Data Slicing Demo - SPAM Example

Jiachang (Ernest) Xu

6/22/2017

```r
## require caret package for machine learning algorithms
require(caret)
```

```
## Loading required package: caret

## Loading required package: lattice

## Loading required package: ggplot2
```

```r
## require kernlab for spam data
require(kernlab)
```

```
## Loading required package: kernlab

##
## Attaching package: 'kernlab'

## The following object is masked from 'package:ggplot2':
##
##     alpha
```

```r
## data loading
data(spam)
```

SPAM Example: Data Splitting

```r
inTrain <- createDataPartition(y = spam$type, p = 0.75, list = FALSE)
training <- spam[inTrain, ]
testing <- spam[-inTrain, ]
dim(training)
```

```
## [1] 3451   58
```

SPAM Example: K-fold (returnTrain = TRUE)

```r
## k-fold
set.seed(32323)
folds.train <- createFolds(y = spam$type, k = 10, list = TRUE, returnTrain = TRUE)
sapply(folds.train, length)
```

```
## Fold01 Fold02 Fold03 Fold04 Fold05 Fold06 Fold07 Fold08 Fold09 Fold10
##   4141   4140   4141   4142   4140   4142   4141   4141   4140   4141
```

```r
folds.train[[1]][1:10]
```

```
##  [1]  1  2  3  4  5  6  7  8  9 10
```

SPAM Example: K-fold (returnTrain = FALSE)

```
## k-fold
set.seed(32323)
folds.test <- createFolds(y = spam$type, k = 10, list = TRUE, returnTrain = FALSE)
sapply(folds.test, length)
```

```
## Fold01 Fold02 Fold03 Fold04 Fold05 Fold06 Fold07 Fold08 Fold09 Fold10
##    460    461    460    459    461    459    460    460    461    460
```

```
folds.test[[1]][1:10]
```

```
##  [1] 24 27 32 40 41 43 55 58 63 68
```

SPAM Example: Resampling

```
set.seed(32323)
folds.resample <- createResample(y = spam$type, times = 10, list = TRUE)
sapply(folds.resample, length)
```

```
## Resample01 Resample02 Resample03 Resample04 Resample05 Resample06
##       4601       4601       4601       4601       4601       4601
## Resample07 Resample08 Resample09 Resample10
##       4601       4601       4601       4601
```

```
folds.resample[[1]][1:10]
```

```
##  [1]  1  2  3  3  3  5  5  7  8 12
```

SPAM Example: Times Slices

```
set.seed(32323)
time <- 1:10000
folds.time <- createTimeSlices(y = time, initialWindow = 20, horizon = 10)
names(folds.time)
```

```
## [1] "train" "test"
```

```
folds.time$train[[1]][1:10]
```

```
##  [1]  1  2  3  4  5  6  7  8  9 10
```

```
folds.time$test[[1]][1:10]
```

```
##  [1] 21 22 23 24 25 26 27 28 29 30
```