The lars Package

September 17, 2004

Version 0.9-5

Date 2004-09-16

Title Least Angle Regression, Lasso and Forward Stagewise

Author Trevor Hastie <hastie@stanford.edu> and Brad Efron <brad@stat.stanford.edu>

Maintainer Trevor Hastie <hastie@stanford.edu>

Depends R

Description Efficient procedures for fitting an entire lasso sequence with the cost of a single least squares

License GPL version 2 or newer

URL http://www-stat.stanford.edu/~hastie/Papers/#LARS

R topics documented:

cv.lars ................................................................. 1
diabetes ............................................................. 2
lars-internal ....................................................... 3
lars ................................................................. 4
plot.lars ........................................................... 5
predict.lars ......................................................... 6

Index 8

---

cv.lars

Computes K-fold cross-validated error curve for lars

Description

Computes the K-fold cross-validated mean squared prediction error for lars, lasso, or forward stagewise.

Usage

cv.lars(x, y, K = 10, fraction = seq(from = 0, to = 1, length = 100),
          trace = FALSE, plot.it = TRUE, se = TRUE, ...)

---

1
Arguments

x  Input to lars
y  Input to lars
K  Number of folds
fraction  Abscissa values at which CV curve should be computed, as a fraction of the saturated |beta|. Default is seq(from = 0, to = 1, length = 100)
trace  Show computations?
plot.it  Plot it?
se  Include standard error bands?
...  Additional arguments to lars

Value

Invisibly returns a list with components (which can be plotted using plotCVlars)
fraction  As above
cv  The CV curve at each value of fraction
cv.error  The standard error of the CV curve

Author(s)

Trevor Hastie

References


Examples

data(diabetes)
attach(diabetes)
cv.lars(x2,y,trace=TRUE,max.steps=80)
detach(diabetes)

---

diabetes  

Blood and other measurements in diabetics

Description

The diabetes data frame has 442 rows and 3 columns. These are the data used in the Efron et al "Least Angle Regression" paper.

Format

This data frame contains the following columns:

x  a matrix with 10 columns
y  a numeric vector
x2  a matrix with 64 columns
Details

The x matrix has been standardized to have unit L2 norm in each column and zero mean. The matrix x2 consists of x plus certain interactions.

Source


References


Description

Internal lars functions

Usage

backsolvet(r,x,k = ncol(r))
cv.folds(n, folds = 10)
delcol(r, z, k = p)
downdateR (R, k = p)
error.bars(x, upper, lower, width = 0.02, ...)
nmls.lars(active, Sign, R, beta, Gram, eps = 1e-10, trace = FALSE, use.Gram = TRUE)
plotCVLars(cv.lars.object, se = TRUE)
updateR(xnew, R = NULL, xold, eps = .Machine$double.eps, Gram = FALSE)
.FIRST.lib(lib, pkg)

Details

These are not to be called by the user. backsolvet is included to make the R code compatible with the Splus code, since backsolve in R has a transpose=TRUE option already.

Author(s)

Trevor Hastie
Fits Least Angle Regression, Lasso and Infinitesimal Forward Stagewise regression models

Description

These are all variants of Lasso, and provide the entire sequence of coefficients and fits, starting from zero, to the least squares fit.

Usage

```r
lars(x, y, type = c("lasso", "lar", "forward.stagewise"),
    trace = FALSE, Gram, eps = .Machine$double.eps, max.steps, use.Gram = TRUE)
```

Arguments

- `x`: matrix of predictors
- `y`: response
- `type`: One of "lasso", "lar" or "forward.stagewise". The names can be abbreviated to any unique substring. Default is "lasso".
- `trace`: If TRUE, lars prints out its progress
- `Gram`: The X'X matrix; useful for repeated runs (bootstrap) where a large X'X stays the same.
- `eps`: An effective zero
- `max.steps`: Limit the number of steps taken; the default is $8 \times \min(m, n-1)$, with $m$ the number of variables, and $n$ the number of samples. For `type="lar"`, the maximum number of steps is $\min(m, n-1)$. For `type="lasso"` and especially `type="forward.stagewise"`, there can be many more terms, because although no more than $\min(m, n-1)$ variables can be active during any step, variables are frequently dropped and added as the algorithm proceeds. Although the default usually guarantees that the algorithm has proceeded to the saturated fit, users should check.
- `use.Gram`: When the number $m$ of variables is very large, i.e. larger than $N$, then you may not want LARS to precompute theGram matrix. Default is `use.Gram=TRUE`

Details

LARS is described in detail in Efron, Hastie, Johnstone and Tibshirani (2002). With the "lasso" option, it computes the complete lasso solution simultaneously for ALL values of the shrinkage parameter in the same computational cost as a least squares fit.

Value

A "lars" object is returned, for which print, plot and predict methods exist.

Author(s)

Brad Efron and Trevor Hastie
References


See Also

print, plot, and predict methods for lars, and cv.lars

Examples

data(diabetes)
par(mfrow=c(2,2))
attach(diabetes)
object <- lars(x,y)
plot(object)
object2 <- lars(x,y,type="lar")
plot(object2)
object3 <- lars(x,y,type="for") # Can use abbreviations
plot(object3)
detach(diabetes)

plot.lars

Plot method for lars objects

Description

Produce a plot of a lars fit. The default is a complete coefficient path.

Usage

plot.lars(x, xvar= c("norm", "df", "arc.length"), breaks = TRUE, plottype = c("coefficients", "Cp"), omit.zeros = TRUE, eps = 1e-10, ...)

Arguments

x lars object
xvar The type of x variable against which to plot. xvar=norm plots against the L1 norm of the coefficient vector (default). xvar=df plots against the step number (which is essentially degrees of freedom for LAR; not for LASSO or Forward Stagewise). xvar=arc.length plots against the arc.length of the fitted vector; this is useful for a LAR object, because the L1 norm of its coefficient vector need not be monotone in the steps.
breaks If TRUE, then vertical lines are drawn at each break point in the piecewise linear coefficient paths
plottype Either coefficients (default) or Cp. The coefficient plot shows the path of each coefficient as a function of the norm fraction or Df. The Cp plot shows the Cp curve.
omit.zeros When the number of variables is much greater than the number of observations, many coefficients will never be nonzero; this logical (default TRUE) avoids plotting these zero coefficients
predict.lars

Make predictions or extract coefficients from a fitted lars model

Description

While lars() produces the entire path of solutions, predict.lars allows one to extract a prediction at a particular point along the path.

Usage

predict.lars(object, newx, s, type = c("fit", "coefficients"), mode = c("step", "fraction", "norm"), ...)  
coef.lars(object, ...)

Arguments

object  
A fitted lars object

newx  
If type="fit", then newx should be the x values at which the fit is required.  
If type="coefficients", then newx can be omitted.

s  
a value, or vector of values, indexing the path. Its values depends on 
the mode argument. By default (mode="step"), s should take on values 
between 0 and p (e.g., a step of 1.3 means .3 of the way between step 1 
and 2.)

type  
If type="fit", predict returns the fitted values. If type="coefficients", predict returns the coefficients. Abbreviations allowed.
predict.lars

mode

Mode="step" means the s= argument indexes the lars step number, and the coefficients will be returned corresponding to the values corresponding to step s. If mode="fraction", then s should be a number between 0 and 1, and it refers to the ratio of the L1 norm of the coefficient vector, relative to the norm at the full LS solution. Mode="norm" means s refers to the L1 norm of the coefficient vector. Abbreviations allowed.

Any arguments for predict.lars should work for coef.lars

Details

LARS is described in detail in Efron, Hastie, Johnstone and Tibshirani (2002). With the "lasso" option, it computes the complete lasso solution simultaneously for ALL values of the shrinkage parameter in the same computational cost as a least squares fit.

Value

Either a vector/matrix of fitted values, or a vector/matrix of coefficients.

Author(s)

Trevor Hastie

References


See Also

print, plot, lars, cv.lars

Examples

data(diabetes)
attach(diabetes)
object <- lars(x,y,type="lasso")
### make predictions at the values in x, at each of the
### steps produced in object
fits <- predict.lars(object, x, type="fit")
### extract the coefficient vector with L1 norm=4.1
coef4.1 <- coef(object, s=4.1, mode="norm") # or
coef4.1 <- predict(object, s=4.1, type="coef", mode="norm")
detach(diabetes)
Index

*Topic datasets
diabetes, 2
*Topic hplot
plot.lars, 5
*Topic internal
lars-internal, 3
*Topic methods
plot.lars, 5
predict.lars, 6
*Topic regression
cv.lars, 1
lars, 4
predict.lars, 6
 .*First.lib (lars-internal), 3
backsolvset (lars-internal), 3
coeff.lars (predict.lars), 6
cv.folds (lars-internal), 3
cv.lars, 1
delcol (lars-internal), 3
diabetes, 2
downdateR (lars-internal), 3
error.bars (lars-internal), 3
lars, 4
lars-internal, 3
mls.lars (lars-internal), 3
plot.lars, 5
plotCVLars (lars-internal), 3
predict.lars, 6
print.lars (lars-internal), 3
updateR (lars-internal), 3