Package: lassopv (via r-universe)

October 23, 2024

Type Package

Title Nonparametric P-Value Estimation for Predictors in Lasso

Version 0.2.1

Date 2019-12-03

Author Lingfei Wang <Lingfei.Wang.github@outlook.com>

Maintainer Lingfei Wang <Lingfei.Wang.github@outlook.com>

Description Estimate the p-values for predictors x against target variable y in lasso regression, using the regularization strength when each predictor enters the active set of regularization path for the first time as the statistic. This is based on the assumption that predictors (of the same variance) that (first) become active earlier tend to be more significant. Three null distributions are supported: normal and spherical, which are computed separately for each predictor and analytically under approximation, which aims at efficiency and accuracy for small p-values.

URL https://github.com/lingfeiwang/lassopv

License GPL-3

Copyright Copyright 2016-2019 Lingfei Wang

Depends R (>= 2.10)

Imports lars, stats

NeedsCompilation no

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

RemoteUrl https://github.com/lingfeiwang/lassopv

RemoteRef HEAD

RemoteSha 0d828c0715e1379d60e536b4fd0c868beccf02c0

Contents

`lassopv-package` *Nonparametric P-Value Estimation for Predictors in Lasso*

Description

Estimate the p-values for predictors x against target variable y in lasso regression, using the regularization strength when each predictor enters the active set of regularization path for the first time as the statistic. This is based on the assumption that predictors (of the same variance) that (first) become active earlier tend to be more significant. Three null distributions are supported: normal and spherical, which are computed separately for each predictor and analytically under approximation, which aims at efficiency and accuracy for small p-values.

Details

This R package provides a simple and efficient method to estimate the p-value of every predictor on a given target variable. The method is based on lasso regression and compares when every predictor enters the active set of the regulatization path against a normally distributed null predictor. The null distribution is computed analytically under approximation, whose errors are small for significant predictors. The whole computation only requires a single lasso regression over the regularization path, and is capable of analyzing high dimensional datasets.

Author(s)

Lingfei Wang <Lingfei.Wang.github@outlook.com>

References

Lingfei Wang and Tom Michoel, Comparable variable selection with lasso, https://arxiv.org/pdf/1701.07011. 2017, 2018.

Examples

```
library(lars)
library(lassopv)
data(diabetes)
attach(diabetes)
pv=lassopv(x,y)
```

lassopv *Estimation of Nonparametric P-Value Estimation for Predictors in Lasso*

Description

This function estimates the p-values for predictors x against target variable y in lasso regression, using the regularization strength when each predictor enters the active set of regularization path for the first time as the statistic. This is based on the assumption that predictors (of the same variance) that (first) become active earlier tend to be more significant. Two null distributions are supported: normal and spherical, which are computed separately for each predictor and analytically under approximation, which aims at efficiency and accuracy for small p-values.

Usage

```
lassopv(x,y,normalize=TRUE,H0=c("spherical","normal"),

log.p=FALSE,max.predictors=NULL,trace = FALSE,Gram,

eps = .Machine$double.eps,max.steps,use.Gram=TRUE)
```

Arguments

x	Input matrix of predictor variables.
y	Input vector of target variable.
normalize	Whether every predictor is scaled to unit variance first. Every predictor is forcefully shifted to zero mean regardless of this argument.
H0	The null distribution for each predictor x. Spherical: uniform distribution on n-1 dimensional sphere S^{n-1}, so the variance is kept the same as sigma_x^2. Normal: i.i.d N(0,sigma_x^2) in R^n, where sigma_x^2 is the variance of the original predictor x and n is the number of rows in x.
log.p	Whether to output log p-values instead.
max.predictors	The number of top predictors to estimate p-values for. Defaults to all predictors.
trace	Whether traces lasso regression. See lars in package lars.
Gram	Optional Gram used by lasso regression in lars.
eps	Precision for lars function.
max.steps	The optional maximum number steps for lasso regression. See lars in package lars.
use.Gram	Whether to use Gram in lasso regression. See lars in package lars.

Value

Vector of p-values for predictors. Predictors never entered the active set of regularization path within the given max.steps or not within the top (max.predictors) predictors have p-value=1. If log.p is set, output log p-values instead.

References

Lingfei Wang and Tom Michoel, Comparable variable selection with lasso, https://arxiv.org/pdf/1701.07011. 2017, 2018.

Examples

```
library(lars)
library(lassopv)
data(diabetes)
attach(diabetes)
pv=lassopv(x,y)
```

Index