Package ‘ORKM’

May 5, 2024

Title The Online Regularized K-Means Clustering Algorithm

Date 2024-5-5

Version 0.8.0.0

Description Algorithm of online regularized k-means to deal with online multi(single) view data.
The philosophy of the package is described in Guo G. (2020)
<doi:10.1080/02331888.2020.1823979>.

License MIT + file LICENSE

Encoding UTF-8

RoxygenNote 7.2.0

Author Guangbao Guo [aut, cre] (<https://orcid.org/0000-0002-4115-6218>),
Miao Yu [aut],
Haoyue Song [aut],
Ruiling Niu [aut]

Maintainer Guangbao Guo <ggb11111111@163.com>

Suggests testthat (>= 3.0.0)

Imports MASS, Matrix, stats,

Config/testthat/edition 3

NeedsCompilation no

Depends R (>= 3.5.0)

Repository CRAN

Date/Publication 2024-05-05 21:50:03 UTC

R topics documented:

`ORKM-package` *The Online Regularized K-Means Clustering Algorithm*

Description

Algorithm of online regularized k-means to deal with online multi(single) view data. The philosophy of the package is described in Guo G. (2020) <doi:10.1080/02331888.2020.1823979>.

Details

The DESCRIPTION file:

Package:	ORKM
Title:	The Online Regularized K-Means Clustering Algorithm
Date:	2024-5-5
Version:	0.8.0.0
Authors@R:	c(person("Guangbao", "Guo",role = c("aut", "cre"), email = "ggb11111111@163.com", comment =
Description:	Algorithm of online regularized k-means to deal with online multi(single) view data. The philosophy
License:	MIT + file LICENSE
Encoding:	UTF-8
Roxygen:	list(markdown = TRUE)
RoxygenNote:	7.2.0
Author:	Guangbao Guo [aut, cre] (0000-0002-4115-6218), Miao Yu [aut], Haoyue Song [aut], Ruiling Niu [a
Maintainer:	Guangbao Guo <ggb11111111@163.com>
Suggests:	testthat (>= 3.0.0)
Imports:	MASS, Matrix, stats,
Config/testthat/edition:	3

Index of help topics:

```
DMC                     Deep matrix clustering algorithm for multi-view
                        data
INDEX                   Caculate the indication on the functions
KMeans                  K-means clustering algorithm for multi/single
                        view data
OGD                     Online gradient descent algorithm for online
                        single-view data clustering
OMU                     Online multiplicative update algorithm for
                        online multi-view data clustering
ORKM-package            The Online Regularized K-Means Clustering
                        Algorithm
ORKMeans                Online regularized K-means clustering algorithm
                        for online multi-view data
PKMeans                 Power K-means clustering algorithm for single
                        view data
QCM                     The QCM data set with K=5.
RKMeans                 Regularized K-means clustering algorithm for
                        multi-view data
Washington_cites        The third view of Washington data set.
Washington_content      The second view of Washington data set.
Washington_inbound      The third view of Washington data set.
Washington_outbound     The fourth view of Washington data set.
Wisconsin_cites         The first view of Wisconsin data set.
Wisconsin_content       The second view of Wisconsin data set.
Wisconsin_inbound       The third view of Wisconsin data set.
Wisconsin_outbound      The fourth view of Wisconsin data set.
```

cora_view1	The first view of Cora data set.
cora_view2	The second view of Cora data set.
cora_view3	The third view of Cora data set.
cora_view4	The fourth view of Cora data set.
cornell_cites	The first view of Cornell data set.
cornell_content	The second view of Cornell data set.
cornell_inbound	The third view of Cornell data set.
cornell_outbound	The fourth view of Cornell data set.
labelTexas	True clustering labels for Texas data set.
labelWashington	True clustering labels for Washington data set.
labelWisconsin	True clustering labels for Wisconsin data set.
labelcora	True clustering labels for Cora data set.
labelcornell	True clustering labels for Cornell data set.
movie_1	The first view of Movie data set.
movie_2	The second view of Movie data set.
seed	A single-view data set named Seeds.
sobar	A single-view data set named Sobar.
texas_cites	The first view of Texas data set.
texas_content	The second view of Texas dataset.
texas_inbound	The third view of Texas data set.
texas_outbound	The fourth view of Texas data set.
turelabel	Ture label of Movie data set.

You can use this package for online multi-view clustering, the dataset and real labels are also provided in the package.

Author(s)

Guangbao Guo [aut, cre] (0000-0002-4115-6218), Miao Yu [aut], Haoyue Song [aut], Ruiling Niu [aut]

Maintainer: Guangbao Guo <ggb11111111@163.com>

References

Guangbao Guo, Miao Yu, Guoqi Qian, (2023), Orkm: Online Regularized k-Means Clustering for Online Multi-View Data.

See Also

https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4484209

Examples

```
library(MASS)
library(Matrix)
 yita=0.5;V=2;chushi=100;K=3;r=0.5;max.iter=10;n1=n2=n3=70;gamma=0.1;alpha=0.98;epsilon=1
  X1<-rnorm(n1,20,2);X2<-rnorm(n2,25,1.5);X3<-rnorm(n3,30,2)
  Xv<-c(X1,X2,X3)
  data<-matrix(Xv,n1+n2+n3,2)
  data[1:70,2]<-1;data[71:140,2]<-2;data[141:210,2]<-3
```

```
  truere=data[,2]
  X<-matrix(data[,1],n1+n2+n3,1)
  lamda1<-0.2;lamda2<-0.8
  lamda<-matrix(c(lamda1,lamda2),nrow=1,ncol=2)
  sol.svd <- svd(lamda)
  U1<-sol.svd$u
  D1<-sol.svd$d
  V1<-sol.svd$v
  C1<-t(U1)
  Y1<-C1/D1
  view<-V1
  view1<-matrix(view[1,])
  view2<-matrix(view[2,])
  X1<-matrix(view1,n1+n2+n3,1)
  X2<-matrix(view2,n1+n2+n3,1)
  ORKMeans(X=X1,K=K,V=V,r=r,chushi=chushi,yita=yita,gamma=gamma,epsilon=epsilon,
max.iter=max.iter,truere=truere,method=0)
```

`cora_view1` *The first view of Cora data set.*

Description

This data matrix is the first view of the multi-view data set called Cora, the keyword view. Cora data set is a multi-view data set of machine learning papers with 4 views, a sample size of nearly 3000 and a number of features of 1500, with a number of clusters of K=4.

Usage

```
data("cora_view1")
```

Format

The format is: num [1:2708, 1:2708] 0 0 0 0 0 0 0 0 1 0 ...

Details

Cora data set includes keyword view, inbound, outbound link view, and citation network view. It takes the form of a sparse matrix. It has 2708 samples and 2708 features.

Source

http://www.cs.umd.edu/projects/linqs/projects/lbc/

References

http://www.cs.umd.edu/projects/linqs/projects/lbc/

Examples

```
data(cora_view1); str(cora_view1)
```

cora_view2 *The second view of Cora data set.*

Description

This data matrix is the second view of Cora data set. It called the citation network view and the form of a sparse matrix. It has 2708 samples and 1433 features. Cora data set is a multi-view data set of machine learning papers with 4 views, a sample size of nearly 3000 and a number of features of 1500, with a number of clusters of K=4.

Usage

```
data("cora_view2")
```

Format

The format is: num [1:2708, 1:1433] 0 0 0 0 0 0 0 0 0 0 ...

Details

The second view of Cora data set.

Source

http://www.cs.umd.edu/projects/linqs/projects/lbc/

References

http://www.cs.umd.edu/projects/linqs/projects/lbc/

Examples

```
data(cora_view2); str(cora_view2)
```

cora_view3 *The third view of Cora data set.*

Description

This data matrix is the third view of Cora data set. It called the inbound link view and the form of a sparse matrix. It has 2708 samples and 2708 features. Cora data set is a multi-view data set of machine learning papers with 4 views, a sample size of nearly 3000 and a number of features of 1500, with a number of clusters of K=4.

Usage

```
data("cora_view3")
```

Format

The format is: num [1:2708, 1:2708] 0 0 0 0 0 0 0 0 0 0 ...

Details

The third view of Cora data set.

Source

http://www.cs.umd.edu/projects/linqs/projects/lbc/

References

http://www.cs.umd.edu/projects/linqs/projects/lbc/

Examples

```
data(cora_view3); str(cora_view3)
```

---

cora_view4 *The fourth view of Cora data set.*

Description

The fourth view(outbound view) of Cora data set. Cora data set is a multi-view data set of machine learning papers with 4 views, a sample size of nearly 3000 and a number of features of 1500, with a number of clusters of K=4.

Usage

```
data("cora_view4")
```

Format

The format is: num [1:2708, 1:2708] 0 0 0 0 0 0 0 0 1 0 ...

Details

The fourth view of Cora data set.

Source

http://www.cs.umd.edu/projects/linqs/projects/lbc/

References

http://www.cs.umd.edu/projects/linqs/projects/lbc/

Examples

```
data(cora_view4); str(cora_view4)
```

`cornell_cites` *The first view of Cornell data set.*

Description

Webkb data set contains web pages from four universities, with the corresponding clusters categorised as Professor, Student, Program, or Other pages. The data set contains four subsets of data, Cornell data set, Texas data set, Washington dataset, and Wisconsin data set.

Usage

```
data("cornell_cites")
```

Format

The format is: num [1:195, 1:195] 0 0 0 0 0 0 0 0 0 0 ...

Details

Cornell data set contains four views with a number of clusters of 5. This data set is the first view with a sample size of 195 and a number of features of 195.

Source

http://www.cs.cmu.edu/~webkb/

References

M. Craven, D. DiPasquo, D. Freitag, A. McCallum, T. Mitchell, K. Nigam and S. Slattery. Learning to Extract Symbolic Knowledge from the World Wide Web. Proceedings of the 15th National Conference on Artificial Intelligence (AAAI-98).

Examples

```
data(cornell_cites)
## maybe str(cornell_cites) ; plot(cornell_cites) ...
```

cornell_content *The second view of Cornell data set.*

Description

Webkb data set contains web pages from four universities, with the corresponding clusters categorised as Professor, Student, Program, or Other pages. The data set contains four subsets of data, Cornell data set, Texas data set, Washington data set, and Wisconsin data set.

Usage

```
data("cornell_content")
```

Format

The format is: num [1:195, 1:1703] 0 0 0 0 0 0 0 0 0 0 ...

Details

Cornell data set contains four views with a number of clusters of 5. This data set is the second view with a sample size of 195 and a number of features of 1703.

Source

http://www.cs.cmu.edu/~webkb/

References

M. Craven, D. DiPasquo, D. Freitag, A. McCallum, T. Mitchell, K. Nigam and S. Slattery. Learning to Extract Symbolic Knowledge from the World Wide Web. Proceedings of the 15th National Conference on Artificial Intelligence (AAAI-98).

Examples

```
data(cornell_content)
## maybe str(cornell_content) ; plot(cornell_content) ...
```

cornell_inbound *The third view of Cornell data set.*

Description

Webkb data set contains web pages from four universities, with the corresponding clusters categorised as Professor, Student, Program, or Other pages. The data set contains four subsets of data, Cornell data set, Texas data set, Washington dataset, and Wisconsin data set.

Usage

```
data("cornell_inbound")
```

Format

The format is: num [1:195, 1:195] 0 0 0 0 0 0 0 0 0 0 ...

Details

Cornell data set contains four views with a number of clusters of 5. This data set is the third view with a sample size of 195 and a number of features of 195.

Source

http://www.cs.cmu.edu/~webkb/

References

M. Craven, D. DiPasquo, D. Freitag, A. McCallum, T. Mitchell, K. Nigam and S. Slattery. Learning to Extract Symbolic Knowledge from the World Wide Web. Proceedings of the 15th National Conference on Artificial Intelligence (AAAI-98).

Examples

```
data(cornell_inbound)
## maybe str(cornell_inbound) ; plot(cornell_inbound) ...
```

cornell_outbound *The fourth view of Cornell data set.*

Description

Webkb data set contains web pages from four universities, with the corresponding clusters categorised as Professor, Student, Program, or Other pages. The data set contains four subsets of data, Cornell data set, Texas data set, Washington data set, and Wisconsin data set.

Usage

```
data("cornell_outbound")
```

Format

The format is: num [1:195, 1:195] 0 0 0 0 0 0 0 0 0 0 ...

Details

Cornell data set contains four views with a number of clusters of 5. This data set is the fourth view with a sample size of 195 and a number of features of 195.

Source

http://www.cs.cmu.edu/~webkb/

References

M. Craven, D. DiPasquo, D. Freitag, A. McCallum, T. Mitchell, K. Nigam and S. Slattery. Learning to Extract Symbolic Knowledge from the World Wide Web. Proceedings of the 15th National Conference on Artificial Intelligence (AAAI-98).

Examples

```
data(cornell_outbound)
## maybe str(cornell_outbound) ; plot(cornell_outbound) ...
```

DMC *Deep matrix clustering algorithm for multi-view data*

Description

This algorithm decomposes the multi-view data matrix into representative subspaces layer by layer, and generates a cluster at each layer. To enhance the diversity between the generated clusters, new redundant quantifiers arising from the proximity between samples in these subspaces are minimised. An iterative optimisation process is further introduced to simultaneously seek multiple clusters with quality and diversity.

Usage

```
DMC(X, K, V, r, lamda, truere, max.iter, method = 0)
```

Arguments

X	data matrix
K	number of cluster
V	number of view
r	first banlance parameter
lamda	second balance parameter
truere	true cluster result
max.iter	max iter
method	caculate the index of NMI

Value

NMI,Alpha1,center,result

Author(s)

Miao Yu

Examples

```
library(MASS)
V=2;lamda=0.5;K=3;r=0.5;max.iter=10;n1=n2=n3=70
X1<-rnorm(n1,20,2);X2<-rnorm(n2,25,1.5);X3<-rnorm(n3,30,2)
Xv<-c(X1,X2,X3)
data<-matrix(Xv,n1+n2+n3,2)
data[1:70,2]<-1;data[71:140,2]<-2;data[141:210,2]<-3
truere=data[,2]
X<-matrix(data[,1],n1+n2+n3,1)
lamda1<-0.2;lamda2<-0.8
lamda0<-matrix(c(lamda1,lamda2),nrow=1,ncol=2)
```

```
sol.svd <- svd(lamda0)
U1<-sol.svd$u
D1<-sol.svd$d
V1<-sol.svd$v
C1<-t(U1)%*%t(X)
Y1<-C1/D1
view<-V1%*%Y1
view1<-matrix(view[1,])
view2<-matrix(view[2,])
X1<-matrix(view1,n1+n2+n3,1)
X2<-matrix(view2,n1+n2+n3,1)
DMC(X=X1,K=K,V=V,lamda=lamda,r=r,max.iter=max.iter,truere=truere,method=0)
```

INDEX *Caculate the indication on the functions*

Description

This function contains the calculation of five clustering effect evaluation metrics, specifically, Purity, NMI, F-score, RI, Precision and Recall, which are used to evaluate the clustering effect of the above functions, method=0 purity;method=1,precision; method=2,recall; method=3, F-score; method=4, RI.

Usage

```
INDEX(vec1, vec2, method = 0, mybeta = 0)
```

Arguments

vec1	algorithm cluster result
vec2	true cluster result
method	Calculate the selection of indicators.
mybeta	caculate the index

Value

accuracy

Examples

```
P1<-c(1,1,1,2,3,2,1);truelabel<-c(1,1,1,2,2,2,3)
INDEX(P1,truelabel,method=0);INDEX(P1,truelabel,method=2)
```

`KMeans` *K-means clustering algorithm for multi/single view data*

Description

The K-means clustering algorithm is a common clustering algorithm that divides a data set into K clusters, with each cluster represented using the mean of all samples within the cluster, referring to that mean as the j-cluster centre. The algorithm is unsupervised learning, where the categories are not known in advance and similar objects are automatically grouped into the same cluster. The K-means algorithm achieves clustering by calculating the distance between each point and the centre of mass of different clusters and assigning it to the nearest cluster. The algorithm is simple and easy to implement, but is susceptible to the initial centre of mass, the possibility of empty clusters, and the possibility of convergence to local minima. Clustering applications can be used to discover different groups of users, allowing for tasks such as precision marketing, document segmentation, finding people in the same circle in social networks, and handling anomalous data.

Usage

```
KMeans(X, K, V, r, max.iter, truere, method = 0)
```

Arguments

`X`	data matrix
`K`	number of cluster
`V`	number of view
`r`	balance parameter
`truere`	true cluster result
`max.iter`	max iter
`method`	caculate the index of NMI

Value

NMI,weight,center,result

Author(s)

Miao Yu

Examples

```
library(MASS)
library(Matrix)
V=2;K=3;r=0.5;max.iter=10;n1=n2=n3=70
X1<-rnorm(n1,20,2);X2<-rnorm(n2,25,1.5);X3<-rnorm(n3,30,2)
Xv<-c(X1,X2,X3)
data<-matrix(Xv,n1+n2+n3,2)
```

```
data[1:70,2]<-1;data[71:140,2]<-2;data[141:210,2]<-3
truere=data[,2]
X<-matrix(data[,1],n1+n2+n3,1)
lamda1<-0.2;lamda2<-0.8
lamda<-matrix(c(lamda1,lamda2),nrow=1,ncol=2)
sol.svd <- svd(lamda)
U1<-sol.svd$u
D1<-sol.svd$d
V1<-sol.svd$v
C1<-t(U1)
Y1<-C1/D1
view<-V1
view1<-matrix(view[1,])
view2<-matrix(view[2,])
X1<-matrix(view1,n1+n2+n3,1)
X2<-matrix(view2,n1+n2+n3,1)
KMeans(X=X1,K=K,V=V,r=r,max.iter=max.iter,truere=truere,method=0)
```

`labelcora` *True clustering labels for Cora data set.*

Description

True clustering labels for the Cora dataset, which can be applied to 4 views.

Usage

```
data("labelcora")
```

Format

The format is: chr [1:2708] "1" "2" "3" "3" "4" "4" "5" "1" "1" "5" "1" "6" "4" "7" ...

Details

True clustering labels for the Cora dataset, which can be applied to 4 views.

Source

http://www.cs.umd.edu/projects/linqs/projects/lbc/

References

http://www.cs.umd.edu/projects/linqs/projects/lbc/

Examples

```
data(labelcora)
```

labelcornell *True clustering labels for Cornell data set.*

Description

Webkb data set contains web pages from four universities, with the corresponding clusters categorised as Professor, Student, Program, or Other pages. The data set contains four subsets of data, Cornell data set, Texas data set, Washington data set, and Wisconsin data set.

Usage

```
data("labelcornell")
```

Format

The format is: int [1:195, 1] 1 1 2 3 3 3 2 4 3 3 ... - attr(*, "dimnames")=List of 2 ..$: NULL ..$: chr "V1"

Details

Cornell dat aset contains four views with a number of clusters of 5. You can use this true label to calculate your clustering accuracy.

Source

http://www.cs.cmu.edu/~webkb/

References

M. Craven, D. DiPasquo, D. Freitag, A. McCallum, T. Mitchell, K. Nigam and S. Slattery. Learning to Extract Symbolic Knowledge from the World Wide Web. Proceedings of the 15th National Conference on Artificial Intelligence (AAAI-98).

Examples

```
data(labelcornell)
## maybe str(labelcornell) ; plot(labelcornell) ...
```

labelTexas *True clustering labels for Texas data set.*

Description

Webkb data set contains web pages from four universities, with the corresponding clusters categorised as Professor, Student, Program, or Other pages. The data set contains four subsets of data, Cornell data set, Texas data set, Washington data set, and Wisconsin data set.

Usage

```
data("labelTexas")
```

Format

The format is: num [1:187] 1 2 3 1 4 3 3 3 4 1 ...

Details

Texas data set contains four views with a number of clusters of 5. You can use this true label to calculate your clustering accuracy.

Source

http://www.cs.cmu.edu/~webkb/

References

M. Craven, D. DiPasquo, D. Freitag, A. McCallum, T. Mitchell, K. Nigam and S. Slattery. Learning to Extract Symbolic Knowledge from the World Wide Web. Proceedings of the 15th National Conference on Artificial Intelligence (AAAI-98).

Examples

```
data(labelTexas)
## maybe str(labelTexas) ; plot(labelTexas) ...
```

labelWashington *True clustering labels for Washington data set.*

Description

Webkb data set contains web pages from four universities, with the corresponding clusters categorised as Professor, Student, Program, or Other pages. The data set contains four subsets of data, Cornell data set, Texas data set, Washington data set, and Wisconsin data set.

Usage

```
data("labelWashington")
```

Format

The format is: num [1:230] 1 2 2 2 2 2 2 2 2 2 ...

Details

Washington data set contains four views with a number of clusters of 5. You can use this true label to calculate your clustering accuracy.

Source

http://www.cs.cmu.edu/~webkb/

References

M. Craven, D. DiPasquo, D. Freitag, A. McCallum, T. Mitchell, K. Nigam and S. Slattery. Learning to Extract Symbolic Knowledge from the World Wide Web. Proceedings of the 15th National Conference on Artificial Intelligence (AAAI-98).

Examples

```
data(labelWashington)
## maybe str(labelWashington) ; plot(labelWashington) ...
```

`labelWisconsin` *True clustering labels for Wisconsin data set.*

Description

Webkb data set contains web pages from four universities, with the corresponding clusters categorised as Professor, Student, Program, or Other pages. The data set contains four subsets of data, Cornell dataset, Texas dataset, Washington dataset, and Wisconsin data set.

Usage

```
data("labelWisconsin")
```

Format

The format is: num [1:265] 1 2 3 3 1 1 1 1 1 1 ...

Details

Wisconsin data set contains four views with a number of clusters of 5. You can use this true label to calculate your clustering accuracy.

Source

http://www.cs.cmu.edu/~webkb/

References

M. Craven, D. DiPasquo, D. Freitag, A. McCallum, T. Mitchell, K. Nigam and S. Slattery. Learning to Extract Symbolic Knowledge from the World Wide Web. Proceedings of the 15th National Conference on Artificial Intelligence (AAAI-98).

Examples

```
data(labelWisconsin)
## maybe str(labelWisconsin) ; plot(labelWisconsin) ...
```

movie_1 *The first view of Movie data set.*

Description

The first view(keyword view) of Movie data set. Movie data set contains 2 views, each containing 1878 variables from 617 instances, and the number of clusters to be clustered is K = 17. The number of clusters is large, so it is difficult to cluster. The data set was extracted from IMDb and the main objective was to to find the movie genres, combined from two view matrices.

Usage

```
data("movie_1")
```

Format

The format is: num [1:617, 1:1878] 1 0 0 0 0 0 0 0 0 0 ...

Details

The first view of Movie dataset.

Source

https://lig-membres.imag.fr/grimal/data.html.

References

C. Grimal. the multi-view movie data set. 2010. URL https://lig-membres.imag.fr/grimal/data.html.

Examples

```
data(movie_1); str(movie_1)
```

movie_2 *The second view of Movie data set.*

Description

The second view(participant view) of Movie data set. Movie data set contains 2 views, each containing 1878 variables from 617 instances, and the number of clusters to be clustered is K = 17. The number of clusters is large, so it is difficult to cluster. The data set was extracted from IMDb and the main objective was to to find the movie genres, combined from two view matrices.

Usage

```
data("movie_2")
```

Format

The format is: num [1:617, 1:1398] 1 0 0 0 0 0 0 0 0 0 ...

Details

The second view of Movie data set.

Source

https://lig-membres.imag.fr/grimal/data.html.

References

C. Grimal. the multi-view movie data set. 2010. URL https://lig-membres.imag.fr/grimal/data.html.

Examples

```
data(movie_2); str(movie_2)
```

OGD — *Online gradient descent algorithm for online single-view data clustering*

Description

Online gradient descent is an optimisation algorithm in machine learning for when the amount of data is too large to process all the data at the same time. In this algorithm, the model parameters are updated based on a single training sample, rather than using the entire training set. The direction of each update is determined by the direction of the gradient of the current sample, and the local or global extremes of the gradient descent algorithm depend on the order of the sampled samples. Compared to Batch Gradient Descent (BGD) algorithm, online gradient descent algorithms can process data streams and update the model as they process the data, and are therefore more efficient for large-scale data. However, online gradient descent algorithm should only be used if the data stream is continuously present and updated.

Usage

```
OGD(X, K, gamma, max.m, chushi, yita, epsilon, truere, method = 0)
```

Arguments

`X`	data matrix
`K`	number of cluster
`gamma`	step size
`yita`	the regularized parameter
`truere`	true cluster result

max.m	max iter
epsilon	epsilon
chushi	the initial value
method	caculate the index of NMI

Value

result,NMI,M

Author(s)

Miao Yu

Examples

```
yita=0.5;V=2;K=3;chushi=100;epsilon=1;gamma=0.1;max.m=10;n1=n2=n3=70
 X1<-rnorm(n1,20,2);X2<-rnorm(n2,25,1.5);X3<-rnorm(n3,30,2)
 Xv<-c(X1,X2,X3)
 data<-matrix(Xv,n1+n2+n3,2)
 data[1:70,2]<-1;data[71:140,2]<-2;data[141:210,2]<-3
 X<-matrix(data[,1],n1+n2+n3,1)
 truere=data[,2]
 lamda1<-0.2;lamda2<-0.8
 lamda<-matrix(c(lamda1,lamda2),nrow=1,ncol=2)
 sol.svd <- svd(lamda)
  U1<-sol.svd$u
 D1<-sol.svd$d
 V1<-sol.svd$v
 C1<-t(U1)
 Y1<-C1/D1
 view<-V1
 view1<-matrix(view[1,])
 view2<-matrix(view[2,])
 X1<-matrix(view1,n1+n2+n3,1)
 X2<-matrix(view2,n1+n2+n3,1)
 OGD(X=X1,K=K,gamma=gamma,max.m=max.m,chushi=chushi,
yita=yita,epsilon=epsilon,truere=truere,method=0)
```

OMU *Online multiplicative update algorithm for online multi-view data clustering*

Description

This algorithm integrates the multiplicative normalization factor as an additional term in the original additivity update rule, which usually has approximately opposite direction. Thus, the improved iteration rule can be easily converted to a multiplicative version. After each iteration After each iteration, non-negativity is maintained.

Usage

```
OMU(X,K,V,chushi,yita,r,max.iter,epsilon,truere,method=0)
```

Arguments

X	data matrix
K	number of cluster
V	number of view
chushi	the initial value
yita	the regularized parameter
r	banlance parameter
max.iter	max iter
epsilon	epsilon
truere	true cluster result
method	caculate the index of NMI

Value

NMI,result,M

Examples

```
yita=0.5;V=2;chushi=100;K=3;r=0.5;max.iter=10;n1=n2=n3=70;epsilon=1
X1<-rnorm(n1,20,2);X2<-rnorm(n2,25,1.5);X3<-rnorm(n3,30,2)
Xv<-c(X1,X2,X3)
data<-matrix(Xv,n1+n2+n3,2)
data[1:70,2]<-1;data[71:140,2]<-2;data[141:210,2]<-3
truere=data[,2]
X<-matrix(data[,1],n1+n2+n3,1)
lamda1<-0.2;lamda2<-0.8
lamda<-matrix(c(lamda1,lamda2),nrow=1,ncol=2)
sol.svd <- svd(lamda)
U1<-sol.svd$u
D1<-sol.svd$d
V1<-sol.svd$v
C1<-t(U1)%*%t(X)
Y1<-C1/D1
view<-V1%*%Y1
view1<-matrix(view[1,])
view2<-matrix(view[2,])
X1<-matrix(view1,n1+n2+n3,1)
X2<-matrix(view2,n1+n2+n3,1)
OMU(X=X1,K=K,V=V,chushi=chushi,yita=yita,r=r,max.iter=max.iter,
epsilon=epsilon,truere=truere,method=0)
```

`ORKMeans` *Online regularized K-means clustering algorithm for online multi-view data*

Description

For the online clustering problem, this function proposes the Online Regularized K-means Clustering (ORKMC) method to deal with online multi-view data. Firstly, for the clustering problem of multi-view data, a non-negative matrix decomposition is used as the starting point of the model to find the indicator matrix and cluster centres of each cluster; for online updating, a projected gradient descent method is proposed to perform online updating to improve the accuracy and speed of data clustering; for the overfitting phenomenon, regularisation is proposed to avoid the above problem. In addition, since the choice of regularization parameters is extremely important to the effectiveness of the ORKMC algorithm, the choice of regularization parameters varies in different datasets. In this paper, a suitable range of regularisation parameters and model parameters is given. The effectiveness of the ORKMC algorithm is tested through an extensive study of multi-view/single-view data. The validity of the ORKMC algorithm is tested through an extensive study of multi-view/single-view data.

Usage

```
ORKMeans(X,K,V,chushi,r,yita,gamma,alpha,epsilon,truere,max.iter,method=0)
```

Arguments

`X`	is the online single/multi-view data matrix
`K`	is the number of cluster
`V`	is the view of X
`chushi`	is the initial value for online
`yita`	is the regularized parameter
`r`	is the banlance parameter
`gamma`	is the step size
`alpha`	is the caculated the weight of view
`epsilon`	is the epsilon
`truere`	is the ture label in data set
`max.iter`	is the max iter
`method`	is the caluate the NMI

Value

NMI,weight,center,result

Author(s)

Miao Yu

Examples

```
library(MASS)
library(Matrix)
 yita=0.5;V=2;chushi=100;K=3;r=0.5;max.iter=10;n1=n2=n3=70;gamma=0.1;alpha=0.98;epsilon=1
  X1<-rnorm(n1,20,2);X2<-rnorm(n2,25,1.5);X3<-rnorm(n3,30,2)
  Xv<-c(X1,X2,X3)
  data<-matrix(Xv,n1+n2+n3,2)
  data[1:70,2]<-1;data[71:140,2]<-2;data[141:210,2]<-3
  truere=data[,2]
  X<-matrix(data[,1],n1+n2+n3,1)
  lamda1<-0.2;lamda2<-0.8
  lamda<-matrix(c(lamda1,lamda2),nrow=1,ncol=2)
  sol.svd <- svd(lamda)
  U1<-sol.svd$u
  D1<-sol.svd$d
  V1<-sol.svd$v
  C1<-t(U1)
  Y1<-C1/D1
  view<-V1
  view1<-matrix(view[1,])
  view2<-matrix(view[2,])
  X1<-matrix(view1,n1+n2+n3,1)
  X2<-matrix(view2,n1+n2+n3,1)
  ORKMeans(X=X1,K=K,V=V,r=r,chushi=chushi,yita=yita,gamma=gamma,epsilon=epsilon,
max.iter=max.iter,truere=truere,method=0)
```

PKMeans *Power K-means clustering algorithm for single view data*

Description

The power K-means algorithm is a generalization of the Lloyd algorithm, which approximates the ordinary K-means algorithm by a majorization-minimization method with the descent properties and lower complexity of the Lloyd algorithm. The power K-means embeds the K-means problem into a series of better performing problems. These smooth intermediate problems have a smoother objective function and tend to guide the clustering to find a global minimum with the K-means as the objective. The method has the same iteration complexity as Lloyd's algorithm, reduces sensitivity to initialization, and greatly improves algorithm performance in the high-dimensional case.

Usage

```
PKMeans(X, K, yitapower, sm, max.m, truere, method = 0)
```

Arguments

X	is the data matrix
K	is the number of cluster
yitapower	is the regularized parameter

sm	is the banlance parameter
max.m	is the max iter
truere	is the ture label in data set
method	is the caluate the NMI

Value

center,NMI,result

Author(s)

Miao Yu

Examples

```
library(MASS)
  yitapower=0.5;K=3;sm=0.5;max.m=100;n1=n2=n3=70
  X1<-rnorm(n1,20,2);X2<-rnorm(n2,25,1.5);X3<-rnorm(n3,30,2)
  Xv<-c(X1,X2,X3)
  data<-matrix(Xv,n1+n2+n3,2)
  data[1:70,2]<-1;data[71:140,2]<-2;data[141:210,2]<-3
  truere=data[,2]
  X11<-matrix(data[,1],n1+n2+n3,1)
  PKMeans(X=X11,K=K,yitapower=yitapower,sm=sm,max.m=max.m,truere=truere,method=0)
```

QCM *The QCM data set with K=5.*

Description

Five different QCM gas sensors were used and five different gas measurements were made for each sensor (1-octanol, 1-propanol, 2-butanol, 2-propanol and 1-isobutanol).

Usage

```
data("QCM")
```

Format

The format is: num [1:125, 1:15] -10.06 -9.69 -12.07 -14.21 -16.57 ...

Details

The QCM data set with K=5.

Source

https://www.sciencedirect.com/science/article/pii/S2215098619303337.

References

M. F. Adak, P. Lieberzeit, P. Jarujamrus, and N. Yumusak. Classification of alcohols obtained by qcm sensors with different characteristics using abc based neural network. Engineering Science and Technology, an International Journal, 23(3):463–469, 2020. ISSN 2215-0986. doi: https://doi.org/10.1016/j.jestch.2019.06.011. URL https://www.sciencedirect.com/science/article/pii/S2215098619303337.

Examples

```
data(QCM); str(QCM)
```

RKMeans *Regularized K-means clustering algorithm for multi-view data*

Description

This function improves the regularized K-means clustering (RKMC) algorithm for the multi-view data clustering problem. Specifically, the regularisation term is added to the K-means algorithm to avoid overfitting of the data. Numerical analysis shows that the RKMC algorithm significantly improves the clustering performance compared to other methods. In addition, in order to reveal the structure of real data as realistically as possible, improve the clustering accuracy of high-dimensional data, and balance the weights of each view, the RKMC algorithm assigns a series of learnable weight values to each view, thus reflecting the relationship and compatibility of each view more flexibly.

Usage

```
RKMeans(X, K, V, yita, r, max.iter, truere, method = 0)
```

Arguments

`X`	is the data matrix
`K`	is the number of cluster
`V`	is the view of X
`yita`	is the regularized parameter
`r`	is the banlance parameter
`max.iter`	is the max iter
`truere`	is the ture label in data set
`method`	is the caluate the NMI

Value

NMI,weight,center,result

Author(s)

Miao Yu

Examples

```
library(MASS)
library(Matrix)
yita=0.5;V=2;K=3;r=0.5;max.iter=10;n1=n2=n3=70
X1<-rnorm(n1,20,2);X2<-rnorm(n2,25,1.5);X3<-rnorm(n3,30,2)
Xv<-c(X1,X2,X3)
data<-matrix(Xv,n1+n2+n3,2)
data[1:70,2]<-1;data[71:140,2]<-2;data[141:210,2]<-3
X<-matrix(data[,1],n1+n2+n3,1)
truere=data[,2]
lamda1<-0.2;lamda2<-0.8
lamda<-matrix(c(lamda1,lamda2),nrow=1,ncol=2)
sol.svd <- svd(lamda)
U1<-sol.svd$u
D1<-sol.svd$d
V1<-sol.svd$v
C1<-t(U1)
Y1<-C1/D1
view<-V1
view1<-matrix(view[1,])
view2<-matrix(view[2,])
X1<-matrix(view1,n1+n2+n3,1)
X2<-matrix(view2,n1+n2+n3,1)
RKMeans(X=X1,K=K,V=V,yita=yita,r=r,max.iter=max.iter,truere=truere,method=0)
```

`seed` *A single-view data set named Seeds.*

Description

The Seeds data set holds data on the area, circumference, compaction, seed length, seed width, asymmetry factor, length of the ventral groove of the seed and category data for different varieties of wheat seeds. The data set contains a total of 210 records, 7 features, and one label, which is divided into 3 categories.

Usage

```
data("seed")
```

Format

The format is: num [1:210, 1:8] 15.3 14.9 14.3 13.8 16.1 ...

Details

A single-view data set named seed.

Source

http://archive.ics.uci.edu/ml/datasets/seeds

References

http://archive.ics.uci.edu/ml/datasets/seeds

Examples

```
data(seed); str(seed)
```

`sobar` *A single-view data set named Sobar.*

Description

A single-view data set named Sobar. Sobar data set is a behavioural risk data set for cervical cancer, which has a number of clusters of 2.

Usage

```
data("sobar")
```

Format

The format is: num [1:72, 1:20] 10 10 10 10 8 10 10 8 10 7 ...

Details

A single-view data set named sobar.

Source

http://archive.ics.uci.edu/ml/datasets/Cervical+Cancer+Behavior+Risk

References

http://archive.ics.uci.edu/ml/datasets/Cervical+Cancer+Behavior+Risk

Examples

```
data(sobar); str(sobar)
```

texas_cites *The first view of Texas data set.*

Description

Webkb data set contains web pages from four universities, with the corresponding clusters categorised as Professor, Student, Program, or Other pages. The data set contains four subsets of data, Cornell data set, Texas data set, Washington data set, and Wisconsin data set.

Usage

```
data("texas_cites")
```

Format

The format is: num [1:187, 1:187] 0 1 1 1 0 1 1 0 1 0 ...

Details

Texas data set contains four views with a number of clusters of 5. This data set is the first view with a sample size of 187 and a number of features of 187.

Source

http://www.cs.cmu.edu/~webkb/

References

M. Craven, D. DiPasquo, D. Freitag, A. McCallum, T. Mitchell, K. Nigam and S. Slattery. Learning to Extract Symbolic Knowledge from the World Wide Web. Proceedings of the 15th National Conference on Artificial Intelligence (AAAI-98).

Examples

```
data(texas_cites)
## maybe str(texas_cites) ; plot(texas_cites) ...
```

texas_content *The second view of Texas dataset.*

Description

Webkb data set contains web pages from four universities, with the corresponding clusters categorised as Professor, Student, Program, or Other pages. The data set contains four subsets of data, Cornell dataset, Texas dataset, Washington dataset, and Wisconsin dataset.

Usage

```
data("texas_content")
```

Format

The format is: num [1:187, 1:1703] 0 0 0 0 0 0 0 0 0 0 ...

Details

Texas data set contains four views with a number of clusters of 5. This data set is the second view with a sample size of 187 and a number of features of 1703.

Source

http://www.cs.cmu.edu/~webkb/

References

M. Craven, D. DiPasquo, D. Freitag, A. McCallum, T. Mitchell, K. Nigam and S. Slattery. Learning to Extract Symbolic Knowledge from the World Wide Web. Proceedings of the 15th National Conference on Artificial Intelligence (AAAI-98).

Examples

```
data(texas_content)
## maybe str(texas_content) ; plot(texas_content) ...
```

texas_inbound *The third view of Texas data set.*

Description

Webkb data set contains web pages from four universities, with the corresponding clusters categorised as Professor, Student, Program, or Other pages. The data set contains four subsets of data, Cornell data set, Texas data set, Washington data set, and Wisconsin data set.

Usage

```
data("texas_inbound")
```

Format

The format is: num [1:187, 1:187] 0 0 0 0 0 0 0 0 0 0 ...

Details

Texas data set contains four views with a number of clusters of 5. This data set is the third view with a sample size of 187 and a number of features of 187.

Source

http://www.cs.cmu.edu/~webkb/

References

M. Craven, D. DiPasquo, D. Freitag, A. McCallum, T. Mitchell, K. Nigam and S. Slattery. Learning to Extract Symbolic Knowledge from the World Wide Web. Proceedings of the 15th National Conference on Artificial Intelligence (AAAI-98).

Examples

```
data(texas_inbound)
## maybe str(texas_inbound) ; plot(texas_inbound) ...
```

texas_outbound *The fourth view of Texas data set.*

Description

Webkb data set contains web pages from four universities, with the corresponding clusters categorised as Professor, Student, Program, or Other pages. The data set contains four subsets of data, Cornell data set, Texas data set, Washington data set, and Wisconsin data set.

Usage

```
data("texas_outbound")
```

Format

The format is: num [1:187, 1:187] 0 1 1 1 0 1 1 0 1 0 ...

Details

Texas data set contains four views with a number of clusters of 5. This data set is the fourth view with a sample size of 187 and a number of features of 187.

Source

http://www.cs.cmu.edu/~webkb/

References

M. Craven, D. DiPasquo, D. Freitag, A. McCallum, T. Mitchell, K. Nigam and S. Slattery. Learning to Extract Symbolic Knowledge from the World Wide Web. Proceedings of the 15th National Conference on Artificial Intelligence (AAAI-98).

Examples

```
data(texas_outbound)
## maybe str(texas_outbound) ; plot(texas_outbound) ...
```

turelabel *Ture label of Movie data set.*

Description

Ture label of Movie data set. You can use it to calculate the accuracy of the clustering results.

Usage

```
data("turelabel")
```

Format

A data frame with 617 observations on the following variable.

V1 a numeric vector

Details

Ture label of Movie data set.

Source

https://lig-membres.imag.fr/grimal/data.html.

References

C. Grimal. the multi-view movie data set. 2010. URL https://lig-membres.imag.fr/grimal/data.html.

Examples

```
data(turelabel)
## maybe str(turelabel) ; plot(turelabel) ...
```

Washington_cites *The third view of Washington data set.*

Description

Webkb data set contains web pages from four universities, with the corresponding clusters categorised as Professor, Student, Program, or Other pages. The data set contains four subsets of data, Cornell data set, Texas data set, Washington data set, and Wisconsin data set.

Usage

```
data("Washington_cites")
```

Format

The format is: num [1:230, 1:230] 2 0 0 0 0 0 0 0 0 0 ...

Details

Washington data set contains four views with a number of clusters of 5. This data set is the third view with a sample size of 230 and a number of features of 230.

Source

http://www.cs.cmu.edu/~webkb/

References

M. Craven, D. DiPasquo, D. Freitag, A. McCallum, T. Mitchell, K. Nigam and S. Slattery. Learning to Extract Symbolic Knowledge from the World Wide Web. Proceedings of the 15th National Conference on Artificial Intelligence (AAAI-98).

Examples

```
data(Washington_cites)
## maybe str(Washington_cites) ; plot(Washington_cites) ...
```

Washington_content *The second view of Washington data set.*

Description

Webkb dataset contains web pages from four universities, with the corresponding clusters categorised as Professor, Student, Program, or Other pages. The data set contains four subsets of data, Cornell data set, Texas data set, Washington data set, and Wisconsin data set.

Usage

```
data("Washington_content")
```

Format

The format is: num [1:230, 1:1703] 0 0 0 0 0 0 0 0 0 0 ...

Details

Washington data set contains four views with a number of clusters of 5. This data set is the second view with a sample size of 230 and a number of features of 1703.

Source

http://www.cs.cmu.edu/~webkb/

References

M. Craven, D. DiPasquo, D. Freitag, A. McCallum, T. Mitchell, K. Nigam and S. Slattery. Learning to Extract Symbolic Knowledge from the World Wide Web. Proceedings of the 15th National Conference on Artificial Intelligence (AAAI-98).

Examples

```
data(Washington_content)
## maybe str(Washington_content) ; plot(Washington_content) ...
```

Washington_inbound *The third view of Washington data set.*

Description

Webkb data set contains web pages from four universities, with the corresponding clusters categorised as Professor, Student, Program, or Other pages. The data set contains four subsets of data, Cornell data set, Texas data set, Washington data set, and Wisconsin data set.

Usage

```
data("Washington_inbound")
```

Format

The format is: num [1:230, 1:230] 1 0 0 0 0 0 0 0 0 0 ...

Details

Washington data set contains four views with a number of clusters of 5. This data set is the third view with a sample size of 230 and a number of features of 230.

Source

http://www.cs.cmu.edu/~webkb/

References

M. Craven, D. DiPasquo, D. Freitag, A. McCallum, T. Mitchell, K. Nigam and S. Slattery. Learning to Extract Symbolic Knowledge from the World Wide Web. Proceedings of the 15th National Conference on Artificial Intelligence (AAAI-98).

Examples

```
data(Washington_inbound)
## maybe str(Washington_inbound) ; plot(Washington_inbound) ...
```

Washington_outbound *The fourth view of Washington data set.*

Description

Webkb data set contains web pages from four universities, with the corresponding clusters categorised as Professor, Student, Program, or Other pages. The data set contains four subsets of data, Cornell data set, Texas data set, Washington data set, and Wisconsin data set.

Usage

```
data("Washington_outbound")
```

Format

The format is: num [1:230, 1:230] 1 0 0 0 0 0 0 0 0 0 ...

Details

Washington data set contains four views with a number of clusters of 5. This data set is the fourth view with a sample size of 230 and a number of features of 230.

Source

http://www.cs.cmu.edu/~webkb/

References

M. Craven, D. DiPasquo, D. Freitag, A. McCallum, T. Mitchell, K. Nigam and S. Slattery. Learning to Extract Symbolic Knowledge from the World Wide Web. Proceedings of the 15th National Conference on Artificial Intelligence (AAAI-98).

Examples

```
data(Washington_outbound)
## maybe str(Washington_outbound) ; plot(Washington_outbound) ...
```

Wisconsin_cites *The first view of Wisconsin data set.*

Description

Webkb data set contains web pages from four universities, with the corresponding clusters categorised as Professor, Student, Program, or Other pages. The data set contains four subsets of data, Cornell data set, Texas data set, Washington data set, and Wisconsin data set.

Usage

```
data("Wisconsin_cites")
```

Format

The format is: num [1:265, 1:265] 0 1 0 1 0 0 0 0 0 0 ...

Details

Wisconsin data set contains four views with a number of clusters of 5. This data set is the first view with a sample size of 265 and a number of features of 265.

Source

http://www.cs.cmu.edu/~webkb/

References

M. Craven, D. DiPasquo, D. Freitag, A. McCallum, T. Mitchell, K. Nigam and S. Slattery. Learning to Extract Symbolic Knowledge from the World Wide Web. Proceedings of the 15th National Conference on Artificial Intelligence (AAAI-98).

Examples

```
data(Wisconsin_cites)
## maybe str(Wisconsin_cites) ; plot(Wisconsin_cites) ...
```

`Wisconsin_content` *The second view of Wisconsin data set.*

Description

Webkb data set contains web pages from four universities, with the corresponding clusters categorised as Professor, Student, Program, or Other pages. The data set contains four subsets of data, Cornell data set, Texas data set, Washington data set, and Wisconsin data set.

Usage

```
data("Wisconsin_content")
```

Format

The format is: num [1:265, 1:1703] 0 0 0 0 0 0 0 0 0 0 ...

Details

Wisconsin data set contains four views with a number of clusters of 5. This data set is the second view with a sample size of 265 and a number of features of 1703.

Source

http://www.cs.cmu.edu/~webkb/

References

M. Craven, D. DiPasquo, D. Freitag, A. McCallum, T. Mitchell, K. Nigam and S. Slattery. Learning to Extract Symbolic Knowledge from the World Wide Web. Proceedings of the 15th National Conference on Artificial Intelligence (AAAI-98).

Examples

```
data(Wisconsin_content)
## maybe str(Wisconsin_content) ; plot(Wisconsin_content) ...
```

`Wisconsin_inbound` *The third view of Wisconsin data set.*

Description

Webkb data set contains web pages from four universities, with the corresponding clusters categorised as Professor, Student, Program, or Other pages. The data set contains four subsets of data, Cornell data set, Texas data set, Washington data set, and Wisconsin data set.

Usage

```
data("Wisconsin_inbound")
```

Format

The format is: num [1:265, 1:265] 0 1 0 1 0 0 0 0 0 0 ...

Details

Wisconsin data set contains four views with a number of clusters of 5. This data set is the third view with a sample size of 265 and a number of features of 265.

Source

http://www.cs.cmu.edu/~webkb/

References

M. Craven, D. DiPasquo, D. Freitag, A. McCallum, T. Mitchell, K. Nigam and S. Slattery. Learning to Extract Symbolic Knowledge from the World Wide Web. Proceedings of the 15th National Conference on Artificial Intelligence (AAAI-98).

Examples

```
data(Wisconsin_inbound)
## maybe str(Wisconsin_inbound) ; plot(Wisconsin_inbound) ...
```

`Wisconsin_outbound` *The fourth view of Wisconsin data set.*

Description

Webkb data set contains web pages from four universities, with the corresponding clusters categorised as Professor, Student, Program, or Other pages. The data set contains four subsets of data, Cornell data set, Texas data set, Washington data set, and Wisconsin data set.

Usage

```
data("Wisconsin_outbound")
```

Format

The format is: num [1:265, 1:265] 0 0 0 0 0 0 0 0 0 0 ...

Details

Wisconsin data set contains four views with a number of clusters of 5. This data set is the fourth view with a sample size of 265 and a number of features of 265.

Source

http://www.cs.cmu.edu/~webkb/

References

M. Craven, D. DiPasquo, D. Freitag, A. McCallum, T. Mitchell, K. Nigam and S. Slattery. Learning to Extract Symbolic Knowledge from the World Wide Web. Proceedings of the 15th National Conference on Artificial Intelligence (AAAI-98).

Examples

```
data(Wisconsin_outbound)
## maybe str(Wisconsin_outbound) ; plot(Wisconsin_outbound) ...
```

Index