



Statistics/Data Analysis
 User: Jhersh
 Project: Rlasso

help Rlasso

Title

> Net regression from within Stata

Syntax

Rlasso [varlist] [weight] [if] [in] , options]

options	Description
Main	
lambda(string)	specify either lambda(min) , to estimate a model using the lambda that minimizes the cross-validated error or lambda(lse) , to use the cross-validated minimum plus 1 standard error of the minimum. Default is lse .
model(string)	choose to fit a linear or logistic model; default is model(linear) .
nfolds(integer)	is the number of folds for cross-validation; default is nfolds(10) , minimum is 3.
penalty	choose penalty(0) to estimate a Ridge model, penalty(1) to estimate a Lasso model, and penalty(#) with $0 < \# < 1$ to estimate an Elastic-Net model; default is penalty(1) .
intercept	specifies whether the estimated model includes an intercept (default=TRUE) or to zero.
rpath	specifies the path name to the R executable file.

Description

Allows estimation of Lasso, Ridge and Elastic-Net regularized regression models using the R package `glmnet`.

Lasso, Ridge and Elastic-Net regularized regression models minimize the OLS sum of squared errors plus an additional penalty term that constrains the size of the regression coefficients. Shrinking the coefficients towards zero balances in-sample and out-of-sample error via the bias-variance tradeoff mechanism. For problems of prediction, which are susceptible to overfitting, this can improve prediction accuracy.

The parameter lambda controls the weight given to the Bayesian shrinkage penalty term, and indicates the severity of Bayesian shrinkage applied. A larger value of lambda results in a more parsimonious model. The optimal lambda is selected through cross-validation and applies the parameterization of lambda for which the cross-validated error is smallest. Finally, the model is re-fit using all of the available observations and the selected parameterization of lambda. For more theoretical detail we refer the reader to Friedman, Hastie and Tibshirani (2008).

Options

Main

- lambda(string)** Specifies how the optimal Bayesian shrinkage parameter lambda should be selected: either **lambda(min)**, the value of lambda that gives the minimum cross-validation error or **lambda(lse)**, the value of lambda that minimizes cross-validated error plus one standard error. Default is **lambda(lse)**. For more on why **lambda(lse)** is the default see Friedman, Hastie and Tibshirani (2008).
- model(string)** Choose to fit a linear or logistic model; default is linear.
- nfolds(integer)** The error measure used for cross-validation is mean squared-error for gaussian regression and deviance for logistic regression.

penalty Choose from `penalty(0)` to estimate a ridge model, `penalty(1)` to estimate a lasso model, `penalty(#)` with $0 < \# < 1$ to estimate an elasticnet model; default is lasso

intercept Fit a model with `intercept(default=TRUE)` or set to zero to estimate a model without an intercept.

intercept(string) Specifies the R executable path. If the `rpath()` option is left unspecified, it is set to the value of the global macro `Rterm_path`, if that macro has been specified. For more discussion, see the technical note in the Rsource help file.

Dependencies

R -- The R software system must be installed on the user's system in order for **Rlasso** to run. You can download R from The Comprehensive R Archive Network (<https://cran.r-project.org/>)

rsource -- Additionally, requires Rsource to be able to run an R source file in batch mode from Stata.

Examples

Example 1: Estimating a model of the price of cars.

```
sysuse auto, replace <---- First, load everyone's favorite dataset.
```

Use **Rlasso** to estimate a Lasso model using 5 fold cross validation.

```
gen lnprice = log(price)
```

```
Rlasso lnprice mpg rep78 headroom trunk weight length turn displacement gear_ratio foreign, nfold(5)
```

Example 2: Predicting low birth weight using a logit model with ridge penalty, reporting coefficient estimates and the MSE minimizing lambda value.

```
use http://www.stata-press.com/data/r13/lbw.dta, clear
```

```
Rlasso low age lwt race smoke ptl ht ui ftv, model(logit) lambda(min) penalty(0)
```

Saved Results

Rlasso displays all results to the results window.

Installation Instructions

1. Ensure you have R executable installed. If you don't have this installed, you can download it from the software from <https://www.r-project.org/>.

2. Install the Stata dependency Rsource in Stata. `ssc install Rsource`

3. If you have a windows machine, the location of the Rterm executable file must be manually set using the global `Rterm_path` or the option `rpath`. First find the location of the Rterm executable file on your machine. On our test PC this is installed in the folder "C:\Program Files\R\R-3.1.2\bin\i386\Rterm.exe". Set the path to this executable using `Rterm_path` using "global Rterm_path [path]" or via the `rpath` option.

Authors

Jonathan Hersh, Poverty Global Practice, World Bank, Washington, D.C. Ana Areias, Poverty Global Practice, World Bank, Washington, D.C.

Email jhersh@worldbank.org Email ana.areias@gmail.com

References

Friedman, J., Hastie, T. and Tibshirani, R. (2008) *Regularization Paths for Generalized Linear Models via Coordinate Descent*

Rsource