

# Package ‘rsdr’

February 1, 2021

**Type** Package

**Title** Re-sampled dimensional reduction (RSDR)

**Version** 0.1.0

**Date** 2021-01-21

**Depends** R (>= 4.0.2)

**Description** This package applies a resampling method to estimate rotated matrix for dimensional reduction. By applying the procedure, a number of new dimensions can be used as feature candidates for predictive modeling; thus, the number of candidates can be optimized depending on the training sample size. This helps to fulfill minimum events per variable (EPV) for a machine learning algorithm while optimizing the proportion of variance explained (PVE). Unlike other packages for dimensional reduction, this package applies resampling methods to prevent overfitting. The PVE optimization takes the predicted outcome into account without using it to represent the features.

**License** GPL-3

**LazyData** true

**Roxygen** list(markdown = TRUE)

**RoxygenNote** 7.1.1

**Imports** devtools,  
tidyverse,  
pbapply,  
parallel,  
BiocGenerics,  
lubridate,  
Biobase,  
caret,  
MLeval

**Suggests** BiocStyle,  
knitr,  
rmarkdown,  
kableExtra,  
dslabs,  
magick,  
testthat

**URL** <https://github.com/herdiantrisufriyana/rsdr>

**BugReports** <https://github.com/herdiantrisufriyana/rsdr/issues>  
**VignetteBuilder** knitr

R topics documented:

composition . . . . .	2
rsdr . . . . .	3
transformation . . . . .	4
<b>Index</b>	<b>7</b>

---

composition	<i>Get table of estimated weights of transformation</i>
-------------	---

---

**Description**

This function get table of estimated weights to transform the input to output dimensions by rsdr() function.

**Usage**

composition(rsdr\_object, dimensions = NULL)

**Arguments**

- rsdr\_object      RSDR object, a list of results and parameters. This is an output from rsdr() function.
- dimensions      Dimension index, a vector of integers for the column indices of the table, indicating the output dimensions transformed by rsdr(). The maximum integer is the minimum number of output dimensions among many dimensional reduction models from differen resampled subsets.

**Value**

RSDR object, a list of results and parameters. Use plot() to visualize weights that are used to transformed all input dimension to each output dimension.

**Examples**

```
## Create input example
library(medhist)
data(medhistdata)
ps_removal=extract_nps_mh(medhistdata)

mh_bin_nps=
  medhistdata[ps_removal_train$key,] %>%
  `exprs<-`(
    exprs(.) %>%
    t() %>%
    as.data.frame() %>%
    rownames_to_column(var='id') %>%
```

```

        column_to_rownames(var='id') %>%
        t()
    ) %>%
    trans_binary(verbose=F)

input=
  mh_bin_nps %>%
  exprs() %>%
  t() %>%
  as.data.frame()

## Fit dimensional reduction models with resampling
rsdr_bin_nps=rsdr(input,'CV',10,'PCA')

## Get table of estimated weights of transformation
composition(rsdr_bin_nps)

```

rsdr

*Fit dimensional reduction models with resampling*

## Description

This function fits a dimensional reduction model with resampling. Dataset is resampled a number of times. A model using the same dimensional reduction algorithm is fitted each time. These models can be used to transform dataset into new dimensions using an estimated weight per a pair of input and output dimension. The output dimensions are sorted from the highest to the lowest proportion of variance explained (PVE). Thus, one can choose dimensions with top PVE as the feature candidates for developing a prediction model.

## Usage

```

rsdr(
  data,
  rs_method = c("BS", "CV"),
  rs_number = c(30, 10),
  dr_method = c("PCA", "SVD"),
  sd_cutoff = 0,
  state = 33,
  cl = 1
)

```

## Arguments

data	Input data, a data frame with rows of samples and columns of variables that will be transformed.
rs_method	Resampling method, a character of BS for bootstrapping or CV for k-fold cross-validation.
rs_number	Resampling time/fold, an integer of any number. A common number for bootstrapping and cross-validation are 30 and 10, respectively.
dr_method	Dimensional reduction method, a character of PCA for principal componen analysis or SVD for singular value decomposition.

<code>sd_cutoff</code>	Standard deviation cutoff, a non-negative numeric of which a variable is excluded if the standard deviation is equal to this number or lower. This number is conceivably 0 if all values in a variable are the same. This situation (i.e. zero variance) is not allowed for dimensional reduction.
<code>state</code>	An integer to set random seed for reproducible results.
<code>cl</code>	Parallel cluster, a non-negative integer of number of CPU cluster used for computation in parallel. Set to 1 if no parallelism is expected.

### Value

RSDR object, a list of results and parameters. Use `plot()` to visualize weights that are used to transformed all input dimension to each output dimension.

### Examples

```
## Create input example
library(medhist)
data(medhistdata)
ps_removal=extract_nps_mh(medhistdata)

mh_bin_nps=
  medhistdata[ps_removal_train$key,] %>%
  `exprs<-`(
    exprs(.) %>%
    t() %>%
    as.data.frame() %>%
    rownames_to_column(var='id') %>%
    column_to_rownames(var='id') %>%
    t()
  ) %>%
  trans_binary(verbose=F)

input=
  mh_bin_nps %>%
  exprs() %>%
  t() %>%
  as.data.frame()

## Fit dimensional reduction models with resampling
rsdr_bin_nps=rsdr(input,'CV',10,'PCA')

## Show fitting results
rsdr_bin_nps

## Plot weights to transform dimensions
plot(rsdr_bin_nps_train)
```

## Description

This function transforms input dimensions using a re-sampled dimensional reduction model fitted by `rsdr()`.

## Usage

```
transformation(
  tidy_set,
  rsdr_object,
  dimensions = NULL,
  input_dim = NULL,
  top_n = NULL,
  output_dim = NULL,
  verbose = T
)
```

## Arguments

<code>tidy_set</code>	A TidySet (i.e. ExpressionSet) containing the visits of subjects in outcome dataset, paid by any payment systems. This TidySet also accomodates outcome dataset. This is an output of <code>compile_mh_outcome()</code> from <code>medhist</code> package, or manually compiled using <code>ExpressionSet()</code> from <code>Biobase</code> package.
<code>rsdr_object</code>	RSDR object, a list of results and parameters. This is an output from <code>rsdr()</code> function.
<code>dimensions</code>	Dimension index, a vector of integers for the column indices of the table, indicating the output dimensions transformed by <code>rsdr()</code> . The maximum integer is the minimum number of output dimensions among many dimensional reduction models from differen resampled subsets.
<code>input_dim</code>	Input dimension, a vector of characters containing the names of input dimensions that are selected to transform into output dimensions.
<code>output_dim</code>	Output dimension, a vector of characters containing the names of output dimensions that are selected after transformation.
<code>verbose</code>	Verbosity, a logical indicating whether progress should be shown.

## Value

A TidySet (i.e. ExpressionSet) containing the transformed table accessed using `exprs()` function from `Biobase` package. Composition of input dimensions as a weight table can be accessed using `fData()` from the same package. RSDR models and proportion of variance explained (PVE) using `input_dim` and `output_dim` can be accessed using `preproc()`.

## Examples

```
## Create input example
library(medhist)
data(medhistdata)
ps_removal=extract_nps_mh(medhistdata)

mh_bin_nps=
  medhistdata[ps_removal_train$key,] %>%
  `exprs<-`(`
```

```
    exprs(.) %>%
      t() %>%
      as.data.frame() %>%
      rownames_to_column(var='id') %>%
      column_to_rownames(var='id') %>%
      t()
  ) %>%
  trans_binary(verbose=F)

input=
  mh_bin_nps %>%
  exprs() %>%
  t() %>%
  as.data.frame()

## Fit dimensional reduction models with resampling
rsdr_bin_nps=rsdr(input,'CV',10,'PCA')

## Transform dimensions using an RSDR model
dr_bin_nps=
  mh_bin_nps %>%
  transformation(
    rsdr_object=rsdr_bin_nps
    ,input_dim=rownames(mh_bin_nps)
    ,top_n=6
    ,verbose=F
  )
```

# Index

- \* **RSDR**
  - transformation, 4
- \* **analysis,**
  - rsdr, 3
- \* **bootstrapping,**
  - rsdr, 3
- \* **component**
  - rsdr, 3
- \* **cross**
  - rsdr, 3
- \* **decomposition,**
  - rsdr, 3
- \* **dimensional**
  - composition, 2
  - rsdr, 3
- \* **matrix,**
  - composition, 2
- \* **principal**
  - rsdr, 3
- \* **reduction,**
  - rsdr, 3
- \* **reduction**
  - composition, 2
- \* **resampling,**
  - rsdr, 3
- \* **rotated**
  - composition, 2
- \* **singular**
  - rsdr, 3
- \* **transformer**
  - transformation, 4
- \* **validation**
  - rsdr, 3
- \* **value**
  - rsdr, 3
- \* **weights,**
  - composition, 2

composition, 2

rsdr, 3

transformation, 4