Package 'iasva'

October 16, 2019

Type Package

Title Iteratively Adjusted Surrogate Variable Analysis

Version 1.2.0

Date 2018-11-29

Maintainer
Donghyung Lee <Donghyung.Lee@jax.org>, Anthony Cheng <Anthony.Cheng@jax.org>

Description Iteratively Adjusted Surrogate Variable Analysis (IA-SVA) is a statistical framework to uncover hidden sources of variation even when these sources are correlated. IA-SVA provides a flexible methodology to i) identify a hidden factor for unwanted heterogeneity while adjusting for all known factors; ii) test the significance of the putative hidden factor for explaining the unmodeled variation in the data; and iii), if significant, use the estimated factor as an additional known factor in the next iteration to uncover further hidden factors.

Depends R (>= 3.5),

Imports irlba, stats, cluster, graphics, SummarizedExperiment, BiocParallel

License GPL-2

biocViews Preprocessing, QualityControl, BatchEffect, RNASeq, Software, StatisticalMethod, FeatureExtraction, ImmunoOncology

Suggests knitr, testthat, rmarkdown, sva, Rtsne, pheatmap, corrplot, DescTools, RColorBrewer

VignetteBuilder knitr

RoxygenNote 6.0.1

git_url https://git.bioconductor.org/packages/iasva

git_branch RELEASE_3_9

git_last_commit 90c0afd

git_last_commit_date 2019-05-02

Date/Publication 2019-10-15

Author Donghyung Lee [aut, cre],
Anthony Cheng [aut],
Nathan Lawlor [aut],
Duygu Ucar [aut]

R topics documented:

`fast_iasva` *A function for fast IA-SVA*

Description

The iterative procedure of fast IA-SVA is implemented in this function (fast_iasva). fast_iasva() iteratively identifies a hidden factor for unwanted variation while accounting for all known factors, and computes its contribution (i.e., the percentage of unmodeled variation explained by the hidden factor) on the unmodeled variation in the data. If the contribution is greater than a user-defined cutoff (pct.cutoff, default = 1 the next iteration to find further hidden factors.

Usage

```
fast_iasva(Y, X, intercept = TRUE, num.sv = NULL, pct.cutoff = 1,
  num.tsv = NULL, tol = 1e-10, verbose = FALSE)
```

Arguments

`Y`	A SummarizedExperiment class containing read counts where rows represent genes and columns represent samples.
`X`	A design matrix of known variables (e.g., patient ID, gender).
`intercept`	If intercept = FALSE, the linear intercept is not included in the model.
`num.sv`	number of surrogate variables to estimate.
`pct.cutoff`	percetage threshold for SV retention. IA-SVA computes the percentage of unmodeled variance explained by the putative hidden factor and compare it with the user-defined threshold. If the percentage is greater than the threshold, SV is retained.
`num.tsv`	num of top singular values to be used in computing the percentage of unmodeled variation explained by the putative hidden factor. If num.tsv = NULL, all singular values are used.
`tol`	stopping tolerance for the augmented implicitly restarted Lanczos bidiagonalization algorithm
`verbose`	If verbose = TRUE, the function outputs detailed messages.

Value

sv matrix of estimated surrogate variables, one column for each surrogate variable.

pct vector of percentages of unmodeled variance explained by each surrogate variable, one value for each surrogate variable.

n.sv number of obtained surrogate variables.

Examples

```
counts_file <- system.file("extdata", "iasva_counts_test.Rds",
 package = "iasva")
counts <- readRDS(counts_file)
anns_file <- system.file("extdata", "iasva_anns_test.Rds",
 package = "iasva")
 anns <- readRDS(anns_file)
Geo_Lib_Size <- colSums(log(counts + 1))
Patient_ID <- anns$Patient_ID
mod <- model.matrix(~Patient_ID + Geo_Lib_Size)
summ_exp <- SummarizedExperiment::SummarizedExperiment(assays = counts)
iasva.res <- fast_iasva(summ_exp, mod[, -1], num.sv = 5)
```

`find_markers` *A function for finding markers for hidden factors*

Description

Function takes a read counts matrix of entire gene set and a matrix of surrogate variables estimated by IA-SVA as input, identifies marker genes highly correlated with each surrogate variable and returns a read counts matrix of the markers.

Usage

```
find_markers(Y, iasva.sv, method = "BH", sig.cutoff = 0.05,
  rsq.cutoff = 0.3, verbose = FALSE)
```

Arguments

`Y`	A SummarizedExperiment class containing read counts where rows represent genes and columns represent samples.
`iasva.sv`	matrix of estimated surrogate variables, one column for each surrogate variable.
`method`	multiple testing adjustment method, default = "BH".
`sig.cutoff`	significance cutoff.
`rsq.cutoff`	R squared cutoff.
`verbose`	If verbose = TRUE, the function outputs detailed messages.

Value

marker.counts read counts matrix of markers, one column for each cell.

Examples

```
counts_file <- system.file("extdata", "iasva_counts_test.Rds",
 package = "iasva")
counts <- readRDS(counts_file)
anns_file <- system.file("extdata", "iasva_anns_test.Rds",
 package = "iasva")
 anns <- readRDS(anns_file)
Geo_Lib_Size <- colSums(log(counts + 1))
Patient_ID <- anns$Patient_ID
```

```
mod <- model.matrix(~Patient_ID + Geo_Lib_Size)
summ_exp <- SummarizedExperiment::SummarizedExperiment(assays = counts)
iasva.res <- iasva(summ_exp, mod[, -1], num.sv = 5, permute = FALSE)
markers <- find_markers(summ_exp, iasva.res$sv)
```

`iasva` *A function for iteratively adjusted surrogate variable analysis (IA-SVA)*

Description

The iterative procedure of IA-SVA is implemented in this function (iasva). iasva() function iteratively runs iasva_unit() function to identify a hidden factor for unwanted variation while accounting for all known factors and test the significance of its contribution on the unmodeled variation in the data. If the test statistic of detected factor is significant, iasva() includes the factor as a known variable in the next iteration to find further hidden factors.

Usage

```
iasva(Y, X, intercept = TRUE, num.sv = NULL, permute = TRUE,
  num.p = 100, sig.cutoff = 0.05, threads = 1, num.sv.permtest = NULL,
  tol = 1e-10, verbose = FALSE)
```

Arguments

`Y`	A SummarizedExperiment class containing read counts where rows represent genes and columns represent samples.
`X`	A model matrix of known variables including the primary variables of interest.
`intercept`	If intercept = FALSE, the linear intercept is not included in the model.
`num.sv`	number of surrogate variables to estimate.
`permute`	If permute = TRUE, a permutation test (Buja and Eyuboglu 1992, Leek and Storey 2008) is conducted to assess the significance of the putative hidden factor.
`num.p`	number of permutations to be used to calculate the permuation test p-value.
`sig.cutoff`	significance threshold for the permutation test
`threads`	number of cores to be used in permutation test.
`num.sv.permtest`	num of top singular values to be used in computing the permutation test statistic. If num.sv.permtest = NULL, all singular values are used.
`tol`	stopping tolerance for the augmented implicitly restarted Lanczos bidiagonalization algorithm
`verbose`	If verbose=TRUE, the function outputs detailed messages.

Value

sv matrix of estimated surrogate variables, one column for each surrogate variable.

pc.stat.obs vector of PC test statistic values, one value for each surrogate variable.

pval vector of permuation p-values, one value for each surrogate variable.

n.sv number of significant/obtained surrogate variables.

Examples

```
counts_file <- system.file("extdata", "iasva_counts_test.Rds",
 package = "iasva")
counts <- readRDS(counts_file)
anns_file <- system.file("extdata", "iasva_anns_test.Rds",
 package = "iasva")
 anns <- readRDS(anns_file)
Geo_Lib_Size <- colSums(log(counts + 1))
Patient_ID <- anns$Patient_ID
mod <- model.matrix(~Patient_ID + Geo_Lib_Size)
summ_exp <- SummarizedExperiment::SummarizedExperiment(assays = counts)
iasva.res<- iasva(summ_exp, mod[, -1],verbose = FALSE,
 permute = FALSE, num.sv = 5)
```

---

study_R2 *A function to study different values of R2*

Description

study_R2() studies how different R2 thresholds is changing: 1) number of marker genes; 2) clustering quality (assuming number of clusters is known). It generated diagnostic plots that shows how selected genes and clustering quality changes as a function of R2 threshold.

Usage

```
study_R2(Y, iasva.sv, selected.svs = 2, no.clusters = 2, verbose = FALSE)
```

Arguments

Y	A SummarizedExperiment class containing read counts where rows represent genes and columns represent samples.
iasva.sv	matrix of estimated surrogate variables, one column for each surrogate variable.
selected.svs	list of SVs that are selected for the analyses. Default is SV2
no.clusters	No of clusters to be used in the analyses. Default is 2.
verbose	If verbose = TRUE, the function outputs detailed messages.

Value

a summary plot that represents silhoutte index and marker gene counts as a function of R2 and corresponding matrices.

Examples

```
counts_file <- system.file("extdata", "iasva_counts_test.Rds",
 package = "iasva")
counts <- readRDS(counts_file)
anns_file <- system.file("extdata", "iasva_anns_test.Rds",
 package = "iasva")
 anns <- readRDS(anns_file)
Geo_Lib_Size <- colSums(log(counts + 1))
Patient_ID <- anns$Patient_ID
mod <- model.matrix(~Patient_ID + Geo_Lib_Size)
```

```
summ_exp <- SummarizedExperiment::SummarizedExperiment(assays = counts)
iasva.res<- iasva(summ_exp, mod[, -1],verbose = FALSE,
permute = FALSE, num.sv = 5)
iasva.sv <- iasva.res$sv
study_res <- study_R2(summ_exp, iasva.sv)
```

Index