Package: ivitr (via r-universe)

October 27, 2024

Type Package

Title Estimate IV-Optimal Individualized Treatment Rules

Version 0.1.0

Author Bo Zhang

Maintainer Bo Zhang <bozhan@wharton.upenn.edu>

Description A method that estimates an IV-optimal individualized treatment rule. An individualized treatment rule is said to be IV-optimal if it minimizes the maximum risk with respect to the putative IV and the set of IV identification assumptions. Please refer to <arXiv:2002.02579> for more details on the methodology and some theory underpinning the method. Function IV-PILE() uses functions in the package 'locClass'. Package 'locClass' can be accessed and installed from the 'R-Forge' repository via the following link: <https://r-forge.r-project.org/projects/locclass/>. Alternatively, one can install the package by entering the following in R: 'install.packages(``locClass'', repos=``<http://R-Forge.R-project.org>'')'.

License GPL-3

Encoding UTF-8

LazyData true

RoxygenNote 7.1.0

Depends R (>= 2.10)

Suggests locClass

Imports stats, nnet, randomForest, dplyr, rlang

NeedsCompilation no

Repository CRAN

Date/Publication 2020-09-11 08:40:03 UTC

Contents

`dt_Rouse` *Rouse (1995) dataset*

Description

Variables of the dataset is as follows:

educ86 Years of education since 1986.

twoyr Attending a two-year college immediately after high school.

female Gender: 1 if female and 0 otherwise.

black Race: 1 if African American and 0 otherwise.

hispanic Race: 1 if Hispanic and 0 otherwise.

bytest Test score.

dadsome Dad's education: some college.

dadcoll Dad's education: college.

momsome Mom's education: some college.

momcoll Mom's education: college.

fincome Family income.

fincmiss Missingness indicator for family income.

tuition2 Average state two-year college tuition.

tuition4 Average state four-year college tuition.

dist2yr Distance to the nearest two-year college.

dist4yr Distance to the nearest four-year college.

Usage

```
data(dt_Rouse)
```

Format

A data frame with 4437 rows and 16 columns.

Source

ss

estimate_BP_bound *Estimate the Balke-Pearl bound for each instance in a dataset*

Description

estimate_BP_bound estimates the Balke-Pearl bound for each instance in the input dataset with a binary IV, observed covariates, a binary treatment indicator, and a binary outcome.

Usage

```
estimate_BP_bound(dt, method = "rf", nodesize = 5)
```

Arguments

dt	A dataframe whose first column is a binary IV 'Z', followed by q columns of observed covariates, followed by a binary treatment indicator 'A', and finally followed by a binary outcome 'Y'. The dataset has q+3 columns in total.
method	A character string indicator the method used to estimate each constituent conditional probability of the Balke-Pearl bound. Users can choose to fit multinomial regression by setting method = 'multinom', and random forest by setting method = 'rf'.
nodesize	Node size to be used in a random forest algorithm if method is set to 'rf'. The default value is set to 5.

Value

The original dataframe with two additional columns: L and U. L indicates the Balke-Pearl lower bound and U is the Balke-Pearl upper bound.

Examples

```
attach(dt_Rouse)
# Construct an IV out of differential distance to two-year versus
# four-year college. Z = 1 if the subject lives not farther from
# a 4-year college compared to a 2-year college.
Z = (dist4yr <= dist2yr) + 0

# Treatment A = 1 if the subject attends a 4-year college and 0
# otherwise.
A = 1 - twoyr

# Outcome Y = 1 if the subject obtained a bachelor's degree
Y = (educ86 >= 16) + 0

# Prepare the dataset
dt = data.frame(Z, female, black, hispanic, bytest, dadsome,
     dadcoll, momsome, momcoll, fincome, fincmiss, A, Y)
```

```
# Calculate the Balke-Pearl bound by estimating each constituent
# conditional probability p(Y = y, A = a | Z, X) with a random
# forest.
dt_with_BP_bound_rf = estimate_BP_bound(dt, method = 'rf', nodesize = 5)

# Calculate the Balke-Pearl bound by estimating each constituent
# conditional probability p(Y = y, A = a | Z, X) with a multinomial
# regression.
dt_with_BP_bound_multinom = estimate_BP_bound(dt, method = 'multinom')
```

`estimate_Sid_bound` *Estimate the partial identification bound as in Siddique (2013, JASA) for each instance in a dataset*

Description

`estimate_Sid_bound` estimates the partial identification bound for each instance in the input dataset with a binary IV, observed covariates, a binary treatment indicator, and a binary outcome according to Siddique (2013, JASA).

Usage

```
estimate_Sid_bound(dt, method = "rf", nodesize = 5)
```

Arguments

`dt`	A dataframe whose first column is a binary IV 'Z', followed by q columns of observed covariates, followed by a binary treatment indicator 'A', and finally followed by a binary outcome 'Y'. The dataset has q+3 columns in total.
`method`	A character string indicator the method used to estimate each constituent conditional probability of the partial identification bound. Users can choose to fit multinomial regression by setting method = 'multinom', and random forest by setting method = 'rf'.
`nodesize`	Node size to be used in a random forest algorithm if method is set to 'rf'. The default value is set to 5.

Value

The original dataframe with two additional columns: L and U. L indicates the lower bound and U the upper bound as in Siddique 2013

Examples

```
attach(dt_Rouse)
# Construct an IV out of differential distance to two-year versus
# four-year college. Z = 1 if the subject lives not farther from
# a 4-year college compared to a 2-year college.
```

```
Z = (dist4yr <= dist2yr) + 0

# Treatment A = 1 if the subject attends a 4-year college and 0
# otherwise.
A = 1 - twoyr

# Outcome Y = 1 if the subject obtained a bachelor's degree
Y = (educ86 >= 16) + 0

# Prepare the dataset
dt = data.frame(Z, female, black, hispanic, bytest, dadsome,
     dadcoll, momsome, momcoll, fincome, fincmiss, A, Y)

# Calculate the Siddique bound by estimating each constituent
# conditional probability p(Y = y, A = a | Z, X) with a random
# forest.
dt_with_Sid_bound_rf = estimate_Sid_bound(dt, method = 'rf', nodesize = 5)

# Calculate the Siddique bound by estimating each constituent
# conditional probability p(Y = y, A = a | Z, X) with a multinomial
# regression.
dt_with_Sid_bound_multinom = estimate_Sid_bound(dt, method = 'multinom')
```

IV_PILE *Estimate an IV-optimal individualized treatment rule*

Description

IV_PILE estimates an IV-optimal individualized treatment rule given a dataset with estimated partial identification intervals for each instance.

Usage

```
IV_PILE(dt, kernel = "linear", C = 1, sig = 1/(ncol(dt) - 5))
```

Arguments

dt	A dataframe whose first column is a binary IV 'Z', followed by q columns of observed covariates, a binary treatment indicator 'A', a binary outcome 'Y', lower endpoint of the partial identification interval 'L', and upper endpoint of the partial identification interval 'U'. The dataset has q+5 columns in total.
kernel	The kernel used in the weighted SVM algorithm. The user may choose between 'linear' (linear kernel) and 'radial' (Gaussian RBF kernel).
C	Cost of violating the constraint. This is the parameter C in the Lagrange formulation.
sig	Sigma in the Gaussian RBF kernel. Default is set to 1/dimension of covariates, i.e., 1/q. This parameter is not relevant for linear kernel.

Value

An object of the type `wsvm`, inheriting from `svm`.

Examples

```
## Not run:
# It is necessary to install the package locClass in order
# to run the following code.

attach(dt_Rouse)
# Construct an IV out of differential distance to two-year versus
# four-year college. Z = 1 if the subject lives not farther from
# a 4-year college compared to a 2-year college.
Z = (dist4yr <= dist2yr) + 0

# Treatment A = 1 if the subject attends a 4-year college and 0
# otherwise.
A = 1 - twoyr

# Outcome Y = 1 if the subject obtained a bachelor's degree
Y = (educ86 >= 16) + 0

# Prepare the dataset
dt = data.frame(Z, female, black, hispanic, bytest, dadsome,
     dadcoll, momsome, momcoll, fincome, fincmiss, A, Y)

# Estimate the Balke-Pearl bound by estimating each constituent
# conditional probability p(Y = y, A = a | Z, X) with a multinomial
# regression.
dt_with_BP_bound_multinom = estimate_BP_bound(dt, method = 'multinom')

# Estimate the IV-optimal individualized treatment rule using a
# linear kernel, under the putative IV and the Balke-Pearl bound.


iv_itr_BP_linear = IV_PILE(dt_with_BP_bound_multinom, kernel = 'linear')

## End(Not run)
```

Index