

Package ‘EMeth’

May 18, 2020

Type Package

Title Cell type decomposition based on methylation data

Version 1.0

Date 2020-01-29

Author Hanyu Zhang

Maintainer Hanyu Zhang <hanyuz6@uw.edu>

Description This is a package for implementing EMeth algorithm. It provide two families of EMeth ("normal" and "laplace") and also has a wrapper for cross-validation of tuning parameters for the penalized likelihood method.

License GPL (>= 2)

R topics documented:

EMeth-package	1
cv.emeth	2
emeth	3

Index	5
--------------	----------

EMeth-package	<i>Cell type decomposition based on methylation data</i>
---------------	--

Description

This is a package for implementing EMeth algorithm. It provide two families of EMeth ("normal" and "laplace") and also has a wrapper for cross-validation of tuning parameters for the penalized likelihood method.

Details

The DESCRIPTION file: This package was not yet installed at build time.

Index: This package was not yet installed at build time.

~~ This package provides implementations of the EMeth algorithm. It contains two important functions. - emeth, which provides implementation of two families of EMeth, normal and laplace, according to what likelihood is used. Generally we recommend the use of laplace family. - cv.emeth, which helps tuning the ridge penalty by cross validation ~~

Author(s)

Hanyu Zhang

Maintainer: Hanyu Zhang <hanyuz6@uw.edu>

Examples

Examples can be found in the "example" subdirectory.

cv.emeth

cv.emeth: cross validation for emeth.

Description

An implementation of cross validation for the ridge penalty parameter in EMeth.

Usage

```
cv.emeth(Y, eta, mu, aber, V, init = "default", nu, family = "laplace", folds = 5, usesubset = TRUE)
```

Arguments

Y	Matrix of size K*I. Methylation of bulk sample for which the cell type decomposition is to be estimated. Y is usually a K*I matrix where K is the number of probes used and I is the number of samples.
eta	Vector of size I. Tumor purity of each sample.
mu	Matrix of size K*Q. Reference matrix, provided by literature. A sample reference data in the example directory is provided, that is obtained from six previous literatures.
aber	Logic variable: if there is unknown aberrant cell type.
V	string: default to be 'c' which stands for constant weight for all probes. It might be 'b' for binomial variance structure or 'w' for specific weight structure of variance.
init	If init is a string 'default', we will adopt the default random initialization of all parameters for the algorithm. Otherwise one can provide a list of initialization of parameters.
family	string: accept 'normal' or 'laplace' to specify what likelihood will be used in the algorithm.
nu	a nonnegative number that stands for the penalty.
folds	Specify the number of folds for the cross validation.
usesubset	Logic variable, if it is true, a random sampled subset of all probes are used to perform cross-validation.
maxiter	max time of iteration of the EM algorithm, default to be 50.
verbose	logic variable. If TRUE then will print additional information in iteration of EMeth.

Value

result	The result of EMeth algorithm after fitting the best selected parameter.
choosenu	The chosen parameter of nu.
losslist	A matrix saving the loss for each fold and each choice of nu.

Author(s)

Hanyu Zhang

Examples

```
## See examples folder.
```

emeth	<i>emeth: cell type decomposition from DNA methylation data based on EM-type algorithm and penalized likelihood maximization.</i>
-------	---

Description

This function is an implementation of EMeth algorithm. It provides estimation of cell type proportion for each sample in bulk tissue samples Y. It needs a reference matrix mu to conduct the estimation. EMeth has two families, normal and laplace, depending on what likelihood is used. It also has a penalization term (ridge penalty).

Usage

```
emeth(Y, eta, mu, aber, V, init = "default", family = "laplace", nu = 0, maxiter = 50, verbose = FALSE)
```

Arguments

Y	Matrix of size K*I. Methylation of bulk sample for which the cell type decomposition is to be estimated. Y is usually a K*I matrix where K is the number of probes used and I is the number of samples.
eta	Vector of size I. Tumor purity of each sample.
mu	Matrix of size K*Q. Reference matrix, provided by literature. A sample reference data in the example directory is provided, that is obtained from six previous literatures.
aber	Logic variable: if there is unknown aberrant cell type.
V	string: default to be 'c' which stands for constant weight for all probes. It might be 'b' for binomial variance structure or 'w' for specific weight structure of variance.
init	If init is a string 'default', we will adopt the default random initialization of all parameters for the algorithm. Otherwise one can provide a list of initialization of parameters.
family	string: accept 'normal' or 'laplace' to specify what likelihood will be used in the algorithm.
nu	a nonnegative number that stands for the penalty.
maxiter	max time of iteration of the EM algorithm, default to be 50.
verbose	logic variable. If TRUE then will print additional information in iteration of EMeth.

Value

rho	estimated proportion of cell types in each sample.
sigma_c	estimated variance.
lambda	estimated ratio between aberrant variance and consistent variance.
nu0	estimated methylation level of unknown cell type.
pi_a	estimated probability in each sample for probes to be aberrant.
gamma	estimated probability of each probe in each sample to be aberrant
weights	the weights used in the iteration of EMeth, this is not meaningful in current LaplaceEM and original Normal EM
iter	Number of iterations

Author(s)

Hanyu Zhang

Examples

See examples folder

Index

`cv.emeth`, [2](#)

EMeth (EMeth-package), [1](#)

`emeth`, [3](#)

EMeth-package, [1](#)