

# lasso

Regularized least-squares regression using lasso or elastic net algorithms

## Syntax

```
B = lasso(X,Y)
[B,FitInfo] = lasso(X,Y)
[B,FitInfo] = lasso(X,Y,Name,Value)
```

## Description

`B = lasso(X,Y)` returns fitted least-squares regression coefficients for a set of regularization coefficients [Lambda](#).

`[B,FitInfo] = lasso(X,Y)` returns a structure containing information about the fits.

`[B,FitInfo] = lasso(X,Y,Name,Value)` fits regularized regressions with additional options specified by one or more `Name,Value` pair arguments.

## Input Arguments

X	Numeric matrix with <i>n</i> rows and <i>p</i> columns. Each row represents one observation, and each column represents one predictor (variable).
Y	Numeric vector of length <i>n</i> , where <i>n</i> is the number of rows of <i>X</i> . <i>Y(i)</i> is the response to row <i>i</i> of <i>X</i> .

## Name-Value Pair Arguments

Specify optional comma-separated pairs of `Name,Value` arguments, where `Name` is the argument name and `Value` is the corresponding value. `Name` must appear inside single quotes ( ' '). You can specify several name and value pair arguments in any order as `Name1,Value1,...,NameN,ValueN`.

'Alpha'	Scalar value from 0 to 1 (excluding 0) representing the weight of lasso ( $L^1$ ) versus ridge ( $L^2$ ) optimization. <code>Alpha = 1</code> represents lasso regression, <code>Alpha</code> close to 0 approaches ridge regression, and other values represent elastic net optimization. See <a href="#">Definitions</a> . <b>Default:</b> 1
'CV'	Method lasso uses to estimate mean squared error: <ul style="list-style-type: none"><li>• <code>K</code>, a positive integer — lasso uses <code>K</code>-fold cross validation.</li><li>• <code>cvp</code>, a <a href="#">cvpartition</a> object — lasso uses the cross-validation method expressed in <code>cvp</code>. You cannot use a 'leaveout' partition with lasso.</li><li>• 'resubstitution' — lasso uses <i>X</i> and <i>Y</i> to fit the model and to estimate the mean squared error, without cross validation.</li></ul> <b>Default:</b> 'resubstitution'
'DFmax'	Maximum number of nonzero coefficients in the model. lasso returns results only for <code>Lambda</code> values that satisfy this criterion. <b>Default:</b> Inf

'Lambda'	<p>Vector of nonnegative Lambda values. See <a href="#">Definitions</a>.</p> <ul style="list-style-type: none"> <li>• If you do not supply Lambda, lasso calculates the largest value of Lambda that gives a nonnull model. In this case, LambdaRatio gives the ratio of the smallest to the largest value of the sequence, and NumLambda gives the length of the vector.</li> <li>• If you supply Lambda, lasso ignores LambdaRatio and NumLambda.</li> </ul> <p><b>Default:</b> Geometric sequence of NumLambda values, the largest just sufficient to produce <math>B = 0</math></p>
'LambdaRatio'	<p>Positive scalar, the ratio of the smallest to the largest Lambda value when you do not set Lambda.</p> <p>If you set LambdaRatio = 0, lasso generates a default sequence of Lambda values, and replaces the smallest one with 0.</p> <p><b>Default:</b> 1e-4</p>
'MCReps'	<p>Positive integer, the number of Monte Carlo repetitions for cross validation.</p> <ul style="list-style-type: none"> <li>• If CV is 'resubstitution' or a cvpartition of type 'resubstitution', MCReps must be 1.</li> <li>• If CV is a cvpartition of type 'holdout', MCReps must be greater than 1.</li> </ul> <p><b>Default:</b> 1</p>
'NumLambda'	<p>Positive integer, the number of Lambda values lasso uses when you do not set Lambda. lasso can return fewer than NumLambda fits if the if the residual error of the fits drops below a threshold fraction of the variance of <math>y</math>.</p> <p><b>Default:</b> 100</p>
'Options'	<p>Structure that specifies whether to cross validate in parallel, and specifies the random stream or streams. Create the Options structure with <a href="#">statset</a>. Option fields:</p> <ul style="list-style-type: none"> <li>• UseParallel — Set to 'always' to compute in parallel. Default is 'never'.</li> <li>• UseSubstreams — Set to 'always' to compute in parallel in a reproducible fashion. To compute reproducibly, set Streams to a type allowing substreams: 'mlfg6331_64' or 'mrg32k3a'. Default is 'never'.</li> <li>• Streams — A <a href="#">RandStream</a> object or cell array consisting of one such object. If you do not specify Streams, lasso uses the default stream.</li> </ul> <p>For details on using parallel computing, see <a href="#">Parallel Statistics</a>.</p>
'PredictorNames'	<p>Cell array of strings representing names of the predictor variables, in the order in which they appear in x.</p> <p><b>Default:</b> {}</p>
'RelTol'	<p>Convergence threshold for the coordinate descent algorithm (see Friedman, Tibshirani, and Hastie <a href="#">[3]</a>). The algorithm terminates when successive estimates of the coefficient vector differ in the <math>L^2</math> norm by a relative amount less than RelTol.</p> <p><b>Default:</b> 1e-4</p>

'Standardize'	Boolean value specifying whether <code>lasso</code> scales <code>x</code> and <code>y</code> before fitting the models. <code>lasso</code> always centers the data before fitting.  <b>Default:</b> <code>true</code>
'Weights'	Observation weights, a nonnegative vector of length <code>n</code> , where <code>n</code> is the number of rows of <code>x</code> . <code>lasso</code> scales <code>Weights</code> to sum to 1.  <b>Default:</b> <code>1/n * ones(n,1)</code>

## Output Arguments

<code>B</code>	Fitted coefficients, a <code>p</code> -by- <code>L</code> matrix, where <code>p</code> is the number of predictors (columns) in <code>x</code> , and <code>L</code> is the number of <code>Lambda</code> values.
<code>FitInfo</code>	Structure containing information about the model fits.

Field in <code>FitInfo</code>	Description
<code>Intercept</code>	Intercept term $\beta_0$ for each linear model, a 1-by- <code>L</code> vector
<code>Lambda</code>	<code>Lambda</code> parameters in ascending order, a 1-by- <code>L</code> vector
<code>Alpha</code>	Value of <code>Alpha</code> parameter, a scalar
<code>DF</code>	Number of nonzero coefficients in <code>B</code> for each value of <code>Lambda</code> , a 1-by- <code>L</code> vector
<code>MSE</code>	Mean squared error (MSE), a 1-by- <code>L</code> vector

If you set the `cv` name-value pair to cross validate, the `FitInfo` structure contains additional fields.

Field in <code>FitInfo</code>	Description
<code>SE</code>	The standard error of MSE for each <code>Lambda</code> , as calculated during cross validation, a 1-by- <code>L</code> vector
<code>LambdaMinMSE</code>	The <code>Lambda</code> value with minimum MSE, a scalar
<code>Lambda1SE</code>	The largest <code>Lambda</code> such that MSE is within one standard error of the minimum, a scalar
<code>IndexMinMSE</code>	The index of <code>Lambda</code> with value <code>LambdaMinMSE</code> , a scalar
<code>Index1SE</code>	The index of <code>Lambda</code> with value <code>Lambda1SE</code> , a scalar

## Definitions

### Lasso

For a given value of  $\lambda$ , a nonnegative parameter, `lasso` solves the problem

$$\min_{\beta_0, \beta} \left( \frac{1}{2N} \sum_{i=1}^N (y_i - \beta_0 - x_i^T \beta)^2 + \lambda \sum_{j=1}^p |\beta_j| \right),$$

where

- $N$  is the number of observations.
- $y_i$  is the response at observation  $i$ .
- $x_i$  is data, a vector of  $p$  values at observation  $i$ .
- $\lambda$  is a nonnegative regularization parameter corresponding to one value of  $\Lambda$ .
- The parameters  $\beta_0$  and  $\beta$  are scalar and  $p$ -vector respectively.

As  $\lambda$  increases, the number of nonzero components of  $\beta$  decreases.

The lasso problem involves the  $L^1$  norm of  $\beta$ , as contrasted with the elastic net algorithm.

### Elastic Net

For an  $\alpha$  strictly between 0 and 1, and a nonnegative  $\lambda$ , elastic net solves the problem

$$\min_{\beta_0, \beta} \left( \frac{1}{2N} \sum_{i=1}^N (y_i - \beta_0 - x_i^T \beta)^2 + \lambda P_\alpha(\beta) \right),$$

where

$$P_\alpha(\beta) = \frac{(1-\alpha)}{2} \|\beta\|_2^2 + \alpha \|\beta\|_1 = \sum_{j=1}^p \left( \frac{(1-\alpha)}{2} \beta_j^2 + \alpha |\beta_j| \right)$$

Elastic net is the same as lasso when  $\alpha = 1$ . As  $\alpha$  shrinks toward 0, elastic net approaches [ridge](#) regression. For other values of  $\alpha$ , the penalty term  $P_\alpha(\beta)$  interpolates between the  $L^1$  norm of  $\beta$  and the squared  $L^2$  norm of  $\beta$ .

## Examples

### Remove Redundant Predictors

Construct a data set with redundant predictors, and identify those predictors using cross-validated [lasso](#).

Create a matrix  $X$  of 100 five-dimensional normal variables and a response vector  $y$  from just two components of  $X$ , with small added noise.

```
X = randn(100,5);  
r = [0;2;0;-3;0]; % only two nonzero coefficients  
Y = X*r + randn(100,1)*.1; % small added noise
```

Construct the default lasso fit.

```
B = lasso(X,Y);
```

Find the coefficient vector for the 25th value in  $B$ .

```
B(:,25)  
  
ans =  
  
         0  
    1.6093  
         0  
   -2.5865  
         0
```

lasso identifies and removes the redundant predictors.

### Plot a Regularized Fit with Cross Validation

Visually examine the cross-validated error of various levels of regularization.

Load the acetylene data and prepare the data with interactions for fitting.

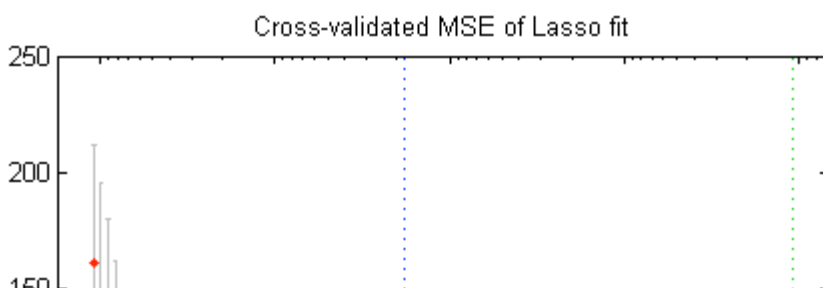
```
load acetylene  
Xs = [x1 x2 x3];  
X = x2fx(Xs, 'interaction');  
X(:,1) = []; % No constant term
```

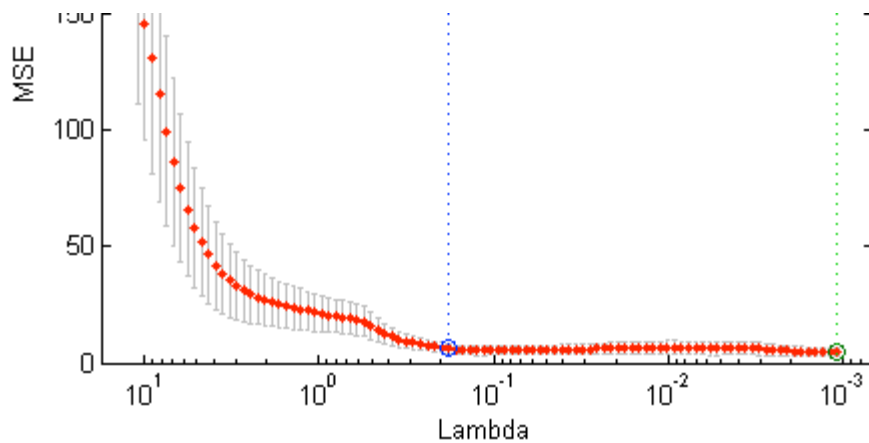
Construct the lasso fit using ten-fold cross validation. Include the `FitInfo` output so you can plot the result.

```
[B FitInfo] = lasso(X,y, 'CV',10);
```

Plot the cross-validated fits.

```
lassoPlot(B,FitInfo, 'PlotType', 'CV');
```





## References

- [1] Tibshirani, R. *Regression shrinkage and selection via the lasso*. Journal of the Royal Statistical Society, Series B, Vol 58, No. 1, pp. 267–288, 1996.
- [2] Zou, H. and T. Hastie. *Regularization and variable selection via the elastic net*. Journal of the Royal Statistical Society, Series B, Vol. 67, No. 2, pp. 301–320, 2005.
- [3] Friedman, J., R. Tibshirani, and T. Hastie. *Regularization paths for generalized linear models via coordinate descent*. Journal of Statistical Software, Vol 33, No. 1, 2010.  
<http://www.jstatsoft.org/v33/i01>
- [4] Hastie, T., R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning*, 2nd edition. Springer, New York, 2008.

## See Also

[lassoPlot](#) | [ridge](#)

## How To

- [Lasso and Elastic Net](#)

Was this topic helpful?

Yes

No