Package ‘MRAM’

March 4, 2026

Type Package

Title Multivariate Regression Association Measure

Version 1.0.1

Description Implementations of an estimator for the multivariate regression association measure (MRAM) proposed in Shih and Chen (2026) <doi:10.1016/j.csda.2025.108288> and its associated variable selection algorithm. The MRAM quantifies the predictability of a random vector Y from a random vector X given a random vector Z. It takes the maximum value 1 if and only if Y is almost surely a measurable function of X and Z, and the minimum value of 0 if Y is conditionally independent of X given Z. The MRAM generalizes the Kendall's tau copula correlation ratio proposed in Shih and Emura (2021) <doi:10.1016/j.jmva.2020.104708> by employing the spatial sign function. The estimator is based on the nearest neighbor method, and the associated variable selection algorithm is adapted from the feature ordering by conditional independence (FOCI) algorithm of Azadkia and Chatterjee (2021) <doi:10.1214/21-AOS2073>. For further details, see the paper Shih and Chen (2026) <doi:10.1016/j.csda.2025.108288>.

Depends RANN

License GPL-2

Encoding UTF-8

RoxygenNote 7.3.3

NeedsCompilation no

Author Jia-Han Shih [aut, cre],
Yi-Hau Chen [aut]

Maintainer Jia-Han Shih <jhshih@math.nsysu.edu.tw>

Repository CRAN

Date/Publication 2026-03-04 07:40:02 UTC

Contents

mram *Estimate the Multivariate Regression Association Measure*

Description

Compute T_n and its standard error estimates using the nearest neighbor method and the m-out-of-n bootstrap.

Usage

```
mram(
  y_data,
  x_data,
  z_data = NULL,
  bootstrap = FALSE,
  B = 1000,
  g_vec = seq(0.4, 0.9, by = 0.05)
)
```

Arguments

`y_data`	A $n \times d$ matrix of responses, where n is the sample size.
`x_data`	A $n \times p$ matrix of predictors.
`z_data`	A $n \times q$ matrix of conditional predictors. The default value is `NULL`.
`bootstrap`	Perform the m-out-of-n bootstrap if `TRUE`. The default value is `FALSE`.
`B`	Number of bootstrap replications. The default value is `1000`.
`g_vec`	A vector of candidate values for γ between 0 and 1, used to generate a collection of rules for the m-out-of-n bootstrap. The default value is `seq(0.4,0.9,by = 0.05)`.

Details

Let $\{(\mathbf{X}_i, \mathbf{Y}_i, \mathbf{Z}_i)\}_{i=1}^n$ be independent and identically distributed data from the population $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$. The estimate $T_n(\mathbf{X}, \mathbf{Y})$ for the unconditional measure (`z_data = NULL`) is given as

$$T_n(\mathbf{X}, \mathbf{Y}) = \binom{n}{2}^{-1} \sum_{i<j} \langle S(\mathbf{Y}_i - \mathbf{Y}_j), S(\mathbf{Y}_{N(i)} - \mathbf{Y}_{N(j)}) \rangle,$$

where $\langle \cdot, \cdot \rangle$ is the dot product, $S(\cdot)$ is the spatial sign function, and $N(i)$ is the index j such that $\mathbf{X}_j$ is the nearest neighbor of $\mathbf{X}_i$ according to the Euclidean distance. The estimate $T_n(\mathbf{X}, \mathbf{Y} \mid \mathbf{Z})$ for the conditional measure is given as

$$T_n(\mathbf{X}, \mathbf{Y} \mid \mathbf{Z}) = \frac{T_n((\mathbf{X}, \mathbf{Z}), \mathbf{Y}) - T_n(\mathbf{Z}, \mathbf{Y})}{1 - T_n(\mathbf{Z}, \mathbf{Y})}.$$

See the paper Shih and Chen (2026) for more details.

For the m-out-of-n bootstrap, the rule (resample size) is set to be $m = \lfloor n^\gamma \rfloor$, where $\lfloor x \rfloor$ denotes the largest integer that is smaller than or equal to x and $0 < \gamma < 1$ takes values from the vector `g_vec`. It is recommended to use `T_se_cluster`, the standard error estimate obtained based on the cluster rule. See Dette and Kroll (2024) for more details.

The `mram` function is used in `vs_mram` function for variable selection.

Value

`T_est`	The estimate of the multivariate regression association measure. The value returned by `T_est` is between -1 and 1. However, it is between 0 and 1 asymptotically. A small value indicates that `x_data` has low predictability for `y_data` condition on `z_data` in the sense of the considered measure. On the other hand, a large value indicates that `x_data` has high predictability for `y_data` condition on `z_data`. If `z_data = NULL`, the returned value indicates the unconditional predictability.
`T_se_cluster`	The standard error estimate based on the cluster rule.
`m_vec`	The vector of m generated by `g_vec`.
`T_se_vec`	The vector of standard error estimates obtained from the m-out-of-n bootstrap, where m is equal to `m_vec`.
`J_cluster`	The index of the best `m_vec` chosen by the cluster rule.

References

Dette and Kroll (2024) A Simple Bootstrap for Chatterjee's Rank Correlation, Biometrika, asae045.

Shih and Chen (2026) Measuring multivariate regression association via spatial sign, Computational Statistics & Data Analysis, 215, 108288.

See Also

`vs_mram`

Examples

```
library(MRAM)

n = 100

set.seed(1)
x_data = matrix(rnorm(n*2),n,2)
y_data = matrix(0,n,2)
y_data[,1] = x_data[,1]*x_data[,2]+x_data[,1]+rnorm(n)
y_data[,2] = x_data[,1]*x_data[,2]-x_data[,1]+rnorm(n)

mram(y_data,x_data[,1],x_data[,2])
mram(y_data,x_data[,2],x_data[,1])
mram(y_data,x_data[,1])
mram(y_data,x_data[,2])
```

```
## Not run:

# perform the m-out-of-n bootstrap
mram(y_data,x_data[,1],x_data[,2],bootstrap = TRUE)
mram(y_data,x_data[,2],x_data[,1],bootstrap = TRUE)
mram(y_data,x_data[,1],bootstrap = TRUE)
mram(y_data,x_data[,2],bootstrap = TRUE)

## End(Not run)
```

`vs_mram` *Variable Selection via the Multivariate Regression Association Measure*

Description

Select a subset of $\mathbf{X}$ which can be used to predict $\mathbf{Y}$ based on T_n.

Usage

```
vs_mram(y_data, x_data)
```

Arguments

`y_data`	A $n \times d$ matrix of responses, where n is the sample size.
`x_data`	A $n \times p$ matrix of predictors.

Details

`vs_mram` performs forward stepwise variable selection based on the multivariate regression association measure proposed in Shih and Chen (2026). At each step, it selects the predictor with the highest conditional predictability for the response given the previously selected predictors. The algorithm is modified from the FOCI algorithm from Azadkia and Chatterjee (2021).

Value

The vector containing the indices of the selected predictors in the order they were chosen.

References

Azadkia and Chatterjee (2021) A simple measure of conditional dependence, Annals of Statistics, 46(6): 3070-3102.

Shih and Chen (2026) Measuring multivariate regression association via spatial sign, Computational Statistics & Data Analysis, 215, 108288.

See Also

`mram`

Examples

```
library(MRAM)

n = 200
p = 10

set.seed(1)
x_data = matrix(rnorm(p*n),n,p)
colnames(x_data) = paste0(rep("X",p),seq(1,p))

y_data = x_data[,1]*x_data[,2]+x_data[,1]-x_data[,3]+rnorm(n)
colnames(x_data)[vs_mram(y_data,x_data)] # selected variables

## Not run:

n = 500
p = 10

set.seed(1)
x_data = matrix(rnorm(p*n),n,p)
colnames(x_data) = paste0(rep("X",p),seq(1,p))

# Linear
y_data = matrix(0,n,2)
y_data[,1] = x_data[,1]*x_data[,2]+x_data[,1]-x_data[,3]+rnorm(n)
y_data[,2] = x_data[,2]*x_data[,4]+x_data[,2]-x_data[,5]+rnorm(n)
colnames(x_data)[vs_mram(y_data,x_data)] # selected variables

# Nonlinear
y_data = matrix(0,n,2)
y_data[,1] = x_data[,1]*x_data[,2]+sin(x_data[,1]*x_data[,3])+0.3*rnorm(n)
y_data[,2] = cos(x_data[,2]*x_data[,4])+x_data[,3]-x_data[,4]+0.3*rnorm(n)
colnames(x_data)[vs_mram(y_data,x_data)] # selected variables

# Non-additive error
y_data = matrix(0,n,2)
y_data[,1] = abs(x_data[,1]+runif(n))^(sin(x_data[,2])-cos(x_data[,3]))
y_data[,2] = abs(x_data[,2]-runif(n))^(sin(x_data[,3])-cos(x_data[,4]))
colnames(x_data)[vs_mram(y_data,x_data)] # selected variables

## End(Not run)
```

Index