector  
 time a numeric vector  
 DEATH\_EVENT a numeric vector

### Details

This dataset contains the medical records of 299 patients who had heart failure, collected during their follow-up period, where each patient profile has 13 clinical features.

### Source

The Heart failure data set comes from the UCI database.

**References**

Davide Chicco, Giuseppe Jurman. (2020). Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone. BMC Medical Informatics and Decision Making.

**Examples**

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

---

Hugging

*Hugging*

---

**Description**

The EMG Physical Action-Hugging data set.

**Usage**

```
data("Hugging")
```

**Format**

A data frame with 9752 observations on the following 8 variables.

A a numeric vector  
B a numeric vector  
C a numeric vector  
D a numeric vector  
E a numeric vector  
F a numeric vector  
G a numeric vector  
H a numeric vector

**Details**

The data set is a body movement data set, including 10 normal and 10 aggressive body movements. The data frame with 9752 observations on the following 8 variables.

**Source**

The Hugging data set comes from the UCI database.

**References**

Demir et al. (2019). Surface emg signals and deep transfer learning-based physical action classification. Neural Computing and Applications.

**Examples**

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

---

IPC	<i>The incremental principal component can handle online data sets with highly correlated.</i>
-----	--

---

**Description**

The incremental principal component can handle online data sets with highly correlated.

**Usage**

```
IPC(data, m, eta)
```

**Arguments**

data	is a highly correlated online data set
m	is the number of principal component
eta	is the proportion of online data to total data

**Value**

Ai,Di

**Examples**

```
IPC(data=PSA,m=3,eta=0.8)
```

---

OPC	<i>The online principal component method refers to the IPC method with the best performance among the IPC, the PPC and the SAPC methods.</i>
-----	--

---

**Description**

The online principal component method refers to the IPC method with the best performance among the IPC, the PPC and the SAPC methods.

**Usage**

```
OPC(data, m, eta)
```

**Arguments**

data is a highly correlated online data set  
 m is the number of principal component  
 eta is the proportion of online data to total data

**Value**

Ao,Do

**Examples**

```
OPC(data=PSA,m=3,eta=0.8)
```

---

PC *The traditional principal component method. This method can estimate the eigen space of the data set.*

---

**Description**

The traditional principal component method. This method can estimate the eigen space of the data set.

**Usage**

```
PC(data, m = m)
```

**Arguments**

data is a highly correlated data set  
 m is the number of principal component

**Value**

Ahat, Dhat

**Examples**

```
PC(data=PSA,m=3)
```

---

PPC	<i>The perturbation principal component can handle online data sets with highly correlated.</i>
-----	---

---

**Description**

The perturbation principal component can handle online data sets with highly correlated.

**Usage**

```
PPC(data, m, eta)
```

**Arguments**

data	is a highly correlated online data set
m	is the number of principal component
eta	is the proportion of online data to total data

**Value**

Ap,Dp

**Examples**

```
PPC(data=PSA,m=3,eta=0.8)
```

---

PSA	<i>Prostate Specific Antigen</i>
-----	----------------------------------

---

**Description**

The prostate specific antigen (PSA) data set.

**Usage**

```
data("PSA")
```

**Format**

lcavol a numeric vector  
lweight a numeric vector  
age a numeric vector  
lbph a numeric vector  
svi a numeric vector  
lcp a numeric vector  
gleason a numeric vector  
pgg45 a numeric vector  
lpsa a numeric vector

**Details**

The data set comes from the prostate specific antigen (PSA) data of 96 patients collected by Stanford University Medical Center. These patients all underwent radical prostatectomy.

**Source**

The Stanford University Medical Center.

**References**

NA

**Examples**

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

---

SAPC

*The stochastic approximation principal component can handle online data sets with highly correlated.*

---

**Description**

The stochastic approximation principal component can handle online data sets with highly correlated.

**Usage**

```
SAPC(data, m, eta)
```

**Arguments**

data	is a highly correlated online data set
m	is the number of principal component
eta	is the proportion of online data to total data

**Value**

Asa,Dsa

**Examples**

SAPC(data=PSA,m=3,eta=0.8)

---

SOPC	<i>The sparse online principal component can not only process online data sets, but also obtain a sparse solution of online data sets.</i>
------	--

---

**Description**

The sparse online principal component can not only process online data sets, but also obtain a sparse solution of online data sets.

**Usage**

SOPC(data, m, gamma, eta)

**Arguments**

data	is a highly correlated online data set
m	is the number of principal component
gamma	is a sparse parameter
eta	is the proportion of online data to total data

**Value**

Aso,Dso

**Examples**

SOPC(data=PSA,m=3,gamma=0.03,eta=0.6)



---

SPC

*The sparse principal component can obtain sparse solutions of the eigenmatrix to better explain the relationship between principal components and original variables.*

---

### **Description**

The sparse principal component can obtain sparse solutions of the eigenmatrix to better explain the relationship between principal components and original variables.

### **Usage**

```
SPC(data, m, gamma)
```

### **Arguments**

data	is a highly correlated data set
m	is the number of principal component
gamma	is a sparse parameter

### **Value**

As,Ds

### **Examples**

```
SPC(data=PSA,m=3,gamma=0.03)
```

# Index

## \* datasets

Heart, [2](#)  
Hugging, [3](#)  
PSA, [6](#)

Heart, [2](#)  
Hugging, [3](#)

IPC, [4](#)

OPC, [4](#)

PC, [5](#)  
PPC, [6](#)  
PSA, [6](#)

SAPC, [7](#)  
SOPC, [8](#)  
SPC, [9](#)