

Package ‘dr’

June 29, 2009

Version 3.0.4

Date 2009/06/29

Title Methods for dimension reduction for regression

Author Sanford Weisberg <sandy@stat.umn.edu>,

Maintainer Sanford Weisberg <sandy@stat.umn.edu>

Depends MASS

Description Functions, methods, and datasets for fitting dimension reduction regression, using slicing (methods SAVE and SIR), Principal Hessian Directions (phd, using residuals and the response), and an iterative IRE. Partial methods, that condition on categorical predictors are also available. A variety of tests, and stepwise deletion of predictors, is also included. Also included is code for computing permutation tests of dimension. Adding additional methods of estimating dimension is straightforward. For documentation, see the vignette in the package. With version 3.0.4, the arguments for dr.step have been modified.

License GPL (>= 2)

URL <http://www.r-project.org>

Repository CRAN

Date/Publication 2009-06-29 18:32:50

R topics documented:

ais	2
banknote	3
dr	3
dr.coordinate.test	7
dr.directions	9
dr.permutation.test	11
dr.pvalue	12
dr.slices	13
dr.weights	14

drop1.dr	16
mussels	18
plot.dr	18

Index	20
--------------	-----------

ais	<i>Australian institute of sport data</i>
-----	---

Description

Data on 102 male and 100 female athletes collected at the Australian Institute of Sport.

Format

This data frame contains the following columns:

Sex (0 = male or 1 = female)
Ht height (cm)
Wt weight (kg)
LBM lean body mass
RCC red cell count
WCC white cell count
Hc Hematocrit
Hg Hemoglobin
Ferr plasma ferritin concentration
BMI body mass index, $\text{weight}/(\text{height})^{**2}$
SSF sum of skin folds
Bfat Percent body fat
Label Case Labels
Sport Sport

Source

Ross Cunningham and Richard Telford

References

S. Weisberg (2005). *Applied Linear Regression*, 3rd edition. New York: Wiley, Section 6.4

Examples

```
data(ais)
```

`banknote`*Swiss banknote data*

Description

Six measurements made on 100 genuine Swiss banknotes and 100 counterfeit ones.

Format

This data frame contains the following columns:

Length Length of bill, mm

Left Width of left edge, mm

Right Width of right edge, mm

Bottom Bottom margin width, mm

Top Top margin width, mm

Diagonal Length of image diagonal, mm

Y 0 = genuine, 1 = counterfeit

Source

Flury, B. and Riedwyl, H. (1988). *Multivariate Statistics: A practical approach*. London: Chapman & Hall.

References

Weisberg, S. (2005). *Applied Linear Regression*, 3rd edition. New York: Wiley, Problem 12.5.

Examples

```
data(banknote)
```

`dr`*Main function for dimension reduction regression*

Description

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

Usage

```
dr (formula, data, subset, group=NULL, na.action = na.fail, weights, ...)

dr.compute (x, y, weights, group=NULL, method = "sir", chi2approx="bx",...)
```

Arguments

formula	<p>a two-sided formula like $y \sim x_1 + x_2 + x_3$, where the left-side variable is a vector or a matrix of the response variable(s), and the right-hand side variables represent the predictors. While any legal formula in the Rogers-Wilkinson notation can appear, dimension reduction methods generally expect the predictors to be numeric, not factors, with no nesting. Full rank models are recommended, although rank deficient models are permitted.</p> <p>The left-hand side of the formula will generally be a single vector, but it can also be a matrix, such as <code>cbind(y1+y2) ~ x1+x2+x3</code> if the method is "save" or "sir". Both of these methods are based on slicing, and for the multivariate case slices are determined by slicing on all the columns of the left-hand side variables.</p>
data	an optional data frame containing the variables in the model. By default the variables are taken from the environment from which 'dr' is called.
subset	an optional vector specifying a subset of observations to be used in the fitting process.
group	If used, this argument specifies a grouping variable so that dimension reduction is done separately for each distinct level. This is implemented only when method is one of "sir", "save", or "ire". This argument must be a one-sided formula. For example, <code>~Location</code> would fit separately for each level of the variable <code>Location</code> . The formula <code>~A:B</code> would fit separately for each combination of A and B, provided that both have been declared factors.
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 <code>dr</code> and then passed to <code>dr.compute</code> , or you can create it yourself.
y	The response vector or matrix
method	This character string specifies the method of fitting. The options include "sir", "save", "phdy", "phdres" and "ire". Each method may have its own additional arguments, or its own defaults; see the details below for more information.
chi2approx	Several dr methods compute significance levels using statistics that are asymptotically distributed as a linear combination of $\chi^2(1)$ random variables. This keyword chooses the method for computing the chi2approx, either "bx", the

default for a method suggested by Bentler and Xie (2000) or "wood" for a method proposed by Wood (1989).

... For `dr`, all additional arguments passed to `dr.compute`. For `dr.compute`, additional arguments may be required for particular dimension reduction method. For example, `nslices` is the number of slices used by "sir" and "save". `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 studies $F(y|x)$, the conditional distribution of a response y given a set of predictors x . This function provides methods for estimating the dimension and central subspace of a general regression problem. That is, we want to find a $p \times d$ matrix B of minimal rank d such that

$$F(y|x) = F(y|B'x)$$

Both the dimension d and the subspace $R(B)$ are unknown. These methods make few assumptions. Many methods are based on the inverse distribution, $F(x|y)$.

For the methods "sir", "save", "phdy" and "phdres", a kernel matrix M is estimated such that the column space of M should be close to the central subspace $R(B)$. The eigenvectors corresponding to the d largest eigenvalues of M provide an estimate of $R(B)$.

For the method "ire", subspaces are estimated by minimizing an objective function.

Categorical predictors can be included using the `groups` argument, with the methods "sir", "save" and "ire", using the ideas from Chiaromonte, Cook and Li (2002).

The primary output from this method is (1) a set of vectors whose span estimates $R(B)$; and various tests concerning the dimension d .

Weights can be used, essentially to specify the relative frequency of each case in the data. Empirical weights that make the contours of the weighted sample closer to elliptical can be computed using `dr.weights`. This will usually result in zero weight for some cases. The function will set zero estimated weights to missing.

Value

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

<code>x</code>	The design matrix
<code>y</code>	The response vector
<code>weights</code>	The weights used, normalized to add to <code>n</code> .
<code>qr</code>	QR factorization of <code>x</code> .
<code>cases</code>	Number of cases used.
<code>call</code>	The initial call to <code>dr</code> .
<code>M</code>	A matrix that depends on the method of computing. The column space of <code>M</code> should be close to the central subspace.
<code>evalues</code>	The eigenvalues of <code>M</code> (or squared singular values if <code>M</code> is not symmetric).

evecors	The eigenvectors of M (or of M'M if M is not square and symmetric) ordered according to the eigenvalues.
chi2approx	Value of the input argument of this name.
numdir	The maximum number of directions to be found. The output value of numdir may be smaller than the input value.
slice.info	output from 'sir.slice', used by sir and save.
method	the dimension reduction method used.
terms	same as terms attribute in lm or glm. Needed to make update work correctly.
A	If method="save", then A is a three dimensional array needed to compute test statistics.

Author(s)

Sanford Weisberg, <sandy@stat.umn.edu>.

References

- Bentler, P. M. and Xie, J. (2000), Corrections to test statistics in principal Hessian directions. *Statistics and Probability Letters*, 47, 381-389. Approximate p-values.
- Cook, R. D. (1998). *Regression Graphics*. New York: Wiley. This book provides the basic results for dimension reduction methods, including detailed discussion of the methods "sir", "phdy" and "phdres".
- Cook, R. D. (2004). Testing predictor contributions in sufficient dimension reduction. *Annals of Statistics*, 32, 1062-1092. Introduced marginal coordinate tests.
- Cook, R. D. and Nachtsheim, C. (1994), Reweighting to achieve elliptically contoured predictors in regression. *Journal of the American Statistical Association*, 89, 592-599. Describes the weighting scheme used by [dr.weights](#).
- Cook, R. D. and Ni, L. (2004). Sufficient dimension reduction via inverse regression: A minimum discrepancy approach, *Journal of the American Statistical Association*, 100, 410-428. The "ire" is described in this paper.
- Cook, R. D. and Weisberg, S. (1999). *Applied Regression Including Computing and Graphics*, New York: Wiley, <http://www.stat.umn.edu/arc>. The program `arc` described in this book also computes most of the dimension reduction methods described here.
- Chiaromonte, F., Cook, R. D. and Li, B. (2002). Sufficient dimension reduction in regressions with categorical predictors. *Ann. Statist.* 30 475-497. Introduced grouping, or conditioning on factors.
- Shao, Y., Cook, R. D. and Weisberg (2007). Marginal tests with sliced average variance estimation. *Biometrika*. Describes the tests used for "save".
- Wen, X. and Cook, R. D. (2007). Optimal Sufficient Dimension Reduction in Regressions with Categorical Predictors, *Journal of Statistical Inference and Planning*. This paper extends the "ire" method to grouping.
- Wood, A. T. A. (1989) An F approximation to the distribution of a linear combination of chi-squared variables. *Communications in Statistics: Simulation and Computation*, 18, 1439-1456. Approximations for p-values.

Examples

```

data(ais)
# default fitting method is "sir"
s0 <- dr(LBM~log(SSF)+log(Wt)+log(Hg)+log(Ht)+log(WCC)+log(RCC)+
  log(Hc)+log(Ferr),data=ais)
# Refit, using a different function for slicing to agree with arc.
summary(s1 <- update(s0,slice.function=dr.slices.arc))
# Refit again, using save, with 10 slices; the default is max(8,ncol+3)
summary(s2<-update(s1,nslices=10,method="save"))
# Refit, using phdres. Tests are different for phd, and not
# Fit using phdres; output is similar for phdy, but tests are not justifiable.
summary(s3<- update(s1,method="phdres"))
# fit using ire:
summary(s4 <- update(s1,method="ire"))
# fit using Sex as a grouping variable.
s5 <- update(s4,group=~Sex)

```

dr.coordinate.test *Dimension reduction tests*

Description

Functions to compute various tests concerning the dimension of a central subspace.

Usage

```

dr.test(object, numdir, ...)

dr.coordinate.test(object, hypothesis,d,chi2approx,...)

## S3 method for class 'ire':
dr.joint.test(object, hypothesis, d = NULL,...)

```

Arguments

object	The name of an object returned by a call to dr.
hypothesis	A specification of the null hypothesis to be tested by the coordinate hypothesis. See details below for options.
d	For conditional coordinate hypotheses, specify the dimension of the central mean subspace, typically 1, 2 or possibly 3. If left at the default, tests are unconditional.
numdir	The maximum dimension to consider. If not set defaults to 4.
chi2approx	Approximation method for p.values of linear combination of $\chi^2(1)$ random variables. Choices are from <code>c("bx", "wood")</code> , for the Bentler-Xie and Wood approximations, respectively. The default is either "bx" or the value set in the call that created the dr object.
...	Additional arguments. None are currently available.

Details

`dr.test` returns marginal dimension tests. `dr.coordinate.test` returns marginal dimension tests (Cook, 2004) if `d=NULL` or conditional dimension tests if `d` is a positive integer giving the assumed dimension of the central subspace. The function `dr.joint.test` tests the coordinate hypothesis and dimension simultaneously. It is defined only for `ire`, and is used to compute the conditional coordinate test.

As an example, suppose we have created a `dr` object using the formula $y \sim x_1 + x_2 + x_3 + x_4$. The marginal coordinate hypothesis defined by Cook (2004) tests the hypothesis that y is independent of some of the predictors given the other predictors. For example, one could test whether x_4 could be dropped from the problem by testing y independent of x_4 given x_1, x_2, x_3 .

The hypothesis to be tested is determined by the argument `hypothesis`. The argument `hypothesis = ~. -x4` would test the hypothesis of the last paragraph. Alternatively, `hypothesis = ~x1+x2+x3` would fit the same hypothesis.

More generally, if H is a $p \times q$ rank q matrix, and $P(H)$ is the projection on the column space of H , then specifying `hypothesis = H` will test the hypothesis that Y is independent of $(I - P(H))X|P(H)X$.

Value

Returns a list giving the value of the test statistic and an asymptotic p.value computed from the test statistic. For SIR objects, the p.value is computed in two ways. The *general test*, indicated by `p.val(Gen)` in the output, assumes only that the predictors are linearly related. The *restricted test*, indicated by `p.val(Res)` in the output, assumes in addition to the linearity condition that a constant covariance condition holds; see Cook (2004) for more information on these assumptions. In either case, the asymptotic distribution is a linear combination of Chi-squared random variables. The function specified by the `chi2approx` approximates this linear combination by a single Chi-squared variable.

For SAVE objects, two p.values are also returned. `p.val(Nor)` assumes predictors are normally distributed, in which case the test statistic is asymptotically Chi-squared with the number of df shown. Assuming general linearly related predictors we again get an asymptotic linear combination of Chi-squares that leads to `p.val(Gen)`.

For IRE and PIRE, the tests statistics have an asymptotic χ^2 distribution, so the value of `chi2approx` is not relevant.

Author(s)

Yongwu Shao for SIR and SAVE and Sanford Weisberg for all methods, <sandy@stat.umn.edu>

References

- Cook, R. D. (2004). Testing predictor contributions in sufficient dimension reduction. *Annals of Statistics*, 32, 1062-1092.
- Cook, R. D. and Ni, L. (2004). Sufficient dimension reduction via inverse regression: A minimum discrepancy approach, *Journal of the American Statistical Association*, 100, 410-428.
- Cook, R. D. and Weisberg, S. (1999). *Applied Regression Including Computing and Graphics*. Hoboken NJ: Wiley.

Shao, Y., Cook, R. D. and Weisberg, S. (2007, in press). Marginal tests with sliced average variance estimation. *Biometrika*.

See Also

[drop1.dr](#), [coord.hyp.basis](#), [dr.step](#), [dr.pvalue](#)

Examples

```
# This will match Table 5 in Cook (2004).
data(ais)
# To make this identical to Arc (Cook and Weisberg, 1999), need to modify slices to match.
summary(s1 <- dr(LBM~log(SSF)+log(Wt)+log(Hg)+log(Ht)+log(WCC)+log(RCC)+log(Hc)+log(Ferr),
  data=ais,method="sir",slice.function=dr.slices.arc,nslices=8))
dr.coordinate.test(s1,~-log(Hg))
#The following nearly reproduces Table 5 in Cook (2004)
drop1(s1,chi2approx="wood",update=FALSE)
drop1(s1,d=2,chi2approx="wood",update=FALSE)
drop1(s1,d=3,chi2approx="wood",update=FALSE)
```

dr.directions

Directions selected by dimension reduction regression

Description

Dimension reduction regression returns a set of up to p orthogonal direction vectors each of length p , the first d of which are estimates a basis of a d dimensional central subspace. The function returns the estimated directions in the original n dimensional space for plotting.

Usage

```
dr.direction(object, which, x)
dr.directions(object, which, x)
## Default S3 method:
dr.direction(object, which=NULL, x=dr.x(object))

dr.basis(object, numdir)

## S3 method for class 'ire':
dr.basis(object, numdir=length(object$result))
```

Arguments

object	a dimension reduction regression object created by <code>dr</code> .
which	select the directions wanted, default is all directions. If method is <code>ire</code> , then the directions depend on the value of the dimension you select. If omitted, select all directions.
numdir	The number of basis vectors to return
x	select the X matrix, the default is <code>dr.x(object)</code>

Details

Dimension reduction regression is used to estimate a basis of the central subspace or mean central subspace of a regression. If there are p predictors, the dimension of the central subspace is less than or equal to p . These two functions, `dr.basis` and `dr.direction`, return vectors that describe the central subspace in various ways.

Consider `dr.basis` first. If you set `numdir=3`, for example, this method will return a p by 3 matrix whose columns span the estimated three dimensional central subspace. For all methods except for `ire`, this simply returns the first three columns of `object$e` vectors. For the `ire` method, this returns the three vectors determined by a three-dimensional solution. Call this matrix C . The basis is determined by back-transforming from centered and scaled predictors to the scale of the original predictors, and then renormalizing the vectors to have length one. These vectors are orthogonal in the inner product determined by $\text{Var}(X)$.

The `dr.direction` method return XC , the same space but now a subspace of the original n -dimensional space. These vectors are appropriate for plotting.

Value

Both functions return a matrix: for `dr.direction`, the matrix has n rows and `numdir` columns, and for `dr.basis` it has p rows and `numdir` columns.

Author(s)

Sanford Weisberg <sandy@stat.umn.edu>

References

See R. D. Cook (1998). Regression Graphics. New York: Wiley.

See Also

[dr](#)

Examples

```
data(ais)
#fit dimension reduction using sir
m1 <- dr(LBM~Wt+Ht+RCC+WCC, method="sir", nslices = 8, data=ais)
summary(m1)
dr.basis(m1)
dr.directions(m1)
```

`dr.permutation.test`*Permutation tests of dimension for dr*

Description

Approximates marginal dimension test significance levels for sir, save, and phd by sampling from the permutation distribution.

Usage

```
dr.permutation.test(object, npermute=50, numdir=object$numdir)
```

Arguments

<code>object</code>	a dimension reduction regression object created by <code>dr</code>
<code>npermute</code>	number of permutations to compute, default is 50
<code>numdir</code>	maximum permitted value of the dimension, with the default from the object

Details

The method approximates significance levels of the marginal dimension tests based on a permutation test. The algorithm: (1) permutes the rows of the predictor but not the response; (2) computes marginal dimension tests for the permuted data; (3) obtains significance levels by comparing the observed statistics to the permutation distribution.

The method is not implemented for `ire`.

Value

Returns an object of type 'dr.permutation.test' that can be printed or summarized to give the summary of the test.

Author(s)

Sanford Weisberg, sandy@stat.umn.edu

References

See www.stat.umn.edu/arc/addons.html, and then select the article on dimension reduction regression or inverse regression.

See Also

[dr](#)

Examples

```
data(ais)
attach(ais) # the Australian athletes data
#fit dimension reduction regression using sir
m1 <- dr(LBM~Wt+Ht+RCC+WCC, method="sir", nslices = 8)
summary(m1)
dr.permutation.test(m1, npermute=100)
plot(m1)
```

dr.pvalue

Compute the Chi-square approximations to a weighted sum of Chi-square(1) random variables.

Description

Returns an approximate quantile for a weighted sum of independent $\chi^2(1)$ random variables.

Usage

```
dr.pvalue(coef, f, chi2approx=c("bx", "wood"), ...)

bentlerxie.pvalue(coef, f)

wood.pvalue(coef, f, tol=0.0, print=FALSE)
```

Arguments

coef	a vector of nonnegative weights
f	Observed value of the statistic
chi2approx	Which approximation should be used?
tol	tolerance for Wood's method.
print	Printed output for Wood's method
...	Arguments passed from dr.pvalue to wood.pvalue.

Details

For Bentler-Xie, we approximate f by $c\chi^2(d)$ for values of c and d computed by the function. The Wood approximation is more complicated.

Value

Returns a data.frame with four named components:

test	The input argument f .
test.adj	For Bentler-Xie, returns cf ; for Wood, returns NA.
df.adj	For Bentler-Xie, returns d ; for Wood, returns NA.
pval.adj	Approximate p.value.

Author(s)

Sanford Weisberg <sandy@stat.umn.edu>

References

Peter M. Bentler and Jun Xie (2000), Corrections to test statistics in principal Hessian directions. *Statistics and Probability Letters*, 47, 381-389.

Wood, Andrew T. A. (1989) An F approximation to the distribution of a linear combination of chi-squared variables. *Communications in Statistics: Simulation and Computation*, 18, 1439-1456.

dr.slices

Divide a vector into slices of approximately equal size

Description

Divides a vector into slices of approximately equal size.

Usage

```
dr.slices(y, nslices)
```

```
dr.slices.arc(y, nslices)
```

Arguments

y	a vector of length n or an $n \times p$ matrix
nslices	the number of slices, no larger than n , or a vector of p numbers giving the number of slices in each direction. If y has p columns and nslices is a number, then the number of slices in each direction is the smallest integer greater than the p -th root of nslices.

Details

If y is an n -vector, order y . The goal for the number of observations per slice is m , the smallest integer in $\text{nslices}/n$. Allocate the first m observations to slice 1. If there are duplicates in y , keep adding observations to the first slice until the next value of y is not equal to the largest value in the first slice. Allocate the next m values to the next slice, and again check for ties. Continue until all values are allocated to a slice. This does not guarantee that nslices will be obtained, nor does it guarantee an equal number of observations per slice. This method of choosing slices is invariant under rescaling, but not under multiplication by -1 , so the slices of y will not be the same as the slices of $-y$. This function was rewritten for Version 2.0.4 of this package, and will no longer give exactly the same results as the program Arc. If you want to duplicate Arc, use the function `dr.slice.arc`, as illustrated in the example below.

If y is a matrix of p columns, slice the first column as described above. Then, within each of the slices determined for the first column, slice based on the second column, so that each of the “cells” has approximately the same number of observations. Continue through all the columns. This method is not invariant under reordering of the columns, or under multiplication by -1 .

Value

Returns a named list with three elements as follows:

<code>slice.indicator</code>	ordered eigenvectors that describe the estimates of the dimension reduction subspace
<code>nslices</code>	Gives the actual number of slices produced, which may be smaller than the number requested.
<code>slice.sizes</code>	The number of observations in each slice.

Author(s)

Sanford Weisberg, <sandy@stat.umn.edu>

References

R. D. Cook and S. Weisberg (1999), *Applied Regression Including Computing and Graphics*, New York: Wiley.

See Also

[dr](#)

Examples

```
data(ais)
summary(s1 <- dr(LBM~log(SSF)+log(Wt)+log(Hg)+log(Ht)+log(WCC)+log(RCC)+
               log(Hc)+log(Ferr), data=ais,method="sir",nslices=8))
# To make this identical to ARC, need to modify slices to match.
summary(s2 <- update(s1,slice.info=dr.slices.arc(ais$LBM,8)))
```

dr.weights

Estimate weights for elliptical symmetry

Description

This function estimate weights to apply to the rows of a data matrix to make the resulting weighted matrix as close to elliptically symmetric as possible.

Usage

```
dr.weights(formula, data = list(), subset, na.action = na.fail,
           sigma=1, nsamples=NULL, ...)
```

Arguments

<code>formula</code>	A one-sided or two-sided formula. The right hand side is used to define the design matrix.
<code>data</code>	An optional data frame.
<code>subset</code>	A list of cases to be used in computing the weights.
<code>na.action</code>	The default is <code>na.fail</code> , to prohibit computations. If set to <code>na.omit</code> , the function will return a list of weights of the wrong length for use with <code>dr</code> .
<code>nsamples</code>	The weights are determined by random sampling from a data-determined normal distribution. This controls the number of samples. The default is 10 times the number of cases.
<code>sigma</code>	Scale factor, set to one by default; see the paper by Cook and Nachtsheim for more information on choosing this parameter.
<code>...</code>	Arguments are passed to <code>cov.rob</code> to compute a robust estimate of the covariance matrix.

Details

The basic outline is: (1) Estimate a mean m and covariance matrix S using a possibly robust method; (2) For each iteration, obtain a random vector from $N(m, \sigma^2 S)$. Add 1 to a counter for observation i if the i -th row of the data matrix is closest to the random vector; (3) return as weights the sample fraction allocated to each observation. If you set the keyword `weights.only` to `T` on the call to `dr`, then only the list of weights will be returned.

Value

Returns a list of n weights, some of which may be zero.

Author(s)

Sanford Weisberg, sandy@stat.umn.edu

References

R. D. Cook and C. Nachtsheim (1994), Reweighting to achieve elliptically contoured predictors in regression. *Journal of the American Statistical Association*, 89, 592–599.

See Also

[dr](#), [cov.rob](#)

Examples

```
data(ais)
w1 <- dr.weights(~ Ht +Wt +RCC, data = ais)
m1 <- dr(LBM~Ht+Wt+RCC,data=ais,weights=w1)
```

 drop1.dr

Sequential fitting of coordinate tests using a dr object

Description

This function implements backward elimination using a `dr` object for which a `dr.coordinate.test` is defined, currently for SIR SAVE, IRE and PIRE.

Usage

```
dr.step(object, scope=NULL, d=NULL, minsize=2, stop=0, trace=1, ...)
```

```
## S3 method for class 'dr':
drop1(object, scope = NULL, update=TRUE,
test="general", trace=1, ...)
```

Arguments

<code>object</code>	A <code>dr</code> object for which <code>dr.coordinate.test</code> is defined, for method equal to one of <code>sir</code> , <code>save</code> or <code>ire</code> .
<code>scope</code>	A one sided formula specifying predictors that will never be removed.
<code>d</code>	To use <i>conditional coordinate tests</i> , specify the dimension of the central (mean) subspace. The default is <code>NULL</code> , meaning no conditioning. This is currently available only for methods <code>sir</code> , <code>save</code> without categorical predictors, or for <code>ire</code> with or without categorical predictors.
<code>minsize</code>	Minimum subset size, must be greater than or equal to 2.
<code>stop</code>	Set stopping criterion: continue removing variables until the p-value for the next variable to be removed is less than <code>stop</code> . The default is <code>stop = 0</code> .
<code>update</code>	If true, the <code>update</code> method is used to return a <code>dr</code> object obtained from <code>object</code> by updating the formula to drop the variable with the largest p.value. This can significantly slow the computations for IRE but has little effect on SAVE and SIR.
<code>test</code>	Type of test to be used for selecting the next predictor to remove for <code>method="save"</code> only. "normal" assumes normal predictors, "general" assumes elliptically contoured predictors. For other methods, this argument is ignored.
<code>trace</code>	If positive, print informative output at each step, the default. If trace is 0 or false, suppress all printing.
<code>...</code>	Additional arguments passed to <code>dr.coordinate.test</code> .

Details

Suppose a `dr` object has $p = a + b$ predictors, with a predictors specified in the `scope` statement. `drop1` will compute either marginal coordinate tests (if `d=NULL`) or conditional marginal coordinate tests (if `d` is positive) for dropping each of the b predictors not in the scope, and return `p.values`. The result is an object created from the original object with the predictor with the largest `p.value` removed.

`dr.step` will call `drop1.dr` repeatedly until $\max(a, d + 1)$ predictors remain.

Value

As a side effect, a data frame of labels, tests, `df`, and `p.values` is printed. If `update=TRUE`, a `dr` object is returned with the predictor with the largest `p.value` removed.

Author(s)

Sanford Weisberg, <sandy@stat.umn.edu>, based on the `drop1` generic function in the base R. The `dr.step` function is also similar to `step` in base R.

References

Cook, R. D. (2004). Testing predictor contributions in sufficient dimension reduction. *Annals of Statistics*, 32, 1062-1092.

Shao, Y., Cook, R. D. and Weisberg (2007). Marginal tests with sliced average variance estimation. *Biometrika*.

See Also

[dr.coordinate.test](#)

Examples

```
data(ais)
# To make this identical to ARC, need to modify slices to match by
# using slice.info=dr.slices.arc() rather than nslices=8
summary(s1 <- dr(LBM~log(SSF)+log(Wt)+log(Hg)+log(Ht)+log(WCC)+log(RCC)+
               log(Hc)+log(Ferr), data=ais,method="sir",
               slice.method=dr.slices.arc,nslices=8))
# The following will almost duplicate information in Table 5 of Cook (2004).
# Slight differences occur because a different approximation for the
# sum of independent chi-square(1) random variables is used:
ans1 <- drop1(s1)
ans2 <- drop1(s1,d=2)
ans3 <- drop1(s1,d=3)
# remove predictors stepwise until we run out of variables to drop.
dr.step(s1,scope=~log(Wt)+log(Ht))
```

mussels

Mussels' muscles data

Description

Data were furnished by Mike Camden, Wellington Polytechnic, Wellington, New Zealand. Horse mussels, (*Atrinia*), were sampled from the Marlborough Sounds. The response is the mussels' Muscle Mass.

Format

This data frame contains the following columns:

H Shell height in mm

L Shell length in mm

M Muscle mass in g

S Shell mass in g

W Shell width in mm

Source

R. D. Cook and S. Weisberg (1999). *Applied Statistics Including Computing and Graphics*. New York: Wiley.

plot.dr

Basic plot of a dr object

Description

Plots selected direction vectors determined by a dimension reduction regression fit. By default, the `pairs` function is used for plotting, but the user can use any other graphics command that is appropriate.

Usage

```
## S3 method for class 'dr':
plot(x, which = 1:x$numdir, mark.by.y = FALSE, plot.method = pairs, ...)
```

Arguments

<code>x</code>	The name of an object of class <code>dr</code> , a dimension reduction regression object
<code>which</code>	selects the directions to be plotted
<code>mark.by.y</code>	if <code>TRUE</code> , color points according to the value of the response, otherwise, do not color points but include the response as a variable in the plot.
<code>plot.method</code>	the name of a function for the plotting. The default is <code>pairs</code> .
<code>...</code>	arguments passed to the <code>plot.method</code> .

Value

Returns a graph.

Author(s)

Sanford Weisberg, <sandy@stat.umn.edu>.

Examples

```
data(ais)
# default fitting method is "sir"
s0 <- dr(LBM~log(SSF)+log(Wt)+log(Hg)+log(Ht)+log(WCC)+log(RCC)+
  log(Hc)+log(Ferr), data=ais)
plot(s0)
plot(s0, mark.by.y=TRUE)
```

Index

*Topic **datasets**

ais, [2](#)
banknote, [3](#)
mussels, [18](#)

*Topic **regression**

dr, [3](#)
dr.coordinate.test, [7](#)
dr.directions, [9](#)
dr.permutation.test, [11](#)
dr.pvalue, [12](#)
dr.slices, [13](#)
dr.weights, [14](#)
drop1.dr, [16](#)
plot.dr, [18](#)

plot.dr, [18](#)

wood.pvalue (*dr.pvalue*), [12](#)

ais, [2](#)

banknote, [3](#)

bentlerxie.pvalue (*dr.pvalue*), [12](#)

coord.hyp.basis, [9](#)

cov.rob, [15](#)

dr, [3](#), [10](#), [11](#), [14](#), [15](#)

dr.basis (*dr.directions*), [9](#)

dr.coordinate.test, [7](#), [17](#)

dr.direction (*dr.directions*), [9](#)

dr.directions, [9](#)

dr.joint.test

(*dr.coordinate.test*), [7](#)

dr.permutation.test, [11](#)

dr.pvalue, [9](#), [12](#)

dr.slices, [13](#)

dr.step, [9](#)

dr.step (*drop1.dr*), [16](#)

dr.test (*dr.coordinate.test*), [7](#)

dr.weights, [6](#), [14](#)

drop1.dr, [9](#), [16](#)

mussels, [18](#)