

Package ‘PPSFS’

October 12, 2022

Type Package

Title Partial Profile Score Feature Selection in High-Dimensional
Generalized Linear Interaction Models

Version 0.1.0

Date 2022-03-19

Maintainer Zengchao Xu <zengc.xu@aliyun.com>

Description This is an implementation of the partial profile score feature
selection (PPSFS) approach to generalized linear (interaction) models.
The PPSFS is highly scalable even for ultra-high-dimensional feature space.
See the paper by Xu, Luo and Chen (2021, <doi:10.4310/21-SII706>).

URL <https://github.com/paradoxical-rhapsody/PPSFS>

BugReports <https://github.com/paradoxical-rhapsody/PPSFS/issues>

Imports Rcpp, brglm2

LinkingTo Rcpp, RcppArmadillo

License GPL-3

Encoding UTF-8

Language en-US

RoxygenNote 7.1.2

NeedsCompilation yes

Author Zengchao Xu [aut, cre],
Shan Luo [aut],
Zehua Chen [aut]

Repository CRAN

Date/Publication 2022-03-21 18:10:05 UTC

R topics documented:

| | |
|----------------------|----------|
| PPS-method | 2 |
| Index | 4 |

Description

ppsfs: PPSFS for **main-effects**.

ppsfsi: PPSFS for **interaction effects**.

Usage

```
ppsfs(
  x,
  y,
  family,
  keep = NULL,
  I0 = NULL,
  ...,
  ebicFlag = 1,
  maxK = min(NROW(x) - 1, NCOL(x) + length(I0)),
  verbose = FALSE
)
```

```
ppsfsi(
  x,
  y,
  family,
  keep = NULL,
  ...,
  ebicFlag = 1,
  maxK = min(NROW(x) - 1, choose(NCOL(x), 2)),
  verbose = FALSE
)
```

Arguments

| | |
|----------|---|
| x | Matrix. |
| y | Vector. |
| family | See glm and family . |
| keep | Initial set of features that are included in model fitting. |
| I0 | Index set of interaction effects to be identified. |
| ... | Additional parameters for glm.fit . |
| ebicFlag | The procedure stops when the EBIC increases after ebicFlag times. |
| maxK | Maximum number of identified features. |
| verbose | Print the procedure path? |

Details

That `ppsfs(x, y, family="gaussian")` is an implementation to *sequential lasso* method proposed by Luo and Chen doi: [10/f6kfr6](https://doi.org/10.4310/21SI1706).

Value

Index set of identified features.

References

Z. Xu, S. Luo and Z. Chen (2022). Partial profile score feature selection in high-dimensional generalized linear interaction models. *Statistics and Its Interface*. doi: [10.4310/21SI1706](https://doi.org/10.4310/21SI1706)

Examples

```
## *****
## Identify main-effect features
## *****
set.seed(2022)
n <- 300
p <- 1000
x <- matrix(rnorm(n*p), n)
eta <- drop( x[, 1:3] %*% runif(3, 1.0, 1.5) )
y <- eta + rnorm(n, sd=sd(eta)/5)
print( A <- ppsfs(x, y, 'gaussian', verbose=TRUE) )

## *****
## Identify interaction effects
## *****
set.seed(2022)
n <- 300
p <- 150
x <- matrix(rnorm(n*p), n)
eta <- drop( cbind(x[, 1:3], x[, 4:6]*x[, 7:9]) %*% runif(6, 1.0, 1.5) )
y <- eta + rnorm(n, sd=sd(eta)/5)
print( group <- ppsfsi(x, y, 'gaussian', verbose=TRUE) )
print( A <- ppsfs(x, y, "gaussian", I0=group, verbose=TRUE) )

print( A <- ppsfs(x, y, "gaussian", keep=c(1, "5:8"),
                 I0=group, verbose=TRUE) )
```

Index

family, [2](#)

glm, [2](#)

glm.fit, [2](#)

PPS-method, [2](#)

ppsfs (PPS-method), [2](#)

ppsfsi (PPS-method), [2](#)