# Package 'FCBF'

April 15, 2019

**Type** Package

**Title** Fast Correlation Based Filter for Feature Selection

**Version** 1.0.1

**Maintainer** Tiago Lubiana <tiago.lubiana.alves@usp.br>

**Description** This package provides a simple R implementation for the Fast Correlation Based Filter described in Yu, L. and Liu, H.; Feature Selection for High-Dimensional Data: A Fast Correlation Based Filter Solution,Proc. 20th Intl. Conf. Mach. Learn. (ICML-2003), Washington DC, 2003
The current package is an intent to make easier for bioinformaticians to use FCBF for feature selection, especially regarding transcriptomic data.This implies discretizing expression (function discretize_exprs) before calculating the features that explain the class, but are not predictable by other features.
The functions are implemented based on the algorithm of Yu and Liu, 2003 and Rajarshi Guha's implementation from 13/05/2005 available (as of 26/08/2018) at http://www.rguha.net/code/R/fcbf.R .

**License** MIT + file LICENSE

**Encoding** UTF-8

**LazyData** False

**RoxygenNote** 6.1.0.9000

**Imports** ggplot2, gridExtra, SummarizedExperiment

**Suggests** caret, mlbench, SingleCellExperiment, knitr, rmarkdown, testthat, BiocManager

**biocViews** ImmunoOncology, GeneTarget, FeatureExtraction, Classification, GeneExpression, SingleCell

**VignetteBuilder** knitr

**Depends** R (>= 3.5.0)

**git_url** https://git.bioconductor.org/packages/FCBF

**git_branch** RELEASE_3_8

**git_last_commit** 2e9e337

**git_last_commit_date** 2019-01-04

**Date/Publication** 2019-04-15

**Author** Tiago Lubiana [aut, cre], Helder Nakaya [aut, ths]

# R topics documented:



---

`discretize_exprs` *discretize_exprs Simple discretizing of gene expression*

---

## Description

This function takes the range of values for each gene in a previously normalized expression table (genes/variables in rows, samples/observations in columns) and uses it for a width-based discretization. Each feature is divide into "n" bins of equal width. The first bin is attributed the class 'low' and the next bins are assigned to "high". It transposes the original expression table.

## Usage

```
discretize_exprs(expression_table, number_of_bins = 3)
```

## Arguments

| | |
|---|---|
| `expression_table` | A previously normalized expression table Note: this might drastically change the number of selected features. |
| `number_of_bins` | Number of equal-width bins for discretization. Note: it is a binary discretization, with the first bin becoming one class ('low') and the other bins, another class ('high'). Defaults to 3. |

## Value

A data frame with the discretized features in the same order as previously

## Examples

```
data(scDengue)
exprs <- SummarizedExperiment::assay(scDengue, 'logcounts')
discrete_expression <- as.data.frame(discretize_exprs(exprs))
head(discrete_expression[,1:4])
```

fcbf *Fast Correlation Based Filter function.*

## Description

This functions allows selection of variables from a feature table of discrete/categorial variables and a target class. The function is based on the algorithm described in Yu, L. and Liu, H.; Feature Selection for High-Dimensional Data A Fast Correlation Based Filter Solution, Proc. 20th Intl. Conf. Mach. Learn. (ICML-2003), Washington DC, 2003

## Usage

```
fcbf(x, y, thresh = 0.25, verbose = FALSE, samples_in_rows = FALSE)
```

## Arguments

| | |
|---|---|
| x | A table of features (samples in rows, variables in columns, and each observation in each cell) |
| y | A target vector, factor containing classes of the observations. Note: the observations must be in the same order as the parameter x |
| thresh | A threshold for the minimum correlation (as determined by symettrical uncertainty) between each variable and the class. Defaults to 0.25. |
| verbose | Adds verbosity. Defaults to FALSE. |
| samples_in_rows | A flag for the case in which samples are in rows and variables/genes in columns. Defaults to FALSE. Note: this might drastically change the number of selected features. |

## Details

Obs: For gene expression, you will need to run discretize_exprs first

## Value

Returns a data frame with the selected features index (first row) and their symmetrical uncertainty values regarding the class (second row). Variable names are present in rownames

## Examples

```
data(scDengue)
exprs <- SummarizedExperiment::assay(scDengue, 'logcounts')
discrete_expression <- as.data.frame(discretize_exprs(exprs))
head(discrete_expression[,1:4])
infection <- SummarizedExperiment::colData(scDengue)
target <- infection$infection
fcbf(discrete_expression,target, thresh = 0.05, verbose = TRUE)
```

scDengue *Dengue infected macrophages; gene expression data from GEO study GSE110496*

## Description

Expression data from single cells, from adengue virus infection study by Zanini et al, #' 2018. The expression was filtered to get cells 12 hours after infection with #' a multiplicity of infection (moi) of 1 (dengue) or uninfected(ctrl). Gene counts were normalized via Bioconductor package "SCNorm".

## Usage

```
data(scDengue)
```

## Format

An object of class `SingleCellExperiment`

## Details

Gene expression has to be discretized for use in FCBF.

## Source

GEO

## References

Zanini, F., Pu, S. Y., Bekerman, E., Einav, S., & Quake, S. R. (2018). Single-cell transcriptional dynamics of flavivirus infection. Elife, 7, e32942. PubMed

## Examples

```
data(scDengue)
data(scDengue)
exprs <- SummarizedExperiment::assay(scDengue, 'logcounts')
infection <- SummarizedExperiment::colData(scDengue)
target <- infection$infection
# Discretize gene expression
discrete_expression <- as.data.frame(discretize_exprs(exprs))
fcbf_features <- fcbf(discrete_expression,
                              target,
                              thresh = 0.05,
                              verbose = TRUE)
```

su_plot *Symmetrical Uncertainty diagnostic*

## Description

This functions runs symmetrical uncertainty for a feature table and a class, returning an histogram of the scores

## Usage

```
su_plot(x, y)
```

## Arguments

| | |
|---|---|
| x | A table of features (observations in rows, variables in columns) |
| y | A target vector, factor containing classes of the observations. Note: the observations must be in the same order as the parameter x. |

## Value

Plots an histogram of symmetrical uncertainty values regarding the class.

## Examples

```
data(scDengue)
exprs <- SummarizedExperiment::assay(scDengue, 'logcounts')
discrete_expression <- as.data.frame(discretize_exprs(exprs))
infection <- SummarizedExperiment::colData(scDengue)
target <- infection$infection
su_plot(discrete_expression,target)
```

# Index