

# Package ‘cdcsis’

February 19, 2015

**Type** Package

**Title** Conditional Distance Correlation and Its Related Feature  
Screening Method

**Version** 1.0

**Date** 2014-09-01

**Author** Canhong Wen, Wenliang Pan, Mian Huang, and Xueqin Wang

**Depends** R(>= 3.0.1), stats

**Imports** ks

**Suggests** MASS, energy

**Maintainer** Canhong Wen <wencanhong@gmail.com>

**Description** Gives conditional distance correlation and performs the conditional distance correlation sure independence screening procedure for ultrahigh dimensional data. The conditional distance correlation is a novel conditional dependence measurement of two random variables given a third variable. The conditional distance correlation sure independence screening is used for screening variables in ultrahigh dimensional setting.

**License** GPL (>= 2)

**Repository** CRAN

**Collate** 'cdcov.R' 'cdcor.R' 'cdcor.ada.R' 'bw.R' 'cdcsis.R'

**NeedsCompilation** yes

**Date/Publication** 2014-10-21 17:48:22

## R topics documented:

cdcsis-package	2
bw	3
cdcor	4
cdcor.ada	5
cdcov	7
cdcsis	8

<b>Index</b>	<b>10</b>
--------------	-----------

---

cdcsis-package

*Conditional Distance Correlation and Its Related Feature Screening Method*

---

## Description

Gives the conditional distance correlation and performs its based sure independence screening method, i.e., CDCSIS of Wen et al.(2014).

## Details

Package: CDCSIS  
Type: Package  
Version: 1.0  
Date: 2014-09-01  
License: GPL(>=2)

The package `cdcsis` is used for calculate the conditional distance correlation and performs the related sure independent screening method. Details of the method can be found in Wen et.al. (2014).

## Author(s)

Canhong Wen, Wenliang Pan, Mian Huang, and Xueqin Wang

Maintainer: Canhong Wen <wencanhong@gmail.com>

## References

Canhong Wen, Wenliang Pan, Mian Huang and Xueqin Wang(2014). Conditional distance correlation sure independence screening for ultrahigh dimensional data. *Submitted to Biostatistics*.

## Examples

```
set.seed(0)
n <- 100 # sample size
p <- 10 # dimensionality
rho <- 0.5 # the correlation between pairwise predictors.
Sigma <- matrix(rho, p, p)
diag(Sigma) <- 1

require(MASS)
x <- mvrnorm(n,rep(0,p), Sigma)

y <- x[,2] + x[,3] + rnorm(n)
z <- x[,1]
cdcsis(x,y,z,2)
```

**Description**

bw is used to select the bandwidth in the conditional distance correlation estimation.

**Usage**

```
bw(x, y, z, index = 1)
```

**Arguments**

x	a numeric vector or matrix
y	a numeric vector or matrix with compatible dimensions to x
z	the variable being conditioned. z is a numeric vector or matrix with compatible dimensions to x
index	exponent on Euclidean distance, in (0,2]

**Details**

For univariate z, the univariate plug-in selector of Wand & Jones (1994) is used. That is, hpi function in the ks package is used.

For multivariate z, an diagonal matrix of the bandwidth is assumed, i.e., select the optimal bandwidth with hpi individually for each column of z.

**Value**

The plug-in bandwidth.

**Author(s)**

Canhong Wen, Wenliang Pan, Mian Huang, and Xueqin Wang.

**References**

Canhong Wen, Wenliang Pan, Mian Huang and Xueqin Wang(2014). Conditional distance correlation sure independence screening for ultrahigh dimensional data. *Submitted to Biostatistics*.

Wand, M.P. & Jones, M.C. (1995) Kernel Smoothing. Chapman & Hall.

**See Also**

[cdc](#), [cdc](#).ada

**Examples**

```

set.seed(1)
# load the distance correlation for comparison
require(energy)

## independent case
x <- rnorm(100)
y <- rnorm(100)
z <- rnorm(100)
bw(x,y,z)
cdcor.ada(x,y,z)
dcor(x,y)

## conditional dependent case
x <- rnorm(100)
y <- x + 0.1*rnorm(100)
z <- rnorm(100)
bw(x,y,z)
cdcor.ada(x,y,z)
dcor(x,y)

## conditional independent case: x and z are correlated
require(MASS)
data <- mvrnorm(100,rep(0,2),matrix(c(1,0.8,0.8,1),2,2))
x <- data[,1]
z <- data[,2]
y <- z + 0.1*rnorm(100)
bw(x,y,z)
cdcor.ada(x,y,z)
dcor(x,y)

```

cdcor

*Conditional Distance Correlation with Given Bandwidth***Description**

Performs conditional distance correlation with given bandwidth.

**Usage**

```
cdcor(x, y, z, width, index = 1)
```

**Arguments**

x	a numeric vector or matrix
y	a numeric vector or matrix with compatible dimensions to x
z	the variable being conditioned. z is a numeric vector or matrix with compatible dimensions to x
width	a positive value, user-specified
index	exponent on Euclidean distance, in (0,2]

**Details**

It performs conditional distance correlation with given bandwidth.

**Value**

cdcov	the conditional distance covariance with given variable z
mcdcov	mean of the conditional distance covariance with given variable z
width	the bandwidth
index	exponent on Euclidean distance, in (0,2]

**Author(s)**

Canhong Wen, Wenliang Pan, Mian Huang, and Xueqin Wang

**References**

Canhong Wen, Wenliang Pan, Mian Huang and Xueqin Wang(2014). Conditional distance correlation sure independence screening for ultrahigh dimensional data. *Submitted to Biostatistics*.

**See Also**

[cdcor.ada](#), [cdcov](#)

**Examples**

```
x <- rnorm(100)
y <- rnorm(100)
z <- rnorm(100)
cdcor(x,y,z,0.25)
```

---

cdcor.ada

*Conditional Distance Correlation with Adaptive Bandwidth*

---

**Description**

Performs conditional distance correlation with adaptive bandwidth.

**Usage**

```
cdcor.ada(x, y, z, tol = 0.1, index = 1)
```

**Arguments**

x	a numeric vector or matrix
y	a numeric vector or matrix with compatible dimensions to x
z	the variable being conditioned. z is a numeric vector or matrix with compatible dimensions to x
tol	the tolerance used in the bandwidth selection
index	exponent on Euclidean distance, in (0,2]

**Details**

It performs conditional distance correlation with adaptive bandwidth. The bandwidth is determined by the bw function.

**Value**

Returns an object with

cdc <code>cor</code>	conditional distance correlation
width	bandwidth used in cor. It is determined by the bw function.

**Author(s)**

Canhong Wen, Wenliang Pan, Mian Huang, and Xueqin Wang

**References**

Canhong Wen, Wenliang Pan, Mian Huang and Xueqin Wang(2014). Conditional distance correlation sure independence screening for ultrahigh dimensional data. *Submitted to Biostatistics*.

**See Also**

[cdc`cor`](#)

**Examples**

```
set.seed(1)
# load the distance correlation for comparison
require(energy)

## independent case
x <- rnorm(100)
y <- rnorm(100)
z <- rnorm(100)
cdccor.ada(x,y,z)
dcor(x,y)

## conditional dependent case
x <- rnorm(100)
y <- x + 0.1*rnorm(100)
```

```

z <- rnorm(100)
cdcov.ada(x,y,z)
dcov(x,y)

## conditional independent case: x and z are correlated
require(MASS)
data <- mvrnorm(100,rep(0,2),matrix(c(1,0.8,0.8,1),2,2))
x <- data[,1]
z <- data[,2]
y <- z + 0.1*rnorm(100)
cdcov.ada(x,y,z)
dcov(x,y)

```

---

cdcov

*Conditional Distance Covariance with Given Bandwidth*


---

### Description

Performs conditional distance covariance with given bandwidth.

### Usage

```
cdcov(x, y, z, width, index = 1)
```

### Arguments

x	a numeric vector or matrix
y	a numeric vector or matrix with compatible dimensions to x
z	the variable being conditioned. z is a numeric vector or matrix with compatible dimensions to x
width	a positive value, user-specified
index	exponent on Euclidean distance, in (0,2]

### Details

Perform the measurement of the conditional independence between x and y given z with given bandwidth.

### Value

cdcov	conditional distance covariance with given variable z; has the same length with z
width	the bandwidth
index	exponent on Euclidean distance, in (0,2]

**Author(s)**

Canhong Wen, Wenliang Pan, and Xueqin Wang

**References**

Canhong Wen, Wenliang Pan, Mian Huang and Xueqin Wang(2014). Conditional distance correlation sure independence screening for ultrahigh dimensional data. *Submitted to Biostatistics*.

**See Also**

[cdcor](#)

**Examples**

```
x <- rnorm(100)
y <- rnorm(100)
z <- rnorm(100)
cdcov(x,y,z,0.25)
```

---

cdcsis

*Conditional Distance Correlation Sure Independence Screening (CD-  
CSIS)*

---

**Description**

Performs conditional distance correlation sure independence screening (CDCSIS).

**Usage**

```
cdcsis(x, y, z, thres)
```

**Arguments**

x	a matrix
y	a numeric vector or matrix with compatible dimensions to x
z	the variable being conditioned. z is a numeric vector or matrix with compatible dimensions to x
thres	the threshold of the number of predictors recruited by CDCSIS. Should be less than or equal than the number of column of x.

**Details**

It performs conditional distance correlation sure independence screening (CDCSIS).



**Value**

CDCSISind	the vector of indices selected by CDCSIS
thres	the threshold of the number of predictors recruited by CDCSIS
DC	the distance correlation for each dimensionality of $x$
DCord	the order of DC for each dimensionality of $x$

**Author(s)**

Canhong Wen, Wenliang Pan, and Xueqin Wang

**References**

Canhong Wen, Wenliang Pan, Mian Huang and Xueqin Wang(2014). Conditional distance correlation sure independence screening for ultrahigh dimensional data. *Submitted to Biostatistics*.

**See Also**

[cdcor.ada](#), [cdcor](#)

**Examples**

```
set.seed(0)
n <- 100
p <- 10
rho <- 0.5
Sigma <- matrix(rho, p, p)
diag(Sigma) <- 1

require(MASS)
x <- mvrnorm(n,rep(0,p), Sigma)
y <- x[,2] + x[,3] + rnorm(n)
z <- x[,1]

cdcsis(x,y,z,2)
```

# Index

- \*Topic **SIS, distance correlation, confounder**
  - [cdcsis-package, 2](#)
- \*Topic **Ultrahigh dimensional**
  - [cdcsis, 8](#)
- \*Topic **bandwidth selection**
  - [bw, 3](#)
- \*Topic **conditional distance correlation**
  - [cdcor, 4](#)
  - [cdcor.ada, 5](#)
  - [cdcsis, 8](#)
- \*Topic **conditional distance covariance**
  - [cdcov, 7](#)
- \*Topic **plug-in**
  - [bw, 3](#)
- \*Topic **sure independence screening**
  - [cdcsis, 8](#)

[bw, 3](#)

[cdcor, 3, 4, 6, 8, 9](#)

[cdcor.ada, 3, 5, 5, 9](#)

[cdcov, 5, 7](#)

[cdcsis, 8](#)

[cdcsis-package, 2](#)