Prediction of Human Activity by Applying Machine Learning Algorithms

Jiachang (Ernest) Xu

6/23/2017

Sectiong 1: Synopsis

The objective of this project is to predict human activity by applying **machine learning** algorithms.

Section 2: Data Processing

First of all, before we do anything, we shall set the seed to 1024 for the purpose of reproducibility. Then, we shall download the training and testing datasets to the **./data** folder.

```
## set the seed for reproducibility
set.seed(1024)
## download training data
if (!file.exists("./data/training.csv")) {
    download.file("https://d396qusza40orc.cloudfront.net/predmachlearn/pml-training.csv",
                  destfile = "./data/training.csv")
}

## download testing data
if (!file.exists("./data/testing.csv")) {
    download.file("https://d396qusza40orc.cloudfront.net/predmachlearn/pml-testing.csv",
                  destfile = "./data/testing.csv")
}
```

After the training and testing datasets are downloaded, we shall read the datasets into R ready for data cleaning.

```
## load the full training data reday for data slicing
data <- read.csv(file = "./data/training.csv")
## loading 20 cases of testing data for validation
validation <- read.csv(file = "./data/testing.csv")
```

Section 3: Data Cleaning

Before actually starting the process of data cleaning, let's take a close look at **data** first.

```
class(data$classe)
```

```
## [1] "factor"
```

```
levels(data$classe)
```

```
## [1] ""  "A"
```

```
dim(data)
```

```
## [1] 2350  160
```

```
head(data)
```

```
##   X user_name raw_timestamp_part_1 raw_timestamp_part_2   cvtd_timestamp
## 1 1  carlitos           1323084231               788290 05/12/2011 11:23
## 2 2  carlitos           1323084231               808298 05/12/2011 11:23
## 3 3  carlitos           1323084231               820366 05/12/2011 11:23
## 4 4  carlitos           1323084232               120339 05/12/2011 11:23
## 5 5  carlitos           1323084232               196328 05/12/2011 11:23
## 6 6  carlitos           1323084232               304277 05/12/2011 11:23
##   new_window num_window roll_belt pitch_belt yaw_belt total_accel_belt
## 1         no         11      1.41       8.07    -94.4                3
## 2         no         11      1.41       8.07    -94.4                3
## 3         no         11      1.42       8.07    -94.4                3
## 4         no         12      1.48       8.05    -94.4                3
## 5         no         12      1.48       8.07    -94.4                3
## 6         no         12      1.45       8.06    -94.4                3
##   kurtosis_roll_belt kurtosis_picth_belt kurtosis_yaw_belt
## 1
## 2
## 3
## 4
## 5
## 6
##   skewness_roll_belt skewness_roll_belt.1 skewness_yaw_belt max_roll_belt
## 1                                                                    <NA>
## 2                                                                    <NA>
## 3                                                                    <NA>
## 4                                                                    <NA>
## 5                                                                    <NA>
## 6                                                                    <NA>
##   max_picth_belt max_yaw_belt min_roll_belt min_pitch_belt min_yaw_belt
## 1             NA                        NA             NA
## 2             NA                        NA             NA
## 3             NA                        NA             NA
## 4             NA                        NA             NA
## 5             NA                        NA             NA
## 6             NA                        NA             NA
##   amplitude_roll_belt amplitude_pitch_belt amplitude_yaw_belt
## 1                  NA                   NA
## 2                  NA                   NA
## 3                  NA                   NA
## 4                  NA                   NA
## 5                  NA                   NA
## 6                  NA                   NA
##   var_total_accel_belt avg_roll_belt stddev_roll_belt var_roll_belt
## 1                   NA            NA               NA            NA
## 2                   NA            NA               NA            NA
## 3                   NA            NA               NA            NA
## 4                   NA            NA               NA            NA
## 5                   NA            NA               NA            NA
## 6                   NA            NA               NA            NA
##   avg_pitch_belt stddev_pitch_belt var_pitch_belt avg_yaw_belt
## 1             NA                NA             NA           NA
## 2             NA                NA             NA           NA
```

```
## 3             NA                NA             NA           NA
## 4             NA                NA             NA           NA
## 5             NA                NA             NA           NA
## 6             NA                NA             NA           NA
##   stddev_yaw_belt var_yaw_belt gyros_belt_x gyros_belt_y gyros_belt_z
## 1              NA           NA         0.00         0.00        -0.02
## 2              NA           NA         0.02         0.00        -0.02
## 3              NA           NA         0.00         0.00        -0.02
## 4              NA           NA         0.02         0.00        -0.03
## 5              NA           NA         0.02         0.02        -0.02
## 6              NA           NA         0.02         0.00        -0.02
##   accel_belt_x accel_belt_y accel_belt_z magnet_belt_x magnet_belt_y
## 1          -21            4           22            -3           599
## 2          -22            4           22            -7           608
## 3          -20            5           23            -2           600
## 4          -22            3           21            -6           604
## 5          -21            2           24            -6           600
## 6          -21            4           21             0           603
##   magnet_belt_z roll_arm pitch_arm yaw_arm total_accel_arm var_accel_arm
## 1          -313     -128      22.5    -161              34            NA
## 2          -311     -128      22.5    -161              34            NA
## 3          -305     -128      22.5    -161              34            NA
## 4          -310     -128      22.1    -161              34            NA
## 5          -302     -128      22.1    -161              34            NA
## 6          -312     -128      22.0    -161              34            NA
##   avg_roll_arm stddev_roll_arm var_roll_arm avg_pitch_arm stddev_pitch_arm
## 1           NA              NA           NA            NA               NA
## 2           NA              NA           NA            NA               NA
## 3           NA              NA           NA            NA               NA
## 4           NA              NA           NA            NA               NA
## 5           NA              NA           NA            NA               NA
## 6           NA              NA           NA            NA               NA
##   var_pitch_arm avg_yaw_arm stddev_yaw_arm var_yaw_arm gyros_arm_x
## 1            NA          NA             NA          NA        0.00
## 2            NA          NA             NA          NA        0.02
## 3            NA          NA             NA          NA        0.02
## 4            NA          NA             NA          NA        0.02
## 5            NA          NA             NA          NA        0.00
## 6            NA          NA             NA          NA        0.02
##   gyros_arm_y gyros_arm_z accel_arm_x accel_arm_y accel_arm_z magnet_arm_x
## 1        0.00       -0.02        -288         109        -123         -368
## 2       -0.02       -0.02        -290         110        -125         -369
## 3       -0.02       -0.02        -289         110        -126         -368
## 4       -0.03        0.02        -289         111        -123         -372
## 5       -0.03        0.00        -289         111        -123         -374
## 6       -0.03        0.00        -289         111        -122         -369
##   magnet_arm_y magnet_arm_z kurtosis_roll_arm kurtosis_picth_arm
## 1          337          516
## 2          337          513
## 3          344          513
## 4          344          512
## 5          337          506
## 6          342          513
##   kurtosis_yaw_arm skewness_roll_arm skewness_pitch_arm skewness_yaw_arm
```

```
## 1
## 2
## 3
## 4
## 5
## 6
##   max_roll_arm max_picth_arm max_yaw_arm min_roll_arm min_pitch_arm
## 1           NA            NA          NA           NA            NA
## 2           NA            NA          NA           NA            NA
## 3           NA            NA          NA           NA            NA
## 4           NA            NA          NA           NA            NA
## 5           NA            NA          NA           NA            NA
## 6           NA            NA          NA           NA            NA
##   min_yaw_arm amplitude_roll_arm amplitude_pitch_arm amplitude_yaw_arm
## 1          NA                 NA                  NA                NA
## 2          NA                 NA                  NA                NA
## 3          NA                 NA                  NA                NA
## 4          NA                 NA                  NA                NA
## 5          NA                 NA                  NA                NA
## 6          NA                 NA                  NA                NA
##   roll_dumbbell pitch_dumbbell yaw_dumbbell kurtosis_roll_dumbbell
## 1      13.05217      -70.49400    -84.87394                     NA
## 2      13.13074      -70.63751    -84.71065                     NA
## 3      12.85075      -70.27812    -85.14078                     NA
## 4      13.43120      -70.39379    -84.87363                     NA
## 5      13.37872      -70.42856    -84.85306                     NA
## 6      13.38246      -70.81759    -84.46500                     NA
##   kurtosis_picth_dumbbell kurtosis_yaw_dumbbell skewness_roll_dumbbell
## 1                      NA                                           NA
## 2                      NA                                           NA
## 3                      NA                                           NA
## 4                      NA                                           NA
## 5                      NA                                           NA
## 6                      NA                                           NA
##   skewness_pitch_dumbbell skewness_yaw_dumbbell max_roll_dumbbell
## 1                      NA                                     NA
## 2                      NA                                     NA
## 3                      NA                                     NA
## 4                      NA                                     NA
## 5                      NA                                     NA
## 6                      NA                                     NA
##   max_picth_dumbbell max_yaw_dumbbell min_roll_dumbbell min_pitch_dumbbell
## 1                 NA               NA                NA                 NA
## 2                 NA               NA                NA                 NA
## 3                 NA               NA                NA                 NA
## 4                 NA               NA                NA                 NA
## 5                 NA               NA                NA                 NA
## 6                 NA               NA                NA                 NA
##   min_yaw_dumbbell amplitude_roll_dumbbell amplitude_pitch_dumbbell
## 1               NA                      NA                       NA
## 2               NA                      NA                       NA
## 3               NA                      NA                       NA
## 4               NA                      NA                       NA
## 5               NA                      NA                       NA
```

```
## 6               NA                      NA                      NA
##   amplitude_yaw_dumbbell total_accel_dumbbell var_accel_dumbbell
## 1                     NA                  37                 NA
## 2                     NA                  37                 NA
## 3                     NA                  37                 NA
## 4                     NA                  37                 NA
## 5                     NA                  37                 NA
## 6                     NA                  37                 NA
##   avg_roll_dumbbell stddev_roll_dumbbell var_roll_dumbbell
## 1                NA                   NA                NA
## 2                NA                   NA                NA
## 3                NA                   NA                NA
## 4                NA                   NA                NA
## 5                NA                   NA                NA
## 6                NA                   NA                NA
##   avg_pitch_dumbbell stddev_pitch_dumbbell var_pitch_dumbbell
## 1                 NA                    NA                 NA
## 2                 NA                    NA                 NA
## 3                 NA                    NA                 NA
## 4                 NA                    NA                 NA
## 5                 NA                    NA                 NA
## 6                 NA                    NA                 NA
##   avg_yaw_dumbbell stddev_yaw_dumbbell var_yaw_dumbbell gyros_dumbbell_x
## 1               NA                  NA               NA                0
## 2               NA                  NA               NA                0
## 3               NA                  NA               NA                0
## 4               NA                  NA               NA                0
## 5               NA                  NA               NA                0
## 6               NA                  NA               NA                0
##   gyros_dumbbell_y gyros_dumbbell_z accel_dumbbell_x accel_dumbbell_y
## 1            -0.02             0.00             -234               47
## 2            -0.02             0.00             -233               47
## 3            -0.02             0.00             -232               46
## 4            -0.02            -0.02             -232               48
## 5            -0.02             0.00             -233               48
## 6            -0.02             0.00             -234               48
##   accel_dumbbell_z magnet_dumbbell_x magnet_dumbbell_y magnet_dumbbell_z
## 1             -271              -559               293               -65
## 2             -269              -555               296               -64
## 3             -270              -561               298               -63
## 4             -269              -552               303               -60
## 5             -270              -554               292               -68
## 6             -269              -558               294               -66
##   roll_forearm pitch_forearm yaw_forearm kurtosis_roll_forearm
## 1         28.4         -63.9       -153
## 2         28.3         -63.9       -153
## 3         28.3         -63.9       -152
## 4         28.1         -63.9       -152
## 5         28.0         -63.9       -152
## 6         27.9         -63.9       -152
##   kurtosis_picth_forearm kurtosis_yaw_forearm skewness_roll_forearm
## 1
## 2
## 3
```

```
## 4
## 5
## 6
##   skewness_pitch_forearm skewness_yaw_forearm max_roll_forearm
## 1                                                           NA
## 2                                                           NA
## 3                                                           NA
## 4                                                           NA
## 5                                                           NA
## 6                                                           NA
##   max_picth_forearm max_yaw_forearm min_roll_forearm min_pitch_forearm
## 1                NA                              NA                NA
## 2                NA                              NA                NA
## 3                NA                              NA                NA
## 4                NA                              NA                NA
## 5                NA                              NA                NA
## 6                NA                              NA                NA
##   min_yaw_forearm amplitude_roll_forearm amplitude_pitch_forearm
## 1                                    NA                      NA
## 2                                    NA                      NA
## 3                                    NA                      NA
## 4                                    NA                      NA
## 5                                    NA                      NA
## 6                                    NA                      NA
##   amplitude_yaw_forearm total_accel_forearm var_accel_forearm
## 1                                        36                NA
## 2                                        36                NA
## 3                                        36                NA
## 4                                        36                NA
## 5                                        36                NA
## 6                                        36                NA
##   avg_roll_forearm stddev_roll_forearm var_roll_forearm avg_pitch_forearm
## 1               NA                  NA                NA                NA
## 2               NA                  NA                NA                NA
## 3               NA                  NA                NA                NA
## 4               NA                  NA                NA                NA
## 5               NA                  NA                NA                NA
## 6               NA                  NA                NA                NA
##   stddev_pitch_forearm var_pitch_forearm avg_yaw_forearm
## 1                   NA                NA               NA
## 2                   NA                NA               NA
## 3                   NA                NA               NA
## 4                   NA                NA               NA
## 5                   NA                NA               NA
## 6                   NA                NA               NA
##   stddev_yaw_forearm var_yaw_forearm gyros_forearm_x gyros_forearm_y
## 1                 NA              NA            0.03            0.00
## 2                 NA              NA            0.02            0.00
## 3                 NA              NA            0.03           -0.02
## 4                 NA              NA            0.02           -0.02
## 5                 NA              NA            0.02            0.00
## 6                 NA              NA            0.02           -0.02
##   gyros_forearm_z accel_forearm_x accel_forearm_y accel_forearm_z
## 1           -0.02             192             203            -215
```

```
## 2           -0.02              192              203            -216
## 3            0.00              196              204            -213
## 4            0.00              189              206            -214
## 5           -0.02              189              206            -214
## 6           -0.03              193              203            -215
##   magnet_forearm_x magnet_forearm_y magnet_forearm_z classe
## 1              -17              654              476      A
## 2              -18              661              473      A
## 3              -18              658              469      A
## 4              -16              658              469      A
## 5              -17              655              473      A
## 6               -9              660              478      A
```

We can see from the output above that there exist a lot of empty spaces and NA values. Let's identify the level of NA value in **data**.

```
## identify NA level
NA.levels <- unique(apply(data, 2, function(x) {sum(is.na(x))} ))
NA.number <- dim(data)[1]-NA.levels[2]
NA.non <- NA.number/dim(data)[1]
sprintf("%1.2f%%", 100*NA.non)
```

```
## [1] "2.30%"
```

Then, we can replace empty spaces and div0 to NA

```
data[data == ""] <- NA
data[data == "#DIV/0!"] <- NA
data[data == "<NA>"] <- NA
```

Now, there are no empty spaces or irregular values in **data**, we shall spitt **data** to **train.data** and **test.data**

```
require(caret)
```

```
## Loading required package: caret
```

```
## Loading required package: lattice
```

```
## Loading required package: ggplot2
```

```
inTrain <- createDataPartition(y = data$classe, p = 0.75, list = FALSE)
```

```
## Warning in createDataPartition(y = data$classe, p = 0.75, list = FALSE):
## Some classes have no records ( ) and these will be ignored
```

```
train.data <- data[inTrain, ]
test.data <- data[-inTrain, ]
dim(train.data)
```

```
## [1] 1763  160
```

We split **train.data** to old window rows (non-aggregated).

```
## select non-aggregated sensor data
train.old.window <- train.data[which(train.data$new_window == "no"), ]
## sensor data without NA columns (summary data)
train.old.window <- train.data[!colSums(is.na(train.data)) > 0]
## test NA purity
sum(is.na(train.old.window))
```

```
## [1] 0
```

We split **train.data** and **test.data** to new window rows (aggregated), and remove NA columns and rows from the new training and testing data frames.

```
## select aggregated sensor data
train.new.window <- train.data[which(train.data$new_window == "yes"), ]
test.new.window <- test.data[which(test.data$new_window == "yes"), ]

## remove full NA columns
train.new.half <- subset(train.new.window,
                         select = -c(kurtosis_picth_belt, kurtosis_yaw_belt, kurtosis_picth_arm,
                                     kurtosis_yaw_arm, skewness_pitch_arm, kurtosis_yaw_dumbbell,
                                     skewness_yaw_dumbbell, skewness_yaw_forearm, kurtosis_yaw_forearm
                                     skewness_yaw_belt, skewness_roll_belt.1))
test.new.half <- subset(test.new.window,
                        select = -c(kurtosis_picth_belt, kurtosis_yaw_belt, kurtosis_picth_arm,
                                    kurtosis_yaw_arm, skewness_pitch_arm, kurtosis_yaw_dumbbell,
                                    skewness_yaw_dumbbell, skewness_yaw_forearm, kurtosis_yaw_forearm,
                                    skewness_yaw_belt, skewness_roll_belt.1))

## remove NA rows from training data
train.new.clean <- train.new.half[complete.cases(train.new.half), ]
sum(is.na(train.new.clean))
```

```
## [1] 0
```

```
## remove NA rows from testing data
test.new.clean <- test.new.half[complete.cases(test.new.half), ]
sum(is.na(test.new.clean))
```

```
## [1] 0
```

Section 4: Machine Learning

Section 5: Conclusion