

Package ‘sigPCA’

August 3, 2026

Title Statistical Significance Testing for Principal Components

Version 0.1.0

Description Identifies principal components whose eigenvalues exceed those expected under noise. Implements analytical thresholds derived from the Marchenko-Pastur distribution (Marchenko and Pastur, 1967) [<doi:10.1070/SM1967v001n04ABEH001994>](https://doi.org/10.1070/SM1967v001n04ABEH001994) and empirical permutation tests, and provides functions for visualizing observed and null eigenvalue spectra.

License MIT + file LICENSE

URL <https://github.com/guillermodeandajauregui/sigPCA>

BugReports <https://github.com/guillermodeandajauregui/sigPCA/issues>

Encoding UTF-8

Imports ggplot2, stats

Suggests knitr, palmerpenguins, rmarkdown, testthat (>= 3.0.0)

VignetteBuilder knitr

Config/testthat/edition 3

Config/roxygen2/markdown TRUE

Config/roxygen2/version 8.0.0

NeedsCompilation no

Author Guillermo de Anda-Jáuregui [aut, cre, cph],
Enrique Hernández-Lemus [aut, cph]

Maintainer Guillermo de Anda-Jáuregui <gdeanda@inmegen.edu.mx>

Repository CRAN

Date/Publication 2026-08-03 18:10:02 UTC

Contents

compute_eigenvalues	2
marchenko_pastur_bounds	2
plot_sigPCA	3

plot_sigPCA_histogram	4
sigPCA	4
sigPCA_mp	5
sigPCA_perm	6

Index	7
--------------	----------

compute_eigenvalues	<i>Compute Eigenvalues of the Covariance Matrix</i>
---------------------	---

Description

Compute Eigenvalues of the Covariance Matrix

Usage

compute_eigenvalues(data)

Arguments

data A numeric matrix or data frame. Observations in rows, features in columns.

Value

A numeric vector of eigenvalues, sorted decreasingly.

Examples

```
set.seed(123)
mat <- matrix(rnorm(50), nrow = 10, ncol = 5)
compute_eigenvalues(mat)
```

marchenko_pastur_bounds	<i>Marchenko–Pastur Theoretical Bounds</i>
-------------------------	--

Description

Computes the lower and upper bounds of the Marchenko–Pastur distribution, based on the aspect ratio of the data matrix.

Usage

marchenko_pastur_bounds(p, n)

Arguments

p	Number of features (columns)
n	Number of observations (rows)

Value

A named list with `lambda_min` and `lambda_max`.

Examples

```
bounds <- marchenko_pastur_bounds(p = 10, n = 100)
print(bounds)
```

plot_sigPCA

Plot PCA Eigenvalue Spectrum with Marchenko–Pastur Bounds

Description

Visualizes the eigenvalue spectrum of a PCA, highlighting significant components based on Marchenko–Pastur bounds.

Usage

```
plot_sigPCA(eigenvalues, mp_bounds, highlight = NULL, add_null = NULL)
```

Arguments

eigenvalues	Numeric vector of eigenvalues from the real data.
mp_bounds	A list with <code>lambda_min</code> and <code>lambda_max</code> (from <code>marchenko_pastur_bounds()</code>).
highlight	Logical vector (same length as <code>eigenvalues</code>). Defaults to <code>eigenvalues > lambda_max</code> .
add_null	Optional numeric vector of null eigenvalues (e.g. from permutation). Default: <code>NULL</code> .

Value

A `ggplot2` object.

Examples

```
set.seed(123)
x <- matrix(rnorm(100 * 10), nrow = 100, ncol = 10)
result <- sigPCA_mp(x)
plot_sigPCA(result$eigenvalues, result$mp_bounds)
```

`plot_sigPCA_histogram` *Plot PCA Eigenvalue Histogram with Marchenko-Pastur Bounds*

Description

PCA eigenvalue histogram, showing the empirical eigenvalue distribution together with the Marchenko-Pastur theoretical bounds. Original design by Enrique Hernandez Lemus

Usage

```
plot_sigPCA_histogram(x, mp_bounds = NULL, bins = 30)
```

Arguments

<code>x</code>	A result from <code>sigPCA_mp()</code> , <code>sigPCA(method = "mp")</code> , or a numeric vector of eigenvalues.
<code>mp_bounds</code>	Optional list with <code>lambda_min</code> and <code>lambda_max</code> . Required if <code>x</code> is a numeric vector.
<code>bins</code>	Number of histogram bins. Default: 30.

Value

A `ggplot2` object.

Examples

```
set.seed(123)
x <- matrix(rnorm(100 * 10), nrow = 100, ncol = 10)
result <- sigPCA_mp(x)
plot_sigPCA_histogram(result)
```

`sigPCA` *Unified PCA Significance Wrapper*

Description

Runs significance testing on PCA eigenvalues using either Marchenko–Pastur bounds, permutation tests, or both. Results are returned as a structured list.

Usage

```
sigPCA(
  data,
  method = c("mp", "perm", "both"),
  num_permutations = 1000,
  center = TRUE,
  scale. = TRUE
)
```

Arguments

data A numeric matrix or data frame.
method One of "mp", "perm", or "both". Default: "mp".
num_permutations Integer. Used only if method is "perm" or "both".
center, scale. Logical. Passed to scale(). Default: TRUE.

Value

A structured list containing method results.

Examples

```

set.seed(123)
x <- matrix(rnorm(100 * 10), nrow = 100, ncol = 10)
sigPCA(x, method = "mp")
sigPCA(x, method = "perm", num_permutations = 100)
sigPCA(x, method = "both", num_permutations = 100)

```

sigPCA_mp

PCA Significance Based on Marchenko–Pastur Distribution

Description

Tests which principal components have eigenvalues beyond the theoretical bounds of the Marchenko–Pastur distribution.

Usage

```
sigPCA_mp(data, center = TRUE, scale. = TRUE)
```

Arguments

data A numeric matrix or data frame. Observations in rows, variables in columns.
center Logical. Whether to center the variables. Default: TRUE.
scale. Logical. Whether to scale variables to unit variance. Default: TRUE.

Value

A list with:

eigenvalues All eigenvalues of the covariance matrix

mp_bounds A list with lambda_min and lambda_max

significant_components Indices of eigenvalues > lambda_max

Examples

```
set.seed(123)
x <- matrix(rnorm(100 * 10), nrow = 100, ncol = 10)
result <- sigPCA_mp(x)
result$significant_components
```

sigPCA_perm

PCA Significance via Permutation Test

Description

Assesses the statistical significance of PCA components using column-wise permutation of the input data to create a null distribution of eigenvalues.

Usage

```
sigPCA_perm(data, num_permutations = 1000, center = TRUE, scale. = TRUE)
```

Arguments

data	A numeric matrix or data frame.
num_permutations	Number of permutations to generate. Default: 1000.
center	Logical. Whether to center variables before analysis. Default: TRUE.
scale.	Logical. Whether to scale variables before analysis. Default: TRUE.

Value

A list with:

- eigenvalues** Observed eigenvalues
- null_eigenvalues** Matrix of eigenvalues from permutations
- pvalues** Empirical p-values per component
- significant_components** Indices where $p < 0.05$

Examples

```
set.seed(123)
x <- matrix(rnorm(100 * 10), nrow = 100, ncol = 10)
result <- sigPCA_perm(x, num_permutations = 100)
result$significant_components
```

Index

`compute_eigenvalues`, [2](#)

`marchenko_pastur_bounds`, [2](#)

`plot_sigPCA`, [3](#)

`plot_sigPCA_histogram`, [4](#)

`sigPCA`, [4](#)

`sigPCA_mp`, [5](#)

`sigPCA_perm`, [6](#)