Package 'do3PCA'

September 24, 2024

Title Probabilistic Phylogenetic Principal Component Analysis

Version 1.0.0

Description Estimates probabilistic phylogenetic Principal Component Analysis (PCA) and non-phylogenetic probabilistic PCA. Provides methods to implement alternative models of trait evolution including Brownian motion (BM), Ornstein-Uhlenbeck (OU), Early Burst (EB), and Pagel's lambda. Also provides flexible biplot functions.

License GPL (>= 2.0)

Encoding UTF-8

Depends R (>= 3.5.0), Rdimtools

Imports ape, phytools, matrixcalc, mclust, nloptr, ratematrix, stats, utils

RoxygenNote 7.3.1

NeedsCompilation no

Author Daniel Caetano [aut, cre] (<https://orcid.org/0000-0002-2803-4064>)

Maintainer Daniel Caetano <caetanods1@gmail.com>

Repository CRAN

Date/Publication 2024-09-24 19:10:10 UTC

Contents

---

doBiplot   *Make biplot for any type of PCA*

Description

Function to make biplots for any kind of PCA. It accepts the outputs from standard PCA (princomp and prcomp). It also works with the "phylProbPCA" and "ProbPCA" functions. It provides more options to the plot than the standard "stats::biplot".

Usage

```
doBiplot(x, choices = 1L:2L, scale = 1, pc.biplot = FALSE, col, ...)
```

Arguments

x	output from PCA analysis.
choices	numeric vector of length 2. Use to choose which of the PC axes to plot. Default plots first and second axes: "choices = c(1,2)".
scale	numeric value between 0 and 1. Same as in "stats::biplot.princomp". See ?biplot.princomp for more information.
pc.biplot	logical. If TRUE it will produce a "principal component biplot" (sensu Gabriel, 1971). Same as in "stats::biplot.princomp". See ?biplot.princomp for more information.
col	character vector of length 3 with the colors of the biplot. First color is used for the score points (or sample sames), second color for arrows and variable names, and third color for the right and top-side ticks (plot axes).
...	extra parameters for the function. Same as "stats::biplot".

Details

Function has the same options as "stats::biplot", with the addition of the following arguments. "plot_dimnames" controls is the names of the samples (species) will be plotted. "add_points" controls if the score points will be plotted. "add_margin" is a numeric value that expands the area of the plot. You can use this to make sure the names of variables and samples (species) fit the plot.

Value

makes a biplot of the PCA results.

Examples

```
phy <- ratematrix::anoles$phy[[1]]
dt <- as.matrix( ratematrix::anoles$data[,1:3] )

## Using probabilistic phylogenetic PCA
phylppca <- phylProbPCA(phy = phy, x = dt, ret_dim = 2)
```

```
doBiplot(x = phylppca, add_margin = 0.3)

## Using standard phylogenetic PCA
phylpca <- phytools::phyl.pca(tree = phy, Y = dt)
doBiplot(x = phylpca, add_margin = 0.3)

## Using probabilistic PCA
ppca <- ProbPCA(x = dt)
doBiplot(x = ppca, add_margin = 0.3)

## Using standard PCA
pca1 <- princomp(x = dt)
doBiplot(x = pca1, add_margin = 0.1)

## Using standard PCA
pca2 <- prcomp(x = dt)
doBiplot(x = pca2, add_margin = 0.1)
```

---

phylProbPCA   *Probabilistic Phylogenetic PCA*

Description

Function to perform probabilistic phylogenetic PCA. Allows for fit of alternative models of trait evolution using branch length transformation.

Usage

```
phylProbPCA(phy, x, ret_dim = 2, model = "BM", quiet = FALSE)
```

Arguments

phy	phylogeny in "phylo" format.
x	a matrix with traits in columns and species values in rows. Rownames must match the tip labels of phylogeny.
ret_dim	number of dimensions (PC axes) to be kept by the model.
model	choice of model of trait evolution. One of "BM", "lambda", "OU", or "EB".
quiet	if function should suppress output to the console while running

Details

The function can be used to estimate the probabilistic phylogenetic PCA (3PCA) using distinct models of trait evolution. Models are implemented using branch length transformation. Model fitting happens in two steps. First the maximum likelihood of the evolutionary covariance matrix (R) and the parameter of the model is estimated. Then the 3PCA model is estimated using the phylogenetic tree with branch lengths transformed following the MLE for the parameter of each trait evolution model.

The function returns a list with the following elements. scores: the scores of the principal components; e_values: eigenvalues; e_vectors: eigenvectors or the projection; model_fit: information about the trait evolution model; loadings: the loadings of the PCs; varnames: the names of the variables; sig: the MLE of the error; mle.W: the MLE of the W matrix; Function also returns AIC, AICc, and BIC for the model.

Value

returns a list of class "phylPPCA". See "Details" for more information.

References

Revell, L. J. 2009. Size-Correction and Principal Components for Interspecific Comparative Studies. Evolution 63:3258–3268. doi: 10.1111/j.1558-5646.2009.00804.x

Revell, L. J. 2024. phytools 2.0: an updated R ecosystem for phylogenetic comparative methods (and other things). PeerJ 12:e16505. doi: 10.7717/peerj.16505

Tipping, M. E., and C. M. Bishop. 1999. Probabilistic Principal Component Analysis. Journal of the Royal Statistical Society Series B: Statistical Methodology 61(3):611–622. doi: 10.1111/1467-9868.00196

Examples

```
phy <- ratematrix::anoles$phy[[1]]
dt <- as.matrix( ratematrix::anoles$data[,1:3] )
ppca <- phylProbPCA(phy = phy, x = dt, ret_dim = 2)
doBiplot(x = ppca, add_margin = 0.3)
```

ProbPCA   *Probabilistic PCA*

Description

Function to perform (non-phylogenetic) probabilistic PCA. This function is a modification (fork) of Rdimtools::do.ppca .

Usage

```
ProbPCA(x, ret_dim = 2)
```

Arguments

x	a matrix with traits in columns and observations in rows.
ret_dim	number of dimensions (PC axes) to be kept by the model.

Details

This function uses the same algorithm as Rdimtools::do.ppca. However, it returns more details from the estimation and computes AIC and AICc.

The function returns a list with the following elements. scores: the scores of the principal components; projection: the eigenvectors; sig: the MLE of the error of the model; mle.W: the MLE of the W matrix: varnames: the names of the traits; loglik: the log-likelihood of the estimate. Function also returns AIC, AICc, and BIC for the model.

Value

returns a list of class "phylPPCA". See "Details" for more information.

References

Tipping, M. E., and C. M. Bishop. 1999. Probabilistic Principal Component Analysis. Journal of the Royal Statistical Society Series B: Statistical Methodology 61(3):611–622. doi: 10.1111/1467-9868.00196

Examples

```
dt <- as.matrix( ratematrix::anoles$data[,1:3] )
ppca <- ProbPCA(x = dt, ret_dim = 2)
doBiplot(x = ppca, add_margin = 0.3)
```

Index