Package 'RefFreeEWAS'

December 14, 2018

Version 2.2

Date 2018-12-14

Title EWAS using Reference-Free DNA Methylation Mixture Deconvolution

Author E. Andres Houseman, Sc.D.

Maintainer E. Andres Houseman <eahouseman@gmail.com>

Depends R (>= 3.2.2), quadprog

Description Reference-free method for conducting EWAS while deconvoluting DNA methylation arising as mixtures of cell types. The older method (Houseman et al., 2014,<doi:10.1093/bioinformatics/btu029>) is similar to surrogate variable analysis (SVA and ISVA), except that it makes additional use of a biological mixture assumption. The newer method (Houseman et al., 2016, <doi:10.1186/s12859-016-1140-4>) is similar to non-negative matrix factorization, with additional constraints and additional utilities.

License GPL (>= 2)

NeedsCompilation no

Repository CRAN

Date/Publication 2018-12-14 06:00:03 UTC

R topics documented:

`BootOneRefFreeEwasModel`

One Bootstrap sample for Reference-Free EWAS Model

Description

Bootstrap generation procedure for reference-free method for conducting EWAS while deconvoluting DNA methylation arising as mixtures of cell types.

Usage

```
BootOneRefFreeEwasModel(mod)
```

Arguments

`mod`	model object of class RefFreeEwasModel (generated with smallOutput=FALSE).

Details

Generates one bootstrapped data set for the reference-free method for conducting EWAS while deconvoluting DNA methylation arising as mixtures of cell types. Typically not run by user.

Value

A matrix representing a bootstrap sample of an DNA methylation assay matrix.

Author(s)

E. Andres Houseman

References

Houseman EA, Molitor J, and Marsit CJ (2013), Reference-Free Cell Mixture Adjustments in Analysis of DNA Methylation Data. Currently a tech report, in revision for publication.

See Also

BootRefFreeEwasModel

BootRefFreeEwasModel *Bootstrap for Reference-Free EWAS Model*

Description

Bootstrap procedure for reference-free method for conducting EWAS while deconvoluting DNA methylation arising as mixtures of cell types.

Usage

```
BootRefFreeEwasModel(mod, nboot)
```

Arguments

mod	model object of class RefFreeEwasModel (generated with smallOutput=FALSE).
nboot	Number of bootstrap samples to generate

Details

Generates the bootstrap samples for the reference-free method for conducting EWAS while deconvoluting DNA methylation arising as mixtures of cell types.

Value

An array object of class "BootRefFreeEwasModel". Bootstraps are generated for both Beta and Bstar.

Author(s)

E. Andres Houseman

References

Houseman EA, Molitor J, and Marsit CJ (2014), Reference-Free Cell Mixture Adjustments in Analysis of DNA Methylation Data. Bioinformatics, doi: 10.1093/bioinformatics/btu029.

See Also

RefFreeEwasModel

Examples

```
data(RefFreeEWAS)

## Not run:
  tmpDesign <- cbind(1, rfEwasExampleCovariate)
  tmpBstar <- (rfEwasExampleBetaValues

  EstDimRMT(rfEwasExampleBetaValues-tmpBstar

## End(Not run)

test <- RefFreeEwasModel(
  rfEwasExampleBetaValues,
  cbind(1,rfEwasExampleCovariate),
  4)

testBoot <- BootRefFreeEwasModel(test,10)
summary(testBoot)
```

---

bootstrapPairs *One Bootstrap Sample for Pairs*

Description

Bootstrap generation procedure for sampling paired data (e.g. twin data)

Usage

```
bootstrapPairs(obs, pairID)
```

Arguments

obs	Observation ids (numeric vector).
pairID	Pair IDs (one unique value per pair).

Details

Generates one bootstrapped set of ids corresponding to pairs for the method for conducting EWAS while deconvoluting DNA methylation arising as mixtures of cell types. Typically not run by user.

Value

A vector of IDs corresponding to bootstrapped pairs

Author(s)

E. Andres Houseman

References

Houseman EA, Molitor J, and Marsit CJ (Bioinformatics, 2014).

See Also

BootRefFreeEwasModel,PairsBootRefFreeEwasModel

`deviance.RefFreeCellMix`
deviance.RefFreeCellMix

Description

Deviance method for objects of type RefFreeCellMix.

Usage

```
## S3 method for class 'RefFreeCellMix'
deviance(object, Y, Y.oob=NULL, EPSILON=1E-9,
  bootstrapIterations=0, bootstrapIndices=NULL, ...)
```

Arguments

`object`	RefFreeCellMix object to summarize
`Y`	Methylation matrix on which x was based
`Y.oob`	Alternate ("out-of-box") methylation matrix for which to calculate deviance, based on x
`EPSILON`	Minimum value of variance (zero variances will be reset to this value)
`bootstrapIterations`	Number of RefFreeCellMix iterations to use in bootstrap (see details)
`bootstrapIndices`	Bootstrap indices (see details)
`...`	(Unused).

Details

Deviance based on normal distribution applied to errors of Y after accounting for cell mixture effect, Mu Omega^T. Since RefFreeCellMix does not save the original data Y in the resulting object x, Y must be supplied here. However, deviance may be calculated for an alternative "out-of-bag" methylation matrix, Y.oob. If bootstrapIterations=0, this is what is done. If bootstrapIterations>0, then x$Mu is used to initialize a new value of x via RefFreeCellMix executed on a bootstrap sample of Y with the number of indicated iterations. If bootstrapIndices is provided, the bootstrap will be based on these indices, otherwise the indices will be sampled randomly with replacement from 1:ncol(Y). See `RefFreeCellMix` for example.

`EstDimIC` *Dimension estimation by AIC and BIC*

Description

Method for estimating latent dimension by AIC and BIC.

Usage

```
EstDimIC(Rmat,Krange=0:25)
```

Arguments

`Rmat`	Residual matrix for which to estimate latent dimension.
`Krange`	Vector of integers representing candidate dimensions to consider

Details

Method for estimating latent dimension by AIC and BIC. Inferior to the RMT method in the isva package, but it appears here because it's mentioned in our paper.

Value

A list containing AIC and BIC for candidate dimensions, as well as the best dimension for each.

Author(s)

E. Andres Houseman

References

HOUSEMAN, Eugene Andres, MOLITOR, John, et MARSIT, Carmen J. Reference-free cell mixture adjustments in analysis of DNA methylation data. Bioinformatics, 2014, vol. 30, no 10, p. 1431-1439.

See Also

`EstDimRMT`

Examples

```
data(RefFreeEWAS)

## Not run:
  tmpDesign <- cbind(1, rfEwasExampleCovariate)
  tmpBstar <- rfEwasExampleBetaValues

  EstDimIC(rfEwasExampleBetaValues-tmpBstar

## End(Not run)
```

`EstDimRMT` *Dimension estimation by Random Matrix Theory*

Description

Method for estimating latent dimension by Random Matrix Theory.

Usage

```
EstDimRMT(Rmat)
```

Arguments

`Rmat`	Residual matrix for which to estimate latent dimension.

Details

Method for estimating latent dimension by Random Matrix Theory. This function originated in the package isva, authored by A. Teschendorff. Previous versions of RefFreeEWAS used the isva version of the function. However, because of dependency issues in that package, the present version of RefFreeEWAS simply reproduces the function found in version 1.9 of isva and removes the dependency on the isva package. Documentation from isva: Given a data matrix, it estimates the number of significant components of variation by comparing the observed distribution of spectral eigenvalues to the theoretical one under a Gaussian Orthogonal Ensemble (GOE). Specifically, a spectral decomposition of the data covariance matrix is performed and the number of eigenvalues larger than the theoretical maximum predicted by the GOE is taken as an estimate of the number of significant components.

Value

A list with following objects:

`cor`	Data covariance matrix.
`dim`	Estimated intrinsic dimensionality of data.
`estdens`	Empirical density of eigenvalues.
`thdens`	Theoretical density of eigenvalues.

Author(s)

E. Andres Houseman

References

1. Random matrix approach to cross correlations in financial data. Plerou et al. Physical Review E (2002), Vol.65.
2. Independent Surrogate Variable Analysis to deconvolve confounding factors in large-scale microarray profiling studies. Teschendorff AE, Zhuang JJ, Widschwendter M. Bioinformatics. 2011 Jun 1;27(11):1496-505.

Examples

```
data(RefFreeEWAS)

## Not run:
  tmpDesign <- cbind(1, rfEwasExampleCovariate)
  tmpBstar <- rfEwasExampleBetaValues
  EstDimRMT(rfEwasExampleBetaValues-tmpBstar

## End(Not run)
```

ImputeByMean *Simple imputation method based on row-mean*

Description

Simple method for imputing missing values by row-mean

Usage

```
ImputeByMean(Y)
```

Arguments

Y	Matrix to impute.

Value

Matrix with missing values replaced by imputed values

omnibusBoot *Bootstrap-based omnibus test of significance across all features*

Description

Support for bootstrap-based omnibus test of significance accounting for correlation.

Usage

```
omnibusBoot(est, boots, denDegFree)
```

Arguments

est	Vector of m estimates, one for each of m features.
boots	Matrix (m x R) of bootstrap samples corresponding to the estimates
denDegFree	Single number representing the denominator degrees-of-freedom for computing p-values

Details

Returns one omnibus p-value based on Kolmogorov-Smirnov distance from a uniform distribution

Value

A single number representing the p-value for the omnibus test over all features.

Author(s)

E. Andres Houseman

References

Houseman EA, Molitor J, and Marsit CJ (2014), Reference-Free Cell Mixture Adjustments in Analysis of DNA Methylation Data. Bioinformatics, doi: 10.1093/bioinformatics/btu029.

See Also

RefFreeEwasModel

Examples

```
data(RefFreeEWAS)

test <- RefFreeEwasModel(
  rfEwasExampleBetaValues,
  cbind(1,rfEwasExampleCovariate),
  4)
```

```
testBoot <- BootRefFreeEwasModel(test,10)
summary(testBoot)
omnibusBoot(test$Beta[,2], testBoot[,2,"B",],-diff(dim(test$X)))
omnibusBoot(test$Bstar[,2], testBoot[,2,"B*",],-diff(dim(test$X)))
```

`PairsBootOneRefFreeEwasModel` *One Bootstrap Sample for Reference-Free EWAS Model, Accounting for Paired Data*

Description

Bootstrap generation procedure for reference-free method for conducting EWAS while deconvoluting DNA methylation arising as mixtures of cell types. This version accounts for paired data (e.g. twin data)

Usage

```
PairsBootOneRefFreeEwasModel(mod, pairID)
```

Arguments

`mod`	model object of class RefFreeEwasModel (generated with smallOutput=FALSE).
`pairID`	Pair IDs (one unique value per pair).

Details

Generates one bootstrapped data set for the reference-free method for conducting EWAS while deconvoluting DNA methylation arising as mixtures of cell types. This version facilitates the estimation of robust standard errors to account for paired data (e.g. twin data) using a strategy similar to that employed by Generalized Estimating Equations (GEEs). Specifically, in bootstrapping the errors, the pairs are sampled rather than individual arrays. Typically not run by user.

Value

A matrix representing a bootstrap sample of an DNA methylation assay matrix.

Author(s)

E. Andres Houseman

References

Houseman EA, Molitor J, and Marsit CJ (Bioinformatics,2014), Reference-Free Cell Mixture Adjustments in Analysis of DNA Methylation Data. Bioinformatics, doi: 10.1093/bioinformatics/btu029.

See Also

BootRefFreeEwasModel,BootOneRefFreeEwasModel

PairsBootRefFreeEwasModel

Bootstrap for Reference-Free EWAS Model, Accounting for Paired Data

---

Description

Bootstrap procedure for reference-free method for conducting EWAS while deconvoluting DNA methylation arising as mixtures of cell types. This version accounts for paired data (e.g. twin data)

Usage

```
PairsBootRefFreeEwasModel(mod, nboot, pairID)
```

Arguments

mod	model object of class RefFreeEwasModel (generated with smallOutput=FALSE).
nboot	Number of bootstrap samples to generate.
pairID	Pair IDs (one unique value per pair).

Details

Generates the bootstrap samples for the reference-free method for conducting EWAS while deconvoluting DNA methylation arising as mixtures of cell types. This paired version facilitates the estimation of robust standard errors to account for paired data (e.g. twin data) using a strategy similar to that employed by Generalized Estimating Equations (GEEs). Specifically, in bootstrapping the errors, the pairs are sampled rather than individual arrays. An error will be generated unless each cluster has exactly two members (i.e. exactly two observations correspond to the same unique ID given in pairID).

Value

An array object of class "BootRefFreeEwasModel". Bootstraps are generated for both Beta and Bstar.

Author(s)

E. Andres Houseman

References

Houseman EA, Molitor J, and Marsit CJ (Bioinformatics, 2014), Reference-Free Cell Mixture Adjustments in Analysis of DNA Methylation Data. Bioinformatics, doi: 10.1093/bioinformatics/btu029.

See Also

RefFreeEwasModel,BootRefFreeEwasModel

Examples

```
data(RefFreeEWAS)

## Not run:
  tmpDesign <- cbind(1, rfEwasExampleCovariate)
  tmpBstar <- (rfEwasExampleBetaValues

  EstDimRMT(rfEwasExampleBetaValues-tmpBstar

## End(Not run)

test <- RefFreeEwasModel(
  rfEwasExampleBetaValues,
  cbind(1,rfEwasExampleCovariate),
  4)

testBoot <- BootRefFreeEwasModel(test,10)
summary(testBoot)
```

---

print.BootRefFreeEwasModel

print.BootRefFreeEwasModel

Description

Print method for objects of type BootRefFreeEwasModel

Usage

```
## S3 method for class 'BootRefFreeEwasModel'
print(x,...)
```

Arguments

x	BootRefFreeEwasModel object to print
...	(Unused).

Details

See RefFreeEwasModel for example.

print.RefFreeCellMix *print.RefFreeCellMix*

Description

Print method for objects of type RefFreeCellMix

Usage

```
## S3 method for class 'RefFreeCellMix'
print(x,...)
```

Arguments

x	RefFreeCellMix object to print
...	(Unused).

Details

See RefFreeCellMix for example.

print.RefFreeEwasModel *print.RefFreeEwasModel*

Description

Print method for objects of type RefFreeEwasModel

Usage

```
## S3 method for class 'RefFreeEwasModel'
print(x,...)
```

Arguments

x	RefFreeEwasModel object to print
...	(Unused).

Details

See RefFreeEwasModel for example.

print.summaryBootRefFreeEwasModel *print.summaryBootRefFreeEwasModel*

Description

Print method for objects of type summaryBootRefFreeEwasModel

Usage

```
## S3 method for class 'summaryBootRefFreeEwasModel'
print(x,...)
```

Arguments

x	summaryBootRefFreeEwasModel object to print
...	(Unused).

Details

See RefFreeEwasModel for example.

projectMix *Cell Mixture Projection (reference-based)*

Description

Constrained linear projection for estimating cell mixture or related coefficients.

Usage

```
projectMix(Y, Xmat, nonnegative=TRUE, sumLessThanOne=TRUE, lessThanOne=!sumLessThanOne)
```

Arguments

Y	Matrix (m CpGs x n Subjects) of DNA methylation beta values
Xmat	Matrix (m CpGs x K cell types) of cell-type specific methylomes
nonnegative	All coefficients >=0?
sumLessThanOne	Coefficient rows should sum to less than one?
lessThanOne	Every value should be less than one (but possibly sum to value greater than one)?

Details

Function for projecting methylation values (Y) onto space of methyomes (Xmat), with various constraints. This is the reference-based method described in Houseman et al. (2012) and also appearing in the minfi package.

Value

Projection coefficients resulting from constrained projection

Author(s)

E. Andres Houseman

References

Houseman EA, Accomando WP et al. DNA methylation arrays as surrogate measures of cell mixture distribution, BMC Bioinformatics, 2012.

`RefFreeCellMix` *Reference-Free Cell Mixture Projection*

Description

Reference-free cell-mixture decomposition of DNA methylation data set

Usage

```
RefFreeCellMix(Y,mu0=NULL,K=NULL,iters=10,Yfinal=NULL,verbose=TRUE)
```

Arguments

`Y`	Matrix (m CpGs x n Subjects) of DNA methylation beta values
`mu0`	Matrix (m CpGs x K cell types) of *initial* cell-type specific methylomes
`K`	Number of cell types (ignored if mu0 is provided)
`iters`	Number of iterations to execute
`Yfinal`	Matrix (m* CpGs x n Subjects) of DNA methylation beta values on which to base final methylomes
`verbose`	Report summary of errors after each iteration?

Details

Reference-free decomposition of DNA methylation matrix into cell-type distributions and cell-type methylomes, Y = Mu Omega^T. Either an initial estimate of Mu must be provided, or else the number of cell types K, in which case RefFreeCellMixInitialize will be used to initialize. Note that the decomposition will be based on Y, but Yfinal (=Y by default) will be used to determine the final value of Mu based on the last iterated value of Omega.

Value

Object of S3 class RefFreeCellMix, containing the last iteration of Mu and Omega.

Author(s)

E. Andres Houseman

References

Houseman, E. Andres, Kile, Molly L., Christiani, David C., et al. Reference-free deconvolution of DNA methylation data and mediation by cell composition effects. BMC bioinformatics, 2016, vol. 17, no 1, p. 259.

See Also

RefFreeCellMixInitialize

Examples

```
data(HNSCC)

# Typical use
Y.shortTest <- Y.HNSCC.averageBetas[1:500,]
Y.shortTest.final <- Y.HNSCC.averageBetas[1:1000,]
testArray1  <- RefFreeCellMixArray(Y.shortTest,Klist=1:3,iters=5,Yfinal=Y.shortTest.final)
testArray1
lapply(testArray1,summary)
sapply(testArray1,deviance,Y=Y.shortTest.final)

# Example with explicit initialization
testKeq2  <- RefFreeCellMix(Y.shortTest, mu0=RefFreeCellMixInitialize(Y.shortTest,K=2))
testKeq2
head(testKeq2$Mu)
head(testKeq2$Omega)
```

RefFreeCellMixArray *Initialize Reference-Free Cell Mixture Projection*

Description

Array of reference-free cell-mixture decompositions of a DNA methylation data set

Usage

```
RefFreeCellMixArray(Y,Klist=1:5,iters=10,Yfinal=NULL,verbose=FALSE,
   dist.method = "euclidean",...)
```

Arguments

Y	Matrix (m CpGs x n Subjects) of DNA methylation beta values
Klist	List of K values (each K = assumed number of cell types)
iters	Number of iterations to execute for each value of K
Yfinal	Matrix (m* CpGs x n Subjects) of DNA methylation beta values on which to base final methylomes
verbose	Report summary of errors after each iteration for each fit?
dist.method	Method for calculating distance matrix for methylome initialization
...	Additional parameters for hclust function for methylome initialization

Details

List of Reference-free decompositions for a range of K values. For each value of K, the decomposition is initialized by hierarchical clutering as specified by the parameters dist.method, etc. Note that for each K, the decomposition will be based on Y, but Yfinal (=Y by default) will be used to determine the final value of Mu based on the last iterated value of Omega.

Value

List, each element is an object of S3 class RefFreeCellMix, containing the last iteration of Mu and Omega.

Author(s)

E. Andres Houseman

References

Houseman, E. Andres, Kile, Molly L., Christiani, David C., et al. Reference-free deconvolution of DNA methylation data and mediation by cell composition effects. BMC bioinformatics, 2016, vol. 17, no 1, p. 259.

See Also

RefFreeCellMix

Examples

```
data(HNSCC)
Y.shortTest <- Y.HNSCC.averageBetas[1:500,]
testArray2  <- RefFreeCellMixArray(Y.shortTest,Klist=1:5,iters=5)
sapply(testArray2,deviance,Y=Y.shortTest)

## Not run:
testBootDevs <- RefFreeCellMixArrayDevianceBoots(testArray2,Y.shortTest,R=10)

testBootDevs
apply(testBootDevs[-1,],2,mean,trim=0.25)
```

```
which.min(apply(testBootDevs[-1,],2,mean,trim=0.25))

## End(Not run)
```

RefFreeCellMixArrayDevianceBoot

RefFreeCellMixArrayDevianceBoot

Description

Vector of bootstrapped deviances corresponding to an array of reference-free cell-mixture decompositions

Usage

```
RefFreeCellMixArrayDevianceBoot(rfArray, Y, EPSILON=1E-9, bootstrapIterations=5)
```

Arguments

rfArray	list of RefFreeCellMix objects (e.g. from RefFreeCellMixArray)
Y	Methylation matrix on which x was based
EPSILON	Minimum value of variance (zero variances will be reset to this value)
bootstrapIterations	Number of RefFreeCellMix iterations to use in bootstrap

Details

Vector of bootstrapped deviances corresponding to an array of reference-free cell-mixture decompositions, used to determine optimal number of cell types. This function returns one bootstrapped vector. See RefFreeCellMixArrayDevianceBoots for more than one bootstrapped vector. The bootstrapped deviance is based on normal distribution applied to errors of Y after accounting for cell mixture effect, Mu Omega^T. See RefFreeCellMixArray for example.

RefFreeCellMixArrayDevianceBoots

RefFreeCellMixArrayDevianceBoots

Description

Matrix of bootstrapped deviances corresponding to an array of reference-free cell-mixture decompositions

Usage

```
RefFreeCellMixArrayDevianceBoots(rfArray, Y, R=5, EPSILON=1E-9, bootstrapIterations=5)
```

Arguments

rfArray	list of RefFreeCellMix objects (e.g. from RefFreeCellMixArray)
Y	Methylation matrix on which x was based
R	Number of bootstrapped vectors to return
EPSILON	Minimum value of variance (zero variances will be reset to this value)
bootstrapIterations	Number of RefFreeCellMix iterations to use in bootstrap

Details

Matrix (multiple vectors) of bootstrapped deviances corresponding to an array of reference-free cell-mixture decompositions, used to determine optimal number of cell types. This function returns one bootstrapped vector. The bootstrapped deviance is based on normal distribution applied to errors of Y after accounting for cell mixture effect, Mu Omega^T. See RefFreeCellMixArray for example.

RefFreeCellMixArrayWithCustomStart

Reference-Free Cell Mixture Projection - Custom Initialization

---

Description

Array of reference-free cell-mixture decompositions of a DNA methylation data set, with custom initialization

Usage

```
RefFreeCellMixArrayWithCustomStart(Y,mu.start,Klist=1:5,iters=10,
   Yfinal=NULL,verbose=FALSE)
```

Arguments

Y	Matrix (m CpGs x n Subjects) of DNA methylation beta values
mu.start	matrix of starting values for Mu: number of columns must be at least the maximum in Klist
Klist	List of K values (each K = assumed number of cell types)
iters	Number of iterations to execute for each value of K
Yfinal	Matrix (m* CpGs x n Subjects) of DNA methylation beta values on which to base final methylomes
verbose	Report summary of errors after each iteration for each fit?

Details

List of Reference-free decompositions for a range of K values. For each value of K, the decomposition is initialized by using the first K columns of mu.start. Note that for each K, the decomposition will be based on Y, but Yfinal (=Y by default) will be used to determine the final value of Mu based on the last iterated value of Omega.

Value

List, each element is an object of S3 class RefFreeCellMix, containing the last iteration of Mu and Omega.

Author(s)

E. Andres Houseman

References

Houseman, E. Andres, Kile, Molly L., Christiani, David C., et al. Reference-free deconvolution of DNA methylation data and mediation by cell composition effects. BMC bioinformatics, 2016, vol. 17, no 1, p. 259.

See Also

RefFreeCellMix, RefFreeCellMixInitializeBySVD

RefFreeCellMixInitialize

Initialize Reference-Free Cell Mixture Projection

Description

Initializes the methylome matrix "Mu" for RefFreeCellMix

Usage

```
RefFreeCellMixInitialize(Y,K=2,Y.Distance=NULL, Y.Cluster=NULL,
    largeOK=FALSE, dist.method = "euclidean", ...)
```

Arguments

Y	Matrix (m CpGs x n Subjects) of DNA methylation beta values
K	Number of cell types
Y.Distance	Distance matrix (object of class "dist") to use for clustering.
Y.Cluster	Hiearchical clustering object (from hclust function)
largeOK	OK to calculate distance matrix for large number of subjects? (See details.)
dist.method	Method for calculating distance matrix
...	Additional parameters for hclust function

Details

Initializes the methylome matrix "Mu" for RefFreeCellMix by computing the mean methylation (from Y) over K clusters of Y, determined by the Y.Cluster object. If Y.Cluster object does not exist, it will be created from Y.Distance (using additional clustering parameters if supplied). If Y.Distance does not exist, it will be created from t(Y). As a protection against attempting to fit a very large distance matrix, the program will stop if the number of columns of Y is > 2500, unless largeOK is explicitly set to TRUE.

Value

An m x K matrix of mean methylation values.

Author(s)

E. Andres Houseman

See Also

RefFreeCellMix

RefFreeCellMixInitializeBySVD

Initialize Reference-Free Cell Mixture Projection by SVD

Description

Initialize Reference-Free Cell Mixture Projection by SVD

Usage

```
RefFreeCellMixInitializeBySVD(Y, type=1)
```

Arguments

Y	Matrix (m CpGs x n Subjects) of DNA methylation beta values
type	See details

Details

This method initializes the reference-free cell mixture deconvolution using an ad-hoc method based on singular value decomposition. Type=1 will attempt to discretize Mu to 0/1, Type=2 will attempt to find a continuous range using column ranks. However, neither of these strategies is guaranteed to result in stable starting values for K larger than the "true" value of K.

Value

Matrix of starting values for Mu.

Author(s)

E. Andres Houseman

See Also

RefFreeCellMix, RefFreeCellMixArrayWithCustomStart

Examples

```
data(HNSCC)
Y.shortTest <- Y.HNSCC.averageBetas[1:500,]
mu.start.svd <- RefFreeCellMixInitializeBySVD(Y.shortTest)
testArray2  <- RefFreeCellMixArrayWithCustomStart(Y.shortTest, mu.start=mu.start.svd,
    Klist=1:3,iters=5)
sapply(testArray2,deviance,Y=Y.shortTest)

## Not run:
testBootDevs <- RefFreeCellMixArrayBySVDDevianceBoots(testArray2,Y.shortTest,R=10)

testBootDevs
apply(testBootDevs[-1,],2,mean,trim=0.25)
which.min(apply(testBootDevs[-1,],2,mean,trim=0.25))

## End(Not run)
```

RefFreeEwasModel *Reference-Free EWAS Model*

Description

Reference-free method for conducting EWAS while deconvoluting DNA methylation arising as mixtures of cell types.

Usage

```
RefFreeEwasModel(Y, X, K, smallOutput=FALSE)
```

Arguments

Y	Matrix of DNA methylation beta values (CpGs x subjects). Missing values *are* supported.
X	Design matrix (subjects x covariates).
K	Latent variable dimension (d in Houseman et al., 2013, technical report)
smallOutput	Smaller output? (Should be FALSE if you intend to run bootstraps.)

Details

Reference-free method for conducting EWAS while deconvoluting DNA methylation arising as mixtures of cell types. This method is similar to surrogate variable analysis (SVA and ISVA), except that it makes additional use of a biological mixture assumption. Returns mixture-adjusted Beta and unadjusted Bstar, as well as estimates of various latent quantities.

Value

A list object of class "RefFreeEwasModel". The most important elements are Beta and Bstar.

Author(s)

E. Andres Houseman

References

Houseman EA, Molitor J, and Marsit CJ (2014), Reference-Free Cell Mixture Adjustments in Analysis of DNA Methylation Data. Bioinformatics, doi: 10.1093/bioinformatics/btu029.

See Also

BootRefFreeEwasModel

Examples

```
data(RefFreeEWAS)

## Not run:
  tmpDesign <- cbind(1, rfEwasExampleCovariate)
  tmpBstar <- (rfEwasExampleBetaValues

  EstDimRMT(rfEwasExampleBetaValues-tmpBstar

## End(Not run)

test <- RefFreeEwasModel(
  rfEwasExampleBetaValues,
  cbind(1,rfEwasExampleCovariate),
  4)

testBoot <- BootRefFreeEwasModel(test,10)
summary(testBoot)
```

rfEwasExampleBetaValues

Simulated mixed-cell DNA methylation data set

Description

1000 CpG sites x 250 subjects. First 250 CpGs are DMRs for the cell types, although the idea is that this would not be known in practice.

Usage

```
rfEwasExampleBetaValues
```

Format

1000 CpG sites x 250 subjects.

rfEwasExampleCovariate

Simulated covariate for mixed-cell DNA methylation data set

Description

Vector of covariates corresponding to 250 subjects.

Usage

```
rfEwasExampleCovariate
```

Format

Numeric vector of length 250

`rfEwasExampleTRUEAlpha` *True alpha intercepts used in simulation*

Description

1000 intercept values; these may not match exactly due to cell mixtures.

Usage

```
rfEwasExampleTRUEAlpha
```

Format

1000 intercept values.

`rfEwasExampleTRUEBeta` *True beta coefficients used in simulation (for comparison purposes)*

Description

1000 coefficient values

Usage

```
rfEwasExampleTRUEBeta
```

Format

1000 coefficient values.

`rfEwasExampleTRUEMethDMR` *True M matrix (cell-specific methylation values) used in simulation (for comparison purposes)*

Description

1000 x 4 matrix of beta values

Usage

```
rfEwasExampleTRUEMethDMR
```

Format

1000 x 4 matrix

rfEwasExampleTRUEOmega

True Omega (cell mixture) coefficients used in simulation (for comparison purposes)

Description

250 x 4 matrix of mixing weights

Usage

```
rfEwasExampleTRUEOmega
```

Format

250 x 4 matrix

summary.BootRefFreeEwasModel

summary.BootRefFreeEwasModel

Description

Summary method for objects of type BootRefFreeEwasModel; calculates bootstrap mean and standard deviation.

Usage

```
## S3 method for class 'BootRefFreeEwasModel'
summary(object,...)
```

Arguments

object	BootRefFreeEwasModel object to summarize
...	(Unused).

Details

See RefFreeEwasModel for example.

summary.RefFreeCellMix *summary.RefFreeCellMix*

Description

Summary method for objects of type RefFreeCellMix.

Usage

```
## S3 method for class 'RefFreeCellMix'
summary(object,...)
```

Arguments

object	RefFreeCellMix object to summarize
...	(Unused).

Details

See RefFreeCellMix for example.

svdSafe *Safe SVD-like matrix decomposition*

Description

SVD that traps errors and switches to QR when necessary

Usage

```
svdSafe(X)
```

Arguments

X	Matrix to decompose

Details

This function traps errors in the svd function due to numerically zero singular values, and replaces the operation with a QR decomposition. Technically, the R component of the decomposition fails the orthogonality constraint required for the SVD decomposition, but this function exists to save bootstraps from rudely failing; since the critical component of the SVD (in this application) is the left orthogonal matrix, this is a reasonable approximation for bootstrap purposes. If there are too many svd failures (which will will be reported by the function) then it is worth looking into the design matrix.

Value

A list as in what svd produces: U and V matrices as well as the d vector of singular values.

Author(s)

E. Andres Houseman

See Also

svd

`SVDwithMissing` *SVD with missing values*

Description

Compute singular value decomposition on a matrix with missing values, using a naive/simple method for imputing missing values by row-mean

Usage

```
SVDwithMissing(Y)
```

Arguments

`Y`	Matrix for which to compute SVD.

Details

Computes singular value decomposition on a matrix with missing values, using a naive/simple method for imputing missing values by row-mean. Not recommended for matrices with very large numbers of missing values.

Value

singular value decomposition (as returned by svd function)

X.HNSCC.caseStatusAge *HNSCC Example - Covariates*

Description

Case status (0=control, 1=case) and age (as Z-score) for HNSCC data set

Usage

```
X.HNSCC.caseStatusAge
```

Format

Numeric matrix of dimension 182 x 3

Y.HNSCC.averageBetas *HNSCC Example - DNA Methylation Average Betas*

Description

Peripheral blood from 92 head and neck squamous cell carcinoma (HNSCC) patients and 92 controls. GEO Accession #GSE32393 with 2 outlier cases removed.

Usage

```
Y.HNSCC.averageBetas
```

Format

Numeric matrix of dimension 26486 by 182

Index