Package ‘mbrdr’

October 13, 2022

Type Package

Title Model-Based Response Dimension Reduction

Version 1.1.1

Date 2022-01-24

Author Jae Keun Yoo

Maintainer Jae Keun Yoo <peter.yoo@ewha.ac.kr>

Description Functions for model-based response dimension reduction. Usual dimension reduction methods in multivariate regression focus on the reduction of predictors, not responses. The response dimension reduction is theoretically founded in Yoo and Cook (2008) <doi:10.1016/j.csda.2008.07.029>. Later, three model-based response dimension reduction approaches are proposed in Yoo (2016) <doi:10.1080/02331888.2017.1410152> and Yoo (2019) <doi:10.1016/j.jkss.2019.02.001>. The metho
parametric ordinary least squares, but the model-based approaches are done through maximum likelihood estimation. For two model-based response dimension reduction methods called principal fitted response reduction and unstructured principal fitted response reduction, chi-squared tests are provided for determining the dimension of the response subspace.

License GPL (>= 2.0)

Depends R(>= 3.5.0)

Imports stats

Repository CRAN

LazyData yes

Encoding UTF-8

RoxygenNote 6.0.1

NeedsCompilation no

Date/Publication 2022-01-24 11:22:43 UTC

R topics documented:

choose.fx *choose fx for principal fitted response reduction and unstructured principal fitted response reduction*

Description

Returns a $n \times q$ matrix used in principal fitted response reduction and unstructured principal fitted response reduction.

Usage

```
choose.fx(X, fx.choice=1, nclust = 5)
```

Arguments

`X`	$n \times p$ predictor matrix
`fx.choice`	four choices for fx; see below
`nclust`	the number of clusters; see below

Details

Both of principal fitted response reduction and unstructured principal fitted response reduction require a choice of fx. The function will return one of four choices of fx, which are popular candidates among many.

`fx.choice=1`: This is default and returns the original predictor matrice X, centered at zero as fx.

`fx.choice=2`: This returns the original predictor matrice X, centered at zero and its squared values.

`fx.choice=3`: This returns the original predictor matrice X, centered at zero and its exponentiated values.

`fx.choice=4`: This clusters X with K-means algoritm with the number of clusters equal to the value in `nclust`. Then, the cluster results are expanded to `nclust` $-$ 1 dummy variables, like factor used in `lm` function. Finally, it returns `nclust-1` categorical basis. The option of `nclust` works only with `fx.choice=4`.

Value

A $n \times q$ matrix for fx.

Author(s)

Jae Keun Yoo, peter.yoo@ewha.ac.kr

Examples

```
data(mps)
X <- mps[,c(5:6,8:14)]
choose.fx(X)

choose.fx(X, fx.choice=2)

choose.fx(X, fx.choice=4, nclust=3)
```

---

matpower — *compute the M^power where M is a symmetric matrix.*

Description

Returns M^power.

Usage

```
matpower(M, pow)
```

Arguments

`M`	symmetric matrix
`pow`	power

Details

The function computes M^power for a symmetric matrix M.

Value

Returns

Author(s)

Jae Keun Yoo, peter.yoo@ewha.ac.kr

Examples

```
X <- matrix(rnorm(100), c(20,5))
matpower(cov(X), -0.5) ## returns cov(X)^-0.5 %*% cov(X)^-0.5 = cov(X)^-1.
```

mbrdr *Main function for model-based response dimension reduction regression*

Description

This is the main function in the mbrdr package. It creates objects of class mbrdr to estimate the response mean subspace and perform tests concerning its dimension. Several helper functions that require a mbrdr object can then be applied to the output from this function.

Usage

```
mbrdr (formula, data, subset, na.action = na.fail, weights, ...)

mbrdr.compute (y, x, weights, method = "upfrr", ...)
```

Arguments

formula	a two-sided formula like `cbind(y1,y2,y3,y4)~x1+x2+x3`, where the left-side variables are a matrix of the response variables, and the right-hand side variables represent the predictors. The left-hand side of the formula must be a matrix, since the package reduces the dimension of the responses variables.
data	an optional data frame containing the variables in the model. By default the variables are taken from the environment from which 'mbrdr' is called.
subset	an optional vector specifying a subset of observations to be used in the fitting process.
weights	an optional vector of weights to be used where appropriate. In the context of dimension reduction methods, weights are used to obtain elliptical symmetry, not constant variance.
na.action	a function which indicates what should happen when the data contain 'NA's. The default is 'na.fail,' which will stop calculations. The option 'na.omit' is also permitted, but it may not work correctly when weights are used.
x	The design matrix. This will be computed from the formula by `dr` and then passed to `dr.compute`, or you can create it yourself.
y	The response vector or matrix
method	This character string specifies the method of fitting. The default is `"upfrr"`. The options include `"yc"`, `"prr"`, `"pfrr"`. Each method may have its own additional arguments, or its own defaults; see the details below for more information.
...	For `mbrdr`, all additional arguments passed to `mbrdr.compute`. For `mbrdr.compute`, additional arguments may be required for particular dimension reduction method. For example, `numdir` is the maximum number of directions to compute, with default equal to 4. Other methods may have other defaults.

Details

The general regression problem mainly focuses on studying $E(y|x)$, the conditional mean of a response y given a set of predictors x, where y is r-dimensional response variables with $rgeq2$ and

This function provides methods for estimating the response dimension subspace of a general regression problem. That is, we want to find a $r \times d$ matrix B of minimal rank d such that

$$E(y|x) = E(P(B)y|x)$$

, where P(B) is an orthogonal projections onto the column space of B. Both the dimension d and the subspace $P(B)$ are unknown. These methods make few assumptions.

For the methods "yc", "prr", "pfrr" and "upfrr", B is estimated and returned. And, only for "pfrr" and "upfrr", chi-squared test results for estimating d is provided.

Weights can be used, essentially to specify the relative frequency of each case in the data.

The option `fx.choice` is required to fit "pfrr" and "upfrr" and has the following four values.

`fx.choice=1`: This is default and returns the original predictor matrice X, centered at zero as fx.

`fx.choice=2`: This returns the original predictor matrice X, centered at zero and its squared values.

`fx.choice=3`: This returns the original predictor matrice X, centered at zero and its exponentiated values.

`fx.choice=4`: This clusters X with K-means algoritm with the number of clusters equal to the value in `nclust`. Then, the cluster results are expanded to `nclust` $- 1$ dummy variables, like factor used in `lm` function. Finally, it returns `nclust-1` categorical basis. The option of `nclust` works only with `fx.choice=4`.

Value

mbrdr returns an object that inherits from mbrdr (the name of the type is the value of the `method` argument), with attributes:

`y`	The response matrix
`x`	The design matrix
`weights`	The weights used, normalized to add to n.
`cases`	Number of cases used.
`call`	The initial call to `mbrdr`.
`evectors`	The eigenvectors from kernel matrices to estimate B computed from each response dimension reduction methods. It is the estimate of B.
`evalues`	The eigenvalues corresponding to the eigenvectors.
`stats`	This is the dimension test statistics for $pfrr$ and "upfrr". It is the cumulatative sum of the eigenvalues for "yc" and "prr"
`fx`	This returns the user-selection of fx for "pfrr" and "upfrr".
`numdir`	The maximum number of directions to be found. The output value of numdir may be smaller than the input value.
`method`	the dimension reduction method used.

Author(s)

Jae Keun Yoo, <peter.yoo@ewhat.ac.kr>.

References

Yoo, JK. (2018). Response dimension reduction: model-based approach. *Statistics : A Journal of Theoretical and Applied Statistic*, 52, 409-425. "prr" and "pfrr"

Yoo, JK. (2019). Unstructured principal fitted response reduction in multivariate regression. *Journal of the Korean Statistical Society*, 48, 561-567. "upfrr"

Yoo, JK. and Cook, R. D. (2008), Response dimension reduction for the conditional mean in multivariate regression. *Statistics and Probability Letters*, 47, 381-389. "yc".

Examples

```
data(mps)
# default fitting method is "upfrr"
s0 <- mbrdr(cbind(A4, B4, A6, B6)~AFDC+Attend+B+Enrol+HS+Minority+Mobility+Poverty+PTR, data=mps)
summary(s0)

# Refit, using different choice of fx.
summary(s1 <- update(s0, fx.choice=2))

# Refit again, using pfrr with fx.choice=2
summary(s2<-update(s1, method="pfrr", fx.choice=1))

# Refit, using prr, which does not require the choice of fx.
summary(s3<- update(s1,method="prr"))

# fit using Yoo-Cook method:
summary(s4 <- update(s1,method="yc"))
```

---

mbrdr.x *Accessor functions for data in dr objects*

Description

Accessor functions for dr objects.

Usage

```
mbrdr.x(object)
mbrdr.y(object)
```

Arguments

object	An object that inherits from mbrdr.

Value

Returns a component of a dr object. `mbrdr.x` returns the predictor matrix reduced to full rank by dropping trailing columns; `mbrdr.y` returns the response vector/matrix.

Author(s)

Jae Keun Yoo, <peter.yoo@ewha.ac.kr>

See Also

`mbrdr`.

`mps` *Minneapolis School dataset*

Description

The Minneapolis school dataset was collected to evaluate the performance of student The percentages of students in 63 Minneapolis schools in 1972. And, The dataset was reported in Star-Tribune in 1973.

Usage

```
data(mps)
```

Format

A data frame of dimension is 63 x 15. Each row represents one elementary school. The first four columns correspond to percentages of students in a school scoring above (A) and below (B) average on standardized fourth and sixth grade reading comprehension tests. Subtracting either pair of grade specific percentages from 100 gives the percentage of students scoring about average on the test. All the other variables are demographic informations for each school.

Details

A4 = percentage of 4th graders scoring ABOVE average on a standard 4th grade vocabulary test in 1972.

B4 = percentage of 4th graders scoring BELOW average on a standard 4th grade vocabulary test in 1972.

A6 = percentage of 6th graders scoring BELOW average on a standard 6th grade comprehension test in 1972.

B6 = percentage of 6th graders scoring BELOW average on a standard 6th grade comprehension test in 1972.

AFDC = percentage of children receiving Aid to Families with Dependent Children

Attend = average percentage of childern in attendance during the year

B = percentage of children in the school not living with Both Parents

BthPts = percentage of children in the school living with Both Parents

Enrol = number of childeren enrolled in the school

HS = percent of adults in the school area who have completed high school

Minority = percent minority children in the area.

Mobility = percentage of children who started in a school, but did not finish there

Poverty = percentage of persons in the school area who are above the federal poverty levels

PTR = pupil-teacher ratio

School = names of school

References

Cook, R. D. and Setodji, C. M. (2003) A model-free test for reduced rank in multivariate regression. Journal of the American Statistical Association, 98, pp. 340-351.

JK. Yoo (2019) Unstructured principal fitted response reduction in multivariate regression. Journal of the Korean Statistical Society, 48, pp. 561-567.

Examples

```
data(mps)
pairs(mps[,1:4])
```

SIGMAS — *compute all required SIGMA matrices for* "pfrr" *and* "upfrr"

Description

Returns Sigmahat, Sigmahat_fit and Sigmahat_res for principal fitted response reduction and unstructured principal fitted response reduction using the choice of fx.

Usage

```
SIGMAS(Y, fx)
```

Arguments

Y	$n \times r$ response matrix
fx	the chosen fx

Details

Both of principal fitted response reduction and unstructured principal fitted response reduction require to compute many SIGMAs. The SIGMAs are as follows: Sigmahat = (Y^T Y)/n; Sigmahat_fit = (Y^T P_fx Y)/n; Sigmahat_res = Sigmahat - Sigmahat_fit.

Value

A list of Sigmahat, Sigmahat_fit and Sigmahat_res.

Author(s)

Jae Keun Yoo, peter.yoo@ewha.ac.kr

Examples

```
data(mps)
X <- mps[,c(5:6,8:14)]
Y <- mps[,c(1:4)]
fx1 <- choose.fx(X)
fx2 <- choose.fx(X, fx.choice=4, nclust=3)

SIGMAS(Y, fx1)
SIGMAS(Y, fx2)
```

Index