

Package ‘DMR’

July 2, 2014

Type Package

Title Delete or Merge Regressors for linear model selection.

Version 2.0

Date 2013-01-30

Author Aleksandra Maj, Agnieszka Prochenka, Piotr Pokarowski

Maintainer Aleksandra Maj <aleksandra.lucja.maj@gmail.com>

Depends R (>= 1.8.0), magic

Description A backward selection procedure called delete or merge regressors (DMR) combines deleting continuous variables with merging levels of factors. The method assumes greedy search among linear models with set of constraints of two types: either a parameter for a continuous variable is set to zero or parameters corresponding to two levels of a factor are compared. DMR is a stepwise regression procedure, where in each step a new constraint is added according to ranking of the hypotheses based on squared t-statistics. As a result a nested family of linear models is obtained and the final decision is made according to minimization of the generalized information criterion (GIC, default BIC). The main function of the package is `DMR`, which is based on hierarchical clustering. Moreover, other functions for extensions of DMR method are given, such as `stepDMR` which is based on recalculation of t-statistics in each step and function `DMR4glm` for generalized linear models.

License GPL-2

LazyLoad yes

NeedsCompilation no

Repository CRAN

Date/Publication 2013-02-21 13:19:28

R topics documented:

DMR-package	2
DMR	3
DMR4glm	4
plot_bf	5
roc	6
stepDMR	7
Index	9

DMR-package	<i>Package for performing simultaneous deleting or merging regressors for linear model.</i>
-------------	---

Description

A backward selection procedure called delete or merge regressors (DMR) combines deleting continuous variables with merging levels of factors. The method assumes greedy search among linear models with set of constraints of two types: either a parameter for a continuous variable is set to zero or parameters corresponding to two levels of a factor are compared. DMR is a stepwise regression procedure, where in each step a new constraint is added according to ranking of the hypotheses based on squared t-statistics. As a result a nested family of linear models is obtained and the final decision is made according to minimization of the generalized information criterion (GIC, default BIC). The main function of the package is DMR, which is based on hierarchical clustering. Moreover, other functions for extensions of DMR method are given, such as stepDMR which is based on recalculation of t-statistics in each step and function DMR4glm for generalized linear models.

Details

Package:	DMR
Type:	Package
Version:	1.0
Date:	2013-02-01
License:	GPL-2
LazyLoad:	yes

Author(s)

Aleksandra Maj, Agnieszka Prochenka, Piotr Pokarowski

Maintainer: <a.maj@phd.ipipan.waw.pl>, <a.prochenka@phd.ipipan.waw.pl>

References

CIAMPI, A., LECHEVALLIER, Y., LIMAS, M.C., MARCOS A.G. (2008). Hierarchical clustering of subpopulations with a dissimilarity based on the likelihood ratio statistic: application to clustering massive data sets. *Pattern Anal Appl* 11 199-220. \ BONDELL, H.D. and REICH, B.J. (2009). Simultaneous factor selection and collapsing levels in anova. *Biometrics* 65 169-177. \ GERTHEISS, J., TUTZ, G. (2010). Sparse modeling of categorical explanatory variables. *The Annals of Applied Statistics* 4 2150-2180. \ ZHENG, X. and LOH, W.-Y. (1995). Consistent variable selection in linear models. *Journal of the American Statistical Association* 90 151-156.

DMR

Delete or Merge Regressors

Description

DMR is a stepwise backward model selection procedure which simultaneously deletes continuous variables and merges levels of factors. It is based on ranking linear hypotheses with squared t-statistics, using hierarchical clustering for each categorical variable. The final model is selected by minimization of generalized information criterion in the nested family of models.

Usage

```
DMR(model, K = log(nrow(model$model)), clust.method = "complete")
```

Arguments

model	initial model of class lm.
K	penalty for the number of parameters in generalized information criterion, default is log(n).
clust.method	method of clustering the same as in hclust .

Value

a list including elements

Partitions	a list of partitions of factors for the models on the nested path searched through
Crit	values of generalized information criterion for the models on the nested path searched through
LogLik	values of log-likelihood for the models on the nested path searched through
Best	a list containing features of the selected model: Partition, Model of class lm, Crit and Hypotheses represented as a matrix of linear hypotheses imposed on the model's parameters

Author(s)

Piotr Pokarowski, Agnieszka Prochenka, Aleksandra Maj

See Also

[stepDMR](#), [DMR4glm](#), [plot_bf](#), [roc](#)

Examples

```
k=4
v1 <- factor(rep(1:8, each = 12*k))
v2 <- factor(rep(1:4, times = 24*k))
v3 <- factor(rep(1:3, times = 32*k))
x1 <- rnorm(96*k)
x2 <- runif(96*k)
y <- rep(c(2, 2, -1, -1, -1, -1, 0, 0), each = 12*k) + rnorm(96*k)
m <- lm(y ~ v1 + v2 + v3 + x1 + x2)
(out <- DMR(m))
```

DMR4glm

Delete or Merge Regressors for Generalized Linear Models

Description

DMR4glm is a backward model selection procedure which simultaneously deletes continuous variables and merges levels of factors. It is a generalization of DMR onto generalized linear models, where instead of squared t-statistics, squared Wald statistics are used. The final model is selected by minimization of generalized information criterion in the nested family of models.

Usage

```
DMR4glm(model, K = log(nrow(model$model)), clust.method = 'complete')
```

Arguments

<code>model</code>	initial model of class <code>glm</code> .
<code>K</code>	penalty for the number of parameters in generalized information criterion, default is $\log(n)$.
<code>clust.method</code>	method of clustering, the same as in hclust .

Value

	a list including elements
<code>Partitions</code>	a list of partitions of factors for the models on the nested path searched through
<code>Crit</code>	values of generalized information criterion for the models on the nested path searched through
<code>LogLik</code>	values of log-likelihood for the models on the nested path searched through
<code>Models</code>	a list of models of class <code>glm</code> on the nested path searched through
<code>Best</code>	a list containing features of the selected model: Partition, Model of class <code>glm</code> , Crit and Hypotheses represented as a matrix of linear hypotheses imposed on the model's parameters

Author(s)

Piotr Pokarowski, Agnieszka Prochenka, Aleksandra Maj

See Also

[DMR](#), [stepDMR](#), [plot_bf](#), [roc](#)

Examples

```
k <- 4
v1 <- factor(rep(1:8, each=12*k))
v2 <- factor(rep(1:4, times = 24*k))
v3 <- factor(rep(1:3, times = 32*k))
x1 <- rnorm(96*k)
x2 <- runif(96*k)
mi <- rep(c(2, 2, -1, -1, -1, -1, 0, 0), each = 12*k)
y <- rbinom(96*k, 1, exp(mi)/(1+exp(mi)))
m <- glm(y ~ x1 + x2 + v1 + v2 + v3, family = binomial)
(out <- DMR4glm(m))
```

plot_bf

Plot Approximate Bayes Factors

Description

plot_bf is used for plotting values of approximate Bayes factors for models on the nested path created by DMR, stepDMR or DMR4glm algorithm with respect to the best model selected by the procedure. Bayes factors are approximated using values of BIC calculated by the function.

Usage

```
plot_bf(m)
```

Arguments

m a list obtained from execution of DMR, stepDMR or DMR4glm.

Value

plot of approximate Bayes factors for the models on the path searched through

Author(s)

Piotr Pokarowski, Agnieszka Prochenka, Aleksandra Maj

See Also

[DMR](#), [stepDMR](#), [DMR4glm](#), [roc](#)

Examples

```

k=4
v1 <- factor(rep(1:8, each = 12*k))
v2 <- factor(rep(1:4, times = 24*k))
v3 <- factor(rep(1:3, times = 32*k))
x1 <- rnorm(96*k)
x2 <- runif(96*k)
y <- rep(c(2, 2, -1, -1, -1, -1, 0, 0), each = 12*k) + rnorm(96*k)
m <- lm(y ~ v1 + v2 + v3 + x1 + x2)
out <- DMR(m)
plot_bf(out)

```

roc

*Measures of Performance***Description**

roc is used for calculating measures of performance such as sensitivity and specificity when the true and predicted models can be described using linear hypotheses.

Usage

```
roc(Y0, X0)
```

Arguments

Y0	matrix of linear constraints for the true model's parameters
X0	matrix of linear constraints for the predicted model's parameters

Value

a list including elements

sensitivity	proportion of the dimension of intersection of linear spaces of parameters for the true and predicted models to the dimension of linear space of parameters for the true model
specificity	proportion of the dimension of intersection of linear spaces that are complements of linear spaces of parameters for the true and predicted models to the dimension of linear space which is a complement of the linear space of parameters for the true model

Author(s)

Piotr Pokarowski, Agnieszka Prochenka, Aleksandra Maj

See Also

[DMR](#), [stepDMR](#), [DMR4glm](#), [plot_bf](#)

Examples

```

v1 <- factor(rep(1:3, times = 16))
set.seed(1)
x1 <- rnorm(48)
set.seed(2)
y <- x1 + rnorm(48, 0, 3)
m <- lm(y ~ v1 + x1)
out <- DMR(m)
x0 <- out$Best$Hypotheses
y0 <- matrix(c(0, 0, -1, 1, 1, 0, 0, 0), 2, 4)
measures <- roc(y0, x0)
sen <- measures$sensitivity
spe <- measures$specificity

```

stepDMR

Stepwise Delete or Merge Regressors

Description

Stepwise DMR is a backward model selection procedure which simultaneously deletes continuous variables and merges levels of factors. It is a stepwise version of DMR, where in every step the values of squared t-statistics are recalculated. The final model is selected by minimization of generalized information criterion in the nested family of models.

Usage

```
stepDMR(model, K = log(nrow(model$model)))
```

Arguments

model	initial model of class lm.
K	penalty for the number of parameters in generalized information criterion, default is log(n).

Value

a list including elements

Partitions	a list of partitions of factors for the models on the nested path searched through
Crit	values of generalized information criterion for the models on the nested path searched through
LogLik	values of log-likelihood for the models on the nested path searched through
Best	a list containing features of the selected model: Partition, Model of class lm, Crit and Hypotheses represented as a matrix of linear hypotheses imposed on the model's parameters

Author(s)

Piotr Pokarowski, Agnieszka Prochenka, Aleksandra Maj

See Also

[DMR](#), [DMR4glm](#), [plot_bf](#), [roc](#)

Examples

```
k=4
v1 <- factor(rep(1:8, each = 12*k))
v2 <- factor(rep(1:4, times = 24*k))
v3 <- factor(rep(1:3, times = 32*k))
x1 <- rnorm(96*k)
x2 <- runif(96*k)
y <- rep(c(2, 2, -1, -1, -1, -1, 0, 0), each = 12*k) + rnorm(96*k)
m <- lm(y ~ v1 + v2 + v3 + x1 + x2)
(out <- stepDMR(m))
```


Index

*Topic **model selection**

DMR, [3](#)

DMR4glm, [4](#)

stepDMR, [7](#)

DMR, [3](#), [5](#), [6](#), [8](#)

DMR-package, [2](#)

DMR4glm, [4](#), [4](#), [5](#), [6](#), [8](#)

hclust, [3](#), [4](#)

plot_bf, [4](#), [5](#), [5](#), [6](#), [8](#)

roc, [4](#), [5](#), [6](#), [8](#)

stepDMR, [4-6](#), [7](#)