Package 'sprintr'

October 14, 2022

Type Package

Title Sparse Reluctant Interaction Modeling

Version 0.9.0

Date 2019-08-08

Description An implementation of a computationally efficient method to fit large-scale interaction models based on the reluctant interaction selection principle. The method and its properties are described in greater depth in Yu, G., Bien, J., and Tibshirani, R.J. (2019) ``Reluctant interaction modeling'', which is available at <arXiv:1907.08414>.

BugReports https://github.com/hugogogo/sprintr/issues

License GPL-3

Imports Rcpp (>= 0.12.16), glmnet

LinkingTo Rcpp, RcppArmadillo

RoxygenNote 6.0.1

Suggests knitr, rmarkdown

VignetteBuilder knitr

NeedsCompilation yes

Author Guo Yu [aut, cre]

Maintainer Guo Yu <gy63@uw.edu>

Repository CRAN

Date/Publication 2019-08-24 10:40:02 UTC

R topics documented:

`cv.sprinter` *Running sprinter with cross-validation*

Description

The main cross-validation function to select the best sprinter fit for a path of tuning parameters.

Usage

```
cv.sprinter(x, y, num_keep = NULL, square = FALSE, lambda = NULL,
  nlam = 100, lam_min_ratio = ifelse(nrow(x) < ncol(x), 0.01, 1e-04),
  nfold = 5, foldid = NULL)
```

Arguments

`x`	An `n` by `p` design matrix of main effects. Each row is an observation of `p` main effects.
`y`	A response vector of size `n`.
`num_keep`	Number of candidate interactions to keep in Step 2. If `num_keep` is not specified (as default), it will be set to `[n / log n]`.
`square`	Indicator of whether squared effects should be fitted in Step 1. Default to be FALSE.
`lambda`	A user specified list of tuning parameter. Default to be NULL, and the program will compute its own `lambda` path based on `nlam` and `lam_min_ratio`.
`nlam`	The number of `lambda` values. Default value is `100`.
`lam_min_ratio`	The ratio of the smallest and the largest values in `lambda`. The largest value in `lambda` is usually the smallest value for which all coefficients are set to zero. Default to be `1e-2` in the `n < p` setting.
`nfold`	Number of folds in cross-validation. Default value is 5. If each fold gets too view observation, a warning is thrown and the minimal `nfold = 3` is used.
`foldid`	A vector of length `n` representing which fold each observation belongs to. Default to be `NULL`, and the program will generate its own randomly.

Value

An object of S3 class "`sprinter`".

`n` The sample size.

`p` The number of main effects.

`a0` estimate of intercept corresponding to the CV-selected model.

`compact` A compact representation of the selected variables. `compact` has three columns, with the first two columns representing the indices of a selected variable (main effects with first index = 0), and the last column representing the estimate of coefficients.

`fit` The whole `glmnet` fit object in Step 3.

`fitted` fitted value of response corresponding to the CV-selected model.

`lambda` The sequence of `lambda` values used.

`cvm` The averaged estimated prediction error on the test sets over K folds.

`cvsd` The standard error of the estimated prediction error on the test sets over K folds.

`foldid` Fold assignment. A vector of length `n`.

`ibest` The index in `lambda` that is chosen by CV.

`call` Function call.

See Also

`predict.cv.sprinter`

Examples

```
n <- 100
p <- 200
x <- matrix(rnorm(n * p), n, p)
y <- x[, 1] - 2 * x[, 2] + 3 * x[, 1] * x[, 3] - 4 * x[, 4] * x[, 5] + rnorm(n)
mod <- cv.sprinter(x = x, y = y)
```

`predict.cv.sprinter` *Calculate prediction from a* `cv.sprinter` *object.*

Description

Calculate prediction from a `cv.sprinter` object.

Usage

```
## S3 method for class 'cv.sprinter'
predict(object, newdata, ...)
```

Arguments

`object`	a fitted `cv.sprinter` object.
`newdata`	a design matrix of all the `p` main effects of some new observations of which predictions are to be made.
`...`	additional argument (not used here, only for S3 generic/method consistency)

Value

The prediction of `newdata` by the cv.sprinter fit `object`.

Examples

```
n <- 100
p <- 200
x <- matrix(rnorm(n * p), n, p)
y <- x[, 1] + 2 * x[, 2] - 3 * x[, 1] * x[, 2] + rnorm(n)
mod <- cv.sprinter(x = x, y = y)
fitted <- predict(mod, newdata = x)
```

`screen_cpp` *Sure Independence Screening in Step 2*

Description

Sure Independence Screening in Step 2

Usage

```
screen_cpp(x, y, num_keep, square = FALSE, main_effect = FALSE)
```

Arguments

`x`	a n-by-p matrix of main effects, with i.i.d rows, and each row represents a vector of observations of p main-effects
`y`	a vector of length n. In sprinter, y is the residual from step 1
`num_keep`	the number of candidate interactions in Step 2. Default to be n / [log n]
`square`	An indicator of whether squared effects should be considered in Step 1 (NOT Step 2!). square == TRUE if squared effects have been considered in Step 1, i.e., squared effects will NOT be considered in Step 2.
`main_effect`	An indicator of whether main effects should also be screened. Default to be false. The functionality of main_effect = true is not used in sprinter, but for SIS_lasso.

Value

an matrix of 2 columns, representing the index pair of the selected interactions.

sprinter *Sparse Reluctant Interaction Modeling*

Description

This is the main function that fits interaction models with a path of tuning parameters (for Step 3).

Usage

```
sprinter(x, y, num_keep = NULL, square = FALSE, lambda = NULL,
  nlam = 100, lam_min_ratio = ifelse(nrow(x) < ncol(x), 0.01, 1e-04))
```

Arguments

x	An n by p design matrix of main effects. Each row is an observation of p main effects.
y	A response vector of size n.
num_keep	Number of candidate interactions to keep in Step 2. If num_keep is not specified (as default), it will be set to [n / log n].
square	Indicator of whether squared effects should be fitted in Step 1. Default to be FALSE.
lambda	A user specified list of tuning parameter. Default to be NULL, and the program will compute its own lambda path based on nlam and lam_min_ratio.
nlam	The number of lambda values. Default value is 100.
lam_min_ratio	The ratio of the smallest and the largest values in lambda. The largest value in lambda is usually the smallest value for which all coefficients are set to zero. Default to be 1e-2 in the n < p setting.

Value

An object of S3 class "sprinter".

n The sample size.

p The number of main effects.

a0 Estimate of intercept.

coef Estimate of regression coefficients.

idx Indices of all main effects and interactions in Step 3.

fitted Fitted response value. It is a n-by-nlam matrix, with each column representing a fitted response vector for a value of lambda.

lambda The sequence of lambda values used.

call Function call.

See Also

cv.sprinter

Examples

```
set.seed(123)
n <- 100
p <- 200
x <- matrix(rnorm(n * p), n, p)
y <- x[, 1] - 2 * x[, 2] + 3 * x[, 1] * x[, 3] - 4 * x[, 4] * x[, 5] + rnorm(n)
mod <- sprinter(x = x, y = y)
```

Index