# Package 'regsel'

March 9, 2016

**Type** Package

**Title** Variable Selection and Regression

**Version** 0.2

**Date** 2016-02-24

**Author** Michal Knut

**Maintainer** Michal Knut <1105406k@student.gla.ac.uk>

**Description** Functions for fitting linear and generalized linear models with variable selection. The functions can automatically do Stepwise Regression, Lasso or Elastic Net as variable selection methods. Lasso and Elastic net are improved and handle factors better (they can either include or exclude all factor levels).

**License** GPL-2

**LazyData** TRUE

**RoxygenNote** 5.0.1

**Depends** glmnet, elasticnet

**NeedsCompilation** no

**Repository** CRAN

**Date/Publication** 2016-03-09 09:13:48

## R topics documented:

## bank *bank dataset*

### Description

the data is related with direct marketing campaigns of a Portuguese banking institution. The marketing campaigns were based on phone calls. Often, more than one contact to the same client was required, in order to access if the product (bank term deposit) would be yes (1) or not (0) subscribed.

### Usage

```
data(bank)
```

### Format

a data frame with 4520 observations on the following 17 variables.

### Details

The data are from Moro (2014) and taken from the UCI ML website https://archive.ics.uci.edu/ml/datasets/Bank+Marketing.

### Value

`bank` data frame of data with predictor columns `age, job, marital, education, default, balance, housing, loan,` and `poutcome` with response column `y` indicating whether the client has subscribed a term deposit.

### Source

S. Moro, P. Cortez and P. Rita. A Data-Driven Approach to Predict the Success of Bank Telemarketing. Decision Support Systems, Elsevier, 62:22-31, June 2014

## concrete *concrete dataset*

### Description

Yeh (1998) describes a collection of data sets from different sources that can be used for modeling the compressive strength of concrete formulations as a functions of their ingredients and age.

### Usage

```
data(concrete)
```

### Format

a `RangedData` instance, 1 row per CpG island.

## Details

The data are from Yeh (1998) and taken from the UCI ML website http://archive.ics.uci.edu/ml/datasets/Concrete+Compressive+Strength. There are 1030 data points from the UCI website, but the paper states that approximately 1,000 samples were made, but only 727 were analyzed in the source material. It is unclear which samples were excluded.

## Value

`concrete` data frame of data with predictor columns `Cement, BlastFurnaceSlag, FlyAsh, Water, Superplasticizer,` and `Age` with response column `CompressiveStrength`.

## Source

Yeh, I. C. (1998). Modeling of strength of high-performance concrete using artificial neural networks. Cement and Concrete Research, 28(12), 1797-1808. Elsevier.

---

# glmsel *Generalised linear models with variable selection*

---

## Description

`glmsel` is used to fit generalised linear models with optionally performing variable selection using stepwise regression, lasso or elastic net methods.

## Usage

```
glmsel(formula, data=environment(), varsel=FALSE, criterion="AIC",
direction="backward", indices=NULL, train=0.3, family, enet.alpha=0.5)
```

## Arguments

| | |
|---|---|
| `formula` | an object of class "formula"; a symbolic description of the model to be fitted. |
| `data` | an optional data frame, list or environment containg the variables in the model. If not specified, the variables are taken from the current environment. |
| `varsel` | a method of variable selection to be used. The default is `"FALSE"`. Available methods include: stepwise regression `"step"`, LASSO `"lasso"`, elastic net`"enet"`. |
| `criterion` | when `varsel="step"`, `criterion` allows to select a method of calculating statistic for model comparison. The default is `"AIC"`. Less liberal, BIC penalty can be used by typing `"BIC"`. |
| `direction` | the mode of stepwise search, can be one of `"both"`, `"backward"`, or `"forward"`, with a default of `"both"`. If the scope argument is missing the default for `direction` is `"backward"`. |
| `indices` | vector of `0` and `1` values indicating which observations are to be used as train and test when `varsel="lasso" or "enet"`. |

| | |
|---|---|
| train | if indices=NULL, the function will randomly assign observations as train and test. train specifies what percentage of data will be used as train observations. Can take values from 0.1 to 0.9. |
| family | a description of the error distribution and link function to be used in the model. This can be a character string naming a family function. |
| enet.alpha | The elastic net mixing parameter, with 0=a= 1. The penalty is defined as $$(1-a)/2\|\|\beta\|\|\_2^2 + a\|\|\beta\|\|_1$$ alpha=1 is the lasso penalty, and alpha=0 the ridge penalty. The default value is 0.5. |

## Value

A "glmsel" object is returned, for which print, plot and summary methods can be used.

## Author(s)

Michal Knut <1105406k@student.gla.ac.uk>.

## Examples

```
data(bank)
glmsel(y~age+as.factor(job)+as.factor(marital)+as.factor(education)+as.factor(default)+balance
+as.factor(housing)+as.factor(loan)+as.factor(contact)+as.factor(day)+month+duration+campaign
+pdays+previous, data=bank, varsel="enet")
```

---

# lmsel *Linear regression with variable selection*

---

## Description

lmsel is used to fit linear models with optionally performing variable selection using stepwise regression, lasso or elastic net methods.

## Usage

```
lmsel(formula, data=environment(), varsel=FALSE, criterion="AIC",
direction="backward", indices=NULL, train=0.3, lambda=1000)
```

## Arguments

| | |
|---|---|
| formula | an object of class "formula"; a symbolic description of the model to be fitted. |
| data | an optional data frame, list or environment containg the variables in the model. If not specified, the variables are taken from the current environment. |
| varsel | a method of variable selection to be used. The default is "FALSE". Available methods include: stepwise regression "step", LASSO "lasso", elastic net"enet". |

| | |
|---|---|
| criterion | when varsel="step", criterion allows to select a method of calculating statistic for model comparison. The default is "AIC". Less liberal, BIC penalty can be used by typing "BIC". |
| direction | the mode of stepwise search, can be one of "both", "backward", or "forward", with a default of "both". If the scope argument is missing the default for direction is "backward". |
| indices | vector of 0 and 1 values indicating which observations are to be used as train and test when varsel="lasso" or "enet". |
| train | if indices=NULL, the function will randomly assign observations as train and test. train specifies what percentage of data will be used as train observations. Can take values from 0.1 to 0.9. |
| lambda | quadratic penalty parameter for elastic net. The default value is 1000. |

## Value

A "lmsel" object is returned, for which print, plot and summary methods can be used.

## Author(s)

Michal Knut <1105406k@student.gla.ac.uk>.

## Examples

```
data(prostate)
set.seed(10)
lmsel(lpsa~lcavol+lweight+age+lbph+svi+lcp+gleason+pgg45, indices=as.numeric(prostate$train),
data=prostate, varsel="lasso")

data(concrete)
lmsel(CompressiveStrength~., data=concrete, varsel="step", criterion="BIC")
```

---

# plot.glmsel *Plot* glmsel *objects*

---

## Description

This function produces diagnostic plots of objects of class glmsel produced by glmsel() function.

## Usage

```
## S3 method for class 'glmsel'
plot(x, ...)
```

## Arguments

| | |
|---|---|
| x | an object of class 'glmsel' |
| ... | arguments to be passed to and from other methods |

## Details

plot will produce residuals versus fitted values and quantile-quantile plots to diagnose the fit of a generalised linear model. If varsel="lasso" or varsel="enet" was selected as arguments in lmsel, plot(object) will produce and additional plot with lasso or elastic net variable selection paths.

## Author(s)

Michal Knut <1105406k@student.gla.ac.uk>.

---

# plot.lmsel *Plot* lmsel *objects*

---

## Description

This function produces diagnostic plots of objects of class lmsel produced by lmsel() function.

## Usage

```
## S3 method for class 'lmsel'
plot(x, ...)
```

## Arguments

| | |
|---|---|
| x | an object of class 'lmsel' |
| ... | arguments to be passed to and from other methods |

## Details

plot will produce residuals versus fitted values and quantile-quantile plots to diagnose the fit of a linear model. If varsel="lasso" or varsel="enet" was selected as arguments in lmsel, plot(object) will produce and additional plot with lasso or elastic net variable selection paths.

## Author(s)

Michal Knut <1105406k@student.gla.ac.uk>.

## prostate *prostate dataset*

### Description

data to examine the correlation between the level of prostate-specific antigen and a number of clinical measures in men who were about to receive a radical prostatectomy.

### Usage

```
data(prostate)
```

### Format

a data frame with 97 observations on the following 10 variables.

### Details

The last column indicates which 67 observations were used as the "training set" and which 30 as the test set, as described on page 48 in the book.

### Value

`concrete` data frame of data with predictor columns `lcavol, lweight, age, lbph, svi, lcp, gleason` and `pgg45` with response column `lpsa` and `train` column indicating "training set".

### Source

Stamey, T., Kabalin, J., McNeal, J., Johnstone, I., Freiha, F., Redwine, E. and Yang, N (1989) Prostate specific antigen in the diagnosis and treatment of adenocarcinoma of the prostate II. Radical prostatectomy treted patients, Journal of Urology 16: 1076-1083.

## summary.glmsel *Summary of* `glmsel` *objects*

### Description

This function produces summary of objects of class `glmsel` produced by `glmsel()` function.

### Usage

```
## S3 method for class 'glmsel'
summary(object, dispersion=NULL, ...)
```

### Arguments

| | |
|---|---|
| object | an object of class 'glmsel' |
| dispersion | argument which allows to include dispersion parameter |
| ... | arguments to be passed to and from other methods |

### Author(s)

Michal Knut <1105406k@student.gla.ac.uk>.

---

## summary.lmsel *Summary of* lmsel *objects*

---

### Description

This function produces summary of objects of class lmsel produced by lmsel() function.

### Usage

```
## S3 method for class 'lmsel'
summary(object, ...)
```

### Arguments

| | |
|---|---|
| object | an object of class 'lmsel' |
| ... | arguments to be passed to and from other methods |

### Author(s)

Michal Knut <1105406k@student.gla.ac.uk>.

# Index