Package 'RKUM'

January 20, 2025

Type Package

Title Robust Kernel Unsupervised Methods

Version 0.1.1.1

Author Md Ashad Alam

Maintainer Md Ashad Alam <malam@tulane.edu>

Description Robust kernel center matrix, robust kernel cross-covariance operator for kernel unsupervised methods, kernel canonical correlation analysis, influence function of identifying significant outliers or atypical objects from multi-modal datasets. Alam, M. A, Fukumizu, K., Wang Y.-P. (2018) <doi:10.1016/j.neucom.2018.04.008>. Alam, M. A, Calhoun, C. D., Wang Y.-P. (2018) <doi:10.1016/j.csda.2018.03.013>.

License GPL-3

Encoding UTF-8

Imports stats, graphics

NeedsCompilation no

Repository CRAN

Date/Publication 2022-06-22 04:50:17 UTC

Contents

1

Index   **[26](#)**

---

gkm   *Kernel Matrix Using Guasian Kernel*

Description

Many radial basis function kernels, such as the Gaussian kernel, map X into a infinte dimensional space. While the Gaussian kernel has a free parameter (bandwidth), it still follows a number of theoretical properties such as boundedness, consistence, universality, robustness etc. It is the most applicable kernel of the positive definite kernel based methods.

Usage

```
gkm(X)
```

Arguments

X   a data matrix.

Details

Many radial basis function kernels, such as the Gaussian kernel, map input sapce into a infinite dimensional space. The Gaussian kernel has a a number of theoretical properties such as boundedness, consistence, universality and robustness, etc.

Value

K   a Gram/ kernel matrix

Author(s)

Md Ashad Alam <malam@tulane.edu>

References

Md. Ashad Alam, Hui-Yi Lin, HOng-Wen Deng, Vince Calhour Yu-Ping Wang (2018), A kernel machine method for detecting higher order interactions in multimodal datasets: Application to schizophrenia, Journal of Neuroscience Methods, Vol. 309, 161-174.

Md Ashad Alam, Kenji Fukumizu and Yu-Ping Wang (2018), Influence Function and Robust Variant of Kernel Canonical Correlation Analysis, Neurocomputing, Vol. 304 (2018) 12-29.

M. Romanazzi (1992), Influence in canonical correlation analysis, Psychometrika vol 57(2) (1992) 237-259.

Examples

```
##Dummy data:
X<-matrix(rnorm(1000),100)
gkm(X)
```

---

gm3edc   *A helper function*

Description

#An matrices dicomposition function

Usage

```
gm3edc(Amat, Bmat, Cmat)
```

Arguments

Amat	a square matrix
Bmat	a square matrix
Cmat	a square matrix

Author(s)

Md Ashad Alam <malam@tulane.edu>

---

gmedc   *A helper function*

Description

#An matrices dicomposition function

Usage

```
gmedc(A, B = diag(nrow(A)))
```

Arguments

A	a square matrix
B	a diagonal matrix

Author(s)

Md Ashad Alam <malam@tulane.edu>

---

gmi   *A helper function*

Description

###An function to adjust

Usage

```
gmi(X, tol = sqrt(.Machine$double.eps))
```

Arguments

X	a square matrix
tol	a real value

Author(s)

Md Ashad Alam <malam@tulane.edu>

hadr   *Hampel's psi function*

Description

##The ratio of the first derivative of the Hampel loss fuction to the argument. Tuning constants are fixed in different quintiles.

Usage

```
hadr(u)
```

Arguments

u   vector values

Value

a real value

Author(s)

Md Ashad Alam <malam@tulane.edu>

References

Md Ashad Alam, Kenji Fukumizu and Yu-Ping Wang (2018), Influence Function and Robust Variant of Kernel Canonical Correlation Analysis, Neurocomputing, Vol. 304 (2018) 12-29.

M. Romanazzi (1992), Influence in canonical correlation analysis, Psychometrika vol 57(2) (1992) 237-259.

See Also

#See Also as gkm, hudr

halfun   *A Hampel loss function*

Description

#Tuning constants of the Hampel loss fuction are fixed in different quintiles of the arguments.

Usage

```
halfun(u)
```

Arguments

u   vector of values.

Value

comp1   a real number

Author(s)

Md Ashad Alam <malam@tulane.edu>

References

Md Ashad Alam, Kenji Fukumizu and Yu-Ping Wang (2018), Influence Function and Robust Variant of Kernel Canonical Correlation Analysis, Neurocomputing, Vol. 304 (2018) 12-29.

M. Romanazzi (1992), Influence in canonical correlation analysis, Psychometrika vol 57(2) (1992) 237-259.

See Also

See Also as hulfun, hadr, hudr

---

halofun   *Objective function*

Description

Objective function of Hampel's loss fucntion

Usage

```
halofun(x)
```

Arguments

x   vector values

Value

a real value

Author(s)

Md Ashad Alam <malam@tulane.edu>

References

Md Ashad Alam, Kenji Fukumizu and Yu-Ping Wang (2018), Influence Function and Robust Variant of Kernel Canonical Correlation Analysis, Neurocomputing, Vol. 304 (2018) 12-29.

M. Romanazzi (1992), Influence in canonical correlation analysis, Psychometrika vol 57(2) (1992) 237-259.

See Also

See also as `hulofun`

| hudr | *Huber's psi function* |

Description

The ratio of the first derivative of the Huber loss fuction to the argument. Tuning constants is fixed as a meadian vlue.

Usage

```
hudr(x)
```

Arguments

x   vector values

Value

y   a real value

Author(s)

Md Ashad Alam <malam@tulane.edu>

References

Md Ashad Alam, Kenji Fukumizu and Yu-Ping Wang (2018), Influence Function and Robust Variant of Kernel Canonical Correlation Analysis, Neurocomputing, Vol. 304 (2018) 12-29.

M. Romanazzi (1992), Influence in canonical correlation analysis, Psychometrika vol 57(2) (1992) 237-259.

See Also

See also as `hadr`

---

hulfun   *A Huber loss function*

Description

Tuning constants of the Huber loss fuction are fixed in different quintiles of the arguments.

Usage

```
hulfun(x)
```

Arguments

x   a vector values

Details

Tuning constants of the Huber fuction is fixed as a median.

Value

a real number

Author(s)

Md Ashad Alam <malam@tulane.edu>

References

Md Ashad Alam, Kenji Fukumizu and Yu-Ping Wang (2018), Influence Function and Robust Variant of Kernel Canonical Correlation Analysis, Neurocomputing, Vol. 304 (2018) 12-29.

M. Romanazzi (1992), Influence in canonical correlation analysis, Psychometrika vol 57(2) (1992) 237-259.

See Also

See also as halfun

---

hulofun   *Objective function*

Description

Objective function of Huber's loss fucntion

Usage

```
hulofun(x)
```

Arguments

x   vector values

Value

a real value

Author(s)

Md Ashad Alam <malam@tulane.edu>

References

Md Ashad Alam, Kenji Fukumizu and Yu-Ping Wang (2018), Influence Function and Robust Variant of Kernel Canonical Correlation Analysis, Neurocomputing, Vol. 304 (2018) 12-29.

M. Romanazzi (1992), Influence in canonical correlation analysis, Psychometrika vol 57(2) (1992) 237-259.

See Also

See Also as `halofun`, ~~~

---

ibskm   *Kernel Matrix Using Identity-by-state Kernel*

Description

For GWASs, a kernel captures the pairwise similarity across a number of SNPs in each gene. Kernel projects the genotype data from original high dimensional space to a feature space. One of the more popular kernels used for genomics similarity is the identity-by-state (IBS) kernel (non- parametric function of the genotypes)

Usage

```
ibskm(Z)
```

Arguments

Z   a data matrix

Details

For genome-wide association study, a kernel captures the pairwise similarity across a number of SNPs in each gene. Kernel projects the genotype data from original high dimensional space to a feature space. One popular kernel used for genomics similarity is the identity-by-state (IBS) kernel, The IBS kernel does not need any assumption on the type of genetic interactions.

Value

K   a Gram/ kernel matrix

Author(s)

Md Ashad Alam <malam@tulane.edu>

References

Md. Ashad Alam, Hui-Yi Lin, HOng-Wen Deng, Vince Calhour Yu-Ping Wang (2018), A kernel machine method for detecting higher order interactions in multimodal datasets: Application to schizophrenia, Journal of Neuroscience Methods, Vol. 309, 161-174.

Md Ashad Alam, Kenji Fukumizu and Yu-Ping Wang (2018), Influence Function and Robust Variant of Kernel Canonical Correlation Analysis, Neurocomputing, Vol. 304 (2018) 12-29.

M. Romanazzi (1992), Influence in canonical correlation analysis, Psychometrika vol 57(2) (1992) 237-259.

See Also

See also as gkm, lkm

Examples

```
##Dummy data:
X <- matrix(rnorm(200),50)
ibskm(X)
```

---

ifcca   *Influence Funciton of Canonical Correlation Analysis*

Description

##To define the robustness in statistics, different approaches have been pro- posed, for example, the minimax approach, the sensitivity curve, the influence function (IF) and the finite sample breakdown point. Due to its simplic- ity, the IF is the most useful approach in statistical machine learning

Usage

```
ifcca(X, Y, gamma = 1e-05, ncomps = 2, jth = 1)
```

Arguments

X	a data matrix index by row
Y	a data matrix index by row
gamma	the hyper-parameters
ncomps	the number of canonical vectors
jth	the influence function of the jth canonical vector

Value

iflccor	Influence value of the data by linear canonical correalation

Author(s)

Md Ashad Alam <malam@tulane.edu>

References

Md Ashad Alam, Kenji Fukumizu and Yu-Ping Wang (2018), Influence Function and Robust Variant of Kernel Canonical Correlation Analysis, Neurocomputing, Vol. 304 (2018) 12-29.

M. Romanazzi (1992), Influence in canonical correlation analysis, Psychometrika vol 57(2) (1992) 237-259.

See Also

See also as rkcca, ifrkcca

Examples

```
##Dummy data:

X <- matrix(rnorm(500),100); Y <- matrix(rnorm(500),100)

ifcca(X,Y,  1e-05,  2, 2)
```

---

ifmkcca   *Influence Function of Multiple Kernel Canonical Analysis*

Description

To define the robustness in statistics, different approaches have been pro- posed, for example, the minimax approach, the sensitivity curve, the influence function (IF) and the finite sample breakdown point. Due to its simplic- ity, the IF is the most useful approach in statistical machine learning.

Usage

```
ifmkcca(xx, yy, zz, kernel = "rbfdot", gamma = 1e-05, ncomps = 1, jth=1)
```

Arguments

xx	a data matrix index by row
yy	a data matrix index by row
zz	a data matrix index by row
kernel	a positive definite kernel
ncomps	the number of canonical vectors
gamma	the hyper-parameters.
jth	the influence function of the jth canonical vector

Value

iflccor	Influence value of the data by multiple kernel canonical correalation

Author(s)

Md Ashad Alam <malam@tulane.edu>

References

Md Ashad Alam, Kenji Fukumizu and Yu-Ping Wang (2018), Influence Function and Robust Variant of Kernel Canonical Correlation Analysis, Neurocomputing, Vol. 304 (2018) 12-29.

M. Romanazzi (1992), Influence in canonical correlation analysis, Psychometrika vol 57(2) (1992) 237-259.

See Also

See also as `ifcca`

Examples

```
##Dummy data:

X <- matrix(rnorm(500),100); Y <- matrix(rnorm(500),100); Z <- matrix(rnorm(500),100)

ifmkcca(X,Y, Z, "rbfdot",  1e-05,  2, 1)
```

---

ifrkcca   *Influence Function of Robust Kernel Canonical Analysis*

Description

##To define the robustness in statistics, different approaches have been pro- posed, for example, the minimax approach, the sensitivity curve, the influence function (IF) and the finite sample breakdown point. Due to its simplic- ity, the IF is the most useful approach in statistical machine learning.

Usage

```
ifrkcca(X, Y, lossfu = "Huber", kernel = "rbfdot", gamma = 0.00001, ncomps = 10, jth = 1)
```

Arguments

X	a data matrix index by row
Y	a data matrix index by row
lossfu	a loss function: square, Hampel's or Huber's loss
kernel	a positive definite kernel
gamma	the hyper-parameters
ncomps	the number of canonical vectors
jth	the influence function of the jth canonical vector

Value

ifrkcor	Influence value of the data by robust kernel canonical correalation
ifrkxcv	Influence value of cnonical vector of X dataset
ifrkycv	Influence value of cnonical vector of Y dataset

Author(s)

Md Ashad Alam <malam@tulane.edu>

References

Md Ashad Alam, Kenji Fukumizu and Yu-Ping Wang (2018), Influence Function and Robust Variant of Kernel Canonical Correlation Analysis, Neurocomputing, Vol. 304 (2018) 12-29.

M. Romanazzi (1992), Influence in canonical correlation analysis, Psychometrika vol 57(2) (1992) 237-259.

See Also

See also as [rkcca](), [ifrkcca]()

Examples

```
##Dummy data:

X <- matrix(rnorm(500),100); Y <- matrix(rnorm(500),100)

ifrkcca(X,Y, lossfu = "Huber", kernel = "rbfdot", gamma = 0.00001, ncomps = 10, jth = 2)
```

---

lcv   *A helper function*

Description

#A function ....

Usage

```
lcv(X, Y, res)
```

Arguments

X	a matrix
Y	a matrix
res	a real value

Author(s)

Md Ashad Alam <malam@tulane.edu>

---

lkm   *Kernel Matrix Using Linear Kernel*

Description

The linear kernel is used by the underlying Euclidean space to define the similarity measure. Whenever the dimensionality is high, it may allow for more complexity in the function class than what we could measure and assess otherwise

Usage

```
lkm(X)
```

Arguments

X   a data matrix

Details

The linear kernel is used by the underlying Euclidean space to define the similarity measure. Whenever the dimensionality of the data is high, it may allow for more complexity in the function class than what we could measure and assess otherwise.

Value

K   a kernel matrix.

Author(s)

Md Ashad Alam <malam@tulane.edu>

References

Md. Ashad Alam, Hui-Yi Lin, HOng-Wen Deng, Vince Calhour Yu-Ping Wang (2018), A kernel machine method for detecting higher order interactions in multimodal datasets: Application to schizophrenia, Journal of Neuroscience Methods, Vol. 309, 161-174.

Md Ashad Alam, Kenji Fukumizu and Yu-Ping Wang (2018), Influence Function and Robust Variant of Kernel Canonical Correlation Analysis, Neurocomputing, Vol. 304 (2018) 12-29.

Md Ashad Alam, Vince D. Calhoun and Yu-Ping Wang (2018), Identifying outliers using multiple kernel canonical correlation analysis with application to imaging genetics, Computational Statistics and Data Analysis, Vol. 125, 70- 85

See Also

See also as gkm, ibskm

Examples

```
##Dummy data:

X <- matrix(rnorm(500),100)
lkm(X)
```

---

mdbw *Bandwidth of the Gaussian kernel*

Description

A median of the pairwise distance of the data

Usage

```
mdbw(X)
```

Arguments

X   a data matrix

Details

While the Gaussian kernel has a free parameter (bandwidth), it still follows a number of theoretical properties such as boundedness, consistenc, universality, robustness, etc. It is the most applicable one. In a Gaussian RBF kernel, we need to select an appropriate a bandwidth. It is well known that the parameter has a strong influence on the result of kernel methods. For the Gaussian kernel, we can use the median of the pairwise distance as a bandwidth.

Value

s   a median of the pairwise distance of the X dataset

Author(s)

Md Ashad Alam <malam@tulane.edu>

References

Md. Ashad Alam, Hui-Yi Lin, HOng-Wen Deng, Vince Calhour Yu-Ping Wang (2018), A kernel machine method for detecting higher order interactions in multimodal datasets: Application to schizophrenia, Journal of Neuroscience Methods, Vol. 309, 161-174.

Md. Ashad Alam, Kenji Fukumizu and Yu-Ping Wang (2018), Influence Function and Robust Variant of Kernel Canonical Correlation Analysis, Neurocomputing, Vol. 304 (2018) 12-29.

Md. Ashad Alam and Kenji Fukumizu (2015), Higher-order regularized kernel canonical correlation analysis, International Journal of Pattern Recognition and Artificial Intelligence, Vol. 29(4) 1551005.

Arthu Gretton, Kenji. Fukumizu, C. H. Teo, L. Song, B. Scholkopf and A. Smola (2008), A Kernel statistical test of independence, in Advances in Neural Information Processing Systems, Vol. 20 585–592.

See Also

See also as lkm, gkm

Examples

```
##Dummy data:

X <- matrix(rnorm(1000),100)

mdbw(X)
```

---

medc	*A helper function*

Description

A function

Usage

```
medc(A, fn = sqrt)
```

Arguments

A	a matrix
fn	a funciton

Author(s)

Md Ashad Alam <malam@tulane.edu>

---

mvnod	*A helper function*

Description

A function

Usage

```
mvnod(n = 1, mu, Sigma, tol = 1e-06, empirical = FALSE, EISPACK = FALSE)
```

Arguments

n	an integer number
mu	a real value
Sigma	a real value
tol	a curection factor
empirical	a logical value
EISPACK	a logical value. TRUE for a complex values.

Author(s)

Md Ashad Alam <malam@tulane.edu>

| ranuf | *A helper function* |

Description

A function

Usage

```
ranuf(p)
```

Arguments

p   a real value

Author(s)

Md Ashad Alam <malam@tulane.edu>

| rkcca | *Robust kernel canonical correlation analysis* |

Description

#A robust correlation

Usage

```
rkcca(X, Y, lossfu = "Huber", kernel = "rbfdot", gamma = 1e-05, ncomps = 10)
```

Arguments

X	a data matrix index by row
Y	a data matrix index by row
lossfu	a loss function: square, Hampel's or Huber's loss
kernel	a positive definite kernel
gamma	the hyper-parameters
ncomps	the number of canonical vectors

Value

An S3 object containing the following slots:

rkcor   Robsut kernel canonical correlation

rxcoef   Robsut kernel canonical coficient of X dataset

rycoef   Robsut kernel canonical coficient of Y dataset

rxcv   Robsut kernel canonical vector of X dataset

rycv   Robsut kernel canonical vector of Y dataset

Author(s)

Md Ashad Alam <malam@tulane.edu>

References

Md Ashad Alam, Kenji Fukumizu and Yu-Ping Wang (2018), Influence Function and Robust Variant of Kernel Canonical Correlation Analysis, Neurocomputing, Vol. 304 (2018) 12-29.

M. Romanazzi (1992), Influence in canonical correlation analysis, Psychometrika vol 57(2) (1992) 237-259.

See Also

See also as `ifcca`, `rkcca`, `ifrkcca`

Examples

```
##Dummy data:

X <- matrix(rnorm(1000),100); Y <- matrix(rnorm(1000),100)

rkcca(X,Y, "Huber",  "rbfdot", 1e-05,  10)
```

---

rkcco   *Robust kernel cross-covariance opetator*

Description

A function

Usage

```
rkcco(X, Y, lossfu = "Huber", kernel = "rbfdot", gamma = 1e-05)
```

Arguments

X	a data matrix index by row
Y	a data matrix index by row
lossfu	a loss function: square, Hampel's or Huber's loss
kernel	a positive definite kernel
gamma	the hyper-parameters

Value

rkcmx	Robust kernel center matrix of X dataset
rkcmy	Robust kernel center matrix of Y dataset
rkcmx	Robust kernel covariacne operator of X dataset
rkcmy	Robust kernel covariacne operator of Y dataset
rkcmx	Robust kernel cross-covariacne operator of X and Y datasets

Author(s)

Md Ashad Alam <malam@tulane.edu>

References

Md Ashad Alam, Kenji Fukumizu and Yu-Ping Wang (2018), Influence Function and Robust Variant of Kernel Canonical Correlation Analysis, Neurocomputing, Vol. 304 (2018) 12-29.

M. Romanazzi (1992), Influence in canonical correlation analysis, Psychometrika vol 57(2) (1992) 237-259.

See Also

See also as rkcca snpfmridata, ifrkcca

Examples

```
##Dummy data:

X <- matrix(rnorm(2000),200); Y <- matrix(rnorm(2000),200)

rkcco(X,Y, "Huber","rbfdot", 1e-05)
```

---

rkcm   *Robsut Kernel Center Matrix*

Description

A functioin

Usage

```
rkcm(X, lossfu = "Huber", kernel = "rbfdot")
```

Arguments

X	a data matrix index by row
lossfu	a loss function: square, Hampel's or Huber's loss
kernel	a positive definite kernel

Value

rkcm	a square robust kernel center matrix

Author(s)

Md Ashad Alam <malam@tulane.edu>

References

Md Ashad Alam, Kenji Fukumizu and Yu-Ping Wang (2018), Influence Function and Robust Variant of Kernel Canonical Correlation Analysis, Neurocomputing, Vol. 304 (2018) 12-29.

Md Ashad Alam, Vince D. Calhoun and Yu-Ping Wang (2018), Identifying outliers using multiple kernel canonical correlation analysis with application to imaging genetics, Computational Statistics and Data Analysis, Vol. 125, 70- 85

See Also

See also as ifcca, rkcca, ifrkcca

Examples

```
##Dummy data:

X <- matrix(rnorm(2000),200); Y <- matrix(rnorm(2000),200)

rkcm(X, "Huber","rbfdot")
```

---

rlogit   *A helper fuction*

Description

#A function to calcualte generalized logit function.

Usage

```
rlogit(x)
```

Arguments

x   a real value to be tranformed

Author(s)

Md Ashad Alam <malam@tulane.edu>

---

snpfmridata   *An example of imaging genetics data to calcualte influential observa-*
                         tions from two view data

Description

#A function

Usage

```
snpfmridata(n = 300, gamma=0.00001, ncomps = 2, jth = 1)
```

Arguments

n   the sample size

gamma   the hyper-parameters

ncomps   the number of canonical vectors

jth   the influence function of the jth canonical vector

Value

IFCCAID	Influence value of canonical correlation analysis for the ideal data
IFCCACD	Influence value of canonical correlation analysis for the contaminated data
IFKCCAID	Influence value of kernel canonical correlation analysis for the ideal data
IFKCCACD	Influence value of kernel canonical correlation analysis for the contaminated data
IFHACCAID	Influence value of robsut (Hampel's loss) canonical correlation analysis for the ideal data
IFHACCACD	Influence value of robsut (Hampel's loss) canonical correlation analysis for the contaminated data
IFHUCCAID	Influence value of robsut (Huber's loss) canonical correlation analysis for the ideal data
IFHUCCACD	Influence value of robsut (Huber's loss) canonical correlation analysis for the contaminated data

Author(s)

Md Ashad Alam <malam@tulane.edu>

References

Md Ashad Alam, Kenji Fukumizu and Yu-Ping Wang (2018), Influence Function and Robust Variant of Kernel Canonical Correlation Analysis, Neurocomputing, Vol. 304 (2018) 12-29.

Md Ashad Alam, Vince D. Calhoun and Yu-Ping Wang (2018), Identifying outliers using multiple kernel canonical correlation analysis with application to imaging genetics, Computational Statistics and Data Analysis, Vol. 125, 70- 85

See Also

See also as rkcca, ifrkcca, snpfmrimth3D

Examples

```
##Dummy data:

n<-100

snpfmridata(n, 0.00001,  10, jth = 1)
```

---

snpfmrimth3D   *An example of imaging genetics and epi-genetics data to calcualte in-*
                        fluential observations from three view data

---

Description

#A function

Usage

```
snpfmrimth3D(n = 500, gamma = 1e-05, ncomps = 1, jth=1)
```

Arguments

n	the sample size
gamma	the hyper-parameters
ncomps	the number of canonical vectors
jth	the influence function of the jth canonical vector

Value

IFim	Influence value of multiple kernel canonical correlation analysis for the ideal data
IFcm	Influence value of multiple kernel canonical correlation analysis for the contaminated data

Author(s)

Md Ashad Alam <malam@tulane.edu>

References

Md Ashad Alam, Kenji Fukumizu and Yu-Ping Wang (2018), Influence Function and Robust Vari-
ant of Kernel Canonical Correlation Analysis, Neurocomputing, Vol. 304 (2018) 12-29.

Md Ashad Alam, Vince D. Calhoun and Yu-Ping Wang (2018), Identifying outliers using multiple
kernel canonical correlation analysis with application to imaging genetics, Computational Statistics
and Data Analysis, Vol. 125, 70- 85

See Also

See also as rkcca, snpfmridata, ifrkcca

Examples

```
##Dummy data:

n<-100

snpfmrimth3D(n, 0.00001,  10, 1)
```

---

udtd   *A helper function*

Description

A function to a measure of a system's real point computing power

Usage

```
 udtd(x)
```

Arguments

x   a real value

Author(s)

Md Ashad Alam <malam@tulane.edu>

Index