Maching Learning in R using mlR

Surag Gupta

August 20, 2017

About me

- Senior Analyst, Assortment Optimization
- MS Business Analytics, University of Minnesota (Class of 2017)
- which means I'm a relatively new team member
- Worked as a Decision Scientist for 3 years prior to my graduate program
- R Enthusiast (This would be a good time for you to notice my t-shirt, if you haven't already)

Disclaimer

Those who use Python for data science tasks shall not throw stuff at me during the session

What is mlR?

mlR is an R package developed to simplify basic and advanced data science tasks, such as:

- Data preparation
- Feature Engineering
- Training Models
- Validating models
- Predictions
- ...

We use the Kaggle credit card dataset to demonstrate the use of `mlR`

Load packages

```
##        Hmisc        caret         kknn          gbm          m
##         TRUE         TRUE         TRUE         TRUE          T
##      xgboost    lightgbm          mlr     parallel paralle
##         TRUE         TRUE         TRUE         TRUE          T
##  data.table   tidyverse
##         TRUE         TRUE
```

Import the dataset

```r
#Use fread() from the data.table package to read
df <- fread("creditcard.csv")

df <- df %>%
  data.frame() %>%
  mutate(Class = as.factor(Class)) #Convert target variable
```

```r
head(df)
```

```
##   Time          V1          V2        V3          V4
## 1    0 -1.3598071 -0.07278117 2.5363467  1.3781552 -0.33
## 2    0  1.1918571  0.26615071 0.1664801  0.4481541  0.06
## 3    1 -1.3583541 -1.34016307 1.7732093  0.3797796 -0.50
## 4    1 -0.9662717 -0.18522601 1.7929933 -0.8632913 -0.01
## 5    2 -1.1582331  0.87773675 1.5487178  0.4030339 -0.40
## 6    2 -0.4259659  0.96052304 1.1411093 -0.1682521  0.42
##            V7          V8         V9          V10
## 1  0.23959855  0.09869790  0.3637870  0.09079417 -0.5515
```