

Package ‘MSML’

December 6, 2023

Title Model selection based on Machine Learning (ML)

Version 1.0.0.0

Description The MSML package is designed to determine the optimal model(s) by leveraging all available features.

License GPL (>=3)

URL <https://github.com/mommy003/MSML>

Encoding UTF-8

Roxygen list(markdown = TRUE)

RoxygenNote 7.1.2

Depends R (>= 2.10)

Imports r2redux,
R2ROC

LazyData true

R topics documented:

data_test	1
data_train	2
data_valid	3
model_configuration	3
model_evaluation	4

Index	5
--------------	----------

data_test	<i>7 sets of PRSs for test dataset and target phenotype</i>
-----------	---

Description

A dataset containing 7 sets of PRSs for test dataset and target phenotype

Usage

```
data_test
```

Format

A data frame for test dataset:

V1 PRS1, for bin1

V2 PRS2, for bin1

V3 PRS3, for bin1

V4 PRS4, for bin1

V5 PRS5, for bin1

V6 PRS6, for bin1

V7 PRS7, for bin1

target Target Phenotype, value

data_train

7 sets of PRSs for training data set and target phenotype

Description

A dataset containing 7 sets of PRSs for training data set and target phenotype

Usage

data_train

Format

A data frame for training dataset:

V1 PRS1, for bin1

V2 PRS2, for bin1

V3 PRS3, for bin1

V4 PRS4, for bin1

V5 PRS5, for bin1

V6 PRS6, for bin1

V7 PRS7, for bin1

target Target Phenotype, value

data_valid

7 sets of PRSs for validation dataset and target phenotype

Description

A dataset containing 7 sets of PRSs for validation dataset and target phenotype

Usage

```
data_valid
```

Format

A data frame for validation dataset:

V1 PRS1, for bin1

V2 PRS2, for bin1

V3 PRS3, for bin1

V4 PRS4, for bin1

V5 PRS5, for bin1

V6 PRS6, for bin1

V7 PRS7, for bin1

target Target Phenotype, value

model_configuration

model_configuration function

Description

This function will generate features (e.g. PRSs) based on all possible combinations of model. The total number of models required to explore the combinations of these 'n' features can be calculated by summing the combinations for each possible number of features, ranging from 1 to 'n' ($C(n,i)$), where $C(n,k)$ represents the binomial coefficient or "n choose k," with n denoting the total number of features and k indicating the number of features to include in each model.

Usage

```
model_configuration(data_train, data_valid, mv)
```

Arguments

data_train This is the matrix for training dataset

data_valid This is the matrix for validation dataset

mv The total number of columns in data_train/data_valid

Value

This function will generate all possible model outcomes for validation and test dataset

Examples

```
#data_train <- data_train
#data_valid <- data_valid
#mv=8
#out=model_configuration(data_train,data_valid,mv)
```

model_evaluation	<i>model_evaluation function</i>
------------------	----------------------------------

Description

This function will identify the best model in the validation and test dataset.

Usage

```
model_evaluation(dat, mv, tn, prev, pthreshold = 0.05, method = "R2ROC")
```

Arguments

dat	This is the matrix for all the combinations of the model
mv	The total number of columns in data_train/data_valid
tn	The total no of best models to be identified
prev	The prevalence of disease in the data
pthreshold	The P value threshold for the significance level
method	The methods to be used to evaluate models

Value

This function will generate all possible model outcomes for validation and test dataset

Examples

```
#dat <- predict_validation
#mv=8
#tn=15
#prev=0.047
#model_evaluation(dat,mv,tn,prev)
```

Index

- *Topic **All**
 - model_configuration, 3
- *Topic **Identify**
 - model_evaluation, 4
- *Topic **best**
 - model_evaluation, 4
- *Topic **combinations**
 - model_configuration, 3
- *Topic **datasets**
 - data_test, 1
 - data_train, 2
 - data_valid, 3
- *Topic **models**
 - model_evaluation, 4
- *Topic **model**
 - model_configuration, 3
- *Topic **possible**
 - model_configuration, 3

data_test, 1
data_train, 2
data_valid, 3

model_configuration, 3
model_evaluation, 4