

# Package ‘L0Learn’

June 20, 2018

**Type** Package

**Title** Fast Algorithms for Best Subset Selection

**Version** 0.2-5

**Date** 2018-05-04

**Author** Hussein Hazimeh  
Rahul Mazumder

**Maintainer** Hussein Hazimeh <hazimeh@mit.edu>

**Description** Highly optimized coordinate descent and local combinatorial search algorithms for (ap-  
proximately) solving L0-regularized learning problems.

**License** MIT + file LICENSE

**Imports** Rcpp (>= 0.12.13), Matrix, methods

**LinkingTo** Rcpp, RcppArmadillo

**RoxygenNote** 6.0.1

## R topics documented:

coef . . . . .	2
L0Learn.cvfit . . . . .	2
L0Learn.fit . . . . .	5
plot.L0Learn . . . . .	7
predict . . . . .	7
print . . . . .	8
<b>Index</b>	<b>9</b>

---

coef	<i>Extract Solutions</i>
------	--------------------------

---

### Description

Extracts a specific solution in the regularization path

### Usage

```
coef(object, ...)

## S3 method for class 'L0Learn'
coef(object, lambda, gamma = 0, ...)
```

### Arguments

object	The output of L0Learn.fit
...	ignore
lambda	The value(s) of lambda at which to extract the solution.
gamma	The value of gamma at which to extract the solution. Note that, unlike lambda, this can only take single values.

---

L0Learn.cvfit	<i>Cross Validation</i>
---------------	-------------------------

---

### Description

Fits an L0 model on the full data and performs K-fold cross-validation.

### Usage

```
L0Learn.cvfit(X, y, Loss = "SquaredError", Penalty = "L0",
  Algorithm = "CD", MaxSuppSize = 100, NLambda = 100, NGamma = 10,
  GammaMax = 10, GammaMin = 1e-04, PartialSort = TRUE, MaxIters = 200,
  Tol = 1e-06, ActiveSet = TRUE, ActiveSetNum = 3, MaxSwaps = 100,
  ScaleDownFactor = 0.8, ScreenSize = 1000, AutoLambda = TRUE,
  LambdaGrid = list(), Nfolds = 10, Seed = 1)
```

**Arguments**

X	The data matrix.
y	The response vector.
Loss	The loss function to be minimized. The currently supported choice is "Squared-Error".
Penalty	The type of regularization. This can take either one of the following choices: "L0", "L0L2", and "L0L1".
Algorithm	The type of algorithm used to minimize the objective. Currently "CD" and "CDPSI" are supported. "CD" is a variant of cyclic coordinate descent and can run very fast. "CDPSI" performs local combinatorial search on top of CD and thus can achieve higher quality solutions (at the expense of increased running time).
MaxSuppSize	The maximum support size to reach in the grid before termination. We recommend setting this to a small fraction of $\min(n,p)$ (e.g. $0.05 * \min(n,p)$ ) as L0 regularization typically selects a small portion of non-zeros.
NLambda	The number of Lambda values to select (recall that Lambda is the regularization parameter corresponding to the L0 norm).
NGamma	The number of Gamma values to select (recall that Gamma is the regularization parameter corresponding to L1 or L2, depending on the chosen penalty).
GammaMax	The maximum value of Gamma when using the L0L2 penalty. For the L0L1 penalty this is automatically selected by the toolkit.
GammaMin	The minimum value of Gamma when using the L0L2 penalty. For the L0L1 penalty, GammaMin specifies the fraction of GammaMax at which the grid ends.
PartialSort	If TRUE partial sorting will be used for sorting the coordinates (see our paper for details). Otherwise, full sorting is used.
MaxIters	The maximum number of iterations (full cycles) for CD per grid point.
Tol	The tolerance which decides when to terminate CD (based on the relative change in the objective).
ActiveSet	If TRUE, performs active set updates.
ActiveSetNum	The number of consecutive times a support should appear before declaring support stabilization.
MaxSwaps	The maximum number of swaps used by CDPSI for each grid point.
ScaleDownFactor	This parameter decides how close the selected Lambda values are. The choice should be between strictly between 0 and 1 (i.e., 0 and 1 are not allowed). For details, see our paper - Section 5 on Adaptive Selection of Tuning Parameters).
ScreenSize	The number of coordinates to cycle over when performing correlation screening.
AutoLambda	If FALSE, the user specifies a grid of Lambda0 values through the Lambda0Grid parameter. Otherwise, if TRUE, the values of Lambda0 are automatically selected based on the data.
LambdaGrid	A vector of Lambda0 values to use in computing the regularization path. This is ignored unless AutoLambda0 = FALSE.
Nfolds	The number of folds for cross-validation.
Seed	The seed used in randomly shuffling the data for cross-validation.

**Value**

An S3 object of type "LOLearn" describing the regularization path. The object has the following members.

`cvmeans`

For L0, this is a sequence of cross-validation errors: `cvmeans[i]` corresponds to the solution indexed by `lambda[i]`. For L0L1 and L0L2, `cvmeans` is a list, where each element is a sequence corresponding to a particular gamma, i.e., `cvmeans[[i]]` is the sequence of cross-validation errors corresponding to `gamma[i]`.

`cvstds`

For L0, this is a sequence of standard deviations for the cross-validation errors. For L0L1 and L0L2, it is a list of sequences: `cvstds[[i]]` corresponds to `cvmeans[[i]]`.

`a0`

For L0, this is a sequence of intercepts. Note for L0L1 and L0L2, `a0` is a list of intercept sequences, where each member of the list corresponds to a single gamma value.

`beta`

For L0, this is a matrix of coefficients of dimensions  $p \times \text{length}(\text{lambda})$ , where each column corresponds to a single lambda value. For L0L1 and L0L2, this is a list of coefficient matrices, where each matrix corresponds to a single gamma value.

`lambda`

For L0, `lambda` is a sequence of lambda values. For L0L1 and L0L2, it is a list of lambda sequences, each corresponding to a single gamma value.

`gamma`

For L0L1 and L0L2, this is a sequence of gamma values.

`suppsize`

For L0, this is a sequence of support sizes (number of non-zero coefficients). For L0L1 and L0L2, it is a list of support size sequences, each representing a single gamma value.

`converged`

For L0, this is a sequence indicating whether the algorithm converged at the current point in the regularization path. For L0L1 and L0L2, this is a list of sequences, each representing a single gamma value.

L0Learn.fit

*Fit an L0-regularized model***Description**

Computes the regularization path for the specified loss function and choice of regularization (which can be a combination of the L0, L1, and L2 (squared) norms).

**Usage**

```
L0Learn.fit(X, y, Loss = "SquaredError", Penalty = "L0", Algorithm = "CD",
  MaxSuppSize = 100, NLambda = 100, NGamma = 10, GammaMax = 10,
  GammaMin = 1e-04, PartialSort = TRUE, MaxIters = 200, Tol = 1e-06,
  ActiveSet = TRUE, ActiveSetNum = 3, MaxSwaps = 100,
  ScaleDownFactor = 0.8, ScreenSize = 1000, AutoLambda = TRUE,
  LambdaGrid = list(0))
```

**Arguments**

X	The data matrix.
y	The response vector.
Loss	The loss function to be minimized. The currently supported choice is "Squared-Error".
Penalty	The type of regularization. This can take either one of the following choices: "L0", "L0L2", and "L0L1".
Algorithm	The type of algorithm used to minimize the objective. Currently "CD" and "CDPSI" are supported. "CD" is a variant of cyclic coordinate descent and can run very fast. "CDPSI" performs local combinatorial search on top of CD and thus can achieve higher quality solutions (at the expense of increased running time).
MaxSuppSize	The maximum support size to reach in the grid before termination. We recommend setting this to a small fraction of $\min(n,p)$ (e.g. $0.05 * \min(n,p)$ ) as L0 regularization typically selects a small portion of non-zeros.
NLambda	The number of Lambda values to select (recall that Lambda is the regularization parameter corresponding to the L0 norm).
NGamma	The number of Gamma values to select (recall that Gamma is the regularization parameter corresponding to L1 or L2, depending on the chosen penalty).
GammaMax	The maximum value of Gamma when using the L0L2 penalty. For the L0L1 penalty this is automatically selected by the toolkit.
GammaMin	The minimum value of Gamma when using the L0L2 penalty. For the L0L1 penalty, GammaMin specifies the fraction of GammaMax at which the grid ends.
PartialSort	If TRUE partial sorting will be used for sorting the coordinates (see our paper for details). Otherwise, full sorting is used.
MaxIters	The maximum number of iterations (full cycles) for CD per grid point.

Tol	The tolerance which decides when to terminate CD (based on the relative change in the objective).
ActiveSet	If TRUE, performs active set updates.
ActiveSetNum	The number of consecutive times a support should appear before declaring support stabilization.
MaxSwaps	The maximum number of swaps used by CDPSI for each grid point.
ScaleDownFactor	This parameter decides how close the selected Lambda values are. The choice should be between strictly between 0 and 1 (i.e., 0 and 1 are not allowed). For details, see our paper - Section 5 on Adaptive Selection of Tuning Parameters).
ScreenSize	The number of coordinates to cycle over when performing correlation screening.
AutoLambda	If FALSE, the user specifies a grid of Lambda0 values through the Lambda0Grid parameter. Otherwise, if TRUE, the values of Lambda0 are automatically selected based on the data.
LambdaGrid	A vector of Lambda0 values to use in computing the regularization path. This is ignored unless AutoLambda0 = FALSE.

### Value

An S3 object of type "L0Learn" describing the regularization path. The object has the following members.

`a0`

For L0, this is a sequence of intercepts. Note for L0L1 and L0L2, `a0` is a list of intercept sequences, where each member of the list corresponds to a single gamma value.

`beta`

For L0, this is a matrix of coefficients of dimensions  $p \times \text{length}(\text{lambda})$ , where each column corresponds to a single lambda value. For L0L1 and L0L2, this is a list of coefficient matrices, where each matrix corresponds to a single gamma value.

`lambda`

For L0, `lambda` is a sequence of lambda values. For L0L1 and L0L2, it is a list of lambda sequences, each corresponding to a single gamma value.

`gamma`

For L0L1 and L0L2, this is a sequence of gamma values.

`suppsize`

For L0, this is a sequence of support sizes (number of non-zero coefficients). For L0L1 and L0L2, it is a list of support size sequences, each representing a single gamma value.

`converged`

For L0, this is a sequence indicating whether the algorithm converged at the current point in the regularization path. For L0L1 and L0L2, this is a list of sequences, each representing a single gamma value.

---

plot.L0Learn	<i>Plot Cross-validation Errors</i>
--------------	-------------------------------------

---

**Description**

Plots cross-validation errors

**Usage**

```
## S3 method for class 'L0Learn'
plot(x, gamma)
```

**Arguments**

x	L0Learn.fit object
gamma	The gamma value for L0L1 and L0L2 models. This is ignored for L0.

---

predict	<i>Predict Response</i>
---------	-------------------------

---

**Description**

Predicts the response for a given sample

**Usage**

```
predict(object, ...)

## S3 method for class 'L0Learn'
predict(object, newx, lambda, gamma = 0, ...)
```

**Arguments**

object	The output of L0Learn.fit
...	ignore
newx	A matrix on which predictions are made. The matrix should have p columns.
lambda	The value(s) of lambda to use for prediction. A summary of the lambdas in the regularization path can be obtained using <code>print(fit)</code> .
gamma	The value of gamma to use for prediction. A summary of the gammas in the regularization path can be obtained using <code>print(fit)</code> .

---

print	<i>Print L0Learn.fit object</i>
-------	---------------------------------

---

**Description**

Prints a summary of L0Learn.fit

**Usage**

```
print(x, ...)  
  
## S3 method for class 'L0Learn'  
print(x, ...)
```

**Arguments**

x	L0Learn.fit object
...	ignore



# Index

`coef`, [2](#)

`L0Learn.cvfit`, [2](#)

`L0Learn.fit`, [5](#)

`plot.L0Learn`, [7](#)

`predict`, [7](#)

`print`, [8](#)