

Package ‘NMF’

February 8, 2010

Type Package

Title Algorithms and framework for Nonnegative Matrix Factorization (NMF).

Version 0.3.1

Date 2010-01-12

Author Renaud Gaujoux

Maintainer Renaud Gaujoux <renaud@cbio.uct.ac.za>

Description This package provides a framework to perform Non-negative Matrix Factorization (NMF). It implements a set of already published algorithms and seeding methods, and provides a framework to test, develop and plug new/custom algorithms.

License GPL (>= 2)

URL <http://nmf.r-forge.r-project.org>

LazyLoad yes

Depends R (>= 2.10), methods, stats

Suggests Biobase, RColorBrewer, fastICA, bigmemory, doMC

Collate utils.R options.R NMF-class.R NMFstd-class.R NMFOffset-class.R NMFns-class.R
NMFfit-class.R NMFSet-class.R registry.R NMFStrategy-class.R NMFStrategyFunction-class.R
NMFStrategyIterative-class.R snmf.R lnmf.R pe-nmf.R ica.R nndsvd.R seed.R nmf.R
Bioc-layer.R nmf-package.R

Repository CRAN

Date/Publication 2010-02-08 09:02:42

R topics documented:

NMF-package	2
advanced	3
basis-methods	5
esGolub	6
Internals	8
NMF-bioc	8
NMF-class	9
NMF-deprecated	13
nmf-methods	14
NMF-utils	20
nmfEstimateRank	27
NMFfit-class	29
NMFns-class	33
NMFOffset-class	36
NMFSet-class	37
NMFstd-class	40
Index	45

NMF-package

NMF Package Overview

Description

The NMF package provides methods to perform Nonnegative Matrix Factorization (NMF) , as well as a framework to develop and test new NMF algorithms.

A number of standard algorithms and seeding methods are implemented. Tuned visualisation and post-analysis methods help in the evaluation of the algorithms' performances or in the interpretation of the results.

Author(s)

Renaud Gaujoux <renaud@cbio.uct.ac.za>

References

Definition of Nonnegative Matrix Factorization in its modern formulation:

Lee D.D. and Seung H.S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*, **401**, 788–791.

Historical first definition and algorithms:

Paatero, P., Tapper, U. (1994). Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics*, **2**, 111–126 , doi:10.1002/env.3170050203.

See Also

[NMF-class](#), [nmf](#), [Biobase](#)

Examples

```
# run default NMF algorithm on a random matrix
V <- matrix(runif(10000), 500, 20)
res <- nmf(V, 3)
res

# compute some quality measures
summary(res)

# Visualize the results as heatmaps
## Not run: metaheatmap(res) # mixture coefficients
## Not run: metaheatmap(res, 'features') # basis vectors

# run default NMF algorithm on a random matrix with actual patterns
set.seed(123456)
V <- syntheticNMF(500, 3, 20, noise=TRUE)
res <- nmf(V, 3)
res

# compute some quality measures
summary(res)

# Visualize the results as heatmaps
## Not run: metaheatmap(res) # mixture coefficients
## Not run: metaheatmap(res, 'features') # basis vectors
```

advanced

Advanced usage of package NMF

Description

Allow the user to get/set/define package NMF specific options in the same way as with base functions [options](#) and [getOption](#).

Usage

```
nmf.options(..., runtime = FALSE)
nmf.getOption(name)
nmf.options.reset()
nmf.options.runtime()

nmf.models(builtin.only=FALSE)
```

Arguments

`...` any options can be defined, using `'name = value'` or by passing a list of such tagged values. However, only the ones below are used in package NMF. Further, `'nmf.options('name') == nmf.options()['name']'`, see the example.

`builtin.only` a single logical. If TRUE only built-in NMF models are returned.

`name` a character string holding an option name.

`runtime` a boolean used to specify if main interface function `nmf` should store the option into the initial `NMF` object before performing the computation.

Details

nmf.getOption get the value of a single option.

nmf.models returns the available NMF models. These are exactly the classes that inherits from class `NMF`. If argument `builtin.only` is TRUE then only the models that are defined within the package are returned.

nmf.options gets/sets/defines options in the way of base function `options`. Invoking `'nmf.options()'` with no arguments returns a list with the current values of the options. To access the value of a single option, one should use `nmf.getOption("error.track")`, e.g., rather than `nmf.options("error.track")` which is a list of length one.

nmf.options.reset Reset all *built-in options* to their default values. Note that only built-in are reset. The options defined by the user during the current session will keep their values.

Value

For `nmf.getOption`, the current value set for option `name`, or NULL if the option is unset.

For `nmf.options()`, a list of all set options sorted by name. For `options(name)`, a list of length one containing the set value, or NULL if it is unset. For uses setting one or more options, a list with the previous values of the options changed (returned invisibly).

For `nmf.models`, a character vector listing the available NMF models.

Options set in package NMF

debug logical. Similar to option `'verbose'` (see below), but reports more information.

error.track logical. Should the estimation error be tracked during the computations? If set to TRUE then the error track can be plotted using method `errorPlot`. The step size of the error track is set via option `track.interval` (see below).

track.interval numeric. The number of iterations performed between two consecutive error points. For performance reason, this value should be too small, as the computation of the estimation error can be time consuming (Default value is 30).

verbose logical. Should R report extra information about the computations?

Author(s)

Renaud Gaujoux <renaud@cbio.uct.ac.za>

See Also[options](#)**Examples**

```
# save all options value
op <- nmf.options();
utils::str(op) # op may contain functions.

nmf.getOption("track.interval") == nmf.options()$track.interval # the latter needs more

x <- matrix(runif(50*10), 50, 10) # create a random target matrix
# or define a synthetic data with a hidden pattern using function syntheticNMF (see ?syn
## Not run: x <- syntheticNMF(50, 5, 10, noise=TRUE)

# perform default NMF computation
res <- nmf(x, 3)

# Toogle on verbose mode
nmf.options(verbose = TRUE)
res <- nmf(x, 3)

# Toogle on debug mode
nmf.options(debug = TRUE)
res <- nmf(x, 3)

# set the error track step size, and save previous value
old.o <- nmf.options(track.interval = 5)
old.o

# check options
utils::str(nmf.options())
# reset to default values
nmf.options.reset()
utils::str(nmf.options())
```

basis-methods

*Extract/Set the matrix of basis vectors from a NMF model***Description**

`basis` and `basis<-` are S4 generic functions which respectively extract and set the matrix of basis vectors (i.e. the first matrix factor) of a NMF model. For example, in the case of the standard NMF model $V \equiv WH$, method `basis` will return matrix W .

`coef` and `coef<-` are S4 methods defined for the associated generic functions from package `stats` (See [coef](#)). They respectively extract and set the matrix of mixture coefficients (i.e. the

second matrix factor) of a NMF model. For example, in the case of the standard NMF model $V \equiv WH$, method `coef` will return matrix H .

Methods `coefficients` and `coefficients` are simple aliases for methods `coef` and `coef<-` respectively.

Methods

basis signature(object = "NMF"): Extracts the matrix of basis vectors from NMF model object.

Note that it is implemented as a pure virtual method, that must be overloaded by sub-classes that implement concrete NMF models. It throws an error if directly called. See [NMF](#) for more details.

basis<- signature(object = "NMF", value = "matrix"): Sets the matrix of basis vectors from NMF model object.

Note that it is implemented as a pure virtual method, that must be overloaded by sub-classes that implement concrete NMF models. It throws an error if directly called. See [NMF](#) for more details.

coef signature(object = "NMF"): Extracts the matrix of mixture coefficients from NMF model object.

Note that it is implemented as a pure virtual method, that must be overloaded by sub-classes that implement concrete NMF models. It throws an error if directly called. See [NMF](#) for more details.

coef<- signature(object = "NMF", value = "matrix"): Sets the matrix of mixture coefficients from NMF model object.

Note that it is implemented as a pure virtual method, that must be overloaded by sub-classes that implement concrete NMF models. It throws an error if directly called. See [NMF](#) for more details.

See Also

NMF, NMFstd, NMFfit

esGolub

Golub ExpressionSet from Brunet et al. Paper

Description

The original data is related to Golub et al., and this version is the one used and referenced in Brunet et al. The samples are from 27 patients with acute lymphoblastic leukemia (ALL) and 11 patients with acute myeloid leukemia (AML).

The samples were assayed using Affymetrix Hgu6800 chips and the original data on the expression of 7129 genes (Affymetrix probes) are available on the Broad Institute web site (see references below).

The data in `esGolub` were obtained from the web site related to Brunet et al.'s publication on an application of Nonnegative Matrix Factorization (see link in section *Source*).

They contain the 5,000 most highly varying genes according to their coefficient of variation, and were installed in an object of class `ExpressionSet-class`.

Usage

```
data(esGolub)
```

Format

There are 3 covariates listed.

- Samples: The original sample labels.
- ALL.AML: Whether the patient had AML or ALL. It is a `factor` with levels `c('ALL', 'AML')`.
- Cell: ALL arises from two different types of lymphocytes (T-cell and B-cell). This specifies which for the ALL patients; There is no such information for the AML samples. It is a `factor` with levels `c('T-cell', 'B-cell', NA)`.

Source

<http://www.broadinstitute.org/publications/broad872>

References

Metagenes and molecular pattern discovery using matrix factorization Brunet, J.-P., Tamayo, P., Golub, T.-R., and Mesirov, J.-P. (2004) Proc Natl Acad Sci U S A 101(12), 4164–4169.

Molecular Classification of Cancer: Class Discovery and Class Prediction by Gene Expression Monitoring, Science, 531-537, 1999, T. R. Golub and D. K. Slonim and P. Tamayo and C. Huard and M. Gaasenbeek and J. P. Mesirov and H. Coller and M.L. Loh and J. R. Downing and M. A. Caligiuri and C. D. Bloomfield and E. S. Lander

Original data from Golub et al.: http://www-genome.wi.mit.edu/mpr/data_set_ALL_AML.html

Examples

```
data(esGolub)
esGolub
## Not run: pData(esGolub)
```

Internals

*Undocumented or Internal methods and functions for package NMF***Description**

Dummy Rd file that gathers undocumented or internal methods and functions for package NMF.

Author(s)

Renaud Gaujoux <renaud@cbio.uct.ac.za>

See Also

[NMF](#), [nmf-methods](#), package's vignette

NMF-bioc

*Layer to use the NMF package within Bioconductor***Description**

The package NMF provides an optional layer for working with common objects defined in the Bioconductor platform.

It provides:

- alias functions whose names are more intuitive when NMF is applied to bioinformatics data.
- specialized vizualization methods that adapt the titles and legend using bioinformatics terminology.
- functions to link the results with annotations, etc...

Methods

metagenes signature(object = "NMF"): returns the metagenes matrix according to the model defined in object. It is a simple alias to method [basis](#).

metagenes<- signature(object = "NMF", value = "matrix"): sets the metagenes matrix in object, and returns the updated object. It is a simple alias to method [basis<-](#).

metaprofiles signature(object = "NMF"): returns the metaprofiles matrix according to the model defined in object. It is a simple alias to method [coef](#).

metaprofiles<- signature(object = "NMF", value = "matrix"): sets the metaprofiles matrix in object, and returns the updated object. It is a simple alias to method [coef<-](#).

nmfmeta signature(object = "NMF"): returns the number of metagenes use in NMF model object. It's an alias to [nbasis](#).

See Also

NMF, NMF-utils

Description

This is a *virtual class* that defines a common interface to handle Nonnegative Matrix Factorisation models (NMF models) in a generic way.

It provides the definition for a minimum set of generic methods that are used in common computations and tasks in the context of Nonnegative Matrix Factorisations.

Class `NMF` makes it easy to develop new models that integrates well into the general framework implemented by the *NMF* package.

Following a few simple guidelines, new models benefit from all the functionalities available to built-in NMF models – that derive themselves from class `NMF`. See section *Defining new NMF models* below.

See section [NMFstd](#), references and links therein for details on the standard NMF model and its – built-in – extensions.

Slots

This class contains a single slot, that is used internally during the computations.

`misc`: A list that is used internally to temporarily store algorithm parameters during the computation.

The purpose of this class is to define a common interface for NMF models as a collection of generic methods. Classes that inherits from class `NMF` are responsible for the management of data storage and the implementation of the interface's pure virtual methods.

Defining new NMF models

The minimum requirement to define a new NMF model that integrates into the framework of the *NMF* package are the followings:

- Define a class that inherits from class `NMF` and implements the new model. Say class `myNMF`.
- Implement the following S4 methods for the new class `myNMF`:
 - fitted** signature(object = "myNMF", value = "matrix"): Must return the estimation of the target matrix as fitted by the NMF model object.
 - basis** signature(object = "myNMF"): Must return the matrix of basis vectors (e.g. the first matrix factor in the standard NMF model).
 - basis<-** signature(object = "myNMF", value = "matrix"): Must return object with the matrix of basis vectors set to value.
 - coef** signature(object = "myNMF"): Must return the matrix of mixture coefficients (e.g. the second matrix factor in the standard NMF model).
 - coef<-** signature(object = "myNMF", value = "matrix"): Must return object with the matrix of mixture coefficients set to value.

The *NMF* package ensures these methods are defined for classes that inherits from class *NMF*, as the methods defined for signatures `(object='NMF', ...)` and `(object='NMF', value='matrix')` throw an error when called.

- Optionally, implement method `rnmf(signature(object="myNMF", target="numeric"))`. This method should fill model object (of class *myNMF*) with random values to fit a target matrix, whose dimension is given by the 2-length numeric vector `target`.

For concrete examples of NMF models implementations, see class *NMFstd* and its extensions (e.g. classes *NMFOffset* or *NMFns*).

Objects from the Class

Strictly speaking, because class *NMF* is virtual, no object of class *NMF* can be instantiated, only objects from its sub-classes. However, those objects are sometimes shortly referred in the documentation as "NMF objects" instead of "objects that inherits from class *NMF*".

For built-in models or for models that inherit from the standard model class *NMFstd*, the factory method `newNMF` enables to easily create valid *NMF* objects in a variety of common situations. See `newNMF` for more details.

Methods

[signature(object = "NMF")]: sub-setting method for object of class *NMF*. Row subsets are applied to the basis matrix rows, while column subsets are applied to the mixture coefficient matrix. It returns an object of class *NMF* whose basis matrix and/or mixture coefficient matrix have been subset accordingly.

basis signature(object = "NMF"): returns the matrix of basis vectors according to the model defined in object. This is a *pure virtual* method that needs to be defined for the sub-classes of class *NMF* that implements concrete models. See also `basis`.

basis<- signature(object = "NMF", value = "matrix"): sets the matrix of basis vectors in object, and returns the updated object. This is a *pure virtual* method that needs to be defined for the sub-classes of class *NMF* that implements concrete models. See also `basis`.

coef signature(object = "NMF"): returns the matrix of mixture coefficients according to the model defined in object. This is a *pure virtual* method that needs to be defined for the sub-classes of class *NMF* that implements concrete models. See also `coef`.

coef<- signature(object = "NMF", value = "matrix"): sets the matrix of mixture coefficients in object, and returns the updated object. This is a *pure virtual* method that needs to be defined for the sub-classes of class *NMF* that implements concrete models. See also `coef<-`.

coefficients signature(object = "NMF"): This is a simple alias to method `coef`. See also `coefficients`.

coefficients<- signature(object = "NMF", value = "matrix"): This is a simple alias to method `coef<-`. See also `coef`.

connectivity signature(x = "NMF"): returns the connectivity matrix associated to the clusters based on NMF factorization *x*. The connectivity matrix *C* of a clustering is the symmetric matrix that shows the shared membership of the samples: entry C_{ij} is 1 if samples *i* and *j* belong to the same cluster, 0 otherwise.

- dim** signature(x = "NMF"): returns a 3-length vector containing the dimension of the target matrix together with the NMF factorization rank. For example `c(2000, 30, 3)` for a NMF object that fits a 2000x30 target matrix using 3 basis vectors.
- distance** signature(target = "matrix", x = "NMF", method, ...): returns the distance between a target matrix and a NMF model, according to a given measure. If both argument `target` and `x` are missing, this function returns the `function` defined by argument `method`. The later can either be a `function` or a `character` string that correspond to a registered distance metric. For the moment only the metric `'KL'` and `'euclidean'` are defined. See function [distance](#) for more details.
- entropy** signature(x = "NMF", class = "factor"): computes the entropy of NMF model `x` given a priori known groups of samples. See generic function [entropy](#) for more details.
- evan** signature(object = "NMF", target): computes the explained variance of NMF model object that approximates `target`. See generic function [evan](#) for more details.
- featureNames** signature(object = "NMF"): returns the row names of the basis matrix. If BioConductor is installed this method is defined for the generic function [featureNames](#) from the Biobase package.
- featureNames<-** signature(object = "NMF", value = "ANY"): sets the row names of the basis matrix. Argument `value` must be in a format accepted by the [rownames](#) method defined for matrices. If BioConductor is installed this method is defined for the generic function [featureNames<-](#) from the Biobase package.
- fitted** signature(object = "NMF"): computes the target matrix estimated by NMF model object. This is a *pure virtual* method that needs to be implemented by the sub-classes of class NMF that implements concrete models.
- featureScore** signature(x = "NMF"): Computes a score for each feature that reflects its specificity to one of the basis vector. The definition of the score follows *Kim and Park (2007)*. See references for more details.
- extractFeatures** signature(x = "NMF"): extract the features that are the most specific to each basis vector. It follows *Kim and Park (2007)*'s methodology. See references for more details.
- is.empty.nmf** signature(object = "NMF"): Tells if `object` is an empty the NMF model, that is it contains no data. It returns `TRUE` if the matrices of basis vectors and mixture coefficients have respectively zero rows and zero columns. It returns `FALSE` otherwise. This means that an empty model can still have a non-zero number of basis vectors. For example, this happens in the case of NMF models created using factory method [newNMF](#) with no initialisation for any factor matrices.
- metaHeatmap** Produces a heatmap of the basis or mixture matrix using function [heatmap.2](#) with parameters tuned for displaying NMF results. See [metaHeatmap](#) for more details.
- nbasis** signature(x = "NMF"): returns the number of basis vectors used in NMF model `x`. It is the number of columns of the matrix of basis vectors.
- predict** signature(object = "NMF"): returns a `factor` that gives the predicted cluster index for each sample (resp. for each feature) based on NMF factorization `object`. The index correspond to the basis vector that most contribute to the sample (resp. to which the feature contribute the most). See [predict](#) details on extra arguments.

purity signature(`x = "NMF"`, `class = "factor"`): computes the purity of NMF model `x` given a priori known groups of samples. The purity definition can be found in *Kim and Park (2007)*. See references for more details.

rnmf signature(`x = "NMF"`, `target`): seeds NMF model `x` with random values drawn from a uniform distribution. The result is a NMF model of the same class as `x` with basis and mixture matrices filled with random values.

Argument `target` can be either:

numeric it must be of length 2 (resp. 1), and give the dimension of the target matrix (resp. symmetric matrix) to fit. The result is a random NMF model.

missing it returns `rnmf(x, c(nrow(x), ncol(x)))`, that is a random NMF model with the same dimensions as defined in model `x`.

matrix it returns `rnmf(x, dim(target))`, that is a random NMF model that fits a matrix of the same dimension as `target`. The values are drawn within the interval $[0, \max(\text{target})]$.

This method's version with signature (`object='NMF'`, `target='numeric'`) might need to be overloaded if the initialisation of the specific NMF model requires setting values for data other than the basis and mixture matrices. The overloading methods must call the generic version using function `callNextMethod`.

rss signature(`object = "NMF"`): returns the Residual Sum of Squares (RSS) between the target matrix and its estimation by the NMF model `object`. *Hutchins et al. (2008)* used the variation of the RSS in combination with *Lee and Seung's* algorithm to estimate the correct number of basis vectors. See `rss` for details on its usage.

sampleNames signature(`object = "NMF"`): returns the column names of the mixture coefficient matrix. If BioConductor is installed this method is defined for the generic function `sampleNames` from the Biobase package.

sampleNames<- signature(`object = "NMF"`, `value = "ANY"`): sets the columns names of the basis matrix. Argument `value` must be in a format accepted by the `colnames` method defined for matrices. If BioConductor is installed this method is defined for the generic function `sampleNames<-` from the Biobase package.

show signature(`object = "NMF"`): standard generic show method for objects of class NMF. It displays the model class (i.e. the name of the sub-class that implements the concrete model), the dimension of the target matrix, and the number of basis vectors.

sparseness signature(`x = "NMF"`): compute the average sparseness of the basis vectors and mixture coefficients. See *Hoyer (2004)* for more details.

summary signature(`x = "NMF"`): standard generic summary method for objects of class NMF. It computes a set of measures to evaluate the quality of the factorization.

Author(s)

Renaud Gaujoux <renaud@cbio.uct.ac.za>

References

Definition of Nonnegative Matrix Factorization in its modern formulation:

Lee D.D. and Seung H.S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*, **401**, 788–791.

Historical first definition and algorithms:

Paatero, P., Tapper, U. (1994). Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics*, **2**, 111–126, doi:10.1002/env.3170050203.

Reference for some utility functions:

Kim, H. and Park, H. (2007). Sparse non-negative matrix factorizations via alternating non-negativity-constrained least squares for microarray data analysis. *Bioinformatics*.

Hoyer (2004). Non-negative matrix factorization with sparseness constraints. *Journal of Machine Learning Research*, **5**, 1457-1469.

See Also

Main interface to perform NMF in [nmf-methods](#).

Built-in NMF models and factory method in [newNMF](#).

Method [seed](#) to set NMF objects with values suitable to start algorithms with.

Examples

```
# show all the NMF models available (i.e. the classes that inherit from class NMF)
nmf.models()
# show all the built-in NMF models available
nmf.models(builtin.only=TRUE)

# class NMF is a virtual class so cannot be instantiated:
# the following generates an error
## Not run: new('NMF')

# To instantiate a NMF model, use factory method newNMF (see ?newNMF)
newNMF(3)
newNMF(3, model='NMFns')
```

NMF-deprecated

Deprecated Functions in NMF Package

Description

These functions are provided for compatibility with older versions of NMF package only, and may be defunct as soon as the next release.

Usage

```
## S4 method for signature 'NMF':
clusters(x, what = c('samples', 'features'))
```

Arguments

what	Specifies on which matrix (basis components or mixture coefficients) the computation should be performed.
x	An object that inherits from class <code>NMF</code> .

Details

`clusters` is a deprecated synonym for `predict`.

nmf-methods	<i>Main Interface to run NMF algorithms</i>
-------------	---

Description

This method implements the main interface to launch NMF algorithms within the framework defined in package `NMF`. It allows to combine NMF algorithms with seeding methods. The returned object can be directly passed to visualisation or comparison methods.

For a tutorial on how to use the interface, please see the package's vignette: `vignette('NMF')`

Usage

```
## S4 method for signature 'matrix,numeric,function':
nmf(x, rank, method, name, objective='euclidean', model='NMFstd', mixed=FALSE, ...)
## S4 method for signature 'matrix,numeric,character':
nmf(x, rank, method, ...)
## S4 method for signature 'matrix,numeric,NMFStrategy':
nmf(x, rank, method, seed=nmf.getOption('default.seed')
, nrun=1, model=NULL, .options=list(), .pbackend = nmf.getOption("parallel.backend"
, ...)
```

Arguments

method	The algorithm to use to perform NMF on <code>x</code> . Different formats are allowed: <code>character</code> , <code>function</code> . If missing, the method to use is retrieved from the <code>NMF</code> package's specific options by <code>nmf.getOption("default.algorithm")</code> (the default built-in option is <code>'brunet'</code>). See section <i>Methods</i> for more details on how each format is used.
mixed	Boolean that states if the algorithm requires a nonnegative input matrix (<code>mixed=FALSE</code> which is the default value) or accepts mixed sign input matrices (<code>mixed=TRUE</code>). An error is thrown if the sign required is not fulfilled. This parameter is useful to plug-in algorithms such as semi-NMF, that typically does not impose nonnegativity constraints on both the input and the basis component matrices. If <code>NULL</code> then the NMF model is used

model	<p>When <code>method</code> is a function, argument <code>model</code> must be either a single character string (default to 'NMFstd') or a list that specifies values for slots in the NMF model. The NMF model to be instantiated can optionally be given by its class name in the first element of the list [note: A NMF model is defined by a S4 class that extends class <code>NMF</code>]. If no class name is specified then the default model is used, see <code>NMFstd</code>.</p> <p>When <code>method</code> is a single character string or a <code>NMFStrategy</code> object, argument <code>model</code> must be <code>NULL</code> (default) or a list. Note that in this case the NMF model is defined by the NMF strategy itself and cannot be changed.</p> <ul style="list-style-type: none"> • If a single character string, argument <code>model</code> must be the name of the class that defines the NMF model to be instantiated. Arguments in ... are handled in the same way as when <code>model</code> is <code>NULL</code>, see below. • If a list <code>all</code> and only its elements are used to initialise the NMF model's slots. • if <code>NULL</code>, the arguments in ... that have the same name as slots in the NMF model associated with the NMF strategy of name <code>method</code> are used to initialise these slots. <p>Important: Values to initialise the NMF model's slots can be passed in However, if argument <code>model</code> is a list – even empty – then all and only its elements are used to initialise the model, those in ... are directly passed to the algorithm.</p> <p>So to pass a parameter to the NMF algorithm, that has the same name as a slot in the NMF model, argument <code>model</code> MUST be a list – possibly empty – and contains all the values one wants to use for the NMF model slots.</p> <p>If a variable appears in both argument <code>model</code> and ..., the former will be used to initialise the NMF model, the latter will be passed to the NMF algorithm. See code examples for an illustration of this situation.</p>
name	A character string to be used as a name for the custom NMF algorithm.
nrun	Used to perform multiple runs of the algorithm. It specifies the number of runs to perform . This argument is useful to achieve stability when using a random seeding method.
objective	Used when <code>method</code> is a function. It must be A character string giving the name of a built-in distance method or a function to be used as the objective function. It is used to compute the residuals between the target matrix and its NMF estimate.
.options	<p>this argument is used to set some runtime options. It can be list containing the named options and their values, or, in the case only boolean options need to be set, a character string that specifies which options are turned on or off. The string must be composed of characters that correspond to a given option. Characters '+' and '-' are used to explicitly specify on and off respectively. E.g. <code>.options='tv'</code> will toggle on options <code>track</code> and <code>verbose</code>, while <code>.options='t-v'</code> will toggle on option <code>track</code> and off option <code>verbose</code>. Note that '+' and '-' apply to all option character found after them. Default behaviour assumes that <code>.options</code> starts with a '+'. The following options are available (note the characters that correspond to each option, to be used when <code>.options</code> is passed as a string):</p>

debug - d	Toogle debug mode. Like option <code>verbose</code> but with more information displayed.
keep.all - k	used when performing multiple runs (<code>nrun > 1</code>): if toogled on, all factorizations are saved and returned, otherwise only the factorization achieving the minimum residuals is returned.
parallel - p	this option is useful on multicore *nix or Mac machine only, when performing multiple runs (<code>nrun > 1</code>). If toggled on, the runs are performed using the parallel backend defined in argument <code>.pbackend</code> . If this is set to <code>'mc'</code> then one tried to perform the runs using multiple cores with package <code>link[package:doMC]{doMC}</code> – which therefore needs to be installed. If the computation cannot be performed in parallel, then it will be carried on sequentially.
parallel.required - P	Same as <code>p</code> , but an error is thrown if the computation cannot be performed in parallel.
track - t	enables (resp. disables) error tracking. When <code>TRUE</code> , the returned object's slot <code>residuals</code> contains the trajectory of the objective values. This tracking functionality is available for all built-in algorithms.
verbose - v	Toogle verbosity. If on, messages about the configuration and the state of the current run(s) are displayed.
<code>.pbackend</code>	define the parallel backend (from the <code>foreach</code> package) to use when running in parallel mode. See options <code>p</code> and <code>P</code> in argument <code>.options</code> . Currently it accepts the following values: <code>'mc'</code> or a number that specifies the number of cores to use, <code>'seq'</code> or <code>NULL</code> to use sequential backend.
<code>rank</code>	The factorization rank to achieve [i.e a single positive numeric]
<code>seed</code>	The seeding method to use to compute the starting point passed to the algorithm. See section <i>Seeding methods</i> for more details on the possible classes and types for argument <code>seed</code> .
<code>x</code>	The target object to estimate. It can be a <code>matrix</code> , a <code>data.frame</code> , an <code>ExpressionSet</code> object (this requires the <code>Biobase</code> package to be installed). See section <i>Methods</i> for more details.
<code>...</code>	Extra parameters passed to the NMF algorithm's <code>run</code> method or used to initialise the NMF model slots. If argument <code>model</code> is not supplied as a list, ANY of the arguments in <code>...</code> that have the same name as slots in the NMF model to be instantiated will be used to initialise these slots. See also the <i>Important</i> paragraph in argument <code>model</code> .

Value

The returned value depends on the run mode:

Single run: An object that inherits from class `NMF`.

Multiple runs:

When `nrun > 1` or when `method` is a list, this method returns an object of class `NMFSet`

Methods

- x = "matrix", rank = "numeric", method = "list"** Performs NMF on matrix `x` for each algorithm defined in the list `method`.
- x = "data.frame", rank = "ANY", method = "ANY"** Performs NMF on a `data.frame`: the target matrix is the converted `data.frame as.matrix(x)`
- x = "ExpressionSet", rank = "ANY", method = "ANY"** Performs NMF on an `ExpressionSet`: the target matrix is the expression matrix `exprs(x)`.
This method requires the `Biobase` package to be installed. Special methods for bioinformatics are provided in an optional layer, which is automatically loaded when the `Biobase` is installed. See [NMF-bioc](#).
- x = "matrix", rank = "numeric", method = "character"** Performs NMF on a `matrix` using an algorithm whose name is given by parameter `method`. The name provided must partially match the name of a registered algorithm. See section *Algorithms* below or the package's vignette.
- x = "matrix", rank = "numeric", method = "function"** Performs NMF using a custom algorithm defined by a `function`. It must have signature `(x=matrix, start=NMF, ...)` and return an object that inherits from class `NMF`. It should use its argument `start` as a starting point.

NMF Algorithms

The following algorithms are available:

- `lee` Standard NMF. Based on euclidean distance, it uses simple multiplicative updates. See *Lee and Seung (2000)*.
- `brunet` Standard NMF. Based on Kullbach-Leibler divergence, it uses simple multiplicative updates from *Lee and Seung (2000)*, enhanced differently to avoid numerical underflow. See *Brunet et al. (2004)*.
- `lnmf` Local Nonnegative Matrix Factorization. Based on a regularized Kullbach-Leibler divergence, it uses a modified version of Lee and Seung's multiplicative updates. See *Li et al. (2001)*.
- `nsNMF` Nonsmooth NMF. Use a modified version of Lee and Seung's multiplicative updates for Kullbach-Leibler divergence to fit a extension of the standard NMF model. See *Pascual-Montatno et al. (2006)*.
- `offset` Use a modified version of Lee and Seung's multiplicative updates for euclidean distance, to fit a NMF model that includes an intercept. See *Badea (2008)*.
- `snmf/r`, `snmf/l` Alternating Least Square (ALS) approach from *Kim and Park (2007)*.

Seeding methods

The purpose of seeding methods is to compute initial values for the factor matrices in a given NMF model. This initial guess will be used as a starting point by the chosen NMF algorithm.

The seeding method to use in combination with the algorithm can be passed to interface `nmf` through argument `seed`. Detailed examples of how to specify the seeding method and its parameters can be found in the *Examples* section of this man page and in the package's vignette.

Argument `seed` accepts the following formats:

- a character string:** giving the name of a *registered* seeding method. The corresponding method will be called to compute the starting point.
- a list:** giving the name of a *registered* seeding method and, optionally, extra parameters to pass to it.
- a single numeric:** that is used to seed the random number generator. The value will be used in a call to the `set.seed` function before computing a starting point with the 'random' seeding method.
- an object that inherits from NMF:** it should contain the data of an initialised NMF model, that is it must contain valid basis and mixture coefficient matrices. It will be directly passed to the algorithm's method – via its argument `seed`.
- a function:** that computes the starting point. It must have signature `(object=NMF, target=matrix, ...)` and return an object that inherits from class `NMF`. Argument `object` should be used as a template for the returned object.

Author(s)

Renaud Gaujoux <renaud@cbio.uct.ac.za>

References

- Lee, D.-D. and Seung, H.-S. (2000). Algorithms for non-negative matrix factorization. In *NIPS*, 556–562.
- Brunet, J.-P., Tamayo, P., Golub, T.-R., and Mesirov, J.-P. (2004). Metagenes and molecular pattern discovery using matrix factorization. *Proc Natl Acad Sci U S A*, **101**(12), 4164–4169.
- Pascual-Montano, A., Carazo, J.-M., Kochi, K., Lehmann, D., and Pascual-Marqui, R.-D. (2006). Nonsmooth nonnegative matrix factorization (nsnmf). *IEEE transactions on pattern analysis and machine intelligence*, **8**(3), 403–415.
- Kim, H. and Park, H. (2007). Sparse non-negative matrix factorizations via alternating non-negativity-constrained least squares for microarray data analysis. *Bioinformatics*.
- Liviu Badea (2008). Extracting Gene Expression Profiles Common To Colon And Pancreatic Adenocarcinoma Using Simultaneous Nonnegative Matrix Factorization. In *Pacific Symposium on Biocomputing*, **13**, 279–290
- S. Li, X. Hou, and H. Zhang (2001). Learning spatially localized, parts-based representation. In *Proc. CVPR*, 2001.

See Also

class `NMF`, `NMF-utils`, package's vignette

Examples

```
## DATA
# generate a synthetic dataset with known classes: 100 features, 23 samples (10+5+8)
n <- 100; counts <- c(10, 5, 8); p <- sum(counts)
V <- syntheticNMF(n, counts, noise=TRUE)
dim(V)
```

```
# build the class factor
groups <- as.factor(do.call('c', lapply(seq(3), function(x) rep(x, counts[x]))))

## RUN NMF ALGORITHMS

# run default algorithm
res <- nmf(V, 3)
res
summary(res, class=groups)

# run default algorithm multiple times (only keep the best fit)
res <- nmf(V, 3, nrun=10)
res
summary(res, class=groups)

# run default algorithm multiple times keeping all the fits
res <- nmf(V, 3, nrun=10, .options='k')
res
summary(res, class=groups)

## Not run:
## Note: one could have equivalently done
res <- nmf(V, 3, nrun=10, .options=list(keep.all=TRUE))

## End(Not run)

# run nonsmooth NMF algorithm
res <- nmf(V, 3, 'nsNMF')
res
summary(res, class=groups)

## Not run:
## Note: partial match also works
nmf(V, 3, 'ns')

## End(Not run)

## Not run:
# Non default values for the algorithm's parameters can be specified in '...'
res <- nmf(V, 3, 'nsNMF', theta=0.8)

## End(Not run)

# compare some NMF algorithms
res <- nmf(V, 3, list('brunet', 'lee', 'nsNMF'))
res
compare(res, class=groups)

# run on an ExpressionSet (requires package Biobase)
## Not run:
data(esGolub)
nmf(esGolub, 3)
```

```
## End(Not run)

## USING SEEDING METHODS

# run default algorithm with the Non-negative Double SVD seeding method ('nndsvd')
nmf(V, 3, seed='nndsvd')

## Not run:
## Note: partial match also works
nmf(V, 3, seed='nn')

## End(Not run)

# run nsNMF algorithm, fixing the seed of the random number generator
nmf(V, 3, 'nsNMF', seed=123456)

# run default algorithm specifying the starting point following the NMF standard model
start.std <- newNMF(W=matrix(0.5, n, 3), H=matrix(0.2, 3, p))
nmf(V, seed=start.std)

# to run nsNMF algorithm with an explicit starting point, this one
# needs to follow the 'NMFns' model:
start.ns <- newNMF(model='NMFns', W=matrix(0.5, n, 3), H=matrix(0.2, 3, p))
nmf(V, seed=start.ns)
# Note: the method name does not need to be specified as it is inferred from the
# when there is only one algorithm defined for the model.

# if the model is not appropriate (as defined by the algorithm) an error is thrown
# [cf. the standard model doesn't include a smoothing parameter used in nsNMF]
## Not run: nmf(V, method='ns', seed=start.std)
```

NMF-utils

Class and Utility Methods for NMF objects

Description

Define generic interface methods for class [NMF](#), which is the base – virtual – class of the results from any NMF algorithms implemented within package NMF's framework.

Usage

```
## S4 method for signature 'NMF':
connectivity(x, ...)

cophcor(object, ...)

dispersion(object, ...)
```

```

## S4 method for signature 'NMF,factor':
entropy(x, class, ...)

## S4 method for signature 'NMF':
evar(object, target)

## S4 method for signature 'NMFfit':
residuals(object, track=FALSE, ...)

rss(object, ...)
## S4 method for signature 'NMF':
rss(object, target)

## S4 method for signature 'NMF':
featureScore(object, method=c('kim', 'max'))

## S4 method for signature 'NMF':
extractFeatures(object, method=c('kim', 'max'), format=c('list', 'combine', 'subset'))

## S4 method for signature 'NMF':
metaHeatmap(object, what=c('samples', 'features'), filter=FALSE, ...)

## S4 method for signature 'NMF':
nmfApply(object, MARGIN, FUN, ...)

## S4 method for signature 'NMF':
predict(object, what = c('samples', 'features'), prob=FALSE)

## S4 method for signature 'NMF,factor':
purity(x, class, ...)

randomize(x, ...)

## S4 method for signature 'NMF':
sparseness(x)

syntheticNMF(n, r, p, offset=NULL, noise=FALSE, return.factors=FALSE)

```

Arguments

<code>class</code>	A factor giving a known class membership for each sample.
<code>filter</code>	if TRUE, only the features that are basis-specific are used. Those features are those returned by function <code>extractFeatures</code> .
<code>format</code>	the output format of the extracted features. Possible values are: <ul style="list-style-type: none"> • <code>list</code> (default) a list with one element per basis vector, each containing the indices of the basis-specific features.

	<ul style="list-style-type: none"> • combine a single integer vector containing the indices of the basis-specific features for ALL the basis. • subset the object <code>object</code> subset to contain only the basis-specific features.
<code>FUN</code>	the function to be applied: see 'Details'. In the case of functions like <code>+</code> , <code>%*%</code> , etc., the function name must be backquoted or quoted. See <code>link[base]{apply}</code> for more details.
<code>MARGIN</code>	a vector giving the subscripts which the function will be applied over. 1 indicates rows, 2 indicates columns, <code>c(1, 2)</code> indicates rows and columns. See <code>link[base]{apply}</code> for more details.
<code>method</code>	<p>Method used to compute the feature scores and selecting the features.</p> <p>Possible values are:</p> <ul style="list-style-type: none"> • <code>kim</code> (default) to use Kim and Park (2007) scoring schema and feature selection method. The features are first scored using the function <code>featureScore</code>. Then only the features that fulfil both following criteria are retained: <ul style="list-style-type: none"> - score greater than $\hat{\mu} + 3\hat{\sigma}$, where $\hat{\mu}$ and $\hat{\sigma}$ are the median and the median absolute deviation (MAD) of the scores respectively; - the maximum contribution to a basis component is greater than the median of all contributions (i.e. of all elements of W) See <i>Kim and Park (2007)</i>. • <code>max</code> where the score is the maximum contribution of each feature to the basis vectors and the selection method is the one described in <i>Carmona-Saez (2006)</i>. Briefly, for each basis vector, the features are first sorted in descending order by their contribution to the basis vector. Then, one selects only the first consecutive features from the sorted list whose highest contribution in the basis matrix is found in the considered basis (see section <i>References</i>).
<code>n</code>	Number of rows of the synthetic target matrix.
<code>noise</code>	if <code>TRUE</code> , a random noise is added the target matrix.
<code>object</code>	A <code>matrix</code> or an object that inherits from class <code>NMF</code> or <code>NMFfit</code> – depending on the method.
<code>offset</code>	a vector giving the offset to add to the synthetic target matrix. Its length should be equal to the number of rows <code>n</code> .
<code>prob</code>	Should the probability associated with each cluster prediction be computed and returned.
<code>p</code>	Number of columns of the synthetic target matrix. Not used if parameter <code>r</code> is a vector (see description of argument <code>r</code>).
<code>r</code>	Underlying factorization rank. If a single <code>numeric</code> is given, the classes are randomly generated from a multinomial distribution. If a numerical vector is given, then it should contain the counts in the different classes (i.e integers). In such a case argument <code>p</code> is not used and the number of columns is forced to be the sum of the counts.
<code>return.factors</code>	If <code>TRUE</code> , the underlying matrices W and H are also returned.

<code>target</code>	the target object estimated by model object. It can be a matrix or an ExpressionSet.
<code>track</code>	if TRUE, the whole residuals track is returned. Otherwise only the last residuals computed is returned.
<code>what</code>	Specifies on which matrix (basis components or mixture coefficients) the computation should be performed.
<code>x</code>	for <code>randomize</code> : the matrix or ExpressionSet object whose entries will be randomised. otherwise: An object that inherits from class <code>NMF</code> .
<code>...</code>	Used to pass extra parameters to subsequent calls: <ul style="list-style-type: none"> • in <code>metaHeatmap</code>: Graphical parameters passed to function <code>heatmap.2</code> • in <code>nmfApply</code>: optional arguments to function <code>FUN</code>. • in <code>randomize</code>: passed to the <code>sample</code> function. • in <code>residuals</code>: not used.

Details

connectivity Computes the connectivity matrix for the samples based on their mixture coefficients. The connectivity matrix of a clustering is a matrix C containing only 0 or 1 entries such that:

$$C_{ij} = \begin{cases} 1 & \text{if sample } i \text{ belongs to the same cluster as sample } j \\ 0 & \text{otherwise} \end{cases}.$$

cophcor Computes the cophenetic correlation coefficient of consensus matrix object, generally obtained from multiple NMF runs.

The cophenetic correlation coefficient is based on the consensus matrix (i.e. the average of connectivity matrices) and was proposed by *Brunet et al. (2004)* to measure the stability of the clusters obtained from NMF.

It is defined as the Pearson correlation between the samples' distances induced by the consensus matrix (seen as a similarity matrix) and their cophenetic distances from a hierarchical clustering based on these very distances (by default an average linkage is used). See *Brunet et al. (2004)*.

Note that argument `...` is not used.

dispersion Computes the dispersion coefficient of consensus matrix object, generally obtained from multiple NMF runs.

The dispersion coefficient is based on the consensus matrix (i.e. the average of connectivity matrices) and was proposed by *Kim and Park (2007)* to measure the reproducibility of the clusters obtained from NMF. It is defined as:

$$\rho = \sum_{i,j=1}^n 4(C_{ij} - \frac{1}{2})^2.$$

, where n is the total number of samples.

We have $0 \leq \rho \leq 1$ and $\rho = 1$ only for a perfect consensus matrix, where all entries 0 or 1. A perfect consensus matrix is obtained only when all the connectivity matrices are the same, meaning that the algorithm gave the same clusters at each run. See *Kim and Park (2007)*

Note that argument `...` is not used.

entropy The entropy is a measure of performance of a clustering method, in recovering classes defined by factor a priori known (i.e. one knows the true class labels). Suppose we are given l categories, while the clustering method generates k clusters. Entropy is given by:

$$Entropy = -\frac{1}{n \log_2 l} \sum_{q=1}^k \sum_{j=1}^l n_q^j \log_2 \frac{n_q^j}{n_q}$$

, where:

- n is the total number of samples;
- n is the total number of samples in cluster q ;
- n_q^j is the number of samples in cluster q that belongs to original class j ($1 \leq j \leq l$).

The smaller the entropy, the better the clustering performance.

See *Kim and Park (2007)*.

evan Computes the explained variance of the NMF model object.

For a target V It is defined as:

$$evan = 1 - \frac{RSS}{\sum_{i,j} v_{ij}^2}$$

$evan = 1 - RSS/\sum v_{ij}^2$,

where RSS is the residual sum of squares.

It is usefull to compare the performance of different models and their ability to accurately reproduce the original target matrix. Note that a possible caveat is that some methods explicitly aim at minimizing the RSS (i.e. maximizing the explained variance), while others do not.

extractFeatures Identify the most basis-specific features, using different methods. See details of argument `method`.

featureScore Computes the feature scores as suggested in *Kim and Park (2007)*.

The score for feature i is defined as:

$$S_i = 1 + \frac{1}{\log_2 k} \sum_{q=1}^k p(i, q) \log_2 p(i, q),$$

where $p(i, q)$ is the probability that the i -th feature contributes to basis q :

$$p(i, q) = \frac{W(i, q)}{\sum_{r=1}^k W(i, r)}$$

The feature scores are real values within the range [0,1]. The higher the feature score the more basis-specific the corresponding feature.

metaHeatmap Produces a heatmap of the basis components or mixture coefficients using a `heatmap`-like custom function, with parameters tuned for displaying NMF results.

The used to draw the heatmap is a mixture of the function `heatmap.2` from the `gplots` package, and the function `heatmap.plus` from the `heatmap.plus` package. It allows to add extra annotation rows using the `ColSideColor` argument. See [heatmap.2](#) and [heatmap.plus](#).

nmfApply apply-like method for objects of class NMF.

When argument `MARGIN=1`, it calls the base method `apply` to apply function `FUN` to the *rows* of the basis component matrix.

When `MARGIN=2`, it calls the base method `apply` to apply function `FUN` on the *columns* of the mixture coefficient matrix.

See [apply](#) for more details on the output format.

predict Computes the dominant basis component for each sample (resp. feature) based on its associated entries in the mixture coefficient matrix (i.e in H) (resp. basis component matrix (i.e in W)).

When `what='samples'` the computation is performed on the mixture coefficient matrix, or on the transposed basis matrix when `what='features'`.

For each column, the dominant basis component is computed as the row index for which the entry is the maximum within the column.

If argument `prob=FALSE` (default), the result is a `factor`. Otherwise it returns a list with two elements: element `predict` contains the computed indexes (as a `factor`) and element `prob` contains the vector of the associated probabilities, that is the relative contribution of the maximum entry within each column.

purity Computes the purity of a clustering given a known factor.

The purity is a measure of performance of a clustering method, in recovering the classes defined by a factor a priori known (i.e. one knows the true class labels). Suppose we are given l categories, while the clustering method generates k clusters. Purity is given by:

$$Purity = \frac{1}{n} \sum_{q=1}^k \max_{1 \leq j \leq l} n_q^j$$

, where:

- n is the total number of samples;

- n_q^j is the number of samples in cluster q that belongs to original class j ($1 \leq j \leq l$).

The purity is therefore a real number in $[0, 1]$. The larger the purity, the better the clustering performance.

See *Kim and Park (2007)*.

randomize permute the row entries within each column of `x`, using a different permutation for each column.

The extra arguments in `...` are passed to the `sample` function, and will be used for each column.

The result is a `matrix` of the same dimension as `x` (or `exprs(x)` in the case `x` is an `ExpressionSet` object.).

residuals returns the – final – residuals between the target matrix and the NMF result `object`. They are computed using the objective function associated to the NMF algorithm that returned `object`. When called with `track=TRUE`, the whole residuals track is returned, if available. Note that method `nmf` does not compute the residuals track, unless explicitly required.

It is a S4 methods defined for the associated generic functions from package `stats` (See [residuals](#))

See `nmf` and `NMFfit`.

rss returns the Residual Sum of Squares (RSS) between the target object `target` and its estimation by the object. *Hutchins et al. (2008)* used the variation of the RSS in combination with *Lee and Seung's* algorithm to estimate the correct number of basis vectors. The optimal rank is chosen where the graph of the RSS first shows an inflexion point. See references.

Note that this way of estimation may not be suitable for all models. Indeed, if the NMF optimization problem is not based on the Frobenius norm, the RSS is not directly linked to the quality of approximation of the NMF model.

sparseness Generic method that computes the sparseness of an object as defined in *Hoyer (2004)*.

This sparseness measure quantifies how much energy of a vector is packed into only few components. It is defined by:

$$Sparseness(x) = \frac{\sqrt{n} - \frac{\sum |x_i|}{\sqrt{\sum x_i^2}}}{\sqrt{n} - 1}$$

, where n is the length of x .

The sparseness is a real number in $[0, 1]$. It is equal to 1 if and only if x contains a single nonzero component, and is equal to 0 if and only if all components of x are equal. It interpolates smoothly between these two extreme values. The closer to 1 is the sparseness the sparser is the vector.

The basic definition is for a `numeric` vector. The sparseness of a `matrix` is the mean sparseness of its column vectors. The sparseness of an object of class `NMF`, is the a 2-length vector that contains the sparseness of the basis and mixture coefficient matrices.

syntheticNMF Generate a synthetic matrix according to an underlying NMF model. It can be used to quickly test NMF algorithms.

Author(s)

Renaud Gaujoux <renaud@cbio.uct.ac.za>

References

Metagenes and molecular pattern discovery using matrix factorization Brunet, J.-P., Tamayo, P., Golub, T.-R., and Mesirov, J.-P. (2004) Proc Natl Acad Sci U S A 101(12), 4164–4169.

Sparse non-negative matrix factorizations via alternating non-negativity-constrained least squares for microarray data analysis Kim, H. & Park, H. (2007) Bioinformatics. <http://dx.doi.org/10.1093/bioinformatics/btm134>.

Non-negative Matrix Factorization with Sparseness Constraints Hoyer, P. O. (2004) *Journal of Machine Learning Research* 5 (2004) 1457–1469

Biclustering of gene expression data by non-smooth non-negative matrix factorization Carmona-Saez, Pedro and Pascual-Marqui, Roberto and Tirado, F and Carazo, Jose and Pascual-Montano, Alberto (2006) BMC Bioinformatics 7(1), 78

See Also

[NMF](#), [summary](#)

Examples

```
# generate a synthetic dataset with known classes: 50 features, 18 samples (5+5+8)
n <- 50; counts <- c(5, 5, 8);
V <- syntheticNMF(n, counts, noise=TRUE)
## Not run: metaHeatmap(V)

# build the class factor
groups <- as.factor(do.call('c', lapply(seq(3), function(x) rep(x, counts[x]))))

# perform default NMF
res <- nmf(V, 2)
res

## Not run: metaHeatmap(res, class=groups)
## Not run: metaHeatmap(res, 'features')
# see the predicted clusters of samples
predict(res)
# compute entropy and purity
entropy(res, class=groups)
purity(res, class=groups)

# perform NMF with the right number of basis components
res <- nmf(V, 3)

## Not run: metaHeatmap(res)
## Not run: metaHeatmap(res, 'features')
entropy(res, class=groups)
purity(res, class=groups)
```

nmfEstimateRank	<i>Estimate optimal rank for Nonnegative Matrix Factorization (NMF) models</i>
-----------------	--

Description

A critical parameter in NMF algorithms is the factorization rank r . It defines the number of basis effects used to approximate the target matrix. Function `nmfEstimateRank` helps in choosing an optimal rank by implementing simple approaches proposed in the literature.

Usage

```
nmfEstimateRank(x, range, method = nmf.getOption("default.algorithm"), nrun = 30, v
plot.NMF.rank(x, what = c('all', 'cophenetic', 'rss', 'residuals'
, 'dispersion', 'evar', 'sparseness'
, 'sparseness.basis', 'sparseness.coef')
, ref=NULL, ...)
```

Arguments

<code>method</code>	A single NMF algorithm, in one of the format accepted by interface <code>nmf</code> .
<code>nrun</code>	a numeric giving the number of run to perform for each value in <code>range</code> .
<code>range</code>	a numeric vector containing the ranks of factorization to try.
<code>ref</code>	reference object of class <code>NMF.rank</code> , as returned by function <code>nmfEstimateRank</code> . The measures contained in <code>ref</code> are used and plotted as a reference. The associated curves are drawn in <i>red</i> , while those from <code>x</code> are drawn in <i>blue</i> .
<code>verbose</code>	toggle verbosity.
<code>what</code>	a character string that partially matches one of the following item: 'all', 'cophenetic', 'rss', 'residuals', 'dispersion'. It specifies which measure must be plotted (<code>what='all'</code> plots all the measures).
<code>x</code>	For <code>nmfEstimateRank</code> a target object to be estimated, in one of the format accepted by interface <code>nmf</code> . For <code>plot.NMF.rank</code> an object of class <code>NMF.rank</code> as returned by function <code>nmfEstimateRank</code> .
<code>...</code>	For <code>nmfEstimateRank</code> , these are extra parameters passed to interface <code>nmf</code> . Note that the same parameters are used for each value of the rank. See <code>nmf</code> . For <code>plot.NMF.rank</code> , these are extra graphical parameter passed to the standard function <code>plot</code> . See <code>plot</code> .

Details

Given a NMF algorithm and the target matrix, a common way of estimating r is to try different values, compute some quality measures of the results, and choose the best value according to this quality criteria. See *Brunet et al. (2004)* and *Hutchins et al. (2008)*.

The function `nmfEstimateRank` allow to launch this estimation procedure. It performs multiple NMF runs for a range of rank of factorization and, for each, returns a set of quality measures together with the associated consensus matrice.

Value

A S3 object (i.e. a list) of class `NMF.rank` with the following slots:

<code>measures</code>	a <code>data.frame</code> containing the quality measures for each rank of factorizations in <code>range</code> . Each row correspond to a measure, each column to a rank.
<code>consensus</code>	a list of consensus matrices, indexed by the rank of factorization (as a character string).

Author(s)

Renaud Gaujoux <renaud@cbio.uct.ac.za>

References

Metagenes and molecular pattern discovery using matrix factorization Brunet, J.-P., Tamayo, P., Golub, T.-R., and Mesirov, J.-P. (2004) Proc Natl Acad Sci U S A 101(12), 4164–4169.

See Also

nmf

Examples

```

set.seed(123456)
n <- 50; r <- 3; m <- 20
V <- syntheticNMF(n, r, m, noise=TRUE)

# Use a seed that will be set before each first run
## Not run: res.estimate <- nmfEstimateRank(V, seq(2,5), method='brunet', nrun=10, seed=123456)

# plot all the measures
## Not run: plot(res.estimate)
# or only one: e.g. the cophenetic correlation coefficient
## Not run: plot(res.estimate, 'cophenetic')

```

NMFfit-class

*Base Class for to store Nonnegative Matrix Factorisation results***Description**

Base class to handle the results of general **Non-negative Matrix Factorisation** algorithms (NMF).

It provides a general structure and generic functions to manage the results of NMF algorithms. It contains a slot with the fitted NMF model (see slot `fit`) as well as data about the methods and parameters used to compute the factorization.

Details

The purpose of this class is to handle in a generic way the results of NMF algorithms. Its slot `fit` contains the fitted NMF model as an object of class [NMF](#).

Other slots contains data about how the factorization has been computed, such as the algorithm and seeding method, the computation time, the final residuals, etc...

Class `NMFfit` acts as a wrapper class for its slot `fit`. It inherits from interface class [NMF](#) defined for generic NMF models. Therefore, all the methods defined by this interface can be called directly on objects of class `NMFfit`. The calls are simply dispatched on slot `fit`, i.e. the results are the same as if calling the methods directly on slot `fit`.

Slots

`fit`: An object that inherits from class "NMF". It contains the fitted NMF model.

Note that class "NMF" is a virtual class. The default class for this slot is `NMFstd`, that implements the standard NMF model.

residuals: A "numeric" vector that contains the final residuals or the residuals track between the target matrix and its NMF estimate(s). Default value is `numeric()`.

See method `residuals` for details on accessor methods and main interface `nmf` for details on how to compute NMF with residuals tracking.

method: A single "character" string that contains the name of the algorithm used to fit the model. Default value is `"`.

seed: A single "character" string that contains the name of the seeding method used to seed the algorithm that computed the NMF. Default value is `"`. See `nmf-methods` for more details.

distance: Either a single "character" string that contains the name of the built-in objective function, or a function that measures the residuals between the target matrix and its NMF estimate.

parameters: A "list" that contains the extra parameters specific to the algorithm used to fit the model.

runtime: Object of class "proc_time" that contains various measures of the time spent to fit the model. See `system.time`

options: A "list" that contains the options used to compute the object.

extra: A "list" that contains extra miscellaneous data set by the algorithm. For example it can be used to store extra parameters, without the need to extend the standard NMF model.

Validity checks

The validity method for class `NMFfit` checks

- slot `fit` calling the suitable validity function on this object of class `NMF` (see `NMF` for more details).
- the validity of slot `objective` that must be either a function definition or a *non-empty* character string.

Objects from the Class

Object of class `NMFfit` using the standard operator `new`.

However, there is usually no need to directly create such an object, as interface methods `nmf` and `seed` take care of it.

Methods

Class-specific methods:

algorithm signature(object = "NMFfit"): Access slot method.

algorithm<- signature(object = "NMFfit", value="ANY"): Set slot method.

basis signature(object = "NMFfit"): Extract the matrix of basis vectors from the fitted NMF model. It returns `basis(fit(object))`.

Note that this method is part of the minimum interface for NMF models, as defined by class `NMF`. See `NMF`.

- basis**`<-` signature(object = "NMFfit", value = "matrix"): Sets the matrix of basis vectors of the fitted NMF model. It calls `basis(fit(object), value)`.
 Note that this method is part of the minimum interface for NMF models, as defined by class NMF. See [NMF](#).
- coef** signature(object = "NMFfit"): Extract the matrix of mixture coefficients from the fitted NMF model. It returns `coef(fit(object))`.
 Note that this method is part of the minimum interface for NMF models, as defined by class NMF. See [NMF](#).
- coef**`<-` signature(object = "NMFfit", value = "matrix"): Sets the matrix of mixture coefficients of the fitted NMF model. It calls `coef(fit(object), value)`.
 Note that this method is part of the minimum interface for NMF models, as defined by class NMF. See [NMF](#).
- distance** signature(target = "matrix", x = "NMFfit"): return the value of the loss function given a target matrix and a NMF fit. It calls method `distance` on slot `fit`. If a distance method is NOT supplied in argument `method`, it uses the objective function from slot `objective`.
- errorPlot** signature(x = "NMFfit"): plot the residuals track of the run that computed object x. See function [nmf](#) for details on how to enable residuals tracking.
- fitted** signature(object = "NMFfit"): compute the estimated target matrix according to the NMF model stored in slot `fit`. It actually dispatches the call to slot `fit`, returning `fitted(fit(object))`.
 Note that this method is part of the minimum interface for NMF models, as defined by class NMF. See [NMF](#).
- fit** signature(object = "NMFfit"): return the NMF object stored in slot `fit`.
- fit**`<-` signature(object = "NMFfit", value = "NMF"): set the value of the NMF object stored in slot `fit`.
- objective** signature(object = "NMF"): return slot distance or compute the objective value if a target is passed in argument x.
- objective**`<-` signature(object = "NMF", value = "character"): sets slot distance to a built-in distance metric identified by a character string.
- objective**`<-` signature(object = "NMF", value = "function"): sets slot distance to a custom function. The function should return a single positive numeric and must take the target (a matrix) as its first parameter, and an object that inherits from class NMF as its second parameter. Extra parameters are passed.
- residuals** signature(object = "NMFfit"): Access slot residuals. See function [residuals](#) for details on extra parameters.
- residuals**`<-` signature(object = "NMFfit", value="numeric"): Set slot residuals to value value.
- runtime** signature(object = "NMF"): return the time spent to fit NMF model object. The time is computed using base function [system.time](#) which returns object of class [proc_time](#).
- seeding** signature(object = "NMF"): returns the seeding method used to seed the algorithm that fitted NMF model object. See section [nmf-methods](#).

seeding<- signature(object = "NMF"): sets the seeding method used to seed the algorithm that fitted NMF model object.

show signature(object = "NMF"): standard generic show method for objects of class NMF.

summary signature(x = "NMF"): standard generic summary method for objects of class NMF. It returns a numeric vector that contains the summary of the fitted NMF model (slot fit), plus the computation time and the final residuals.

\$ signature(x = "NMFFit", name): Access element name in slot extra (which is a list) partially matching argument name. It is equivalent to `x@extra[[name, exact=FALSE]]`.

\$<- signature(x = "NMFFit", name, value): Set the value in of element name in slot extra (which is a list). Note that it does not partially match argument name.

Class NMFFit inherits from all the methods defined on class NMF to manipulate and interpret NMF models. For those methods, class NMFFit act as a wrapper class, dispatching the calls to slot fit. Some useful methods are: dim, nbasis, predict, sparseness. See [NMF](#) for more details.

Author(s)

Renaud Gaujoux <renaud@cbio.uct.ac.za>

See Also

Main interface to perform NMF in [nmf-methods](#).

Method [seed](#) to set NMF objects with values suitable to start algorithms with.

Examples

```
# run default NMF algorithm on a random matrix
n <- 50; r <- 3; p <- 20
V <- matrix(runif(n*p), n, p)
res <- nmf(V, r)

# result class is NMFFit
class(res)

# show result
res

# compute summary measures
summary(res)
```


Description

Class that implements the *Nonsmooth Nonnegative Matrix Factorization* (nsNMF) model, required by the Nonsmooth NMF algorithm.

The Nonsmooth NMF algorithm is defined by Pascual-Montano et al. (2006) as a modification of the standard divergence based NMF algorithm (see section Details and references below). It aims at obtaining sparser factor matrices, by the introduction of a smoothing matrix.

Usage

```
## S4 method for signature 'NMFns':
fitted(object, W, H, S, ...)
## S4 method for signature 'NMFns':
smoothing(x, theta)
```

Arguments

<code>x</code>	an object of class <code>NMFns</code>
<code>object</code>	an object of class <code>NMFns</code>
<code>W</code>	the matrix of basis vectors, i.e. the first matrix factor in the non-smooth NMF model.
<code>H</code>	the matrix of mixture coefficients, i.e. the third matrix factor the non-smooth NMF model.
<code>S</code>	the smoothing matrix, i.e. the middle matrix factor in the non-smooth NMF model.
<code>theta</code>	a single numeric to be used as smoothing parameter (see section Details).
<code>...</code>	extra parameters passed to method <code>smoothing</code> . So typically used to pass a value for <code>theta</code> .

Details

The Nonsmooth NMF algorithm is a modification of the standard divergence based NMF algorithm (see [NMF](#)). Given a non-negative $n \times p$ matrix V and a factorization rank r , it fits the following model:

$$V \equiv WS(\theta)H,$$

where:

- W and H are such as in the standard model, that is non-negative matrices of dimension $n \times r$ and $r \times p$ respectively;

- S is a $r \times r$ square matrix whose entries depends on an extra parameter $0 \leq \theta \leq 1$ in the following way:

$$S = (1 - \theta)I + \frac{\theta}{r}11^T,$$

where I is the identity matrix and 1 is a vector of ones.

The interpretation of S as a smoothing matrix can be explained as follows: Let X be a positive, nonzero, vector. Consider the transformed vector $Y = SX$. If $\theta = 0$, then $Y = X$ and no smoothing on X has occurred. However, as $\theta \rightarrow 1$, the vector Y tends to the constant vector with all elements almost equal to the average of the elements of X . This is the smoothest possible vector in the sense of non-sparseness because all entries are equal to the same nonzero value, instead of having some values close to zero and others clearly nonzero.

Algorithm

The Nonsmooth NMF algorithm uses a modified version of the multiplicative update equations in Lee & Seung's method for Kullbach-Leibler divergence minimization. The update equations are modified to take into account the – constant – smoothing matrix. The modification reduces to using matrix WS instead of matrix W in the update of matrix H , and similarly using matrix SH instead of matrix H in the update of matrix W .

After matrix W have been updated, each of its columns is scaled so that it sums up to 1.

Objects from the Class

Object of class `NMFns` can be created using the standard way with operator `new`

However, as for all the classes that extend class `NMFstd`, objects of class `NMFns` should be created using factory method `newNMF` :

```
new('NMFns', theta=0.8)
newNMF(model='NMFns')
newNMF(model='NMFns', theta=0.8)
```

See `newNMF` for more details on how to use the factory method.

Slots

Class `NMFns` extends `NMF` adding a single slot:

```
theta: Single "numeric" that contains the smoothing parameter. Default prototype value is
0.5.
```

Extends

Class "`NMF`", directly.

Methods

fitted signature(object = "NMFns"): returns the estimated target matrix according to the Nonsmooth-NMF model object:

$$\hat{V} = \hat{V}(\theta) = WS(\theta)H$$

Note that this method is part of the minimum interface for NMF model, as defined by class [NMF](#).

smoothing returns the smoothing matrix $S(\theta)$. See section *Details*.

show signature(object = "NMFns"): standard generic show method for objects of class NMFns. It calls the parent class show method (i.e. for class NMF) and add the value of parameter `theta` to the display.

Author(s)

Renaud Gaujoux <renaud@cbio.uct.ac.za>

References

Alberto Pascual-Montano et al. (2006). Nonsmooth Nonnegative Matrix Factorization (nsNMF). *IEEE Transactions On Pattern Analysis And Machine Intelligence* , Vol. 28, No. 3, March 2006 403

See Also

[NMF](#) , [nmf-methods](#)

Examples

```
# create a completely empty NMF object
new('NMFns')

# create a NMF object based on random (compatible) matrices
n <- 50; r <- 3; p <- 20
w <- matrix(runif(n*r), n, r)
h <- matrix(runif(r*p), r, p)
newNMF(model='NMFns', W=w, H=h)

# apply Nonsmooth NMF algorithm to a random target matrix
V <- matrix(runif(n*p), n, p)
## Not run: nmf(V, r, 'ns')
```

NMFOffset-class *Nonnegative Matrix Factorization with Offset*

Description

Class that implements the *Nonnegative Matrix Factorization with Offset* model, required by the NMF with Offset algorithm.

The NMF with Offset algorithm is defined by Badea (2008) as a modification of Lee & Seung's euclidean based NMF algorithm (see section Details and references below). It aims at obtaining 'cleaner' factor matrices, by the introduction of an offset matrix, explicitly modelling a feature specific baseline – constant across samples.

Objects from the Class

Object of class `NMFOffset` can be created using the standard way with operator `new`

However, as for all the classes that extend class `NMFstd`, objects of class `NMFOffset` should be created using factory method `newNMF` :

```
new('NMFOffset')
newNMF(model='NMFOffset')
newNMF(model='NMFOffset', W=w, offset=rep(1, nrow(w)))
```

See `newNMF` for more details on how to use the factory method.

Slots

Class `NMFOffset` extends `NMF` adding a single slot:

offset: A "numeric" vector to handle the common baseline for each feature. Its length will always be equal to the number of features, i.e. the number of rows in slot `W`.

Extends

Class "`NMF`", directly.

Methods

fitted signature(object = "NMFOffset"): returns the estimated target matrix according to the NMF with Offset model object:

$$\hat{V} = WH + offset$$

Note that this method is part of the minimum interface for NMF model, as defined by class `NMF`.

initialize initialize method for class `NMFOffset`. It ensures consistency between slots `W` and slot `offset`.

offset signature(object = "NMFOffset"): return the value of slot `offset`.

rmnf returns the object with slots `W`, `H` and `offset` filled with random values drawn from a uniform distribution. This method first calls method `rmnf` for NMF object to set the entries of slots `W` and `H`, then sets the entries in slot `offset` within the interval $[0, \max(\max(W), \max(H))]$.

show `signature(object = "NMFOffset")`: standard generic `show` method for objects of class `NMFOffset`. It calls the parent class `show` method (i.e. for class `NMF`) and add the value of vector `offset` to the display (only the first five elements are displayed).

Author(s)

Renaud Gaujoux <renaud@cbio.uct.ac.za>

References

Badea (2008). Extracting Gene Expression Profiles Common To Colon And Pancreatic Adenocarcinoma Using Simultaneous Nonnegative Matrix Factorization. In *Pacific Symposium on Biocomputing*, **13**, 279-290

See Also

[NMF](#), [nmf-methods](#)

Examples

```
# create a completely empty NMF object
new('NMFOffset')

# create a NMF object based on random (compatible) matrices
n <- 50; r <- 3; p <- 20
w <- matrix(runif(n*r), n, r)
h <- matrix(runif(r*p), r, p)
newNMF(model='NMFOffset', W=w, H=h, offset=rep(0.5, nrow(w)))

# apply Nonsmooth NMF algorithm to a random target matrix
V <- matrix(runif(n*p), n, p)
## Not run: nmf(V, r, 'offset')
```

NMFSet-class

Class to store results from multiple runs of NMF algorithms

Description

Class that extends base class `list` to store the result from a multiple run of NMF algorithms. The elements are of class `NMF`.

Slots

- consensus:** Object of class "matrix" used to store the consensus matrix when multiple runs have been performed with option `keep.all=FALSE`. In this case, only the best factorization is returned, so the object is of length 1. However the consensus matrix across all runs is still computed and stored in this slot.
- nrun:** an integer that contains the number of runs when NMF is performed with option `keep.all=FALSE`. See [nmf](#).
- runtime:** Object of class "proc_time" that contains various measures of the time spent to perform all the runs.
- .Data:** standard slot that contains the S3 list object data. See R documentation on S4 classes for more details.

Methods

- cophcor** signature(object = "NMFSet"):
Computes the cophenetic correlation coefficient of the consensus matrix associated to the multiple NMF runs described by object. It's been proposed by *Brunet et al. (2004)* to measure the stability of the clusters obtained by NMF approaches. See [cophcor](#) for more details.
- compare** signature(x = "NMFSet"): compute summary measures for each element in the list and return them in rows in a data.frame.
- connectivity** signature(x = "NMFSet"): compute the consensus matrix of the set of results. That is the mean of the connectivity matrices of each element. See [connectivity](#).
- dispersion** signature(object = "NMFSet"):
Computes the dispersion coefficient of the consensus matrix associated to the multiple NMF runs described by object. It's been proposed by *Kim and Park (2007)* to measure the reproducibility of the clusters. See [dispersion](#) for more details.
- entropy** signature(x = "NMFSet", class = "ANY"): computes the mean entropy of the set of NMF results. See [entropy](#).
- featureNames** signature(object = "NMFSet"): returns the row names of the basis matrix from the best fit of the set of results. If BioConductor is installed this method is defined for the generic function [featureNames](#) from the Biobase package.
- fit** signature(object = "NMFSet"): returns the element that achieves the lowest residual approximation error.
- residuals** signature(object = "NMFSet"): returns the average residuals of the set of results.
- errorPlot** signature(x = "NMFSet"): plot on a single graphic the residuals tracks for each of the run. The tracks are normalized against their maximum value, to generate tracks that lies in range [0, 1].
- join** signature(x = "list"): method to create a NMFSet object from a list of NMFfit or NMFSet objects.
- nrun** signature(object = "NMFSet"): returns the number of runs performed to create object. This is not necessarily the length of object that is the number of NMF fits contained in object, nor always the value of slot nrun.

When performing multiple NMF runs with option `keep.all=FALSE`, only one fit is kept so `length(object)=1`, but the number of runs is stored in slot `nrun`. When `keep.all=TRUE`, slot `nrun` is not set at all.

See [nmf](#).

metaHeatmap Produces a heatmap of the consensus matrix using function [heatmap.2](#). See [metaHeatmap](#).

predict `signature(object = "NMFSet")`: returns a factor that gives the predicted cluster index for each sample (resp. for each feature) based on the *best* NMF factorization in the set of results `object`. The index correspond to the basis vector that most contribute to the sample (resp. to which the feature contribute the most). See [predict](#) details on extra arguments.

purity `signature(x = "NMFSet", class = "ANY")`: computes the mean purity of the set of NMF results. See [purity](#).

rss `signature(object = "NMFSet")`: computes the Residual Sum of Squares (RSS) of the best factorization in the set. See [rss](#).

runtime `signature(object = "NMFSet")`: returns the time spent to compute all the runs.

sampleNames `signature(object = "NMFSet")`: returns the column names of the mixture coefficient matrix from the best fit of the set of results. If BioConductor is installed this method is defined for the generic function [featureNames](#) from the Biobase package.

show `signature(object = "NMFSet")`: show method for class NMFSet.

summary `signature(x = "NMFSet")`: standard generic summary method for objects of class NMFSet. It computes a set of measures to evaluate the quality of the best factorization of the set. The result is similar to the result from the `summary` method of `NMFfit` objects. See [NMFfit](#) for details on the computed measures. In addition, the cophenetic correlation coefficient and the dispersion coefficient of the consensus matrix are returned. See methods `cophcor` and `dispersion` above.

Author(s)

Renaud Gaujoux <renaud@cbio.uct.ac.za>

References

Metagenes and molecular pattern discovery using matrix factorization Brunet, J.-P., Tamayo, P., Golub, T.-R., and Mesirov, J.-P. (2004) Proc Natl Acad Sci U S A 101(12), 4164–4169.

Sparse non-negative matrix factorizations via alternating non-negativity-constrained least squares for microarray data analysis Kim, H. & Park, H. (2007) Bioinformatics. <http://dx.doi.org/10.1093/bioinformatics/btm134>.

See Also

[NMF](#), [nmf-methods](#), [NMF-utils](#)

Examples

```
# generate a synthetic dataset with known classes
n <- 50; counts <- c(5, 5, 8);
V <- syntheticNMF(n, counts, noise=TRUE)

# build the class factor
groups <- as.factor(do.call('c', lapply(seq(3), function(x) rep(x, counts[x]))))

# perform multiple runs of one algorithm
res <- nmf(V, 3, nrun=20)
res

# plot a heatmap of the consensus matrix
## Not run: metaHeatmap(res)

# run multiple methods
res <- nmf(V, 3, list('brunet', 'lee', 'nsNMF'))
# compare the results passing prior knowledge on classes
compare(res, class=groups)
```

NMFstd-class

Implement of the standard NMF model

Description

Class that implements the standard model of Nonnegative Matrix Factorisation.

It provides a general structure and generic functions to manage factorizations that follow NMF standard model.

Details

Let V be a $n \times m$ non-negative matrix and r a positive integer. In its standard form (see references below), a NMF of V is commonly defined as a pair of matrices (W, H) such that:

$$V \equiv WH,$$

where:

- W and H are $n \times r$ and $r \times m$ matrices respectively with non-negative entries;
- \equiv is to be understood with respect to some loss function. Common choices of loss functions are based on Frobenius norm or Kullback-Leibler divergence.

Integer r is called the *factorization rank*. Depending on the context of application of NMF, the columns of W and H take different names:

columns of W basis vector, metagenes, factors, source, image basis

columns of H mixture coefficients, metagenes expression profiles, weights

NMF approach has been successfully applied to several fields. Package NMF was implemented trying to use names as generic as possible for objects and methods. The following terminology is used:

samples the columns of the target matrix V

features the rows of the target matrix V

basis matrix the first matrix factor W

basis vectors the columns of first matrix factor W

mixture matrix the second matrix factor H

mixtures coefficients the columns of second matrix factor H

However, because package NMF was primarily implemented to work with gene expression microarray data, it also provides a layer to easily and intuitively work with objects from the Bioconductor base framework. See [NMF-bioc](#) for more details.

Slots

W: A "matrix" that contains the *first* matrix factor of the factorisation

H: A "matrix" that contains the *second* matrix factor of the factorisation

Validity checks

The validity method for class NMF checks for compatibility of slots W and H, as those matrices must be compatible with respect to the matrix product. It also checks the relevance of the factorisation, and throws a warning when the factorisation rank is greater than the number of columns in H.

Objects from the Class

Factory method

The more convenient way of creating NMF objects is to use factory method `newNMF`:

```
newNMF(rank, target=0, model='NMFstd', ...)
```

It provides a unique interface to create NMF objects that can follow different NMF models. Using this interface allows to automatically resolve some inconsistencies in the matrices dimensions. The standard model is the default model, so the following calls will create objects of class NMFstd:

```
newNMF()
```

This creates a completely empty NMFstd object, where both slots W and H are of dimension 0x0.

```
newNMF(3)
```

This creates an empty NMFstd object with factorization rank of 3. Slots W and H are of dimension 0x3 and 3x0 respectively.

```
newNMF(3, c(50,10))
```

This creates a NMFstd object to fit a target matrix of dimension 50x10, with factorization rank of 3. Slots W and H are set to dimension 50x3 and 3x10 respectively, and filled with NAs.

```
newNMF(W=w)
```

```
newNMF (H=h)
newNMF (W=w, H=h)
```

This creates a `NMFstd` object, where slots `W` and/or `H` are provided as matrices. When only one of the matrices is provided, a compatible NA-filled matrix is created for the missing slot.

When `W` and `H` are both provided, the `NMFstd` object created is suitable to seed a NMF algorithm. Note that it implicitly sets the factorisation rank to the number of columns in `W`.

Note that when not created as results of algorithm methods, only slots `W` and `H` are usually set. The remaining slots would be automatically set by method `nmf` with data about the way the factorisation was computed.

Standard way

Objects can still be created by calls of the usual form:

```
new ("NMF")
new ("NMF", W=w, H=h)
```

Methods

distance signature(target = "matrix", x = "NMF"): return the value of the loss function given a target matrix and a NMF fit.

fitted signature(object = "NMF"): compute the estimated target matrix according to the standard NMF model object, i.e. as the matrix product of slots `W` and `H`.

Note that this method is part of the minimum interface for NMF model, as defined by class `NMF`.

basis signature(object = "NMF"): Returns slot `W`, the matrix of basis vectors in NMF model object.

Note that this method is part of the minimum interface for NMF models, as defined by class `NMF`. See `NMF`.

basis<- signature(object = "NMF", value = "matrix"): Sets the value of slot `W`, the matrix of basis vectors in NMF model object.

Note that this method is part of the minimum interface for NMF models, as defined by class `NMF`. See `NMF`.

coef signature(object = "NMF"): Returns slot `H`, the matrix of mixture coefficients in the NMF model object.

Note that this method is part of the minimum interface for NMF models, as defined by class `NMF`. See `NMF`.

coef<- signature(object = "NMF", value = "matrix"): Set the value of slot `H`, the matrix of mixture coefficients in the NMF model object.

Note that this method is part of the minimum interface for NMF models, as defined by class `NMF`. See `NMF`.

nmf signature(x = "NMF", target): seed NMF model `x` with random values drawn from a random distribution. If a target is specified as a matrix, then the values are drawn within the interval `[0, max(target)]`.

show signature(object = "NMF"): standard generic show method for objects of class `NMF`.

summary signature(x = "NMF"): standard generic summary method for objects of class NMF.

Class NMFstd inherits from all the methods defined on class NMF to manipulate and interpret NMF models. Some useful are: dim, nbasis, predict, sparseness. See [NMF](#) for more details.

Author(s)

Renaud Gaujoux <renaud@cbio.uct.ac.za>

References

Definition of Nonnegative Matrix Factorization in its modern formulation:

Lee D.D. and Seung H.S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*, **401**, 788–791.

Historical first definition and algorithms:

Paatero, P., Tapper, U. (1994). Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics*, **2**, 111–126, doi:10.1002/env.3170050203.

Reference for some utility functions:

Kim, H. and Park, H. (2007). Sparse non-negative matrix factorizations via alternating non-negativity-constrained least squares for microarray data analysis. *Bioinformatics*.

Hoyer (2004). Non-negative matrix factorization with sparseness constraints. *Journal of Machine Learning Research*, **5**, 1457-1469.

See Also

Main interface to perform NMF in [nmf-methods](#).

Method [seed](#) to set NMF objects with values suitable to start algorithms with.

Examples

```
# create a completely empty NMF object (i.e. 0 features, 0 basis components, 0 samples)
new('NMFstd')

# create a NMF object based on one random matrix: the missing matrix is deduced
# Note this only works when using factory method NMF
n <- 50; r <- 3;
w <- matrix(runif(n*r), n, r)
newNMF(W=w)

# create a NMF object based on random (compatible) matrices
p <- 20
h <- matrix(runif(r*p), r, p)
newNMF(W=w, H=h)

# create a NMF object based on incompatible matrices: generate an error
h <- matrix(runif((r+1)*p), r+1, p)
## Not run: new('NMFstd', W=w, H=h)
```

```
# same thing using the factory method: dimensions are corrected and a warning
# is thrown saying that the dimensions used are reduced
newNMF(W=w, H=h)

# apply default NMF algorithm to a random target matrix
V <- matrix(runif(n*p), n, p)
## Not run: nmf(V, r)
```

Index

*Topic **classes**

- NMF-class, 8
- NMFfit-class, 29
- NMFns-class, 32
- NMFOffset-class, 35
- NMFSet-class, 37
- NMFstd-class, 40

*Topic **cluster**

- nmf-methods, 14

*Topic **datasets**

- esGolub, 6

*Topic **math**

- nmf-methods, 14

*Topic **methods**

- basis-methods, 5
- NMF-bioc, 8
- nmf-methods, 14

*Topic **misc**

- NMF-deprecated, 13

*Topic **optimize**

- nmf-methods, 14

*Topic **package**

- NMF-package, 2

- [, NMF-method (NMF-class), 8
- \$ (NMFfit-class), 29
- \$, NMF-method (Internals), 7
- \$, NMFfit-method (NMFfit-class), 29
- \$<- (NMFfit-class), 29
- \$<-, NMF-method (Internals), 7
- \$<-, NMFfit-method (NMFfit-class), 29

advanced, 3

algorithm (NMFfit-class), 29

algorithm, NMFfit-method
(NMFfit-class), 29

algorithm, NMFSet-method
(NMFSet-class), 37

algorithm, NMFStrategyFunction-method
(Internals), 7

algorithm<- (NMFfit-class), 29

algorithm<-, NMFfit, ANY-method
(NMFfit-class), 29

algorithm<-, NMFStrategyFunction, character-method
(Internals), 7

algorithm<-, NMFStrategyFunction, function-method
(Internals), 7

apply, 24

basis, 8, 10

basis (basis-methods), 5

basis, NMF-method (basis-methods),
5

basis, NMFfit-method
(NMFfit-class), 29

basis, NMFstd-method
(NMFstd-class), 40

basis-methods, 5

basis<- (basis-methods), 5

basis<-, NMF, matrix-method
(basis-methods), 5

basis<-, NMFfit, matrix-method
(NMFfit-class), 29

basis<-, NMFstd, matrix-method
(NMFstd-class), 40

basis<-methods (basis-methods), 5

basis<-, 8

Biobase, 2

callNextMethod, 12

clusters (NMF-deprecated), 13

clusters, matrix-method
(NMF-deprecated), 13

clusters, NMF-method
(NMF-deprecated), 13

clusters-methods
(NMF-deprecated), 13

coef, 5, 8, 10

coef (basis-methods), 5

coef, NMF-method (basis-methods), 5

- coef, NMFfit-method
(NMFfit-class), 29
- coef, NMFstd-method
(NMFstd-class), 40
- coef<- (basis-methods), 5
- coef<-, NMF, matrix-method
(basis-methods), 5
- coef<-, NMFfit, matrix-method
(NMFfit-class), 29
- coef<-, NMFstd, matrix-method
(NMFstd-class), 40
- coef<-methods (basis-methods), 5
- coef<-, 8, 10
- coefficients, 10
- coefficients (basis-methods), 5
- coefficients, NMF-method
(basis-methods), 5
- coefficients<-, NMF, matrix-method
(basis-methods), 5
- colnames, 12
- compare (NMFSet-class), 37
- compare, list-method
(NMFSet-class), 37
- compare, NMFSet-method
(NMFSet-class), 37
- connectivity, 38
- connectivity (NMF-utils), 20
- connectivity, NMF-method
(NMF-utils), 20
- connectivity, NMFSet-method
(NMFSet-class), 37
- connectivity-methods (NMF-utils), 20
- cophcor, 37
- cophcor (NMF-utils), 20
- cophcor, matrix-method
(NMF-utils), 20
- cophcor, NMFSet-method
(NMFSet-class), 37
- cophcor-methods (NMF-utils), 20
- dim, NMF-method (NMF-class), 8
- dispersion, 38
- dispersion (NMF-utils), 20
- dispersion, matrix-method
(NMF-utils), 20
- dispersion, NMFSet-method
(NMFSet-class), 37
- dispersion-methods (NMF-utils), 20
- distance, 10
- distance (NMF-class), 8
- distance, matrix, NMF-method
(NMF-class), 8
- distance, matrix, NMFfit-method
(NMFfit-class), 29
- distance, missing, missing-method
(NMF-class), 8
- entropy, 11, 38
- entropy (NMF-utils), 20
- entropy, factor, factor-method
(NMF-utils), 20
- entropy, NMF, factor-method
(NMF-utils), 20
- entropy, NMFSet, ANY-method
(NMFSet-class), 37
- entropy, table, missing-method
(NMF-utils), 20
- entropy-methods (NMF-utils), 20
- errorPlot, 4
- errorPlot (NMFfit-class), 29
- errorPlot, NMFfit-method
(NMFfit-class), 29
- errorPlot, NMFSet-method
(NMFSet-class), 37
- esGolub, 6
- evan, 11
- evan (NMF-utils), 20
- evan, NMF-method (NMF-utils), 20
- evan-methods (NMF-utils), 20
- ExpressionSet, 16
- ExpressionSet-class, 6
- extractFeatures (NMF-utils), 20
- extractFeatures, NMF-method
(NMF-utils), 20
- factor, 6
- featureNames, 11, 38
- featureNames (NMF-class), 8
- featureNames, NMF-method
(NMF-class), 8
- featureNames, NMFSet-method
(NMFSet-class), 37
- featureNames<- (NMF-class), 8
- featureNames<-, NMF-method
(NMF-class), 8
- featureNames<-, 11
- featureScore (NMF-utils), 20

- featureScore, matrix-method
(*NMF-utils*), 20
- featureScore, NMF-method
(*NMF-utils*), 20
- fit (*NMFfit-class*), 29
- fit, NMFfit-method (*NMFfit-class*),
29
- fit, NMFSet-method (*NMFSet-class*),
37
- fit-methods (*NMFfit-class*), 29
- fit<- (*NMFfit-class*), 29
- fit<-, NMFfit, NMF-method
(*NMFfit-class*), 29
- fitted, NMF-method (*NMF-class*), 8
- fitted, NMFfit-method
(*NMFfit-class*), 29
- fitted, NMFns-method
(*NMFns-class*), 32
- fitted, NMFOffset-method
(*NMFOffset-class*), 35
- fitted, NMFstd-method
(*NMFstd-class*), 40
- foreach, 16
- getOption, 3
- heatmap.2, 11, 23, 24, 38
- heatmap.plus, 24
- initialize, NMFOffset-method
(*NMFOffset-class*), 35
- Internals, 7
- is.empty.nmf (*NMF-class*), 8
- is.empty.nmf, NMF-method
(*NMF-class*), 8
- join (*NMFSet-class*), 37
- join, list-method (*NMFSet-class*),
37
- metagenes (*NMF-bioc*), 8
- metagenes, NMF-method (*NMF-bioc*), 8
- metagenes-methods (*NMF-bioc*), 8
- metagenes<- (*NMF-bioc*), 8
- metagenes<-, NMF, matrix-method
(*NMF-bioc*), 8
- metaHeatmap, 11, 38
- metaHeatmap (*NMF-utils*), 20
- metaHeatmap, matrix-method
(*NMF-utils*), 20
- metaHeatmap, NMF-method
(*NMF-utils*), 20
- metaHeatmap, NMFSet-method
(*NMFSet-class*), 37
- metaHeatmap-methods (*NMF-utils*),
20
- metaprofiles (*NMF-bioc*), 8
- metaprofiles, NMF-method
(*NMF-bioc*), 8
- metaprofiles-methods (*NMF-bioc*), 8
- metaprofiles<- (*NMF-bioc*), 8
- metaprofiles<-, NMF, matrix-method
(*NMF-bioc*), 8
- name, NMFSeed-method (*Internals*), 7
- name, NMFStrategy-method
(*Internals*), 7
- name<-, NMFSeed, character-method
(*Internals*), 7
- name<-, NMFStrategy, character-method
(*Internals*), 7
- nbasis, 8
- nbasis (*NMF-class*), 8
- nbasis, NMF-method (*NMF-class*), 8
- new, 30, 34, 35
- newNMF, 10, 11, 13, 34, 35
- newNMF (*NMFstd-class*), 40
- newNMF, missing, ANY-method
(*NMFstd-class*), 40
- newNMF, numeric, matrix-method
(*NMFstd-class*), 40
- newNMF, numeric, missing-method
(*NMFstd-class*), 40
- newNMF, numeric, numeric-method
(*NMFstd-class*), 40
- newNMF-methods (*NMFstd-class*), 40
- nmeta (*NMF-bioc*), 8
- nmeta, NMF-method (*NMF-bioc*), 8
- nmeta-methods (*NMF-bioc*), 8
- NMF, 3, 5–7, 13, 14, 16, 18, 20, 22, 23, 26,
29–36, 39, 42
- NMF (*NMF-package*), 2
- nmf, 2, 3, 25, 27–31, 37, 38, 41
- nmf (*nmf-methods*), 14
- NMF Algorithms (*nmf-methods*), 14
- nmf, data.frame, ANY, ANY-method
(*nmf-methods*), 14
- nmf, ExpressionSet, ANY, ANY-method
(*nmf-methods*), 14

- nmf, matrix, ANY, ANY-method
(*nmf-methods*), 14
- nmf, matrix, numeric, character-method
(*nmf-methods*), 14
- nmf, matrix, numeric, function-method
(*nmf-methods*), 14
- nmf, matrix, numeric, list-method
(*nmf-methods*), 14
- nmf, matrix, numeric, NMFStrategy-method
(*nmf-methods*), 14
- NMF-bioc, 17
- NMF-class, 2
- nmf-methods, 7, 13, 30–32, 35, 36, 39, 43
- NMF-utils, 18, 39
- NMF-advanced (*advanced*), 3
- NMF-bioc, 8, 40
- NMF-class, 8
- NMF-deprecated, 13
- NMF-internals (*Internals*), 7
- nmf-methods, 14
- NMF-package, 2
- NMF-utils, 20
- nmf.getOption (*advanced*), 3
- nmf.models (*advanced*), 3
- nmf.options (*advanced*), 3
- nmfAlgorithm (*Internals*), 7
- nmfApply (*NMF-utils*), 20
- nmfApply, NMF-method (*NMF-utils*),
20
- nmfEstimateRank, 27
- NMFfit, 22, 25, 39
- NMFfit-class, 29
- NMFns, 9
- NMFns-class, 32
- NMFOffset, 9
- NMFOffset-class, 35
- nmfSeed (*Internals*), 7
- NMFSet, 16
- NMFSet-class, 37
- NMFstd, 9, 14, 34, 35
- NMFstd-class, 40
- nrun (*NMFSet-class*), 37
- nrun, NMFSet-method
(*NMFSet-class*), 37
- objective (*NMFfit-class*), 29
- objective, NMFfit-method
(*NMFfit-class*), 29
- objective, NMFStrategy-method
(*Internals*), 7
- objective<- (*NMFfit-class*), 29
- objective<-, NMFfit, character-method
(*NMFfit-class*), 29
- objective<-, NMFfit, function-method
(*NMFfit-class*), 29
- objective<-, NMFStrategy, character-method
(*Internals*), 7
- objective<-, NMFStrategy, function-method
(*Internals*), 7
- offset, NMFOffset-method
(*NMFOffset-class*), 35
- options, 3, 4
- plot, 28
- plot.NMF.rank (*nmfEstimateRank*),
27
- predict, 11, 13, 38
- predict (*NMF-utils*), 20
- predict, matrix-method
(*NMF-utils*), 20
- predict, NMF-method (*NMF-utils*), 20
- predict, NMFSet-method
(*NMFSet-class*), 37
- predict-methods (*NMF-utils*), 20
- predict.nmf (*NMF-utils*), 20
- proc_time, 31
- purity, 38
- purity (*NMF-utils*), 20
- purity, factor, factor-method
(*NMF-utils*), 20
- purity, NMF, factor-method
(*NMF-utils*), 20
- purity, NMFSet, ANY-method
(*NMFSet-class*), 37
- purity, table, missing-method
(*NMF-utils*), 20
- purity-methods (*NMF-utils*), 20
- randomize (*NMF-utils*), 20
- residuals, 25, 29, 31
- residuals (*NMF-utils*), 20
- residuals, NMFfit-method
(*NMF-utils*), 20
- residuals, NMFSet-method
(*NMFSet-class*), 37
- residuals<- (*NMFfit-class*), 29

`residuals<-`, `NMFfit`-method
 (`NMFfit-class`), 29
`residuals<-`methods
 (`NMFfit-class`), 29
`rnmf` (`NMF-class`), 8
`rnmf`, `NMF`, matrix-method
 (`NMF-class`), 8
`rnmf`, `NMF`, missing-method
 (`NMF-class`), 8
`rnmf`, `NMF`, numeric-method
 (`NMF-class`), 8
`rnmf`, `NMFOffset`, numeric-method
 (`NMFOffset-class`), 35
`rnmf`-methods (`NMF-class`), 8
`rownames`, 11
`rss`, 12, 38
`rss` (`NMF-utils`), 20
`rss`, `NMF`-method (`NMF-utils`), 20
`rss`, `NMFSet`-method (`NMFSet-class`),
 37
`runtime` (`NMFfit-class`), 29
`runtime`, `NMFfit`-method
 (`NMFfit-class`), 29
`runtime`, `NMFSet`-method
 (`NMFSet-class`), 37

`sampleNames`, 12
`sampleNames` (`NMF-class`), 8
`sampleNames`, `NMF`-method
 (`NMF-class`), 8
`sampleNames`, `NMFSet`-method
 (`NMFSet-class`), 37
`sampleNames<-` (`NMF-class`), 8
`sampleNames<-`, `NMF`, ANY-method
 (`NMF-class`), 8
`sampleNames<-`, `NMF`-method
 (`NMF-class`), 8
`sampleNames<-`, 12
`seed`, 13, 30, 32, 43
`seed` (`nmf-methods`), 14
`seeding` (`NMFfit-class`), 29
`seeding`, `NMFfit`-method
 (`NMFfit-class`), 29
`seeding`-methods (`NMFfit-class`), 29
`seeding<-` (`NMFfit-class`), 29
`seeding<-`, `NMFfit`-method
 (`NMFfit-class`), 29
`set.seed`, 17
`show`, `NMF`-method (`NMF-class`), 8

`show`, `NMFfit`-method
 (`NMFfit-class`), 29
`show`, `NMFns`-method (`NMFns-class`),
 32
`show`, `NMFOffset`-method
 (`NMFOffset-class`), 35
`show`, `NMFSeed`-method (`Internals`), 7
`show`, `NMFSet`-method
 (`NMFSet-class`), 37
`show`, `NMFStrategy`-method
 (`Internals`), 7
`show`, `NMFStrategyIterative`-method
 (`Internals`), 7
`smoothing` (`NMFns-class`), 32
`smoothing`, `NMFns`-method
 (`NMFns-class`), 32
`smoothing`-methods (`NMFns-class`),
 32
`sparseness` (`NMF-utils`), 20
`sparseness`, matrix-method
 (`NMF-utils`), 20
`sparseness`, `NMF`-method
 (`NMF-utils`), 20
`sparseness`, numeric-method
 (`NMF-utils`), 20
`sparseness`-methods (`NMF-utils`), 20
`summary`, 26
`summary`, `NMF`-method (`NMF-class`), 8
`summary`, `NMFfit`-method
 (`NMFfit-class`), 29
`summary`, `NMFSet`-method
 (`NMFSet-class`), 37
`syntheticNMF` (`NMF-utils`), 20
`system.time`, 30, 31