Package 'pcaBootPlot'

August 29, 2016

Title Create 2D Principal Component Plots with Bootstrapping

Version 0.2.0

Author Joshua Starmer

Maintainer Joshua Starmer <starmer@unc.edu>

Description Draws a 2D principal component plot using the first 2 principal components from the original and bootstrapped data to give some sense of variability.

Depends R (>= 3.0.2)

License GPL-2

LazyData true

Imports FactoMineR, RColorBrewer

Suggests knitr

VignetteBuilder knitr

URL https://github.com/starmerj/pcaBootPlot

NeedsCompilation no

Repository CRAN

Date/Publication 2015-06-23 00:34:33

R topics documented:

pcaBootPlot *Create 2D PCA Plots with Bootstrapping*

Description

pcaBootPlot draws a 2D PCA plot using the first 2 principal components using the original and bootstrapped data to give some sense of variability.

Usage

```
pcaBootPlot(data = NULL, groups = NULL, min.value = 1,
  all.min.value = FALSE, num.boot.samples = 100, log2.transform = TRUE,
  pdf.filename = NULL, pdf.width = 6, pdf.height = 6,
  draw.legend = FALSE, legend.names = NULL, legend.x = NULL,
  legend.y = NULL, transparency = 77, min.x = NULL, max.x = NULL,
  min.y = NULL, max.y = NULL, correct.inversions = TRUE,
  confidence.regions = FALSE, confidence.size = 0.95, step.size = 0.1,
  trim.proportion = 0, return.samples = FALSE, use.prcomp = FALSE)
```

Arguments

`data` A data.frame where the first column is named "ID" and contains IDs for each item measured. Measurements for each sample are in subsequent columns.

`groups` The default value is `NULL`.

If you want use different colors and shapes to deliniate the samples into groups, you can specify the grouping with this argument. Currently there is a limit of 9 different groups.

For example, if you have three consecutive columns of "untreated" samples followed by three consecutive columns of "treated" samples, you can set this argument to c(1,1,1,2,2,2), and the untreated samples will be red circles and the treated samples will be blue triangles.

`min.value` The default value is **1**.

This allows you to filter out rows (entries) that will not conribute to the PCA. For example, if you are performing PCA on RNA-seq data, you may wish to filter out genes with less than 1 read per sample, 1 read per group or 1 read overall. If you set `all.min.value` to `TRUE`, it will filter entries where at least one sample has less than `min.value`. If you do not set `all.min.value` to `TRUE`, then filtering will be performed by group if `groups` are specified. In this case, an entry will be filtered out if one or more groups have less than `min.value`.

If `groups` are not specified, then only entries where all samples have less than `min.value` will be removed from the analysis.

groups will also effect filtering based on min.value. See that part of the documentation for details.

all.min.value This parameter, set to either TRUE or FALSE, affects min.value. See the documentation for min.value for more details.

num.boot.samples
The default value is **100**. The number of bootstrap iterations to be performed.

log2.transform The default value is TRUE. Should the data be log2 transformed or not?

pdf.filename If you wish to save the the graph as a PDF, you may use this argument to specify the filename.

pdf.width If you specify a value for pdf.filename, you can specify a width for the saved graph. The default value is **6** inches.

pdf.height If you specify a value for pdf.filename, you can specify a height for the saved graph. The default value is **6** inches.

draw.legend The default value is FALSE. Should there be a legend in the graph?

legend.names If draw.legend is TRUE, you can specify the names of the groups listed in the legend.

legend.x,legend.y
If draw.legend is TRUE, you can specify the x and y axis coordinate for its top left corner.

transparency The default value is **77**. This allow you to set how transparent the bootstrapped symbols are in the graph. Values range from 00 to FF.

min.x,min.y,max.x,max.y
By default, pcaBootPlot automatically determines limits for the x and y axes. Use this option to override this behavior.

correct.inversions
The default value is TRUE. Some of the boostrapped PCAs may have their axes inverted. pcaBootPlot can try to correct for this by ensuring that the PCA loading values are positively correlated with the orginal dataset.

confidence.regions
The default value is FALSE. This option will draw circles that contain confidence.size of the bootstrapped values.

confidence.size
The default value is **0.95**. A value betweeo 0 and 1 - the proportion of bootstrapped points that need to be within the confidence regions.

step.size The default value is **0.1**. This option determines how the radii for confidence regions are increased each iteration when trying to contain confidence.size of the bootstrapped samples.

trim.proportion
The default value is **0.0**. This is the proportion of entries that should be removed from the plot based on the size of the the confidence regions. This should be a value between 0 and 1. For example, if you set it to 0.1, then the top 10 regions will be removed from the plot.

return.samples The default value is FALSE. If this is set to TRUE then the program will return the names of the samples that were included in the plot. This can be useful if trim.proportion > 0.

use.prcomp The default value is FALSE. Usually, pcaBootPlot uses FactoMineR to process samples. However, this can be unnecessarily slow if there are less than 50 samples. By setting use.prcomp to TRUE, it will use prcomp() to process samples and will, most likely, run much faster.

Examples

```
sample1=rnorm(n=100, mean=100, sd=10)
sample2=jitter(sample1, factor=10, amount=10)
sample3=rnorm(n=100, mean=100, sd=10)

data <- data.frame(ID=c(1:100), sample1, sample2, sample3)

pcaBootPlot(data, log2.transform = FALSE)
```

Index