Package 'DMLLZU'

May 21, 2021

Type Package

Title Double Machine Learning

Version 0.1.1

Description Yang(2020,<doi:10.1016/j.jeconom.2020.01.018>) come up with Double Machine Learning model ,it is based on this model, using four machine learning methods-- bagging, Boosting, random forest and neural network, and then based on the four models for two different combinations of the integrated model -- linear model combination and random forest .

Author Lixiong Yang [aut],
Junchang Zhao [aut, cre]

Maintainer Junchang Zhao <zhaojch19@lzu.edu.cn>

Depends R (>= 3.2)

Imports rlang, gbm ,lmtest ,nnet , sandwich,randomForest,caret,ISLR

License GPL (>= 2)

Encoding UTF-8

RoxygenNote 7.1.1

NeedsCompilation no

Repository CRAN

Date/Publication 2021-05-21 12:12:10 UTC

R topics documented:

dml_bagging *Double Machine Learning based on bagging*

Description

The most famous representative of parallel ensemble learning. This method uses the self-help method to repeatedly sample from a single training set and generate several different self-help sampling training sets. Then, the self-help sampling training sets are used to fit the model and then the predicted values are obtained.

Usage

```
dml_bagging(y,x,d,data,sed)
```

Arguments

```
y,x,d,data,sed
```

Value

y Dependent variable;

d Independent variable;

x Control variables;

sed A random seed;

data Data

Author(s)

Lixiong Yang<ylx@lzu.edu.cn>

References

Leo Breiman. (1996). Bagging Predictors. Machine Learning, 24(2), pp. 123-140. doi: 10.1023/A:1018054314350

Jui-Chung Yang,,Hui-Ching Chuang & Chung-Ming Kuan.(2020).Double machine learning with gradient boosting and its application to the Big N audit quality effect. Journal of Econometrics(1),.doi:10.1016/j.jeconom.2020.01.018

Victor Chernozhukov,,Denis Chetverikov,,Mert Demirer,... & James Robins.(2018).Double/debiased machine learning for treatment and structural parameters. The Econometrics Journal(1),. doi:10.3386/w23564.

See Also

https://github.com/lixiongyang

Examples

```
library(ISLR)
attach(Auto)
data<- Auto
y <- data$mpg    #Dependent variable
d <- data$origin   #Independent variable
x="weight+year +horsepower"      #Control variables;

dml_bagging(y,x,d,data,sed=123)
```

`dml_boosting` *Double Machine Learning based on boosting*

Description

The biggest difference with other method, the trees of this method are generated sequentially. Each tree is constructed using the information of the previous generated trees. Each tree is generated according to a modified version of the original data set, and finally these trees are combined to establish a prediction model

Usage

```
dml_boosting(y,x,d,data,sed)
```

Arguments

`y,x,d,data,sed`

Value

y Dependent variable;

d Independent variable;

x Control variable;

sed A random seed;

data Data

Author(s)

Lixiong Yang<ylx@lzu.edu.cn>

References

Jui-Chung Yang,,Hui-Ching Chuang & Chung-Ming Kuan.(2020).Double machine learning with gradient boosting and its application to the Big N audit quality effect. Journal of Econometrics(1),. doi:10.1016/j.jeconom.2020.01.018 Victor Chernozhukov,,Denis Chetverikov,,Mert Demirer,... & James Robins.(2018).Double/debiased machine learning for treatment and structural parameters. The Econometrics Journal(1),. doi:10.3386/w23564.

See Also

https://github.com/lixiongyang

Examples

```
library(ISLR)
attach(Auto)
data<- Auto
y <- data$mpg    #Dependent variable
d <- data$origin   #Independent variable
x="weight+year +horsepower"      #Control variables;

dml_boosting(y,x,d,data,sed=123)
```

`dml_ensemble_lm` *dml_ensemble_lm*

Description

As an important integrated learning method, stacking consists of at least two layers of structure, including a primary learner and a secondary learner or a meta-learner used to combine the learner.Stacking first trained the primary learner from the initial data set, and then generated a new data set used to train the secondary learner, in this data set, the output of the primary learner is taken as the sample input characteristics, and the initial sample mark is still taken as the sample mark. Integrate the four basic model through linear model.

Usage

```
dml_ensemble_lm(y,x,d,data,sed)
```

Arguments

`y,x,d,data,sed`

Value

y Dependent variable;

d Independent variable;

x Control variable;

sed A random seed;

data Data

Author(s)

Lixiong Yang<ylx@lzu.edu.cn>; Junchang Zhao <zhaojch19@lzu.edu.cn>

References

Wolpert David H.. (1992). Stacked generalization. 5(2), pp. 241-259. doi: 10.1016/S0893-6080(05)80023-1

Jui-Chung Yang,,Hui-Ching Chuang & Chung-Ming Kuan.(2020).Double machine learning with gradient boosting and its application to the Big N audit quality effect. Journal of Econometrics(1),.doi:10.1016/j.jeconom.2020.01.018

Victor Chernozhukov,,Denis Chetverikov,,Mert Demirer,... & James Robins.(2018).Double/debiased machine learning for treatment and structural parameters. The Econometrics Journal(1),. doi:10.3386/w23564.

See Also

help

Examples

```
library(ISLR)
attach(Auto)
data<- Auto
y <- data$mpg    #Dependent variable
d <- data$origin   #Independent variable
x="weight+year +horsepower"      #Control variables;

dml_ensemble_lm(y,x,d,data,sed=123)
```

`dml_ensemble_rf` *dml_ensemble_rf*

Description

As an important integrated learning method, stacking consists of at least two layers of structure, including a primary learner and a secondary learner or a meta-learner used to combine the learner. Stacking first trained the primary learner from the initial data set, and then generated a new data set used to train the secondary learner, in this data set, the output of the primary learner is taken as the sample input characteristics, and the initial sample mark is still taken as the sample mark.Integrate the four basic model through random forest

Usage

```
dml_ensemble_rf(y,x,d,data,sed)
```

Arguments

`y,x,d,data,sed`

Value

y Dependent variable;

d Independent variable;

x Control variable;

sed A random seed;

data Data

Author(s)

Lixiong Yang<ylx@lzu.edu.cn>; Junchang Zhao <zhaojch19@lzu.edu.cn>

References

Jui-Chung Yang,,Hui-Ching Chuang & Chung-Ming Kuan.(2020).Double machine learning with gradient boosting and its application to the Big N audit quality effect. Journal of Econometrics(1),.doi:10.1016/j.jeconom.2020.01.018 Victor Chernozhukov,,Denis Chetverikov,,Mert Demirer,... & James Robins.(2018).Double/debiased machine learning for treatment and structural parameters. The Econometrics Journal(1),. doi:10.3386/w23564.

See Also

help

Examples

```
library(ISLR)
attach(Auto)
data<- Auto
y <- data$mpg    #Dependent variable
d <- data$origin   #Independent variable
x="weight+year +horsepower"      #Control variables;

dml_ensemble_rf(y,x,d,data,sed=123)
```

---

dml_neural_network *Double Machine Learning based on neural network*

Description

Each node represents a specific output function, known as the excitation function.It is a mathematical model or a computational model that imitates the structure and function of biological net. It is calculated by the connection of a large number of artificial neurons, mainly composed of nodes and the mutual connections between nodes. Each connection between two nodes represents a weighted value for the signal passing through the connection, known as the weight. The output of the network is different according to the connection mode of the network, the weight value and the excitation function.

Usage

```
dml_neural_network(y,x,d,data,sed)
```

Arguments

```
y,x,d,data,sed
```

Value

y Dependent variable;

d Independent variable;

x Control variable;

sed A random seed;

data Data

Author(s)

Yang Lixiong

References

Jui-Chung Yang,,Hui-Ching Chuang & Chung-Ming Kuan.(2020).Double machine learning with gradient boosting and its application to the Big N audit quality effect. Journal of Econometrics(1),. doi:10.1016/j.jeconom.2020.01.018

Victor Chernozhukov,,Denis Chetverikov,,Mert Demirer,... & James Robins.(2018).Double/debiased machine learning for treatment and structural parameters. The Econometrics Journal(1),. doi:10.3386/w23564.

See Also

help

Examples

```
library(ISLR)
attach(Auto)
data<- Auto
y <- data$mpg    #Dependent variable
d <- data$origin   #Independent variable
x="weight+year +horsepower"      #Control variables;

dml_neural_network(y,x,d,data,sed=123)
```

`dml_random_forest` *Double Machine Learning based on random forest*

Description

To establish a series of decision tree, the difference is the this method for each division point considering the decision tree, should be chosen from among all variables contain random sample with some of the variables as candidate variables, the explanatory variables can only be used in the split point from the selected part of the explanation variable selection, and then to make predictions.

Usage

```
dml_random_forest(y,x,d,data,sed)
```

Arguments

`y,x,d,data,sed`

Value

y Dependent variable;

d Independent variable;

x Control variable;

sed A random seed;

data Data

Author(s)

Lixiong Yang<ylx@lzu.edu.cn>

References

Leo Breiman. (2001). Random Forests. Machine Learning, 45(1), pp. 5-32. doi:10.1023/A:1010933404324

Jui-Chung Yang,,Hui-Ching Chuang & Chung-Ming Kuan.(2020).Double machine learning with gradient boosting and its application to the Big N audit quality effect. Journal of Econometrics(1).doi:10.1016/j.jeconom.2020.01.018 Victor Chernozhukov,,Denis Chetverikov,,Mert Demirer,... & James Robins.(2018).Double/debiased machine learning for treatment and structural parameters. The Econometrics Journal(1). doi:10.3386/w23564.

Examples

```
library(ISLR)
attach(Auto)
data<- Auto
y <- data$mpg    #Dependent variable
d <- data$origin   #Independent variable
x="weight+year +horsepower"      #Control variables;
```

```
dml_random_forest(y,x,d,data,sed=123)
```

Index