Package 'scTSSR2'

January 27, 2021

Type Package

Title Accurate gene expression recovery for single-cell RNA sequencing via fast two-side self-representation

Version 1.1

Author Ke Jin, Xiao-Fei Zhang

Maintainer Ke Jin <kej13@mails.ccnu.edu.cn>

Description An implementation of a fast two-side self-representation prediction and empirical Bayes method to recover the true gene expression profile in noisy and sparse single-cell RNA-seq data.

License GPL-2

Encoding UTF-8

LazyData true

RoxygenNote 6.1.1

Depends R (>= 3.1)

Imports SAVER, keras, tensorflow

Suggests knitr, rmarkdown

VignetteBuilder knitr

R topics documented:

`baron` *Human pancreatic islet data*

Description

This is the Human pancreatic islet dataset (GSM2230757). The raw data contains 20,125 genes and 1,937 cells. Here we use the reference and downsampled datasets generated by Huang et al (2018) which contain 2,284 genes and 1,076 cells (available at https://github.com/mohuangx/SAVER-paper/tree/master/SAVER-data). For details about the approach to generate the reference and downsampled datasets, please refer to Huang et al (2018). This data is an object of class list of length two. count.ref is the reference count matrix and count.samp is the downsampled count matrix.

Usage

```
baron
```

Format

An object of class `list` of length 2.

Author(s)

Ke Jin, <kej13@mails.ccnu.edu.cn>

References

Baron, Maayan, et al (2016). A single-cell transcriptomic map of the human and mouse pancreas reveals inter-and intra-cell population structure. *Cell systems*, 3(4):346-360.

Huang, M. et al. (2018). Saver: gene expression recovery for single-cell rna sequencing. Nat Methods, 15, 539–542.

Zhang, X. F. et al. (2019) EnImpute: imputing dropout events in single cell RNA sequencing data via ensemble learning.

Jin, K. et al. (2020) scTSSR: gene expression recovery for single-cell RNA sequencing using two-side sparse self-representation.

Examples

```
data("baron")
```

`scTSSR2` *use scTSSR2 to impute dropout values in scRNA-seq data*

Description

use scTSSR2 to impute dropout values in scRNA-seq data

Usage

```
scTSSR2(X.count, k.gene = NULL, k.cell = NULL, W = NULL,
  lambda = 256, percent = 0.05, ncores = 1, MAX.ITER = 4,
  ABSTOL = 0.001, learning.rate = 1e-04, epochs = 100,
  verbose = TRUE, estimates.only = FALSE)
```

Arguments

`X.count`	Raw read count matrix. The rows correspond to genes and the columns correspond to cells. Can be sparse.
`k.gene`	A hyper-parameter that controls the sparsity level of the estimated coefficient matrices, A1 and A2. Default is k_gene = min(100, m/30).
`k.cell`	A hyper-parameter that controls the sparsity level of the estimated coefficient matrices, B1 and B2. Default is k_cell = min(100, n/30).
`W`	A weight matrix with element W_gc denotes the non-dropout probability of the expression level of gene g in cell c. Default is W_gc=X_gc/max(X_gc).

lambda	Ridge penalty parameter. Default is 256.
percent	The expression count matrix is preprocessed by filtering out the genes expressed in at most percent*100% of the cells. Default is 0.05.
ncores	Number of cores to use. Default is 1.
MAX.ITER	Maximum iteration of the external circulation of scTSSR2. Default is 4.
ABSTOL	Absolute tolerance of the external circulation. Default is 1e-3.
learning.rate	A hyper-parameter that controls the speed of adjusting the weights of the network with respect to the loss gradient. Default is 0.0001.
epochs	The number of the entire training set going through the entire network. Default is 100.
verbose	Whether to output the value of metrics at the end of each epoch. Default is TRUE.
estimates.only	logical item. A logical flag to determine whether to output only imputed estimates. Default is FALSE.

Value

If 'estimates.only = TRUE', then a matrix of scTSSR2 estimates.

If 'estimates.only = FALSE', a list with the following components

estimate	Recovered (normalized) expression.
se	Standard error of estimates.
info	Information about dataset.

The info element is a list with the following components:

size.factor	Size factor used for normalization.
pred.time	Time taken to generate predictions.
posterior.time	Time taken to compute the posterior distribution.
total.time	Total time for scTSSR2 estimation.

Author(s)

Ke Jin, <kej13@mails.ccnu.edu.cn>

Examples

```
data("baron")

baron_imputation_result = scTSSR2(baron$count.samp)
```

Index