Package 'scTSSR'

September 16, 2019

Type Package

Title Accurate gene expression recovery for single-cell RNA sequencing via bilinear regression

Version 1.1

Author Ke Jin, Xiao-Fei Zhang

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

Description An implementation of a regularized bilinear regression 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

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.

Examples

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

---

scTSSR *use scTSSR to impute dropout values in scRNA-seq data*

Description

use scTSSR to impute dropout values in scRNA-seq data

Usage

```
scTSSR(X_count, lambda = NULL, initA = NULL, initB = NULL,
  percent = 0.1, ncores = 1, MAX_ITER = 4, ABSTOL = 0.001,
  gamma = 1, beta = 0.5, max_iter = 100, abstol = 1e-04,
  penalize_diagonal = FALSE, estimates.only = FALSE)
```

Arguments

`X_count`	An expression count matrix. The rows correspond to genes and the columns correspond to cells. Can be sparse.
`lambda`	Tuning parameter to facilitate feature selection and regularization.
`initA`	The initionlization of A. The elements of A represent the similarities between genes.
`initB`	The initionlization of B. The elements of B represent the similarities between cells.
`percent`	The expression count matrix is preprocessed by filtering out the genes expressed in at most percent*100% of the cells.

ncores	Number of cores to use. Default is 1.
MAX_ITER	Maximum iteration of the external circulation of scTSSR.
ABSTOL	Absolute tolerance of the external circulation.
gamma	The step size.
beta	The line search parameter.
max_iter	Maximum iteration of the accelerated proximal gradient descent algorithm.
abstol	Absolute tolerance of the accelerated proximal gradient descent algorithm.
penalize_diagonal	Whether penalize the diagonal elements of the regression coefficient matrix. Default is FALSE.

Value

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

Author(s)

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

Examples

```
 # data("baron")
# baron_imputation_result = scTSSR(baron$count.samp)
```

Index