Package 'prescience'

July 24, 2018

Type Package

Title Approximate Best Subset Maximum Binary Prediction Rule (PRESCIENCE)

Version 0.2.0

Date 2018-07-24

Author Yankang (Bennie) Chen [aut, cre],
Le-Yu Chen [aut],
Sokbae Lee [aut]

Maintainer Yankang (Bennie) Chen <yankang.chen@columbia.edu>

Description This package employs the Approximate Best Subset Maximum Binary Prediction Rule (PRESCIENCE) on the given variable selection problem.

Depends R (>= 3.4.0)

Imports gurobi, slam, stats

License GPL-3

Encoding UTF-8

LazyData true

RoxygenNote 6.0.1

R topics documented:

prescience-package 1
coef.select 2
select 2
summary.select 4
transportation 4

Index **6**

`prescience-package` *Functions for the Approximate Best Subset Maximum Binary Prediction Rule (PRESCIENCE).*

Description

`prescience` provides tools for best subset binary selection.

coef.select *PRESCIENCE estimates for coefficients*

Description

`coef.select` produces the estimated coefficients of the PRESCIENCE produced by the `select` function.

Usage

```
## S3 method for class 'select'
coef(object, ...)
```

Arguments

`object`	an object of class `select`.
`...`	additional parameters.

Value

the input object is returned silently.

Author(s)

Yankang (Bennie) Chen <yankang.chen@columbia.edu>

Examples

```
results <- select(auto ~ dcost + cars + dovtt + divtt,
data = transportation, nfoc = 1, q = 1, bound = 10)

coef(results)
```

select *Approximate Best Subset Maximum Binary Prediction Rule (PRESCIENCE)*

Description

`select` employs the Approximate Best Subset Maximum Binary Prediction Rule (PRESCIENCE) on the given variable selection problem.

Usage

```
select(formula, data, nfoc, q, bound, beta0 = 1, tol = NULL,
  warmstart = TRUE, tau = 1.5, mio = 1, tlim = 86400)
```

Arguments

formula	an object of class formula of the format: (binary dependent variable) ~ (normalized focus variable) + (remaining focus variables) + (auxiliary variables).
data	a data frame containing the variables in the model.
nfoc	integer. The number of focus variable(s) excluding the intercept.
q	integer. The cardinality constraint for the covariate selection.
bound	numeric. The maximum absolute value of the bounds for all variables.
beta0	integer. The coefficient taking value either 1 or -1 to normalize the scale for the first focus variable.
tol	numeric. Tolerance level. If NULL, use the default tolerance level.
warmstart	logical. If TRUE, use the warm start strategy.
tau	the tuning parameter for enlarging the estimated bounds.
mio	integer. 1 for MIO method 1 and 2 for method 2 in the paper.
tlim	time limit (in seconds) specified for the MIO solver.

Value

a list with 7 elements:

tolerance	tolerance level
status	optimization status
score	Gurobi score
gap	the MIO optimization gap value in case of early termination (0 if optimal solution is found within the time limit)
rtime	time used by the MIO solver in the estimation procedure
ncount	the number of Branch-and-bound nodes
bhat	maximum score estimates for coefficients

Author(s)

Yankang (Bennie) Chen <yankang.chen@columbia.edu>

References

Best Subset Binary Prediction by Le-Yu Chen and Sokbae Lee (2018). https://arxiv.org/abs/1610.02738

Examples

```
results <- select(auto ~ dcost + cars + dovtt + divtt,
data = transportation, nfoc = 1, q = 1, bound = 10)

summary(results)
coef(results)
```

summary.select *Summarize the PRESCIENCE results*

Description

summary.select summarizes the results of PRESCIENCE produced by the select function.

Usage

```
## S3 method for class 'select'
summary(object, ...)
```

Arguments

object	an object of class select.
...	additional parameters.

Value

the input object is returned silently.

Author(s)

Yankang (Bennie) Chen <yankang.chen@columbia.edu>

Examples

```
results <- select(auto ~ dcost + cars + dovtt + divtt,
data = transportation, nfoc = 1, q = 1, bound = 10)

summary(results)
```

transportation *Transportation Mode Choice by Horowitz (1993)*

Description

A dataset of the transportation mode choice containing 842 observations sampled randomly from Washington, D.C., area transportation study.

Usage

```
transportation
```

Format

A data frame with 842 rows and 5 variables. The variables are as follows:

auto Transportation mode choice with 1 for automobile and 0 otherwise

dcost Transit fare minus automobile travel cost (in dollars)

cars Number of cars owned by the traveler's household

dovtt Transit out-of-vehicle travel time minus automobile out-of-vehicle travel time (in minutes)

divtt Transit in-vehicle travel time minus automobile in-vehicle travel time (in minutes)

Source

The `transportation` data were obtained from *Semiparametric estimation of a work-trip mode choice model* by Joel L.Horowitz (1993). https://www.sciencedirect.com/science/article/pii/030440769390113J.

Index

∗Topic **datasets**
transportation, 4

coef.select, 2

prescience *(*prescience-package*)*, 1
prescience-package, 1

select, *2*, 2, *4*
summary.select, 4

transportation, 4