Practical Machine Learning CourseProject

Nicole

2016/05/23

Step1. Download & assign data

The 1st step is to load the data into R dataset. And assign the training & testing data into variables final_training, final_testing respectively.

Then, for the final_training dataset, we seperate it pre_training & pre_testing (70% & 30%) for modeling.

```
final_training <- read.csv("D:\\Coursera\\Material\\08. Practical Machine Learning\\CourseProject\\pml-
final_testing <- read.csv("D:\\Coursera\\Material\\08. Practical Machine Learning\\CourseProject\\pml-te

library(lattice);library(ggplot2);library(caret);
set.seed(33833)
inTrain <- createDataPartition(y=final_training$classe,p=0.7,list=FALSE)
pre_training <- final_training[inTrain,]
pre_testing <- final_testing[-inTrain,]
```

Step2. Data preprocess & Variable selection

Let's do the briefly data explorer. As you can see, there are too many NA or Null variables in the dataset. We remove it from our modeling.

```
head(pre_training)
```

```
##   X user_name raw_timestamp_part_1 raw_timestamp_part_2   cvtd_timestamp
## 1 1  carlitos           1323084231               788290 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
## 7 7  carlitos           1323084232               368296 05/12/2011 11:23
## 8 8  carlitos           1323084232               440390 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
## 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
## 7         no         12      1.42       8.09    -94.4                3
## 8         no         12      1.42       8.13    -94.4                3
##   kurtosis_roll_belt kurtosis_picth_belt kurtosis_yaw_belt
## 1
## 3
## 4
## 5
## 7
## 8
##   skewness_roll_belt skewness_roll_belt.1 skewness_yaw_belt max_roll_belt
## 1                                                                      NA
```

```
## 3                                                                      NA
## 4                                                                      NA
## 5                                                                      NA
## 7                                                                      NA
## 8                                                                      NA
##   max_picth_belt max_yaw_belt min_roll_belt min_pitch_belt min_yaw_belt
## 1             NA                         NA             NA
## 3             NA                         NA             NA
## 4             NA                         NA             NA
## 5             NA                         NA             NA
## 7             NA                         NA             NA
## 8             NA                         NA             NA
##   amplitude_roll_belt amplitude_pitch_belt amplitude_yaw_belt
## 1                  NA                   NA
## 3                  NA                   NA
## 4                  NA                   NA
## 5                  NA                   NA
## 7                  NA                   NA
## 8                  NA                   NA
##   var_total_accel_belt avg_roll_belt stddev_roll_belt var_roll_belt
## 1                   NA            NA               NA            NA
## 3                   NA            NA               NA            NA
## 4                   NA            NA               NA            NA
## 5                   NA            NA               NA            NA
## 7                   NA            NA               NA            NA
## 8                   NA            NA               NA            NA
##   avg_pitch_belt stddev_pitch_belt var_pitch_belt avg_yaw_belt
## 1             NA                NA             NA           NA
## 3             NA                NA             NA           NA
## 4             NA                NA             NA           NA
## 5             NA                NA             NA           NA
## 7             NA                NA             NA           NA
## 8             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
## 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
## 7              NA           NA         0.02         0.00        -0.02
## 8              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
## 3          -20            5           23            -2           600
## 4          -22            3           21            -6           604
## 5          -21            2           24            -6           600
## 7          -22            3           21            -4           599
## 8          -22            4           21            -2           603
##   magnet_belt_z roll_arm pitch_arm yaw_arm total_accel_arm var_accel_arm
## 1          -313     -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
## 7          -311     -128      21.9    -161              34            NA
## 8          -313     -128      21.8    -161              34            NA
```

```
##   avg_roll_arm stddev_roll_arm var_roll_arm avg_pitch_arm stddev_pitch_arm
## 1           NA              NA           NA            NA               NA
## 3           NA              NA           NA            NA               NA
## 4           NA              NA           NA            NA               NA
## 5           NA              NA           NA            NA               NA
## 7           NA              NA           NA            NA               NA
## 8           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
## 3            NA          NA             NA          NA        0.02
## 4            NA          NA             NA          NA        0.02
## 5            NA          NA             NA          NA        0.00
## 7            NA          NA             NA          NA        0.00
## 8            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
## 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
## 7       -0.03        0.00        -289         111        -125         -373
## 8       -0.02        0.00        -289         111        -124         -372
##   magnet_arm_y magnet_arm_z kurtosis_roll_arm kurtosis_picth_arm
## 1          337          516
## 3          344          513
## 4          344          512
## 5          337          506
## 7          336          509
## 8          338          510
##   kurtosis_yaw_arm skewness_roll_arm skewness_pitch_arm skewness_yaw_arm
## 1
## 3
## 4
## 5
## 7
## 8
##   max_roll_arm max_picth_arm max_yaw_arm min_roll_arm min_pitch_arm
## 1           NA            NA          NA           NA            NA
## 3           NA            NA          NA           NA            NA
## 4           NA            NA          NA           NA            NA
## 5           NA            NA          NA           NA            NA
## 7           NA            NA          NA           NA            NA
## 8           NA            NA          NA           NA            NA
##   min_yaw_arm amplitude_roll_arm amplitude_pitch_arm amplitude_yaw_arm
## 1          NA                 NA                  NA                NA
## 3          NA                 NA                  NA                NA
## 4          NA                 NA                  NA                NA
## 5          NA                 NA                  NA                NA
## 7          NA                 NA                  NA                NA
## 8          NA                 NA                  NA                NA
##   roll_dumbbell pitch_dumbbell yaw_dumbbell kurtosis_roll_dumbbell
## 1      13.05217      -70.49400    -84.87394
## 3      12.85075      -70.27812    -85.14078
## 4      13.43120      -70.39379    -84.87363
## 5      13.37872      -70.42856    -84.85306
```

```
## 7      13.12695      -70.24757    -85.09961
## 8      12.75083      -70.34768    -85.09708
##   kurtosis_picth_dumbbell kurtosis_yaw_dumbbell skewness_roll_dumbbell
## 1
## 3
## 4
## 5
## 7
## 8
##   skewness_pitch_dumbbell skewness_yaw_dumbbell max_roll_dumbbell
## 1                                                              NA
## 3                                                              NA
## 4                                                              NA
## 5                                                              NA
## 7                                                              NA
## 8                                                              NA
##   max_picth_dumbbell max_yaw_dumbbell min_roll_dumbbell min_pitch_dumbbell
## 1                 NA                                NA                 NA
## 3                 NA                                NA                 NA
## 4                 NA                                NA                 NA
## 5                 NA                                NA                 NA
## 7                 NA                                NA                 NA
## 8                 NA                                NA                 NA
##   min_yaw_dumbbell amplitude_roll_dumbbell amplitude_pitch_dumbbell
## 1                                       NA                      NA
## 3                                       NA                      NA
## 4                                       NA                      NA
## 5                                       NA                      NA
## 7                                       NA                      NA
## 8                                       NA                      NA
##   amplitude_yaw_dumbbell total_accel_dumbbell var_accel_dumbbell
## 1                                         37                NA
## 3                                         37                NA
## 4                                         37                NA
## 5                                         37                NA
## 7                                         37                NA
## 8                                         37                NA
##   avg_roll_dumbbell stddev_roll_dumbbell var_roll_dumbbell
## 1                NA                   NA                NA
## 3                NA                   NA                NA
## 4                NA                   NA                NA
## 5                NA                   NA                NA
## 7                NA                   NA                NA
## 8                NA                   NA                NA
##   avg_pitch_dumbbell stddev_pitch_dumbbell var_pitch_dumbbell
## 1                 NA                    NA                 NA
## 3                 NA                    NA                 NA
## 4                 NA                    NA                 NA
## 5                 NA                    NA                 NA
## 7                 NA                    NA                 NA
## 8                 NA                    NA                 NA
##   avg_yaw_dumbbell stddev_yaw_dumbbell var_yaw_dumbbell gyros_dumbbell_x
## 1               NA                  NA               NA                0
## 3               NA                  NA               NA                0
```

```
## 4               NA                  NA               NA                0
## 5               NA                  NA               NA                0
## 7               NA                  NA               NA                0
## 8               NA                  NA               NA                0
##   gyros_dumbbell_y gyros_dumbbell_z accel_dumbbell_x accel_dumbbell_y
## 1            -0.02             0.00             -234               47
## 3            -0.02             0.00             -232               46
## 4            -0.02            -0.02             -232               48
## 5            -0.02             0.00             -233               48
## 7            -0.02             0.00             -232               47
## 8            -0.02             0.00             -234               46
##   accel_dumbbell_z magnet_dumbbell_x magnet_dumbbell_y magnet_dumbbell_z
## 1             -271              -559               293               -65
## 3             -270              -561               298               -63
## 4             -269              -552               303               -60
## 5             -270              -554               292               -68
## 7             -270              -551               295               -70
## 8             -272              -555               300               -74
##   roll_forearm pitch_forearm yaw_forearm kurtosis_roll_forearm
## 1         28.4         -63.9        -153
## 3         28.3         -63.9        -152
## 4         28.1         -63.9        -152
## 5         28.0         -63.9        -152
## 7         27.9         -63.9        -152
## 8         27.8         -63.8        -152
##   kurtosis_picth_forearm kurtosis_yaw_forearm skewness_roll_forearm
## 1
## 3
## 4
## 5
## 7
## 8
##   skewness_pitch_forearm skewness_yaw_forearm max_roll_forearm
## 1                                                           NA
## 3                                                           NA
## 4                                                           NA
## 5                                                           NA
## 7                                                           NA
## 8                                                           NA
##   max_picth_forearm max_yaw_forearm min_roll_forearm min_pitch_forearm
## 1                NA                                NA                NA
## 3                NA                                NA                NA
## 4                NA                                NA                NA
## 5                NA                                NA                NA
## 7                NA                                NA                NA
## 8                NA                                NA                NA
##   min_yaw_forearm amplitude_roll_forearm amplitude_pitch_forearm
## 1                                      NA                     NA
## 3                                      NA                     NA
## 4                                      NA                     NA
## 5                                      NA                     NA
## 7                                      NA                     NA
## 8                                      NA                     NA
##   amplitude_yaw_forearm total_accel_forearm var_accel_forearm
```

```
## 1                                     36               NA
## 3                                     36               NA
## 4                                     36               NA
## 5                                     36               NA
## 7                                     36               NA
## 8                                     36               NA
##   avg_roll_forearm stddev_roll_forearm var_roll_forearm avg_pitch_forearm
## 1               NA                  NA               NA                NA
## 3               NA                  NA               NA                NA
## 4               NA                  NA               NA                NA
## 5               NA                  NA               NA                NA
## 7               NA                  NA               NA                NA
## 8               NA                  NA               NA                NA
##   stddev_pitch_forearm var_pitch_forearm avg_yaw_forearm
## 1                   NA                NA              NA
## 3                   NA                NA              NA
## 4                   NA                NA              NA
## 5                   NA                NA              NA
## 7                   NA                NA              NA
## 8                   NA                NA              NA
##   stddev_yaw_forearm var_yaw_forearm gyros_forearm_x gyros_forearm_y
## 1                 NA              NA            0.03            0.00
## 3                 NA              NA            0.03           -0.02
## 4                 NA              NA            0.02           -0.02
## 5                 NA              NA            0.02            0.00
## 7                 NA              NA            0.02            0.00
## 8                 NA              NA            0.02           -0.02
##   gyros_forearm_z accel_forearm_x accel_forearm_y accel_forearm_z
## 1           -0.02             192             203            -215
## 3            0.00             196             204            -213
## 4            0.00             189             206            -214
## 5           -0.02             189             206            -214
## 7           -0.02             195             205            -215
## 8            0.00             193             205            -213
##   magnet_forearm_x magnet_forearm_y magnet_forearm_z classe
## 1              -17              654              476      A
## 3              -18              658              469      A
## 4              -16              658              469      A
## 5              -17              655              473      A
## 7              -18              659              470      A
## 8               -9              660              474      A
```

```
colIdx <- c(7:11,37:49,60:68,84:86,102,113:124,140,151:159,160)
training <- final_training[inTrain,colIdx]
testing <- final_training[-inTrain,colIdx]
```

Step3. Start modeling

Due to performance consideration. Here, I only choose two modeling “rpart” & lda“.

```
library(rpart);library(MASS);
memory.limit(60000)
```

```
## [1] 60000
```

```
set.seed(33833)
rpart <- train(classe~., data=training[,-1],method="rpart")
lda <- train(classe~., data=training[,-1],method="lda")
rpart
```

```
## CART
##
## 13737 samples
##    52 predictor
##     5 classes: 'A', 'B', 'C', 'D', 'E'
##
## No pre-processing
## Resampling: Bootstrapped (25 reps)
## Summary of sample sizes: 13737, 13737, 13737, 13737, 13737, 13737, ...
## Resampling results across tuning parameters:
##
##   cp          Accuracy   Kappa
##   0.03722917  0.5017537  0.35068186
##   0.06133659  0.4229226  0.22020767
##   0.11484081  0.3323672  0.07387813
##
## Accuracy was used to select the optimal model using  the largest value.
## The final value used for the model was cp = 0.03722917.
```

```
lda
```

```
## Linear Discriminant Analysis
##
## 13737 samples
##    52 predictor
##     5 classes: 'A', 'B', 'C', 'D', 'E'
##
## No pre-processing
## Resampling: Bootstrapped (25 reps)
## Summary of sample sizes: 13737, 13737, 13737, 13737, 13737, 13737, ...
## Resampling results:
##
##   Accuracy   Kappa
##   0.6998909  0.6204494
##
##
```

Step4. Out of Sample Error Estimation (Model selection by accurance)

Select to better model by accurance. Here, the rpart get 49% score & lda get 70% score. Hence, I choose lda as my final model.

```
pred.rpart <- predict(rpart,testing)
pred.lda <- predict(lda,testing)
sum(pred.rpart == testing$classe) / length(testing$classe)
```

```
## [1] 0.4895497
```

```
sum(pred.lda == testing$classe) / length(testing$classe)
```

```
## [1] 0.6987256
```

Step5. The prediction results in

```
pred.rpart <- predict(rpart,final_testing)
pred.rpart
```

```
##  [1] C A C A A C C A A A C C C A C A A A A C
## Levels: A B C D E
```