Package: fasterElasticNet (via r-universe)

September 13, 2024

Type Package

Title An Amazing Fast Way to Fit Elastic Net

Version 1.1.2

Date 2018-08-01

Description Fit Elastic Net, Lasso, and Ridge regression and do
cross-validation in a fast way. We build the algorithm based on
Least Angle Regression by Bradley Efron, Trevor Hastie, Iain
Johnstone, etc. (2004)(<doi:10.1214/009053604000000067 >) and
some algorithms like Givens rotation and Forward/Back
Substitution. In this way, many matrices to be computed are
retained as triangular matrices which can eventually speed up
the computation. The fitting algorithm for Elastic Net is
written in C++ using Armadillo linear algebra library.

Depends R (>= 3.1.0)

License GPL (>= 2)

Imports Rcpp (>= 0.12.16)

LinkingTo Rcpp, RcppArmadillo

Suggests knitr, rmarkdown

URL https://github.com/CUFESAM/Elastic-Net

BugReports https://github.com/CUFESAM/Elastic-Net/issues

Repository https://cufesam.r-universe.dev

RemoteUrl https://github.com/cufesam/elastic-net

RemoteRef HEAD

RemoteSha e4d85dd7a617073545e9d62957367d6443c8ca8d

Contents

---

fasterElasticNet-package

Fitting ElasticNet in a fast way.

Description

FasterElasticNet uses some math algorithm such as cholesky decomposition and forward solve etc. to reduce the amount of computation. We also use Rcpp with Armadillo to improve our algorithm by speeding up almost 5 times compared by the R version.

Details

To use fasterElasticNet, dataset x(mxn) and y(mx1) should be put into the function to fit the model. Then, a completely trace of lambda1 and lambda2 can be computed if no lambda1 and lambda2 were input by using ElasticNet. Using cv.choosemodel with the number of folds will returns a best model with smallest MSE after cross-validation. Using output to print the output and predict function will return the prediction based on a new dataset.

Author(s)

Jingyi Ma

Maintainer: Linyu Zuo <zuozhe5959@gmail.com>

References

BRADLEY, EFRON, TREVOR, HASTIE, IAIN, JOHNSTONE, AND, ROBERT, TIBSHIRANI. LEAST ANGLE REGRESSION[J]. The Annals of Statistics, 2004, 32(2): 407-499

See Also

https://github.com/CUFESAM/Elastic-Net

Examples

```
#Use R built-in datasets mtcars for a model fitting
x <- mtcars[,-1]
y <- mtcars[, 1]

#fit model
model <- ElasticNetCV(x,y)

#fit a elastic net with lambda2 = 1
model$Elasticnet_(lambda2 = 1)

#choose model using cv
model$cv.choosemodel(k = 31)     #Leave-one-out cross validation
model$output()       #See the output
```

```
#predict
pre <- mtcars[1:3,-1]
model$predict(pre)
```

---

elasticnet   *A fast way fitting elastic net using RcppArmadillo*

Description

Elastic net is a regularization and variable selection method which linearly combines the L1 penalty of the lasso and L2 penalty of ridge methods. Based on this method, elastic- net is designed to return the trace of finding the best linear regression model. Compared with the existed R version of ElasticNet, our version speeds up the algorithm by using Cholesky decomposition, Givens rotation and RcppArmadillo.

Usage

```
elasticnet(XTX, XTY, lam2, lam1 = -1)
```

Arguments

XTX	The product of the transpose of independent variable X and itself.
XTY	The product of the transpose of independent variable X and response variable Y
lam1	Penalty of L1-norm. No L1 penalty when lam1 = -1
lam2	Penalty of L2-norm, a hyper-paramater

Details

When only lambda2 is given, elasticnet will return the trace of variable selection with lambda1 decreasing from lambda1_0 to zero. lambda1_0 is a value for lambda1 when there is only one predictor (the one most correlated with the response variable) in the model.

If lambda1 and lambda2 are both given, it will also return a trace. But in this case, the trace will stop when lambda1 and lambda2 reach the given ones.

To speed up the algorithm, we use some calculational tricks:

In the consideration of the low efficiency of R dealing with high-dimensional matrix, we use lower triangular matrices during the iteration of the algorithm to avoid massive matrix calculations. When adding one predictor into the model, we update XTX by recalcuting the lower triangular matrix in the Cholesky decomposition of it. While re- moving one predictor from the model, we update the lower triangular matrix with the help of Givens rotations.

Furthermore, due to the low efficiency of R dealing with loops, we rewrite the entire algorithm with RcppArmadillo, a C++ linear algebra library.

Value

A list will be returned. When only lambda2 is given, the returned list contains the trace of lambda1
(relamb) and the corresponding coefficients of the predictors (reb). If both lambda1 and lambda2
are given, the corresponding coefficients of the predictors will be returned.

Examples

```
#Use R built-in datasets mtcars for a model fitting
x <- as.matrix(mtcars[,-1])
y <- as.matrix(mtcars[, 1])

XTX <- t(x) %*% x
XTY <- t(x) %*% y

#Prints the output of elastic net model with lambda2 = 0
res <- elasticnet(XTX,XTY,lam2 = 0)
```

ElasticNetCV   *Cross validation*

Description

Computes k-fold cross-validation for elastic net.

Usage

```
ElasticNetCV(x, y)
```

Arguments

x	A data.frame or matrix of predictors
y	A vector of response variables

Details

This function reads data into its environment and returns a list of three outcomes. To perform
elastic net or cross-validation of elastic net, use the corresponding element of the returned list.
See examples below. The penalty of L1-norm and L2-norm is denoted by lambda1 and lambda2
respectively.

Value

cv.choosemodel Given the parameter k folds and lambda2 (optional), cv.choosemodel performs
cross-validation to select the opti- mal lambda1 and computes the corresponding
coefficient of each variable. If lambda2 is NULL, cv.choosemodel selects the
optimal lambda2 from a sequence going from 0 to 1 in steps of 0.1 and the
corresponding optimal lambda1, then it returns the coefficient of each variable.

A list of three outcomes will be returned:

Elasticnet	Given lambda1 (optional) and lambda2, Elasticnet_ calculates an elastic net-regularized regression and returns the coefficients of each variable. If lambda1 is NULL, Elasticnet_ prints out the trace of lambda1 and the corresponding coefficient of each variable.
output	Prints the cross-validation outputs, including the minimum MSE, the coefficient of each variable, lambda1 and lambda2.
predict	Reads a data.frame of the testing data set and returns predictions using the trained model.

Examples

```
#Use R built-in datasets mtcars for a model fitting
x <- mtcars[,-1]
y <- mtcars[, 1]

#fit model
model <- ElasticNetCV(x,y)

#fit a elastic net with lambda2 = 1
model$Elasticnet_(lambda2 = 1)

#choose model using cv
model$cv.choosemodel(k = 31)    #Leave-one-out cross validation
model$output()      #See the output

#predict
pre <- mtcars[1:3,-1]
model$predict(pre)
```

---

housing	*Housing data from kaggle*

Description

A subdata from kaggle "Get start" competition

Usage

```
data("housing")
```

Format

A data frame with 10153 observations on the following 140 variables.

floor for apartments, floor of the building

area_m Area, sq.m.

green_zone_part Proportion of area of greenery in the total area

indust_part Share of industrial zones in area of the total area

`preschool_quota` Number of seats in pre-school organizations

`preschool_education_centers_raion` Number of pre-school institutions

`school_quota` Number of high school seats in area

`school_education_centers_raion` Number of high school institutions

`school_education_centers_top_20_raion` Number of high schools of the top 20 best schools
in Moscow

`healthcare_centers_raion` Number of healthcare centers in district

`university_top_20_raion` Number of higher education institutions in the top ten ranking of the
Federal rank

`sport_objects_raion` Number of higher education institutions

`additional_education_raion` Number of additional education organizations

`culture_objects_top_25_raion` Number of objects of cultural heritage

`shopping_centers_raion` Number of malls and shopping centres in district

`office_raion` Number of malls and shopping centres in district

`build_count_block` Share of block buildings

`build_count_wood` Share of wood buildings

`build_count_frame` Share of frame buildings

`build_count_brick` Share of brick buildings

`build_count_monolith` Share of monolith buildings

`build_count_panel` Share of panel buildings

`build_count_foam` Share of foam buildings

`build_count_slag` Share of slag buildings

`build_count_before_1920` Share of before_1920 buildings

`build_count_1921.1945` Share of 1921-1945 buildings

`build_count_1946.1970` Share of 1946-1970 buildings

`build_count_1971.1995` Share of 1971-1995 buildings

`build_count_after_1995` Share of after_1995 buildings

`kindergarten_km` Distance to kindergarten

`school_km` Distance to high school

`park_km` Distance to park

`green_zone_km` Distance to green zone

`industrial_km` Distance to industrial zone

`water_treatment_km` Distance to water treatment

`cemetery_km` Distance to the cemetery

`incineration_km` Distance to the incineration

`railroad_station_walk_min` Time to the railroad station (walk)

`railroad_station_avto_km` Distance to the railroad station (avto)

`railroad_station_avto_min` Time to the railroad station (avto)

`public_transport_station_min_walk` Time to the public transport station (walk)

`water_km` Distance to the water reservoir / river

`mkad_km` Distance to MKAD (Moscow Circle Auto Road)

`big_road1_km` Distance to Nearest major road

`big_road2_km` The distance to next distant major road

`railroad_km` Distance to the railway / Moscow Central Ring / open areas Underground

`bus_terminal_avto_km` Distance to bus terminal (avto)

`oil_chemistry_km` Distance to dirty industries

`nuclear_reactor_km` Distance to nuclear reactor

`radiation_km` Distance to burial of radioactive waste

`power_transmission_line_km` Distance to power transmission line

`thermal_power_plant_km` Distance to thermal power plant

`ts_km` Distance to power station

`big_market_km` Distance to grocery / wholesale markets

`market_shop_km` Distance to markets and department stores

`fitness_km` Distance to fitness

`swim_pool_km` Distance to swimming pool

`ice_rink_km` Distance to ice palace

`stadium_km` Distance to stadium

`basketball_km` Distance to the basketball courts

`hospice_morgue_km` Distance to hospice/morgue

`detention_facility_km` Distance to detention facility

`public_healthcare_km` Distance to public healthcare

`university_km` Distance to universities

`workplaces_km` Distance to workplaces

`shopping_centers_km` Distance to shopping centers

`office_km` Distance to business centers/ offices

`additional_education_km` Distance to additional education

`preschool_km` Distance to preschool education organizations

`big_church_km` Distance to large church

`church_synagogue_km` Distance to Christian chirches and Synagogues

`mosque_km` Distance to mosques

`theater_km` Distance to theater

`museum_km` Distance to museums

`exhibition_km` Distance to exhibition

`catering_km` Distance to catering

`green_part_500` The share of green zones in 500 meters zone

`prom_part_500` The share of industrial zones in 500 meters zone

`office_count_500` The number of office space in 500 meters zone

`office_sqm_500` The square of office space in 500 meters zone

`trc_count_500` The number of shopping malls in 500 meters zone

`trc_sqm_500` The square of shopping malls in 500 meters zone

`cafe_count_500_na_price` Cafes and restaurant bill N/A in 500 meters zone

`cafe_count_500_price_500` Cafes and restaurant bill, average under 500 in 500 meters zone

`cafe_count_500_price_1000` Cafes and restaurant bill, average 500-1000 in 500 meters zone

`cafe_count_500_price_1500` Cafes and restaurant bill, average 1000-1500 in 500 meters zone

`cafe_count_500_price_2500` Cafes and restaurant bill, average 1500-2500 in 500 meters zone

`cafe_count_500_price_4000` Cafes and restaurant bill, average 2500-4000 in 500 meters zone

`cafe_count_500_price_high` Cafes and restaurant bill, average over 4000 in 500 meters zone

`big_church_count_500` The number of big churchs in 500 meters zone

`church_count_500` The number of churchs in 500 meters zone

`mosque_count_500` The number of mosques in 500 meters zone

`leisure_count_500` The number of leisure facilities in 500 meters zone

`sport_count_500` The number of sport facilities in 500 meters zone

`market_count_500` The number of markets in 500 meters zone

`green_part_1000` The share of green zones in 1000 meters zone

`prom_part_1000` The share of industrial zones in 1000 meters zone

`office_sqm_1000` The square of office space in 1000 meters zone

`trc_count_1000` The number of shopping malls in 1000 meters zone

`trc_sqm_1000` The square of shopping malls in 1000 meters zone

`cafe_count_1000_na_price` Cafes and restaurant bill N/A in 1000 meters zone

`cafe_count_1000_price_high` Cafes and restaurant bill, average over 4000 in 1000 meters zone

`big_church_count_1000` The number of big churchs in 1000 meters zone

`mosque_count_1000` The number of mosques in 1000 meters zone

`leisure_count_1000` The number of leisure facilities in 1000 meters zone

`sport_count_1000` The number of sport facilities in 1000 meters zone

`market_count_1000` The number of markets in 1000 meters zone

`green_part_1500` The share of green zones in 1500 meters zone

`prom_part_1500` The share of industrial zones in 1500 meters zone

`office_sqm_1500` The square of office space in 1500 meters zone

`trc_count_1500` The number of shopping malls in 1500 meters zone

`trc_sqm_1500` The square of shopping malls in 1500 meters zone

`cafe_count_1500_price_high` Cafes and restaurant bill, average over 4000 in 1500 meters zone

`mosque_count_1500` The number of mosques in 1500 meters zone

`sport_count_1500` The number of sport facilities in 1500 meters zone

`market_count_1500` The number of markets in 1500 meters zone

`green_part_2000` The share of green zones in 2000 meters zone

`prom_part_2000` The share of industrial zones in 2000 meters zone

`office_sqm_2000` The square of office space in 2000 meters zone

`trc_count_2000` The number of shopping malls in 2000 meters zone

`trc_sqm_2000` The square of shopping malls in 2000 meters zone

`mosque_count_2000` The number of mosques in 2000 meters zone

`sport_count_2000` The number of sport facilities in 2000 meters zone

`market_count_2000` The number of markets in 2000 meters zone

`green_part_3000` The share of green zones in 3000 meters zone

`prom_part_3000` The share of industrial zones in 3000 meters zone

`office_sqm_3000` The square of office space in 3000 meters zone

`trc_count_3000` The number of shopping malls in 3000 meters zone

`trc_sqm_3000` The square of shopping malls in 3000 meters zone

`mosque_count_3000` The number of mosques in 3000 meters zone

`sport_count_3000` The number of sport facilities in 3000 meters zone

`market_count_3000` The number of markets in 3000 meters zone

`green_part_5000` The share of green zones in 5000 meters zone

`prom_part_5000` The share of industrial zones in 5000 meters zone

`trc_count_5000` The number of shopping malls in 5000 meters zone

`trc_sqm_5000` The square of shopping malls in 5000 meters zone

`mosque_count_5000` The number of mosques in 5000 meters zone

`sport_count_5000` The number of sport facilities in 5000 meters zone

`market_count_5000` The number of markets in 5000 meters zone

`price_doc` I don't know

Source

www.kaggle.com

Examples

```
data(housing)
```

Index