Package ‘scDHA’

July 21, 2020

Type Package

Title single-cell Decomposition using Hierarchical Autoencoder

Version 0.1.0

Author Duc Tran, Hung Nguyen, Bang Tran, Carlo La Vecchia, Hung N. Luu and Tin Nguyen

Maintainer Duc Tran <duct@nevada.unr.edu>

Description Provide a fast and accurate pipeline for single-cell analyses.
The scDHA software package conducts cell segregation through unsupervised learning, dimension reduction and visualization, cell classification, and time-trajectory inference.
scDHA currently supports Linux.

License LGPL

Encoding UTF-8

LazyData true

Depends R (>= 3.4)

Imports matrixStats, foreach, keras, tensorflow, doParallel, igraph, Matrix, caret, cluster, clusterCrit, Rcpp, RcppParallel, RcppAnnoy, methods

LinkingTo Rcpp, RcppArmadillo, RcppParallel, RcppAnnoy

RoxygenNote 7.1.0

R topics documented:

scDHA 2
scDHA.class 3
scDHA.pt 3
scDHA.vis 4
scDHA.w 5

Index **6**

scDHA *scDHA*

Description

The main function to perform dimension deduction and clustering.

Usage

```
scDHA(
  data = data,
  k = NULL,
  sparse = F,
  n = 5000,
  ncores = 15L,
  gen_fil = T,
  do.clus = T,
  sample.prob = NULL,
  seed = NULL
)
```

Arguments

data	Gene expression matrix, with rows represent samples and columns represent genes.
k	Number of clusters, leave as default for auto detection. Has no effect when `do.clus = False`.
sparse	Boolen variable indicating whether data is a sparse matrix. The input must be a non negative sparse matrix.
ncores	Number of processor cores to use.
gen_fil	Boolean variable indicating whether to perform scDHA gene filtering before performing dimension deduction and clustering.
do.clus	Boolean variable indicating whether to perform scDHA clustering. If `do.clus = False`, only dimension deduction is performed.
sample.prob	Probability used for classification application only. Leave this parameter as default, no user input is required.
seed	Seed for reproducibility.

Value

List with the following keys:

- cluster - A numeric vector containing cluster assignment for each sample. If `do.clus = False`, this values is always `NULL`.
- latent - A matrix representing compressed data from the input data, with rows represent samples and columns represent latent variables.

scDHA.class — *scDHA classification*

Description

Perform classification of new data based on available data.

Usage

```
scDHA.class(
  train = train,
  train.label = train.label,
  test = test,
  ncores = 15L,
  seed = NULL
)
```

Arguments

`train`	Expression matrix of available data, with rows represent samples and columns represent genes.
`train.label`	A vector containing label for each sample in training data.
`test`	Expression matrix new data for classification, with rows represent samples and columns represent genes.
`ncores`	Number of processor cores to use.
`seed`	Seed for reproducibility.

Value

A vector contain classified labels for new data.

scDHA.pt — *scDHA pseudo time inference*

Description

Inferring pseudo-time data.

Usage

```
scDHA.pt(sc = sc, start.point = 1, ncores = 15L, seed = NULL)
```

Arguments

sc	Embedding object, produced by scDHA function.
start.point	Starting point of the trajectory.
ncores	Number of processor cores to use.
seed	Seed for reproducibility.

Value

List with the following keys:

- pt - Pseudo-time values for each sample.

---

scDHA.vis *scDHA visulization*

Description

Generating 2D embeded data for visulation.

Usage

```
scDHA.vis(sc = sc, ncores = 15L, seed = NULL)
```

Arguments

sc	Embedding object produced by the scDHA function.
ncores	number of processor cores to use.
seed	Seed for reproducibility.

Value

a list with the following keys:

- pred - A matrix representing the 2D projection of single-cell data, where rows represent samples and columns represent latent components.

scDHA.w *scDHA.w*

Description

This function will plot a graph with normalized weights of all genes so user can select the appropriate number of genes to keep.

Usage

```
scDHA.w(data = data, sparse = F, ncores = 15L, seed = NULL)
```

Arguments

data	Gene expression matrix, with rows represent samples and columns represent genes.
sparse	Boolen variable indicating whether data is a sparse matrix. The input must be a non negative sparse matrix.
ncores	Number of processor cores to use.
seed	Seed for reproducibility.

Value

A plot with normalized weights of all genes.

Examples

```
#Generate weight variances for each genes
weight_variance <- scDHA.w(data, seed = 1)

#Plot weight variance for top 5,000 genes
plot(weight_variance, xlab = "Genes", ylab = "Normalized Weight Variance", xlim=c(1, 5000))

#Plot the change of weight variances of top 5,000 genes
weight_variance_change <- weight_variance[-length(weight_variance)] - weight_variance[-1]
plot(weight_variance_change, xlab = "Genes", ylab = "Weight Variance Change", xlim=c(1, 5000))
```

Index

scDHA, 2
scDHA.class, 3
scDHA.pt, 3
scDHA.vis, 4
scDHA.w, 5