Package 'conformal'

October 4, 2014

Type Package

Title Implementation of conformal prediction using caret models

Version 0.1

Date 04/10/2014

Author Isidro Cortes <isidrolauscher@gmail.com>

Maintainer Isidro Cortes <isidrolauscher@gmail.com>

Depends R (>= 2.12.0), caret, ggplot2, grid

Description Conformal prediction in R

License GPL

LazyLoad yes

R topics documented:

`ConformalReg` *Conformal Prediction for Regression*

Description

R class to create and visualize confidence intervals for individual predictions accorfing to the conformal prediction framework.

Usage

```
ConformalReg(...)
```

Arguments

...

Author(s)

Isidro Cort<c3><a9>s-Ciriano <isidrolauscher@gmail.com>

Examples

```
##########################################
### Example
##########################################

# Optional for parallel training
#library(doMC)
#registerDoMC(cores=4)

data(LogS)

algorithm <- "svmRadial"
tune.grid <- expand.grid(.sigma = expGrid(power.from=-10, power.to=-6, power.by=1, base=2),
                         .C = expGrid(power.from=4, power.to=10, power.by=2, base=2))
trControl <- trainControl(method = "cv",  number=5,savePredictions=TRUE)
set.seed(3)
model <- train(LogSDescsTrain, LogSTrain, algorithm,
               tuneGrid=tune.grid,
               trControl=trControl)


# Train an error model
ErrorModel <- error_model(PointPredictionModel=model,x.train=,LogSDescsTrain,
                          savePredictions=TRUE,algorithm=algorithm,
                          trControl=trControl, tune.grid=tune.grid)

# Instantiate the class and get the confidence intervals
example <- ConformalReg$new()
example$CalculateAlphas(model=model,error_model=ErrorModel,ConformityMeasure=StandardMeasure)
example$GetConfidenceIntervals(new.data=LogSDescsTest)
example$CorrelationPlot(obs=LogSTest)
example$plot
```

LogS *Small Molecule Solubility (LogS) Data*

Usage

```
data(LogS)
```

Details

describe these data: "Bakken and Jurs studied a set of compounds originally discussed by Klopman et al., who were interested in multidrug resistance reversal (MDRR) agents. The original response variable is a ratio measuring the ability of a compound to reverse a leukemia cell's resistance to adriamycin. However, the problem was treated as a classification problem, and compounds with

the ratio >4.2 were considered active, and those with the ratio <= 2.0 were considered inactive. Compounds with the ratio between these two cutoffs were called moderate and removed from the data for two class classification, leaving a set of 528 compounds (298 actives and 230 inactives). (Various other arrangements of these data were examined by Bakken and Jurs, but we will focus on this particular one.) We did not have access to the original descriptors, but we generated a set of 342 descriptors of three different types that should be similar to the original descriptors, using the DRAGON software." The data and R code are in the Supplemental Data file for the article.

Using \textttdata(LogS) exposes 4 objects: (i) \textttmdrrDescr is a data frame with the descriptor data and \textttmdrrClass is a factor vector with the activity results.

References

Wang et al (2007)

Examples

```
data(LogS)
## maybe str(LogS) ; plot(LogS) ...
```

Index