Package 'pesel'

May 9, 2026

Type Package

Title Automatic Estimation of Number of Principal Components in PCA

Version 0.7.5

Date 2023-08-14

Author Piotr Sobczyk, Julie Josse, Malgorzata Bogdan

Maintainer Piotr Sobczyk <pj.sobczyk@gmail.com>

Description Automatic estimation of number of principal components in PCA
with PEnalized SEmi-
integrated Likelihood (PESEL). See Piotr Sobczyk, Malgorzata Bogdan, Julie Josse
``Bayesian dimensionality reduction with PCA using penalized semi-integrated likelihood"
(2017) <doi:10.1080/10618600.2017.1340302>.

License GPL-3

Encoding UTF-8

URL https://github.com/psobczyk/pesel

BugReports https://github.com/psobczyk/pesel/issues

Depends R (>= 3.1.3),

Imports stats, graphics

RoxygenNote 7.2.3

NeedsCompilation no

Repository CRAN

Date/Publication 2023-10-17 13:20:02 UTC

Contents

pesel-package *Automatic estimation of number of principal components in PCA*

Description

Automatic estimation of number of principal components in PCA with PEnalized SEmi-integrated Likelihood (PESEL).

Details

Version: 0.7.5

Author(s)

Piotr Sobczyk, Julie Josse, Malgorzata Bogdan

Maintainer: Piotr Sobczyk <pj.sobczyk@gmail.com>

References

Piotr Sobczyk, Malgorzata Bogdan, Julie Josse "Bayesian dimensionality reduction with PCA using penalized semi-integrated likelihood", Journal of Computational and Graphical Statistics 2017

Examples

```
# EXAMPLE 1 - noise
with(set.seed(23), pesel(matrix(rnorm(10000), ncol = 100), npc.min = 0))

# EXAMPLE 2 - fixed effects PCA model
sigma <- 0.5
k <-  5
n <- 100
numb.vars <- 10
# factors are drawn from normal distribution
factors <- replicate(k, rnorm(n, 0, 1))
# coefficients are drawn from uniform distribution
coeff <- replicate(numb.vars, rnorm(k, 0, 1))
SIGNAL <- scale(factors %*% coeff)
X <- SIGNAL + replicate(numb.vars, sigma * rnorm(n))
pesel(X)
```

`pesel` *Automatic estimation of number of principal components in PCA with PEnalized SEmi-integrated Likelihood (PESEL)*

Description

Underlying assumption is that only small number of principal components, associated with largest singular values, is relevent, while the rest of them is noise. For a given numeric data set, function estimates the number of PCs according to penalized likelihood criterion. Function adjusts the model used to the case when number of variables is larger than the number of observations.

Usage

```
pesel(
  X,
  npc.min = 0,
  npc.max = 10,
  prior = NULL,
  scale = TRUE,
  method = c("heterogenous", "homogenous"),
  asymptotics = NULL
)
```

Arguments

`X`	a data frame or a matrix contatining only continuous variables
`npc.min`	minimal number of principal components, for all the possible number of PCs between npc.min and npc.max criterion is computed
`npc.max`	maximal number of principal components, if greater than dimensions of X, min(ncol(X), nrow(X))-1 is used, for all the possible number of PCs between npc.min and npc.max criterion is computed
`prior`	a numeric positive vector of length npc.max-ncp.min+1. Prior distribution on number of principal components. Defaults to uniform distibution
`scale`	a boolean, if TRUE (default value) then data is scaled before applying criterion
`method`	name of criterion to be used
`asymptotics`	a character, asymptotics ('n' or 'p') to be used. Default is NULL for which asymptotics is selected based on dimensions of X

Details

Please note that no categorical variables and missing values are allowed.

Value

number of components

Examples

```
# EXAMPLE 1 - noise
with(set.seed(23), pesel(matrix(rnorm(10000), ncol = 100), npc.min = 0))

# EXAMPLE 2 - fixed effects PCA model
sigma <- 0.5
k <-  5
n <- 100
numb.vars <- 10
# factors are drawn from normal distribution
factors <- replicate(k, rnorm(n, 0, 1))
# coefficients are drawn from uniform distribution
coeff <- replicate(numb.vars, rnorm(k, 0, 1))
SIGNAL <- scale(factors %*% coeff)
X <- SIGNAL + replicate(numb.vars, sigma * rnorm(n))
pesel(X)
```

---

pesel_heterogeneous *PEnalized SEmi-integrated Likelihood for heterogeneous singular values and large number of variables*

Description

Derived under assumption that number of variables tends to infinity while number of observations is limited.

Usage

```
pesel_heterogeneous(X, minK, maxK)
```

Arguments

X	a matrix containing only continuous variables
minK	minimal number of principal components fitted
maxK	maximal number of principal components fitted

Value

numeric vector, PESEL criterion for each k in range [minK, maxK]

`pesel_homogeneous` *PEnalized SEmi-integrated Likelihood for homogeneous singular values and large number of variables*

Description

Derived under assumption that number of variables tends to infinity while number of observations is limited.

Usage

```
pesel_homogeneous(X, minK, maxK)
```

Arguments

`X`	a matrix containing only continuous variables
`minK`	minimal number of principal components fitted
`maxK`	maximal number of principal components fitted

Value

numeric vector, PESEL criterion for each k in range [minK, maxK]

Index