WhiBoClustering

November 14, 2018

Type Package

Version 0.1.2

Title White Box Clustering Algorithm Design

Author Sandro Radovanovic, Milan Vukicevic

Maintain	er Sandro Radovanovic <sandro.radovanovic@gmail.com></sandro.radovanovic@gmail.com>
allo by u avai also	ws you to create Representative based cluster algorithm using reusable components. This way one can recreate already lable cluster algorithms (i.e. K-Means, K-Means++, PAM) but create new cluster algorithms not available in the literature ny other software.
License (GPL-3 + file LICENSE
Depends	graphics, stats, clusterCrit, cluster
Suggests	methods, testthat
Encoding	UTF-8
LazyData	true
Roxygenl	Note 6.0.1
R topi	es documented:
1	olot.whibo_cluster
_	plot_pairs
1	print.whibo_cluster
•	wc_assign_bhattacharyya_numerical
•	wc_assign_canberra
	wc_assign_chebyshev 8 wc_assign_correlation 9
	wc_assign_cosine
•	wc_assign_euclidean

wc_assign_gower11wc_assign_hellinger_numerical11wc_assign_inner_product12

wc_assign_intersection	13
wc_assign_kulczynski	13
wc_assign_lorentzian	14
wc_assign_manhattan	14
wc_assign_sorensen	15
wc_assign_squared_euclidean	15
wc_assign_tanimoto	16
wc_assign_types	16
wc_assign_whittaker	17
wc_eval_ball_hall	17
wc_eval_banfeld_raftery	18
wc_eval_between_sum_of_squares	18
wc_eval_calinski_harabasz	19
wc_eval_c_index	19
	20
	20
	21
	21
<u> </u>	22
* ·	 22
	23
•	23
	24
	24
	25
C	25
	26
	26
	27
	27
	28
– –	28
•	29
	29
= = 71	30
	30
-	30 31
_	31
<i>E</i>	31 32
	32 32
	32 33
E	33
	34
	34
	35
	35
6	36
	36
	37
	37
-	38
wc_recalculate_types	38

1.4. 1.7 1	2
plot.whibo cluster	1
piot. wineo_craster	J

Index		51
	whibo_clustering	49
	whibo_cluster-class	
	wc_recalc_trimmed_mean	
	wc_recalc_trimean	
	wc_recalc_quadratic_mean	
	wc_recalc_online_trimmed_mean	
	wc_recalc_online_trimean	
	wc_recalc_online_quadratic_mean	
	wc_recalc_online_midrange	
	wc_recalc_online_midhinge	
	wc_recalc_online_median	
	wc_recalc_online_mean	
	wc_recalc_online_harmonic_mean	
	wc_recalc_online_geometric_mean	
	wc_recalc_midrange	
	wc_recalc_midhinge	
	wc_recalc_median	
	wc_recalc_mean	
	wc_recalc_harmonic_mean	
	wc_recalc_geometric_mean	

plot.whibo_cluster

Plot WhiBo Cluster Representatives

Description

Plot WhiBo Cluster Representatives

Usage

```
## S3 method for class 'whibo_cluster' plot(x, ...)
```

Arguments

x WhiBo Cluster model.

... None of those will be used.

Value

Line plot with Cluster representatives

Author(s)

plot_pairs

Examples

```
data <- iris[, 1:4] #Numerical data only
model <- whibo_clustering(data = data, k = 3)
plot(model)</pre>
```

plot_pairs

Plot WhiBo Cluster Representatives

Description

Plot WhiBo Cluster Representatives

Usage

```
plot_pairs(model, data)
```

Arguments

model WhiBo Cluster model.

data Data used for clustering (optional).

Value

Ploting pairs plot where Cluster representatives are presented with data (if provided).

Author(s)

Sandro Radovanovic <sandro.radovanovic@gmail.com>

Examples

```
data <- iris[, 1:4] #Numerical data only
model <- whibo_clustering(data = data, k = 3)
plot_pairs(model) #Ploting Cluster Representatives only
plot_pairs(model, data) #Ploting Cluster Representatives and Data</pre>
```

predict.whibo_cluster 5

predict.whibo_cluster Predict to which Cluster new data belongs

Description

Predict to which Cluster new data belongs

Usage

```
## S3 method for class 'whibo_cluster'
predict(object, data, ...)
```

Arguments

object WhiBo Cluster model.

data Data for which Cluster should be obtained.

... None of those will be used.

Value

Vector of assignments.

Author(s)

Sandro Radovanovic <sandro.radovanovic@gmail.com>

Examples

```
data <- iris[1:100, 1:4] #Numerical data only and first 100 rows model <- whibo_clustering(data = data, k = 3) predict(object = model, data = iris[101:150, 1:4])
```

print.whibo_cluster

Show White-Box Cluster Algorithm model

Description

Show White-Box Cluster Algorithm model

Usage

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

Arguments

x WhiBo Cluster model.

... None of those will be used.

Value

Summary text about Cluster model.

Author(s)

Sandro Radovanovic <sandro.radovanovic@gmail.com>

See Also

```
summary.whibo_cluster
```

Examples

```
data <- iris[, 1:4] #Numerical data only
model <- whibo_clustering(data = data, k = 3)
print(model)</pre>
```

summary.whibo_cluster Show White-Box Cluster Algorithm model

Description

Show White-Box Cluster Algorithm model

Usage

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

Arguments

object WhiBo Cluster model.
... None of those will be used.

Value

Summary text about Cluster model.

Author(s)

 $Sandro\ Radovanovic < \verb|sandro.radovanovic@gmail.com|| > \\$

See Also

```
print.whibo_cluster
```

wc_assignment 7

Examples

```
data <- iris[, 1:4] #Numerical data only
model <- whibo_clustering(data = data, k = 3)
summary(model)</pre>
```

wc_assignment

General Component for Assignment of data points to Cluster Representatives.

Description

General Component for Assignment of data points to Cluster Representatives.

Usage

```
wc_assignment(data, centroids, assignment_type)
```

Arguments

data A dataset for which data points needs to be assigned to Cluster Representatives.

centroids Cluster representatives.

assignment_type

String which signal which assignment type to be used. Check wc_assign_types for possible values.

Value

A vector of assignments.

Author(s)

Sandro Radovanovic <sandro.radovanovic@gmail.com>

```
wc_assign_bhattacharyya_numerical
```

Assign data points using Bhattacharyya distance.

Description

Assign data points using Bhattacharyya distance.

```
wc_assign_bhattacharyya_numerical(data, centroids)
```

Arguments

data A dataset for which data points needs to be assigned to Cluster Representatives.

centroids Cluster representatives.

Value

A vector of assignments.

Author(s)

Sandro Radovanovic <sandro.radovanovic@gmail.com>

wc_assign_canberra

Assign data points using Canberra distance.

Description

Assign data points using Canberra distance.

Usage

```
wc_assign_canberra(data, centroids)
```

Arguments

data A dataset for which data points needs to be assigned to Cluster Representatives.

centroids Cluster representatives.

Value

A vector of assignments.

Author(s)

Sandro Radovanovic <sandro.radovanovic@gmail.com>

wc_assign_chebyshev

Assign data points using Chebyshev distance.

Description

Assign data points using Chebyshev distance.

```
wc_assign_chebyshev(data, centroids)
```

wc_assign_correlation 9

Arguments

data A dataset for which data points needs to be assigned to Cluster Representatives.

centroids Cluster representatives.

Value

A vector of assignments.

Author(s)

Sandro Radovanovic <sandro.radovanovic@gmail.com>

wc_assign_correlation Assign data points using Correlation distance.

Description

Assign data points using Correlation distance.

Usage

```
wc_assign_correlation(data, centroids)
```

Arguments

data A dataset for which data points needs to be assigned to Cluster Representatives.

centroids Cluster representatives.

Value

A vector of assignments.

Author(s)

Sandro Radovanovic <sandro.radovanovic@gmail.com>

wc_assign_cosine

Assign data points using Cosine distance.

Description

Assign data points using Cosine distance.

```
wc_assign_cosine(data, centroids)
```

Arguments

data A dataset for which data points needs to be assigned to Cluster Representatives.

centroids Cluster representatives.

Value

A vector of assignments.

Author(s)

Sandro Radovanovic <sandro.radovanovic@gmail.com>

wc_assign_euclidean

Assign data points using Euclidean distance.

Description

Assign data points using Euclidean distance.

Usage

```
wc_assign_euclidean(data, centroids)
```

Arguments

data A dataset for which data points needs to be assigned to Cluster Representatives.

centroids Cluster representatives.

Value

A vector of assignments.

Author(s)

Sandro Radovanovic <sandro.radovanovic@gmail.com>

 $wc_assign_fidelity_numerical$

Assign data points using Fidelity (numerical version) distance.

Description

Assign data points using Fidelity (numerical version) distance.

```
wc_assign_fidelity_numerical(data, centroids)
```

wc_assign_gower 11

Arguments

data A dataset for which data points needs to be assigned to Cluster Representatives.

centroids Cluster representatives.

Value

A vector of assignments.

Author(s)

Sandro Radovanovic <sandro.radovanovic@gmail.com>

wc_assign_gower

Assign data points using Gower distance.

Description

Assign data points using Gower distance.

Usage

```
wc_assign_gower(data, centroids)
```

Arguments

data A dataset for which data points needs to be assigned to Cluster Representatives.

centroids Cluster representatives.

Value

A vector of assignments.

Author(s)

Sandro Radovanovic <sandro.radovanovic@gmail.com>

```
wc_assign_hellinger_numerical
```

Assign data points using Hellinger (numerical version) distance.

Description

Assign data points using Hellinger (numerical version) distance.

```
wc_assign_hellinger_numerical(data, centroids)
```

Arguments

data A dataset for which data points needs to be assigned to Cluster Representatives.

centroids Cluster representatives.

Value

A vector of assignments.

Author(s)

Sandro Radovanovic <sandro.radovanovic@gmail.com>

wc_assign_inner_product

Assign data points using Inner product distance.

Description

Assign data points using Inner product distance.

Usage

wc_assign_inner_product(data, centroids)

Arguments

data A dataset for which data points needs to be assigned to Cluster Representatives.

centroids Cluster representatives.

Value

A vector of assignments.

Author(s)

wc_assign_intersection 13

wc_assign_intersection

Assign data points using intersection distance.

Description

Assign data points using intersection distance.

Usage

```
wc_assign_intersection(data, centroids)
```

Arguments

data A dataset for which data points needs to be assigned to Cluster Representatives.

centroids Cluster representatives.

Value

A vector of assignments.

Author(s)

Sandro Radovanovic <sandro.radovanovic@gmail.com>

wc_assign_kulczynski Assign data points using Kulczynski distance.

Description

Assign data points using Kulczynski distance.

Usage

```
wc_assign_kulczynski(data, centroids)
```

Arguments

data A dataset for which data points needs to be assigned to Cluster Representatives.

centroids Cluster representatives.

Value

A vector of assignments.

Author(s)

14 wc_assign_manhattan

Description

Assign data points using Lorentzian distance.

Usage

```
wc_assign_lorentzian(data, centroids)
```

Arguments

data A dataset for which data points needs to be assigned to Cluster Representatives.

centroids Cluster representatives.

Value

A vector of assignments.

Author(s)

Sandro Radovanovic <sandro.radovanovic@gmail.com>

Description

Assign data points using Manhattan distance.

Usage

```
wc_assign_manhattan(data, centroids)
```

Arguments

data A dataset for which data points needs to be assigned to Cluster Representatives.

centroids Cluster representatives.

Value

A vector of assignments.

Author(s)

wc_assign_sorensen 15

wc_assign_sorensen

Assign data points using Sorensen distance.

Description

Assign data points using Sorensen distance.

Usage

```
wc_assign_sorensen(data, centroids)
```

Arguments

data A dataset for which data points needs to be assigned to Cluster Representatives.

centroids Cluster representatives.

Value

A vector of assignments.

Author(s)

Sandro Radovanovic <sandro.radovanovic@gmail.com>

wc_assign_squared_euclidean

Assign data points using squared Euclidean distance.

Description

Assign data points using squared Euclidean distance.

Usage

```
wc_assign_squared_euclidean(data, centroids)
```

Arguments

data A dataset for which data points needs to be assigned to Cluster Representatives.

centroids Cluster representatives.

Value

A vector of assignments.

Author(s)

16 wc_assign_types

wc_assign_tanimoto

Assign data points using Tanimoto distance.

Description

Assign data points using Tanimoto distance.

Usage

```
wc_assign_tanimoto(data, centroids)
```

Arguments

data A dataset for which data points needs to be assigned to Cluster Representatives.

centroids Cluster representatives.

Value

A vector of assignments.

Author(s)

Sandro Radovanovic <sandro.radovanovic@gmail.com>

wc_assign_types

Data frame for possible values of assignment types.

Description

Data frame for possible values of assignment types.

Usage

```
wc_assign_types
```

Format

An object of class data. frame with $18\ rows$ and $2\ columns$.

Author(s)

wc_assign_whittaker 17

wc_assign_whittaker

Assign data points using Whittaker distance.

Description

Assign data points using Whittaker distance.

Usage

```
wc_assign_whittaker(data, centroids)
```

Arguments

data A dataset for which data points needs to be assigned to Cluster Representatives.

centroids Cluster representatives.

Value

A vector of assignments.

Author(s)

Sandro Radovanovic <sandro.radovanovic@gmail.com>

wc_eval_ball_hall

Calculate Ball-Hall internal Cluster evaluation measure

Description

Calculate Ball-Hall internal Cluster evaluation measure

Usage

```
wc_eval_ball_hall(data, assignment)
```

Arguments

data A dataset for which internal cluster quality should be calculated.

assignment Vector of assignments.

Value

A value of internal cluster quality evaluation measure.

Author(s)

wc_eval_banfeld_raftery

Calculate Banfeld-Raftery internal Cluster evaluation measure

Description

Calculate Banfeld-Raftery internal Cluster evaluation measure

Usage

```
wc_eval_banfeld_raftery(data, assignment)
```

Arguments

data A dataset for which internal cluster quality should be calculated.

assignment Vector of assignments.

Value

A value of internal cluster quality evaluation measure.

Author(s)

Sandro Radovanovic <sandro.radovanovic@gmail.com>

Description

Calculate between (Clusters) sum of squares

Usage

```
wc_eval_between_sum_of_squares(data, centroids, assignment)
```

Arguments

data A dataset for which between sum of squared should be calculated.

centroids A data frame of cluster representatives.

assignment Vector of assignments.

Value

A vector of number which shows between (clusters) sum of squares.

Author(s)

wc_eval_calinski_harabasz

Calculate Calinski-Harabasz internal Cluster evaluation measure

Description

Calculate Calinski-Harabasz internal Cluster evaluation measure

Usage

```
wc_eval_calinski_harabasz(data, assignment)
```

Arguments

data A dataset for which internal cluster quality should be calculated.

assignment Vector of assignments.

Value

A value of internal cluster quality evaluation measure.

Author(s)

Sandro Radovanovic <sandro.radovanovic@gmail.com>

wc_eval_c_index

Calculate C index internal Cluster evaluation measure

Description

Calculate C index internal Cluster evaluation measure

Usage

```
wc_eval_c_index(data, assignment)
```

Arguments

data A dataset for which internal cluster quality should be calculated.

assignment Vector of assignments.

Value

A value of internal cluster quality evaluation measure.

Author(s)

20 wc_eval_det_ratio

```
wc_eval_davies_bouldin
```

Calculate Davies-Bouldin internal Cluster evaluation measure

Description

Calculate Davies-Bouldin internal Cluster evaluation measure

Usage

```
wc_eval_davies_bouldin(data, assignment)
```

Arguments

data A dataset for which internal cluster quality should be calculated.

assignment Vector of assignments.

Value

A value of internal cluster quality evaluation measure.

Author(s)

Sandro Radovanovic <sandro.radovanovic@gmail.com>

wc_eval_det_ratio

Calculate Det ratio internal Cluster evaluation measure

Description

Calculate Det ratio internal Cluster evaluation measure

Usage

```
wc_eval_det_ratio(data, assignment)
```

Arguments

data A dataset for which internal cluster quality should be calculated.

assignment Vector of assignments.

Value

A value of internal cluster quality evaluation measure.

Author(s)

wc_eval_dunn 21

wc_eval_dunn

Calculate Dunn index internal Cluster evaluation measure

Description

Calculate Dunn index internal Cluster evaluation measure

Usage

```
wc_eval_dunn(data, assignment)
```

Arguments

data A dataset for which internal cluster quality should be calculated.

assignment Vector of assignments.

Value

A value of internal cluster quality evaluation measure.

Author(s)

Sandro Radovanovic <sandro.radovanovic@gmail.com>

wc_eval_gamma

Calculate Gamma index internal Cluster evaluation measure

Description

Calculate Gamma index internal Cluster evaluation measure

Usage

```
wc_eval_gamma(data, assignment)
```

Arguments

data A dataset for which internal cluster quality should be calculated.

assignment Vector of assignments.

Value

A value of internal cluster quality evaluation measure.

Author(s)

22 wc_eval_silhouette

wc_eval_g_plus

Calculate G+ index internal Cluster evaluation measure

Description

Calculate G+ index internal Cluster evaluation measure

Usage

```
wc_eval_g_plus(data, assignment)
```

Arguments

data A dataset for which internal cluster quality should be calculated.

assignment Vector of assignments.

Value

A value of internal cluster quality evaluation measure.

Author(s)

Sandro Radovanovic <sandro.radovanovic@gmail.com>

wc_eval_silhouette

Calculate Silhouette score internal Cluster evaluation measure

Description

Calculate Silhouette score internal Cluster evaluation measure

Usage

```
wc_eval_silhouette(data, assignment)
```

Arguments

data A dataset for which internal cluster quality should be calculated.

assignment Vector of assignments.

Value

A value of internal cluster quality evaluation measure.

Author(s)

wc_eval_total_sum_of_squares

Calculate total sum of squares

Description

Calculate total sum of squares

Usage

```
wc_eval_total_sum_of_squares(data)
```

Arguments

data

A dataset for which total sum of squared should be calculated.

Value

A number which shows total sum of squares.

Author(s)

Sandro Radovanovic <sandro.radovanovic@gmail.com>

wc_eval_within_sum_of_squares

Calculate within (Cluster) sum of squares

Description

Calculate within (Cluster) sum of squares

Usage

```
wc_eval_within_sum_of_squares(data, centroids, assignment)
```

Arguments

data A dataset for which within sum of squared should be calculated.

centroids A data frame of cluster representatives.

assignment Vector of assignments.

Value

A vector of number which shows within (cluster) sum of squares.

Author(s)

24 wc_initialize

wc_eval_xie_beni

Calculate Xie-Beni internal Cluster evaluation measure

Description

Calculate Xie-Beni internal Cluster evaluation measure

Usage

```
wc_eval_xie_beni(data, assignment)
```

Arguments

data A dataset for which internal cluster quality should be calculated.

assignment Vector of assignments.

Value

A value of internal cluster quality evaluation measure.

Author(s)

Sandro Radovanovic <sandro.radovanovic@gmail.com>

wc_initialize

General Component for Initialization of Cluster Representatives.

Description

General Component for Initialization of Cluster Representatives.

Usage

```
wc_initialize(data, k = 3, initialization_type)
```

Arguments

data A dataset for which Cluster Representatives needs to be initialized.

k A number of Cluster Representatives.

initialization_type

String which signal which initialization type to be used. Check wc_init_types

for possible values.

Value

As a result initial Cluster Representatives are obtained. Result is in for of data.frame or data.matrix.

Author(s)

wc_init_agnes 25

wc_init_agnes

AGNES Cluster Representatives initialization.

Description

AGNES Cluster Representatives initialization.

Usage

```
wc_init_agnes(data, k)
```

Arguments

data A dataset for which Cluster Representatives needs to be initialized.

k A number of Cluster Representatives to be initialized.

Value

As a result initial Cluster Representatives are obtained. Result is in for of data.frame or data.matrix.

Author(s)

Sandro Radovanovic <sandro.radovanovic@gmail.com>

wc_init_ccia

CCIA Cluster Representatives initialization.

Description

CCIA Cluster Representatives initialization.

Usage

```
wc_init_ccia(data, k)
```

Arguments

data A dataset for which Cluster Representatives needs to be initialized.

k A number of Cluster Representatives to be initialized.

Value

As a result initial Cluster Representatives are obtained. Result is in for of data.frame or data.matrix.

Author(s)

26 wc_init_forgy

wc_init_diana

DIANA Cluster Representatives initialization.

Description

DIANA Cluster Representatives initialization.

Usage

```
wc_init_diana(data, k)
```

Arguments

data A dataset for which Cluster Representatives needs to be initialized.

k A number of Cluster Representatives to be initialized.

Value

As a result initial Cluster Representatives are obtained. Result is in for of data.frame or data.matrix.

Author(s)

Sandro Radovanovic <sandro.radovanovic@gmail.com>

wc_init_forgy

Forgy algorithm Cluster Representatives initialization.

Description

Forgy algorithm Cluster Representatives initialization.

Usage

```
wc_init_forgy(data, k = 3)
```

Arguments

data A dataset for which Cluster Representatives needs to be initialized.

k A number of Cluster Representatives to be initialized.

Value

As a result initial Cluster Representatives are obtained. Result is in for of data.frame or data.matrix.

Author(s)

wc_init_kkz 27

wc_init_kkz

KKZ Cluster Representatives initialization.

Description

KKZ Cluster Representatives initialization.

Usage

```
wc_init_kkz(data, k)
```

Arguments

data A dataset for which Cluster Representatives needs to be initialized.

k A number of Cluster Representatives to be initialized.

Value

As a result initial Cluster Representatives are obtained. Result is in for of data.frame or data.matrix.

Author(s)

Sandro Radovanovic <sandro.radovanovic@gmail.com>

```
wc_init_kmeansplusplus
```

K-Means++ Cluster Representatives initialization.

Description

K-Means++ Cluster Representatives initialization.

Usage

```
wc_init_kmeansplusplus(data, k = 3)
```

Arguments

data A dataset for which Cluster Representatives needs to be initialized.

k A number of Cluster Representatives to be initialized.

Value

As a result initial Cluster Representatives are obtained. Result is in for of data.frame or data.matrix.

Author(s)

28 wc_init_quantile

wc_init_pca

PCA Cluster Representatives initialization.

Description

PCA Cluster Representatives initialization.

Usage

```
wc_init_pca(data, k)
```

Arguments

data A dataset for which Cluster Representatives needs to be initialized.

k A number of Cluster Representatives to be initialized.

Value

As a result initial Cluster Representatives are obtained. Result is in for of data.frame or data.matrix.

Author(s)

Sandro Radovanovic <sandro.radovanovic@gmail.com>

wc_init_quantile

Quantile Cluster Representatives initialization.

Description

Quantile Cluster Representatives initialization.

Usage

```
wc_init_quantile(data, k)
```

Arguments

data A dataset for which Cluster Representatives needs to be initialized.

k A number of Cluster Representatives to be initialized.

Value

As a result initial Cluster Representatives are obtained. Result is in for of data.frame or data.matrix.

Author(s)

wc_init_random 29

wc_init_random

Random Cluster Representatives initialization.

Description

Random Cluster Representatives initialization.

Usage

```
wc_init_random(data, k = 3)
```

Arguments

data A dataset for which Cluster Representatives needs to be initialized.

k A number of Cluster Representatives to be initialized.

Value

As a result initial Cluster Representatives are obtained. Result is in for of data.frame or data.matrix.

Author(s)

Sandro Radovanovic <sandro.radovanovic@gmail.com>

wc_init_types

Data frame for possible values of initialization types.

Description

Data frame for possible values of initialization types.

Usage

```
wc_init_types
```

Format

An object of class data. frame with $10 \ \text{rows}$ and $2 \ \text{columns}$.

Author(s)

30 wc_normalize

wc_init_ward

Ward algorithm Cluster Representatives initialization.

Description

Ward algorithm Cluster Representatives initialization.

Usage

```
wc_init_ward(data, k)
```

Arguments

data A dataset for which Cluster Representatives needs to be initialized.

k A number of Cluster Representatives to be initialized.

Value

As a result initial Cluster Representatives are obtained. Result is in for of data.frame or data.matrix.

Author(s)

Sandro Radovanovic <sandro.radovanovic@gmail.com>

wc_normalize

General (Optional) Component for Normalization of data.

Description

General (Optional) Component for Normalization of data.

Usage

```
wc_normalize(data, normalization_type)
```

Arguments

data A dataset which needs to be normalized.

 $normalization_type$

String which signal which normalization type to be used. Check wc_norm_types for possible values.

Value

As a result normalized data are obtained. Result is in for of data.frame or data.matrix.

Author(s)

wc_norm_comprehensive Comprehensive normalization of data.

Description

Comprehensive normalization of data.

Usage

```
wc_norm_comprehensive(data, model = NULL)
```

Arguments

data A dataset which needs to be normalized.

model Additional data needed for future data to be normalized using same normaliza-

tion technique.

Value

As a result normalized data are obtained. Result is in for of data.frame or data.matrix.

Author(s)

Sandro Radovanovic <sandro.radovanovic@gmail.com>

```
wc_norm_decimal_scaling
```

Decimal scaling of data.

Description

Decimal scaling of data.

Usage

```
wc_norm_decimal_scaling(data, model = NULL)
```

Arguments

data A dataset which needs to be normalized.

model Additional data needed for future data to be normalized using same normaliza-

tion technique.

Value

As a result normalized data are obtained. Result is in for of data.frame or data.matrix.

Author(s)

32 wc_norm_12

wc_norm_11

 l_1 normalization of data.

Description

 l_1 normalization of data.

Usage

```
wc_norm_l1(data, model = NULL)
```

Arguments

data A dataset which needs to be normalized.

model Additional data needed for future data to be normalized using same normaliza-

tion technique.

Value

As a result normalized data are obtained. Result is in for of data.frame or data.matrix.

Author(s)

Sandro Radovanovic <sandro.radovanovic@gmail.com>

wc_norm_12

 l_2 normalization of data.

Description

 l_2 normalization of data.

Usage

```
wc_norm_12(data, model = NULL)
```

Arguments

data A dataset which needs to be normalized.

model Additional data needed for future data to be normalized using same normaliza-

tion technique.

Value

As a result normalized data are obtained. Result is in for of data.frame or data.matrix.

Author(s)

wc_norm_linf 33

wc_norm_linf

 l_∞ normalization of data.

Description

 l_{∞} normalization of data.

Usage

```
wc_norm_linf(data, model = NULL)
```

Arguments

data A dataset which needs to be normalized.

model Additional data needed for future data to be normalized using same normaliza-

tion technique.

Value

As a result normalized data are obtained. Result is in for of data.frame or data.matrix.

Author(s)

Sandro Radovanovic <sandro.radovanovic@gmail.com>

wc_norm_log

Logarithmic normalization of data.

Description

Logarithmic normalization of data.

Usage

```
wc_norm_log(data, model = NULL)
```

Arguments

data A dataset which needs to be normalized.

model Additional data needed for future data to be normalized using same normaliza-

tion technique.

Value

As a result normalized data are obtained. Result is in for of data.frame or data.matrix.

Author(s)

34 wc_norm_mean

wc_norm_max_min

Max-Min normalization of data.

Description

Max-Min normalization of data.

Usage

```
wc_norm_max_min(data, model = NULL)
```

Arguments

data A dataset which needs to be normalized.

model Additional data needed for future data to be normalized using same normaliza-

tion technique.

Value

As a result normalized data are obtained. Result is in for of data.frame or data.matrix.

Author(s)

Sandro Radovanovic <sandro.radovanovic@gmail.com>

wc_norm_mean

Mean (Max-Min like) normalization of data.

Description

Mean (Max-Min like) normalization of data.

Usage

```
wc_norm_mean(data, model = NULL)
```

Arguments

data A dataset which needs to be normalized.

model Additional data needed for future data to be normalized using same normaliza-

tion technique.

Value

As a result normalized data are obtained. Result is in for of data.frame or data.matrix.

Author(s)

wc_norm_no 35

wc_norm_no

Ignore normalization of data.

Description

Ignore normalization of data.

Usage

```
wc_norm_no(data, model = NULL)
```

Arguments

data A dataset which needs to be normalized.

model Additional data needed for future data to be normalized using same normaliza-

tion technique.

Value

As a result normalized data are obtained. Result is in for of data.frame or data.matrix.

Author(s)

Sandro Radovanovic <sandro.radovanovic@gmail.com>

wc_norm_non_monotonic Non-monotonic normalization of data.

Description

Non-monotonic normalization of data.

Usage

```
wc_norm_non_monotonic(data, model = NULL)
```

Arguments

data A dataset which needs to be normalized.

model Additional data needed for future data to be normalized using same normaliza-

tion technique.

Value

As a result normalized data are obtained. Result is in for of data.frame or data.matrix.

Author(s)

36 wc_norm_softmax

wc_norm_sigmoid

Sigmoid normalization of data.

Description

Sigmoid normalization of data.

Usage

```
wc_norm_sigmoid(data, model = NULL)
```

Arguments

data A dataset which needs to be normalized.

model Additional data needed for future data to be normalized using same normaliza-

tion technique.

Value

As a result normalized data are obtained. Result is in for of data.frame or data.matrix.

Author(s)

Sandro Radovanovic <sandro.radovanovic@gmail.com>

wc_norm_softmax

Softmax normalization of data.

Description

Softmax normalization of data.

Usage

```
wc_norm_softmax(data, model = NULL)
```

Arguments

data A dataset which needs to be normalized.

model Additional data needed for future data to be normalized using same normaliza-

tion technique.

Value

As a result normalized data are obtained. Result is in for of data.frame or data.matrix.

Author(s)

wc_norm_types 37

wc_norm_types

Data frame for possible values of normalization types.

Description

Data frame for possible values of normalization types.

Usage

```
wc_norm_types
```

Format

An object of class data. frame with 13 rows and 2 columns.

Author(s)

Sandro Radovanovic <sandro.radovanovic@gmail.com>

wc_norm_z

Z transformation of data.

Description

Z transformation of data.

Usage

```
wc_norm_z(data, model = NULL)
```

Arguments

data A dataset which needs to be normalized.

model Additional data needed for future data to be normalized using same normaliza-

tion technique.

Value

As a result normalized data are obtained. Result is in for of data.frame or data.matrix.

Author(s)

 $Sandro\ Radovanovic < \verb|sandro.radovanovic@gmail.com|| > \\$

38 wc_recalculate_types

wc_recalculate

General Component for Cluster Representative update.

Description

General Component for Cluster Representative update.

Usage

```
wc_recalculate(data, assignment, recalculate_type, assignment_type = NULL,
   old_centroids = NULL)
```

Arguments

data A dataset for which Cluster Representatives needs to be updated.

assignment Vector of Cluster assignments.

recalculate_type

String which signal which update type to be used. Check wc_recalculate_types

for possible values.

assignment_type

Assignment type (Optional).

old_centroids Old centroids (Optional).

Value

As a result new Cluster Representatives are obtained. Result is in for of data.frame or data.matrix.

Author(s)

Sandro Radovanovic <sandro.radovanovic@gmail.com>

 $wc_recalculate_types \quad \textit{Data frame for possible values of recalculate types}.$

Description

Data frame for possible values of recalculate types.

Usage

```
wc_recalculate_types
```

Format

An object of class data. frame with 18 rows and 2 columns.

Author(s)

wc_recalc_geometric_mean

Solution for Cluster Representative update which uses Geometric mean.

Description

Solution for Cluster Representative update which uses Geometric mean.

Usage

```
wc_recalc_geometric_mean(data, assignment, old_centroids = NULL)
```

Arguments

data A dataset for which Cluster Representatives needs to be updated.

assignment Vector of Cluster assignments.

old_centroids Old centroids.

Value

As a result new Cluster Representatives are obtained. Result is in for of data.frame or data.matrix.

Author(s)

Sandro Radovanovic <sandro.radovanovic@gmail.com>

```
wc_recalc_harmonic_mean
```

Solution for Cluster Representative update which uses Harmonic mean.

Description

Solution for Cluster Representative update which uses Harmonic mean.

Usage

```
wc_recalc_harmonic_mean(data, assignment, old_centroids = NULL)
```

Arguments

data A dataset for which Cluster Representatives needs to be updated.

assignment Vector of Cluster assignments.

old_centroids Old centroids.

Value

As a result new Cluster Representatives are obtained. Result is in for of data.frame or data.matrix.

40 wc_recalc_median

Author(s)

Sandro Radovanovic <sandro.radovanovic@gmail.com>

wc_recalc_mean

Solution for Cluster Representative update which uses Mean.

Description

Solution for Cluster Representative update which uses Mean.

Usage

```
wc_recalc_mean(data, assignment, old_centroids = NULL)
```

Arguments

data A dataset for which Cluster Representatives needs to be updated.

assignment Vector of Cluster assignments.

old_centroids Old centroids.

Value

As a result new Cluster Representatives are obtained. Result is in for of data.frame or data.matrix.

Author(s)

Sandro Radovanovic <sandro.radovanovic@gmail.com>

wc_recalc_median

Solution for Cluster Representative update which uses Median.

Description

Solution for Cluster Representative update which uses Median.

Usage

```
wc_recalc_median(data, assignment, old_centroids = NULL)
```

Arguments

data A dataset for which Cluster Representatives needs to be updated.

assignment Vector of Cluster assignments.

old_centroids Old centroids.

Value

As a result new Cluster Representatives are obtained. Result is in for of data.frame or data.matrix.

Author(s)

wc_recalc_midhinge 41

wc_recalc_midhinge

Solution for Cluster Representative update which uses Midhinge.

Description

Solution for Cluster Representative update which uses Midhinge.

Usage

```
wc_recalc_midhinge(data, assignment, old_centroids = NULL)
```

Arguments

data A dataset for which Cluster Representatives needs to be updated.

assignment Vector of Cluster assignments.

old_centroids Old centroids.

Value

As a result new Cluster Representatives are obtained. Result is in for of data.frame or data.matrix.

Author(s)

Sandro Radovanovic <sandro.radovanovic@gmail.com>

wc_recalc_midrange

Solution for Cluster Representative update which uses Midrange.

Description

Solution for Cluster Representative update which uses Midrange.

Usage

```
wc_recalc_midrange(data, assignment, old_centroids = NULL)
```

Arguments

data A dataset for which Cluster Representatives needs to be updated.

assignment Vector of Cluster assignments.

old_centroids Old centroids.

Value

As a result new Cluster Representatives are obtained. Result is in for of data.frame or data.matrix.

Author(s)

```
wc_recalc_online_geometric_mean
```

Solution for Cluster Representative update which uses Online Geometric mean.

Description

Solution for Cluster Representative update which uses Online Geometric mean.

Usage

```
wc_recalc_online_geometric_mean(data, assignment, assignment_type,
  old_centroids = NULL)
```

Arguments

data A dataset for which Cluster Representatives needs to be updated.

assignment Vector of Cluster assignments.

assignment_type

Assignment type to be used.

old_centroids Old centroids.

Value

As a result new Cluster Representatives are obtained. Result is in for of data.frame or data.matrix.

Author(s)

Sandro Radovanovic <sandro.radovanovic@gmail.com>

```
wc_recalc_online_harmonic_mean
```

Solution for Cluster Representative update which uses Online Harmonic mean.

Description

Solution for Cluster Representative update which uses Online Harmonic mean.

Usage

```
wc_recalc_online_harmonic_mean(data, assignment, assignment_type,
  old_centroids = NULL)
```

Arguments

data A dataset for which Cluster Representatives needs to be updated.

assignment Vector of Cluster assignments.

 ${\tt assignment_type}$

Assignment type to be used.

wc_recalc_online_mean

Value

As a result new Cluster Representatives are obtained. Result is in for of data.frame or data.matrix.

43

Author(s)

Sandro Radovanovic <sandro.radovanovic@gmail.com>

wc_recalc_online_mean Solution for Cluster Representative update which uses Online mean.

Description

Solution for Cluster Representative update which uses Online mean.

Usage

```
wc_recalc_online_mean(data, assignment, assignment_type, old_centroids = NULL)
```

Arguments

data A dataset for which Cluster Representatives needs to be updated.

assignment Vector of Cluster assignments.

assignment_type

Assignment type to be used.

old_centroids Old centroids.

Value

As a result new Cluster Representatives are obtained. Result is in for of data.frame or data.matrix.

Author(s)

Sandro Radovanovic <sandro.radovanovic@gmail.com>

```
wc_recalc_online_median
```

 $Solution\ for\ Cluster\ Representative\ update\ which\ uses\ Online\ median.$

Description

Solution for Cluster Representative update which uses Online median.

Usage

```
wc_recalc_online_median(data, assignment, assignment_type,
  old_centroids = NULL)
```

Arguments

data A dataset for which Cluster Representatives needs to be updated.

assignment Vector of Cluster assignments.

assignment_type

Assignment type to be used.

old_centroids Old centroids.

Value

As a result new Cluster Representatives are obtained. Result is in for of data.frame or data.matrix.

Author(s)

Sandro Radovanovic <sandro.radovanovic@gmail.com>

wc_recalc_online_midhinge

Solution for Cluster Representative update which uses Online Midhindge.

Description

Solution for Cluster Representative update which uses Online Midhindge.

Usage

```
wc_recalc_online_midhinge(data, assignment, assignment_type,
  old_centroids = NULL)
```

Arguments

data A dataset for which Cluster Representatives needs to be updated.

assignment Vector of Cluster assignments.

assignment_type

Assignment type to be used.

old_centroids Old centroids.

Value

As a result new Cluster Representatives are obtained. Result is in for of data.frame or data.matrix.

Author(s)

```
wc_recalc_online_midrange
```

Solution for Cluster Representative update which uses Online Midrange.

Description

Solution for Cluster Representative update which uses Online Midrange.

Usage

```
wc_recalc_online_midrange(data, assignment, assignment_type,
  old_centroids = NULL)
```

Arguments

data A dataset for which Cluster Representatives needs to be updated.

assignment Vector of Cluster assignments.

assignment_type

Assignment type to be used.

old_centroids Old centroids.

Value

As a result new Cluster Representatives are obtained. Result is in for of data.frame or data.matrix.

Author(s)

Sandro Radovanovic <sandro.radovanovic@gmail.com>

```
wc_recalc_online_quadratic_mean
```

Solution for Cluster Representative update which uses Online Quadratic mean.

Description

Solution for Cluster Representative update which uses Online Quadratic mean.

Usage

```
wc_recalc_online_quadratic_mean(data, assignment, assignment_type,
  old_centroids = NULL)
```

Arguments

data A dataset for which Cluster Representatives needs to be updated.

assignment Vector of Cluster assignments.

assignment_type

Assignment type to be used.

Value

As a result new Cluster Representatives are obtained. Result is in for of data.frame or data.matrix.

Author(s)

Sandro Radovanovic <sandro.radovanovic@gmail.com>

```
wc_recalc_online_trimean
```

Solution for Cluster Representative update which uses Online Trimean.

Description

Solution for Cluster Representative update which uses Online Trimean.

Usage

```
wc_recalc_online_trimean(data, assignment, assignment_type,
  old_centroids = NULL)
```

Arguments

data A dataset for which Cluster Representatives needs to be updated.

assignment Vector of Cluster assignments.

assignment_type

Assignment type to be used.

old_centroids Old centroids.

Value

As a result new Cluster Representatives are obtained. Result is in for of data.frame or data.matrix.

Author(s)

```
wc_recalc_online_trimmed_mean
```

Solution for Cluster Representative update which uses Online Trimmed mean.

Description

Solution for Cluster Representative update which uses Online Trimmed mean.

Usage

```
wc_recalc_online_trimmed_mean(data, assignment, assignment_type,
  old_centroids = NULL)
```

Arguments

data A dataset for which Cluster Representatives needs to be updated.

assignment Vector of Cluster assignments.

assignment_type

Assignment type to be used.

old_centroids Old centroids.

Value

As a result new Cluster Representatives are obtained. Result is in for of data.frame or data.matrix.

Author(s)

Sandro Radovanovic <sandro.radovanovic@gmail.com>

```
wc_recalc_quadratic_mean
```

Solution for Cluster Representative update which uses Quadratic mean.

Description

Solution for Cluster Representative update which uses Quadratic mean.

Usage

```
wc_recalc_quadratic_mean(data, assignment, old_centroids = NULL)
```

Arguments

data A dataset for which Cluster Representatives needs to be updated.

assignment Vector of Cluster assignments.

Value

As a result new Cluster Representatives are obtained. Result is in for of data.frame or data.matrix.

Author(s)

Sandro Radovanovic <sandro.radovanovic@gmail.com>

wc_recalc_trimean

Solution for Cluster Representative update which uses Trimean.

Description

Solution for Cluster Representative update which uses Trimean.

Usage

```
wc_recalc_trimean(data, assignment, old_centroids = NULL)
```

Arguments

data A dataset for which Cluster Representatives needs to be updated.

assignment Vector of Cluster assignments.

old_centroids Old centroids.

Value

As a result new Cluster Representatives are obtained. Result is in for of data.frame or data.matrix.

Author(s)

Sandro Radovanovic <sandro.radovanovic@gmail.com>

```
wc_recalc_trimmed_mean
```

Solution for Cluster Representative update which uses Trimmed mean.

Description

Solution for Cluster Representative update which uses Trimmed mean.

Usage

```
wc_recalc_trimmed_mean(data, assignment, old_centroids = NULL)
```

Arguments

data A dataset for which Cluster Representatives needs to be updated.

assignment Vector of Cluster assignments.

whibo_cluster-class 49

Value

As a result new Cluster Representatives are obtained. Result is in for of data.frame or data.matrix.

Author(s)

Sandro Radovanovic <sandro.radovanovic@gmail.com>

whibo_cluster-class

As S4 class to represent WhiBo Cluster model

Description

As S4 class to represent WhiBo Cluster model

Slots

whibo_cluster Whibo Clustering object - list of objects for White-Box Clustering

Author(s)

Sandro Radovanovic

whibo_clustering

Find Cluster model using White-Box Cluster Algorithm Design.

Description

Find Cluster model using White-Box Cluster Algorithm Design.

Usage

```
whibo_clustering(data, k = 3, normalization_type = "No",
  cluster_initialization_type = "Random", assignment_type = "Euclidean",
  recalculation_type = "Mean", max_iteration = 20, no_of_restarts = 1)
```

Arguments

data Data on which clustering should be performed.

k Number of Cluster Representatives.

normalization_type

Which normalization should be used (look at wc_norm_types for possible values). Default value is No.

cluster_initialization_type

Which initialization of Cluster Representatives should be used (look at wc_init_types for possible values). Default value is Random.

assignment_type

Which assignment function should be used (look at wc_assign_types for possible values). Default value is Euclidean.

50 whibo_clustering

```
recalculation_type

Which function for updating Cluster Representatives should be used (look at wc_recalculate_types for possible values). Default value is Mean.

max_iteration Number of iterations. Default value is 20.

no_of_restarts Number of restarts of whole clustering procedure. Default value is 1.
```

Value

Object of type whibo_cluster which include Cluster Representatives (centroids), number of elements per cluster (elements_per_cluster), assignments (assignments), measures of cluster quality (within_sum_of_squares, between_ss_div_total_ss and internal_measures_of_quality), cluster models per iterations (model_history), iterations (iterations) and parameters used (params)

Author(s)

Sandro Radovanovic <sandro.radovanovic@gmail.com>

See Also

```
plot.whibo_cluster, predict.whibo_cluster
```

Examples

Index

*Topic datasets	<pre>wc_eval_total_sum_of_squares, 23</pre>
wc_assign_types, 16	wc_eval_within_sum_of_squares, 23
wc_assign_types, 10 wc_init_types, 29	wc_eval_xie_beni, 24
wc_norm_types, 37	wc_init_agnes, 25
wc_norm_types, 37 wc_recalculate_types, 38	_
wc_recarculate_types, 50	<pre>wc_init_ccia, 25 wc_init_diana, 26</pre>
plot.whibo_cluster, 3	
plot_pairs, 4	wc_init_forgy, 26
predict.whibo_cluster, 5	wc_init_kkz,27
print.whibo_cluster, 5	wc_init_kmeansplusplus, 27
print: miso_cluster, s	wc_init_pca, 28
<pre>summary.whibo_cluster, 6</pre>	wc_init_quantile, 28
	wc_init_random, 29
<pre>wc_assign_bhattacharyya_numerical, 7</pre>	wc_init_types, 29
wc_assign_canberra, 8	wc_init_ward, 30
wc_assign_chebyshev, 8	wc_initialize, 24
wc_assign_correlation, 9	wc_norm_comprehensive, 31
wc_assign_cosine, 9	wc_norm_decimal_scaling, 31
wc_assign_euclidean, 10	wc_norm_11, 32
wc_assign_fidelity_numerical, 10	wc_norm_12, 32
wc_assign_gower, 11	wc_norm_linf, 33
<pre>wc_assign_hellinger_numerical, 11</pre>	wc_norm_log, 33
<pre>wc_assign_inner_product, 12</pre>	wc_norm_max_min, 34
wc_assign_intersection, 13	wc_norm_mean, 34
wc_assign_kulczynski, 13	wc_norm_no, 35
wc_assign_lorentzian, 14	wc_norm_non_monotonic, 35
wc_assign_manhattan, 14	wc_norm_sigmoid, 36
wc_assign_sorensen, 15	wc_norm_softmax, 36
<pre>wc_assign_squared_euclidean, 15</pre>	wc_norm_types, 37
wc_assign_tanimoto, 16	wc_norm_z, 37
<pre>wc_assign_types, 16</pre>	wc_normalize, 30
wc_assign_whittaker, 17	wc_recalc_geometric_mean,39
<pre>wc_assignment, 7</pre>	wc_recalc_harmonic_mean,39
wc_eval_ball_hall, 17	wc_recalc_mean, 40
<pre>wc_eval_banfeld_raftery, 18</pre>	$wc_recalc_median, 40$
<pre>wc_eval_between_sum_of_squares, 18</pre>	wc_recalc_midhinge,41
<pre>wc_eval_c_index, 19</pre>	wc_recalc_midrange,41
wc_eval_calinski_harabasz,19	wc_recalc_online_geometric_mean, 42
<pre>wc_eval_davies_bouldin, 20</pre>	<pre>wc_recalc_online_harmonic_mean, 42</pre>
wc_eval_det_ratio,20	wc_recalc_online_mean,43
wc_eval_dunn, 21	wc_recalc_online_median, 43
<pre>wc_eval_g_plus, 22</pre>	wc_recalc_online_midhinge,44
wc_eval_gamma, 21	wc_recalc_online_midrange, 45
wc_eval_silhouette, 22	wc_recalc_online_quadratic_mean, 45

52 INDEX

```
wc_recalc_online_trimean, 46
wc_recalc_online_trimmed_mean, 47
wc_recalc_quadratic_mean, 47
wc_recalc_trimean, 48
wc_recalc_trimmed_mean, 48
wc_recalculate, 38
wc_recalculate_types, 38
whibo_cluster-class, 49
whibo_clustering, 49
```