Package ‘l0ara’

July 24, 2017

**Type** Package

**Title** Sparse Generalized Linear Model with L0 Approximation for Feature Selection

**Version** 0.1.5

**Date** 2017-07-23

**Author** Wenchuan Guo, Shujie Ma, Zhenqiu Liu

**Maintainer** Wenchuan Guo <wguo007@ucr.edu>

**Description** An efficient procedure for feature selection for generalized linear models with L0 penalty, including linear, logistic, Poisson, gamma, inverse Gaussian regression. Adaptive ridge algorithms are used to fit the models.

**License** GPL-2

**Imports** Rcpp (>= 0.12.6)

**LinkingTo** Rcpp, RcppArmadillo

**LazyData** TRUE

**RoxygenNote** 6.0.1

**NeedsCompilation** yes

**Repository** CRAN

**Date/Publication** 2017-07-24 03:09:30 UTC

**R topics documented:**

<table>
<thead>
<tr>
<th>Function</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>coef.cv.l0ara</td>
<td>2</td>
</tr>
<tr>
<td>coef.l0ara</td>
<td>2</td>
</tr>
<tr>
<td>cv.l0ara</td>
<td>3</td>
</tr>
<tr>
<td>l0ara</td>
<td>5</td>
</tr>
<tr>
<td>plot.cv.l0ara</td>
<td>7</td>
</tr>
<tr>
<td>plot.l0ara</td>
<td>8</td>
</tr>
<tr>
<td>predict.l0ara</td>
<td>8</td>
</tr>
<tr>
<td>print.cv.l0ara</td>
<td>9</td>
</tr>
<tr>
<td>print.l0ara</td>
<td>10</td>
</tr>
</tbody>
</table>

**Index** 11
 coef.cv.l0ara

Description

Print the coefficients from the model with the optimal lambda.

Usage

## S3 method for class 'cv.l0ara'
coef(object, ...)

Arguments

- object Fitted "cv.l0ara" object.
- ... Not used argument.

Details

This function fits the model with the optimal lambda first and then print the coefficients. This function makes it easier to use the results to make a prediction or to see the fitted model.

Value

The object returns the coefficients.

Author(s)

Wenchuan Guo <wguo007@ucr.edu>, Shujie Ma <shujie.ma@ucr.edu>, Zhenqiu Liu <Zhenqiu.Liu@cshs.org>

See Also

predict method and l0ara function.

 coef.l0ara

Description

Print the coefficients from the model.

Usage

## S3 method for class 'l0ara'
coef(object, ...)

cv.l0ara

Arguments

object  Fitted "l0ara" object.
...

Details

This function makes it easier to use the results to make a prediction or to see the fitted model.

Value

The object returns the coefficients.

Author(s)

Wenchuan Guo <wguo007@ucr.edu>, Shujie Ma <shujie.ma@ucr.edu>, Zhenqiu Liu <Zhenqiu.Liu@cshs.org>

See Also

predict method and lPara function.

cv.l0ara  cross-validation for l0ara

Description

Does k-fold cross-validation for l0ara, produces a plot, and returns the optimal lambda

Usage

cv.l0ara(x, y, family, lam, measure, nfolds, maxit, eps, seed)

Arguments

x  Input matrix as in l0ara.
y  Response variable as in l0ara.
family  Response type as in l0ara.
lam  A user supplied lambda sequence in descending or ascending order. This function does not fit models. To fit a model with given lam value, use l0ara.
measure  Loss function used for cross validation. measure="mse" or "mae" for all models. "measure"="class" or "measure"="auc" only for logistic regression.
folds  Number of folds. Default value is 10. Smallest value is 3.
maxit  Maximum number of passes over the data for lambda. Default value is 1e3.
eps  Convergence threshold. Default value is 1e-4.
seed  Seed of random number generator.
Details

This function calls l0ara nfolds times, each time leaving out 1/nfolds of the data. The cross-validation error is based on either mean square error (mse) or mean absolute error (mae).

Value

An object with S3 class "cv.l0ara" containing:

- cv.error: The mean cross validated error for given lambda sequence
- cv.std: The estimates of standard error of cv.error
- lam.min: The lambda gives min cv.error
- lambda: The lambda used
- measure: Type of measure
- family: Model used
- x: Design matrix
- y: Response variable
- name: Full name of the measure

Author(s)

Wenchuan Guo <wguo007@ucr.edu>, Shujie Ma <shujie.ma@ucr.edu>, Zhenqiu Liu <Zhenqiu.Liu@cshs.org>

See Also

l0ara, coef.cv.l0ara, plot.cv.l0ara methods.

Examples

```r
# Linear regression
# Generate design matrix and response variable
n <- 100
p <- 40
x <- matrix(rnorm(n*p), n, p)
beta <- c(1, 0, 2, 3, rep(0, p-4))
noise <- rnorm(n)
y <- x %*% beta + noise
lam <- c(0.1, 0.3, 0.5)
fit <- cv.l0ara(x, y, family="gaussian", lam, measure = "mse")```
Description
An adaptive ridge algorithm for feature selection with L0 penalty.

Usage
l0ara(x, y, family, lam, standardize, maxit, eps)

Arguments
- **x**: Input matrix, of dimension nobs x nvars; each row is an observation vector.
- **y**: Response variable. Quantitative for family="gaussian"; positive quantitative for family="gamma" or family="inv.gaussian"; a factor with two levels for family="logit"; non-negative counts for family="poisson".
- **family**: Response type (see above).
- **lam**: A user supplied lambda value. If you have a lam sequence, use cv.l0ara first to select optimal tuning and then refit with lam.min. To use AIC, set lam=2; to use BIC, set lam=log(n).
- **standardize**: Logical flag for data normalization. If standardize=TRUE (default), independent variables in the design matrix x will be standardized with mean 0 and standard deviation 1.
- **maxit**: Maximum number of passes over the data for lambda. Default value is 1e5.
- **eps**: Convergence threshold. Default value is 1e-3.

Details
The sequence of models indexed by the parameter lambda is fit using adaptive ridge algorithm. The objective function for generalized linear models (including family above) is defined to be

\[-(loglikelihood) + (\lambda/2) \times |\beta|_0\]

$|\beta|_0$ is the number of non-zero elements in $\beta$. To select the "best" model with AIC or BIC criterion, let lambda to be 2 or log(n). This adaptive ridge algorithm is developed to approximate L0 penalized generalized linear models with sequential optimization and is efficient for high-dimensional data.

Value
An object with S3 class "l0ara" containing:

- **beta**: A vector of coefficients
- **df**: Number of nonzero coefficients
- **iter**: Number of iterations
The lambda used

x Design matrix

y Response variable

Author(s)

Wenchuan Guo <wguo007@ucr.edu>, Shujie Ma <shujie.ma@ucr.edu>, Zhenqiu Liu <Zhenqiu.Liu@chshs.org>

See Also


Examples

```r
# Linear regression
# Generate design matrix and response variable
n <- 100
p <- 40
x <- matrix(rnorm(n*p), n, p)
beta <- c(1, 0, 2, 3, rep(0, p-4))
noise <- rnorm(n)
y <- x %*% beta + noise
# fit sparse linear regression using BIC
res.gaussian <- lPara(x, y, family="gaussian", log(n))

# predict for new observations
print(res.gaussian)
predict(res.gaussian, newx=matrix(rnorm(3, p, 3, p)))
coef(res.gaussian)

# Logistic regression
# Generate design matrix and response variable
n <- 100
p <- 40
x <- matrix(rnorm(n*p), n, p)
beta <- c(1, 0, 2, 3, rep(0, p-4))
prob <- exp(x %*% beta) / (1 + exp(x %*% beta))
y <- rbinom(n, rep(1, n), prob)
# fit sparse logistic regression
res.logit <- lPara(x, y, family="logit", 0.7)

# predict for new observations
print(res.logit)
predict(res.logit, newx=matrix(rnorm(3, p, 3, p)))
coef(res.logit)

# Poisson regression
# Generate design matrix and response variable
n <- 100
p <- 40
x <- matrix(rnorm(n*p), n, p)
beta <- c(1, 0, 0.5, 0.3, rep(0, p-4))
```
mu <- exp(xE*beta)
y <- rpois(nL mu)
# fit sparse Poisson regression using AIC
res.pois <- l0ara(x, y, family="poisson", 2)

# predict for new observations
print(res.pois)
predict(res.pois, newx=matrix(rnorm(3*p,3,p))
coef(res.pois)

plot.cv.l0ara  plot for an "cv.l0ara" object

Description

Produces curves from a fitted "cv.l0ara" object.

Usage

## S3 method for class 'cv.l0ara'
plot(x, col = 3, ...)

Arguments

x  Fitted "cv.l0ara" object.
col  color of the dots.
...  Not used argument.

Author(s)

Wenchuan Guo <wguo007@ucr.edu>, Shujie Ma <shujie.ma@ucr.edu>, Zhenqiu Liu <Zhenqiu.Liu@eshs.org>

See Also

predict, coef methods, cv.l0ara and l0ara function.
plot.l0ara

plot for an "l0ara" object

Description

Two plots are available: a plot of fitted value against linear predictor; roc(auc) curve for family="logit".

Usage

## S3 method for class 'l0ara'
plot(x, auc = FALSE, split = FALSE, col = 4, ...)

Arguments

x  Fitted "l0ara" object.
auc  logical; if TRUE, produces auc curve for family=logit.
split  logical; if TRUE, produces separate plots.
col  color of the dots.
...  Not used argument.

Author(s)

Wenchuan Guo <wguo007@ucr.edu>, Shujie Ma <shujie.ma@ucr.edu>, Zhenqiu Liu <Zhenqiu.Liu@csus.org>

See Also

predict, coef methods and l0ara function.

predict.l0ara

make predictions from a "l0ara" object.

Description

Make predictions from the model.

Usage

## S3 method for class 'l0ara'
predict(object, newx, type = c("link", "response", "coefficients", "class"), ...)
Arguments

object Fitted "l0ara" object.
newx Matrix of new values for x at which predictions are to be made. Must be a matrix.
type Type of prediction required. "link" gives the linear predictors (for "gaussian" models it gives the fitted values). "response" gives the fitted probabilities for "logit" and fitted mean for "poisson". "coefficients" gives the coefficients which is same as "coef" function. "class" (applies only to "logit") produces the class label corresponding to the maximum probability.

Details

This function makes it easier to use the results to make a prediction or to see the fitted model.

Value

The object returned depends the functions.

Author(s)

Wenchuan Guo <wguo007@ucr.edu>, Shujie Ma <shujie.ma@ucr.edu>, Zhenqiu Liu <Zhenqiu.Liu@cshs.org>

See Also

coeff method and l0ara function.

Description

Print the general information of the cross validated fit.

Usage

```r
## S3 method for class 'cv.l0ara'
print(x, ...)
```

Arguments

x Fitted "cv.l0ara" object.

Details

This function makes it easier to see the cross-validation results.
Author(s)
Wenchuan Guo <wguo007@ucr.edu>, Shujie Ma <shujie.ma@ucr.edu>, Zhenqiu Liu <Zhenqiu.Liu@cshs.org>

See Also
predict, coef methods and l0ara function.

print.l0ara

summarizing the fits from a "l0ara" object.

Description
Print the general information of the fit.

Usage

## S3 method for class 'l0ara'
print(x, ...)

Arguments

x  Fitted "l0ara" object.
... Not used argument.

Details
This function makes it easier to see the fitted model.

Author(s)
Wenchuan Guo <wguo007@ucr.edu>, Shujie Ma <shujie.ma@ucr.edu>, Zhenqiu Liu <Zhenqiu.Liu@cshs.org>

See Also
predict, coef methods and l0ara function.
Index

coef, 7–10
coef.cv.10ara, 2, 4
coef.10ara, 2, 6
cv.10ara, 3, 6, 7
l0ara, 2–4, 5, 7–10
plot.cv.10ara, 4, 7
plot.10ara, 6, 8
predict, 2, 3, 7, 8, 10
predict.10ara, 6, 8
print.cv.10ara, 9
print.10ara, 10