

Package ‘LrSClust’

October 27, 2020

Type Package

Title Graph-based clustering method with various types of distances
based on L-infinity, L-1, and L-2 norms.

Version 1.0

Date 2020-09-30

Author Hao Wu, Disheng Mao, Yuping Zhang, Zhiyi Chi, Michael Stitzel, Zhengqing Ouyang

Maintainer Hao Wu <hao.5.wu@uconn.edu>, Yuping Zhang <yuping.zhang@uconn.edu> and Zhengqing Ouyang <ouyang@schoolph.umass.edu>

Description R-package for the paper (NAR Genomics and Bioinformatics, lqaa087) entitled “A new graph-based clustering method with application to single-cell RNA-seq data from human pancreatic islets” by Hao Wu, Disheng Mao, Yuping Zhang, Zhiyi Chi, Michael Stitzel, Zhengqing Ouyang

Imports methods, ggplot2, knitr

Depends R (>= 3.4.0)

VignetteBuilder knitr

License GPL (>= 3)

RoxygenNote 7.1.1

NeedsCompilation no

LazyData True

R topics documented:

GAP_statistic	2
LD_SC_clu	3
SelectClu2	3
Index	5

GAP_statistic	<i>Calculate a Gap-statistic type of criterion</i>
---------------	--

Description

This function returns the Gap-statistic type of criterion.

Usage

```
GAP_statistic(
  data,
  clusterF,
  pre_output,
  distType = c("maximum"),
  K = 1:10,
  P = seq(5, 50, 5),
  B = 20,
  spaceH0 = c("scaledPCA", "original"),
  d_X = NULL,
  seed = 1205
)
```

Arguments

data	cell-by-gene expression matrix
clusterF	clustering function
pre_output	result from LD_SC_clu
distType	distance types including "maximum" (based on L-infinity norm), "euclidean" (based on L-2 norm), "manhattan" (based on L-1 norm)
K	range of cluster numbers
P	perplexity
B	Number of bootstrap samples
spaceH0	reference distribution, can be selected from scaledPCA and original
d_X	distance matrix, if NULL, will be computed from data
seed	random seed

Value

An array of Gap statistics

LD_SC_clu	<i>Graph-based clustering method with various types of distances based on L-infinity, L-1, and L-2 norms.</i>
-----------	---

Description

This function returns the cluster assignment

Usage

```
LD_SC_clu(
  X,
  krange = 1:10,
  perp = seq(5, 50, 5),
  distType = "maximum",
  seed = 1205
)
```

Arguments

X	cell-by-gene expression matrix
krange	number of clusters range, default 1:10
perp	perplexity
distType	distance type, default L-infinity distance, including (maximum, euclidean, manhattan)
seed	random seed

Value

An array of all cluster assignments corresponding to different perplexity and number of clusters

SelectClu2	<i>Select the tuning parameters through Gap-statistic type of criterion.</i>
------------	--

Description

This function returns the selected cluster membership

Usage

```
SelectClu2(clumat, gapmat, method = "SE", prefix, K, perp)
```

Arguments

clumat	Cluster assignments generated by LD_SC_clu
gapmat	Gap statistics and standard deviations generated by GAP_statistic
method	Suggest to be "SE" (default)
prefix	path for plots
K	range of cluster numbers
perp	perplexity

Value

Cluster membership

Examples

```
set.seed(721)
dat <- matrix(rnorm(10000), 100, 10)
dat[1:10, ] <- dat[1:10, ] + 5
dat[51:100, ] <- dat[51:100, ] - 5
input_prefix <- "./"
output_pathLinf <- paste(input_prefix, "Linf", sep = '')
K <- 1:10
perp <- c(seq(5, 55, 10), c(70, 100, 150, 200, 300, 500))

SeedLinf <- 1205
res <- LD_SC_clu(dat, krange = K, perp = perp, distType = "maximum", SeedLinf)
gap <- GAP_statistic(dat, LD_SC_clu, res, distType = "maximum", K, perp, spaceH0 = "original")

clu <- as.factor(SelectClu2(res, gap, method = "SE", output_pathLinf, K, perp)[[1]])
clu
```

Index

GAP_statistic, [2](#)

LD_SC_clu, [3](#)

SelectClu2, [3](#)