

Package ‘OPL’

February 27, 2025

Type Package

Title Optimal Policy Learning

Version 1.0.2

Author Federico Brogi [aut, cre],
Barbara Guardabascio [aut],
Giovanni Cerulli [aut]

Maintainer Federico Brogi <federicobrogi@gmail.com>

Description Provides functions for optimal policy learning in socioeconomic applications helping users to learn the most effective policies based on data in order to maximize empirical welfare. Specifically, 'OPL' allows to find "treatment assignment rules" that maximize the overall welfare, defined as the sum of the policy effects estimated over all the policy beneficiaries. Documentation about 'OPL' is provided by several international articles via Athey et al (2021, <doi:10.3982/ECTA15732>), Kitagawa et al (2018, <doi:10.3982/ECTA13288>), Cerulli (2022, <doi:10.1080/13504851.2022.2032577>), the paper by Cerulli (2021, <doi:10.1080/13504851.2020.1820939>) and the book by Gareth et al (2013, <doi:10.1007/978-1-4614-7138-7>).

License GPL-3

Encoding UTF-8

RoxygenNote 7.3.1

Imports stats, dplyr, ggplot2, pander, randomForest, tidyr

Suggests knitr, rmarkdown

VignetteBuilder knitr

NeedsCompilation no

Repository CRAN

Date/Publication 2025-02-27 13:00:06 UTC

Contents

make_cate	2
---------------------	---

opl_dt_c	3
opl_dt_max_choice	4
opl_lc_c	5
opl_tb_c	6
overlapping	7

Index	9
--------------	----------

make_cate	<i>Function to calculate the Causal Treatment Effect</i>
-----------	--

Description

Predicting conditional average treatment effect (CATE) on a new policy based on the training over an old policy

Usage

```
make_cate(
  model,
  train_data,
  test_data,
  w,
  x,
  y,
  family = gaussian(),
  ntree = 100,
  mtry = 2,
  verbose = TRUE
)
```

Arguments

model	A model object used for estimation.
train_data	The training dataset.
test_data	The test dataset.
w	Set the treatment variable.
x	set Independent variables for the model.
y	Set the outcome variable.
family	The family type for the model (e.g., 'binomial').
ntree	Number of trees for the Random Forest model.
mtry	Number of variables to consider at each tree split in the Random Forest model.
verbose	Set TRUE to print the output on the console.

Value

An object containing the estimated causal treatment effect results.

References

- Athey, S., and Wager S. 2021. Policy Learning with Observational Data, *Econometrica*, 89, 1, 133–161.
- Cerulli, G. 2021. Improving econometric prediction by machine learning, *Applied Economics Letters*, 28, 16, 1419-1425.
- Cerulli, G. 2022. Optimal treatment assignment of a threshold-based policy: empirical protocol and related issues, *Applied Economics Letters*, DOI: 10.1080/13504851.2022.2032577.
- Gareth, J., Witten, D., Hastie, D.T., Tibshirani, R. 2013. *An Introduction to Statistical Learning : with Applications in R*. New York, Springer.
- Kitagawa, T., and A. Tetenov. 2018. Who Should Be Treated? Empirical Welfare Maximization Methods for Treatment Choice, *Econometrica*, 86, 2, 591–616.

opl_dt_c

Optimal Policy Learning with Decision Tree

Description

Implementing ex-ante treatment assignment using as policy class a 2-layer fixed-depth decision-tree at specific splitting variables and threshold values.

Usage

```
opl_dt_c(make_cate_result, z, w, c1 = NA, c2 = NA, c3 = NA, verbose = TRUE)
```

Arguments

make_cate_result	A data frame resulting from the make_cate function, containing the predicted treatment effects (my_cate) and other variables for treatment assignment.
z	A character vector containing the names of the variables used for treatment assignment.
w	A string representing the treatment indicator variable name.
c1	Value of the threshold value c1 for the first splitting variable. This number must be chosen between 0 and 1.
c2	Value of the threshold value c2 for the second splitting variable. This number must be chosen between 0 and 1.
c3	Value of the threshold value c3 for the third splitting variable. This number must be chosen between 0 and 1.
verbose	Set TRUE to print the output on the console.

Value

A list containing:

- `W_opt_constr`: The maximum average constrained welfare.
- `W_opt_unconstr`: The average unconstrained welfare.
- `units_to_be_treated`: A data frame of the units to be treated based on the optimal policy.
- A plot showing the optimal policy assignment.

References

- Athey, S., and Wager S. 2021. Policy Learning with Observational Data, *Econometrica*, 89, 1, 133–161.
- Cerulli, G. 2021. Improving econometric prediction by machine learning, *Applied Economics Letters*, 28, 16, 1419-1425.
- Cerulli, G. 2022. Optimal treatment assignment of a threshold-based policy: empirical protocol and related issues, *Applied Economics Letters*, DOI: 10.1080/13504851.2022.2032577.
- Gareth, J., Witten, D., Hastie, D.T., Tibshirani, R. 2013. *An Introduction to Statistical Learning : with Applications in R*. New York, Springer.
- Kitagawa, T., and A. Tetenov. 2018. Who Should Be Treated? Empirical Welfare Maximization Methods for Treatment Choice, *Econometrica*, 86, 2, 591–616.

`opl_dt_max_choice` *User selection on multiple choice*

Description

Function that allows the user to select a row of maximum welfare among the rows with maximum welfare constrained. The function prints out the result and requires user input to select the row.

Usage

```
opl_dt_max_choice(nc, col_max, verbose = TRUE)
```

Arguments

<code>nc</code>	Numeber of max welfare.
<code>col_max</code>	Row index for max constrained welfare.
<code>verbose</code>	Set TRUE to print the output on the console.

Value

Return the user's selection as an input.

opl_lc_c

*Linear Combination Based Policy Learning***Description**

Implementing ex-ante treatment assignment using as policy class a linear-combination approach at specific parameters' values c_1 , c_2 , and c_3 for the linear-combination of variables var_1 and var_2 : $c_1var_1+c_2var_2 \geq c_3$.

Usage

```
opl_lc_c(make_cate_result, z, w, c1 = NA, c2 = NA, c3 = NA, verbose = TRUE)
```

Arguments

<code>make_cate_result</code>	A data frame containing the input data. It must include a column named <code>my_cate</code> representing conditional average treatment effects (CATE) generated using <code>make_cate</code> function.
<code>z</code>	A character vector of length 2 specifying the column names of the two threshold variables to be standardized.
<code>w</code>	A character string specifying the column name indicating treatment assignment (binary variable).
<code>c1</code>	Threshold for <code>var1</code> given by the user or optimized by the the function. This number must be chosen between 0 and 1.
<code>c2</code>	Threshold for <code>var2</code> given by the user or optimized by the the function. This number must be chosen between 0 and 1.
<code>c3</code>	Third parameter of the linear-combination. This number must be chosen between 0 and 1.
<code>verbose</code>	Set TRUE to print the output on the console.

Details

The function performs the following steps:

- Standardizes the threshold variables using a min-max scaling technique.
- Determines the optimal treatment assignment based on the linear combination of the threshold variables.
- Performs a grid search to estimate the optimal policy.
- Outputs a plot visualizing the optimal treatment assignments.
- Prints the main results, including the percentage of treated units, the unconstrained and constrained welfare, and the policy parameters.

Value

The function returns a data frame containing the standardized variables and treatment assignments, and prints a summary of the results and a plot showing the optimal policy assignment.

References

- Athey, S., and Wager S. 2021. Policy Learning with Observational Data, *Econometrica*, 89, 1, 133–161.
- Cerulli, G. 2021. Improving econometric prediction by machine learning, *Applied Economics Letters*, 28, 16, 1419-1425.
- Cerulli, G. 2022. Optimal treatment assignment of a threshold-based policy: empirical protocol and related issues, *Applied Economics Letters*, DOI: 10.1080/13504851.2022.2032577.
- Gareth, J., Witten, D., Hastie, D.T., Tibshirani, R. 2013. *An Introduction to Statistical Learning : with Applications in R*. New York, Springer.
- Kitagawa, T., and A. Tetenov. 2018. Who Should Be Treated? Empirical Welfare Maximization Methods for Treatment Choice, *Econometrica*, 86, 2, 591–616.

opl_tb_c

Threshold-based policy learning at specific values

Description

Implementing ex-ante treatment assignment using as policy class a threshold-based (or quadrant) approach at specific threshold values `c1` and `c2` for respectively the selection variables `var1` and `var2`.

Usage

```
opl_tb_c(make_cate_result, z, w, c1 = NA, c2 = NA, verbose = TRUE)
```

Arguments

<code>make_cate_result</code>	A data frame containing the input data. It must include a column named <code>my_cate</code> representing conditional average treatment effects (CATE) generated using <code>make_cate</code> function.
<code>z</code>	A character vector of length 2 specifying the column names of the two threshold variables to be standardized.
<code>w</code>	A character string specifying the column name indicating treatment assignment (binary variable).
<code>c1</code>	Threshold for <code>var1</code> given by the user or optimized by the the function. This number must be chosen between 0 and 1.
<code>c2</code>	Threshold for <code>var2</code> given by the user or optimized by the the function. This number must be chosen between 0 and 1.
<code>verbose</code>	Set TRUE to print the output on the console.

Details

The function:

1. Standardizes the threshold variables to a 0-1 range.
2. Identifies the optimal thresholds based on grid search for maximizing constrained welfare.
3. Computes and displays key statistics, including average welfare measures and the percentage of treated units.

Value

The function invisibly returns the input data frame augmented with the following columns:

- `z[1]_std`: Standardized version of the first threshold variable.
- `z[2]_std`: Standardized version of the second threshold variable.
- `units_to_be_treated`: Binary indicator for whether a unit should be treated based on the optimal policy.

Additionally, the function:

- Prints the main results summary, including optimal threshold values, average constrained and unconstrained welfare, and treatment proportions.
- Displays a scatter plot visualizing the policy assignment.

References

- Athey, S., and Wager S. 2021. Policy Learning with Observational Data, *Econometrica*, 89, 1, 133–161.
- Cerulli, G. 2021. Improving econometric prediction by machine learning, *Applied Economics Letters*, 28, 16, 1419-1425.
- Cerulli, G. 2022. Optimal treatment assignment of a threshold-based policy: empirical protocol and related issues, *Applied Economics Letters*, DOI: 10.1080/13504851.2022.2032577.
- Gareth, J., Witten, D., Hastie, D.T., Tibshirani, R. 2013. *An Introduction to Statistical Learning : with Applications in R*. New York, Springer.
- Kitagawa, T., and A. Tetenov. 2018. Who Should Be Treated? Empirical Welfare Maximization Methods for Treatment Choice, *Econometrica*, 86, 2, 591–616.

overlapping

Testing overlap between old and new policy sample

Description

Function to perform overlap analysis between train and test datasets. The function performs principal component analysis (PCA) on the covariates for both sets and calculates the Kolmogorov-Smirnov test for overlap.

Usage

```
overlapping(train_data, test_data, x)
```

Arguments

<code>train_data</code>	Train Dataset indicating the old policy sample.
<code>test_data</code>	Test Dataset indicating the new policy sample.
<code>x</code>	Vector of predictor variables.

Value

The function prints the superposition graph and the results of the Kolmogorov-Smirnov test.

Index

make_cate, 2

opl_dt_c, 3

opl_dt_max_choice, 4

opl_lc_c, 5

opl_tb_c, 6

overlapping, 7