Package 'partitionUcurve'

June 1, 2022

Type Package

Title U-curve Algorithm for Model Selection on the Partition Lattice Learning Space

Version 0.1

Author Diego Marcondes

Maintainer Diego Marcondes <dmarcondes@ime.usp.br>

Description U-curve Algorithm for Model Selection on the Partition Lattice Learning Space.

Imports plyr (>= 1.8.6),
fst,
ggplot2,
ggthemes,
numbers,
rlist,
stringi,
data.table

Depends R (>= 3.5.0)

License GPL (>= 3)

URL https://github.com/dmarcondes/PhDthesis

BugReports https://github.com/dmarcondes/PhDthesis/issues

Encoding UTF-8

LazyData true

RoxygenNote 7.1.1

References Diego Marcondes. A data-driven systematic, consistent and feasible approach to Model Selection. PhD thesis. Universidade de São Paulo, 2022

R topics documented:

ucurve *U-curve algorithm on the Partition Lattice Learning Space*

Description

Optimal and suboptimal U-curve algorithm to perfom Model Selection on the Partition Lattice Learning Space.

Usage

```
ucurve(
  xtrain,
  ytrain,
  xval,
  yval,
  X,
  Y,
  optimal = T,
  exhaust = 1000,
  sampleNeigh = F,
  increasing = F,
  earlystop = F,
  verbose = T,
  stop = 0,
  path = "~",
  Lh = NULL,
  cores = 4,
  save = F
)
```

Arguments

xtrain	Vector with the training sample of x values.
ytrain	Vector with the training sample of y values.
xval	Vector with the validation sample of x values.
yval	Vector with the validation sample of y values.
X	Domain of variable x.
Y	Domain of variable y.
optimal	Logical indicating if the algorithm should return an optimal solution.
exhaust	Number of points to exhaust before stopping the suboptimal algorithm.
sampleNeigh	Either false to consider all neighboors, or the maximum number of neighboors to sample at each exhaustion. If a number, then optimal should be false.
increasing	Logical indicating if target function is increasing.
earlystop	Logical indicating if it should be possible to early stop the algorithm.

verbose	Logical to print a trace of the algorithm.
stop	Number of nodes yet to evaluate to trigger exhaustive search.
path	Path to preprocessed partition files.
Lh	A data frame with the partition lattice.
cores	Number of cores for parallel computing.
save	Logical. Whether to save the nodes exhausted.

Details

The optimal U-curve algorithm works for at most 12 points in the domain, if the respetive fst file is available. The suboptimal algorithm in principle works for any number of points in the domain. The algorithm also performs a suboptimal search considering only increasing classifiers. More details about the algorithm may be found at the reference.

Value

hypotheses	The estimated hypothesis of the global minimums with least VC dimension.
partitions	Partitions of the global minimums with least VC dimension.
error	Validation error of the global minimums.
exhausted	Number of nodes exhausted during algorithm.
remain	Number of nodes remaining after optimal algorithm stopped.
finished	If the algorithm was finished or ended after not finding any Strong Local Minimum.
SLMvis	Number of nodes exhausted until the last Strong Local Minimum was found.
remain_after_prune	Number of nodes remaining after finding each Strong Local Minimum.
exhausted_until_prune	Number of nodes exhausted until finding each Strong Local Minimum.
optimal	Wheter an optimal solution was returned.
plot	Plot with the error of each Strong Local Minimum.

Examples

```
set.seed(1)
x <- sample(x = c("01","02","03","04","05","06","07","08","09","10"),size = 50,replace = TRUE)
y <- as.factor(ifelse(as.numeric(x)-5+rnorm(50,0,5/3) > 0,1,0))
x <- factor(x)
train <- sample(1:50,35,FALSE)
xtrain <- x[train]
ytrain <- y[train]
xval <- x[!(c(1:50) %in% train)]
yval <- y[!(c(1:50) %in% train)]
u <- ucurve(xtrain,ytrain,xval,yval,
X = c("01","02","03","04","05","06","07","08","09","10"),Y = c(0,1),
optimal = FALSE,sampleNeigh = 5000,exhaust = 10,cores = 1)
```

Index