Practical Machine Learning Assignment

Toshiyuki Tachibana

2017-1-18

INTRODUCTION

This is document describing the analysis I conducted for my final project for the Johns Hopkins' Coursera course "Practical Machine Learning" in the Data Science specialization. One thing that people regularly do is quantify how much of a particular activity they do, but they rarely quantify how well they do it. In this project, your goal will be to use data from accelerometers on the belt, forearm, arm, and dumbell of 6 participants.

The goal of your project is to predict the manner in which they did the exercise. I describe how to built my models, how I used cross validation, what I think the expected out of sample error is, and why I made the choices I did. I also used my prediction model to predict 20 different test cases.

Data

The training data for this project are available here:

https://d396qusza40orc.cloudfront.net/predmachlearn/pml-training.csv

The test data are available here:

https://d396qusza40orc.cloudfront.net/predmachlearn/pml-testing.csv

Load Data and conduct initial data exploration

Set working directory

```
setwd("C:/Users/Tachibana/Documents/GitHub/Practical_Machine_learning")
```

Load Data

```
trainUrl <- "http://d396qusza40orc.cloudfront.net/predmachlearn/pml-training.csv"

testUrl <- "http://d396qusza40orc.cloudfront.net/predmachlearn/pml-testing.csv"

Rawtraining <- data.frame(read.csv(trainUrl, header=TRUE))

Rawtesting <- data.frame(read.csv(testUrl, header=TRUE))

head(Rawtraining)
```

```
##   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
## 2      13.13074      -70.63751    -84.71065
## 3      12.85075      -70.27812    -85.14078
## 4      13.43120      -70.39379    -84.87363
## 5      13.37872      -70.42856    -84.85306
## 6      13.38246      -70.81759    -84.46500
##   kurtosis_picth_dumbbell kurtosis_yaw_dumbbell skewness_roll_dumbbell
## 1
## 2
## 3
## 4
## 5
## 6
##   skewness_pitch_dumbbell skewness_yaw_dumbbell max_roll_dumbbell
## 1                                                              NA
## 2                                                              NA
## 3                                                              NA
## 4                                                              NA
## 5                                                              NA
## 6                                                              NA
##   max_picth_dumbbell max_yaw_dumbbell min_roll_dumbbell min_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
##   min_yaw_dumbbell amplitude_roll_dumbbell amplitude_pitch_dumbbell
## 1                                       NA                       NA
## 2                                       NA                       NA
## 3                                       NA                       NA
## 4                                       NA                       NA
## 5                                       NA                       NA
## 6                                       NA                       NA
##   amplitude_yaw_dumbbell total_accel_dumbbell var_accel_dumbbell
## 1                                         37                NA
## 2                                         37                NA
## 3                                         37                NA
## 4                                         37                NA
## 5                                         37                NA
## 6                                         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
```

Partioning the training set into two

```
library(caret)
```

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

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

```
set.seed(975)

inTrain = createDataPartition(Rawtraining$classe, p = 0.7)[[1]]

training = Rawtraining[ inTrain,]

testing = Rawtraining[-inTrain,]
```

Clean Data

Remove categorical variables, leaving only the sensor readings

```
df <- training[,8:ncol(training)]
```

Remove Columns near to Zero

```
df_nzv <- nearZeroVar(df, saveMetrics=TRUE)
remaining <- df_nzv[which(df_nzv$nzv==FALSE),]

df_all_var <- subset(df , select=rownames(remaining))
```

Remove Columns with NAs

```
df_rm_na <- df_all_var[ , colSums(is.na(df_all_var)) == 0]
apply(df_rm_na,2,function(x) {all(is.na(df_all_var))})
```

```
##            roll_belt           pitch_belt             yaw_belt
##                FALSE                FALSE                FALSE
##     total_accel_belt         gyros_belt_x         gyros_belt_y
##                FALSE                FALSE                FALSE
##         gyros_belt_z         accel_belt_x         accel_belt_y
##                FALSE                FALSE                FALSE
##         accel_belt_z        magnet_belt_x        magnet_belt_y
##                FALSE                FALSE                FALSE
##        magnet_belt_z             roll_arm            pitch_arm
##                FALSE                FALSE                FALSE
##              yaw_arm      total_accel_arm          gyros_arm_x
##                FALSE                FALSE                FALSE
```

```
##          gyros_arm_y          gyros_arm_z          accel_arm_x
##                FALSE                FALSE                FALSE
##          accel_arm_y          accel_arm_z         magnet_arm_x
##                FALSE                FALSE                FALSE
##         magnet_arm_y         magnet_arm_z        roll_dumbbell
##                FALSE                FALSE                FALSE
##       pitch_dumbbell         yaw_dumbbell total_accel_dumbbell
##                FALSE                FALSE                FALSE
##     gyros_dumbbell_x     gyros_dumbbell_y     gyros_dumbbell_z
##                FALSE                FALSE                FALSE
##     accel_dumbbell_x     accel_dumbbell_y     accel_dumbbell_z
##                FALSE                FALSE                FALSE
##    magnet_dumbbell_x    magnet_dumbbell_y    magnet_dumbbell_z
##                FALSE                FALSE                FALSE
##         roll_forearm        pitch_forearm          yaw_forearm
##                FALSE                FALSE                FALSE
##  total_accel_forearm      gyros_forearm_x      gyros_forearm_y
##                FALSE                FALSE                FALSE
##      gyros_forearm_z      accel_forearm_x      accel_forearm_y
##                FALSE                FALSE                FALSE
##      accel_forearm_z     magnet_forearm_x     magnet_forearm_y
##                FALSE                FALSE                FALSE
##     magnet_forearm_z               classe
##                FALSE                FALSE
```

```
table(complete.cases(df_rm_na))
```

```
##
##  TRUE
## 13737
```

```
table(sapply(df_rm_na[1,], class))
```

```
##
##  factor integer numeric
##       1      25      27
```

Train Model

We used randomForest function in R to fit the predictors to the training set.

```
library(randomForest)
```

```
## randomForest 4.6-12
```

```
## Type rfNews() to see new features/changes/bug fixes.
```

```
##
## Attaching package: 'randomForest'
```

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

```
library(e1071)
set.seed(123)
modFit<- train(classe~ . , data = df_rm_na, method='rf')
```

```
save(modFit, file="modFit.RData")
print(modFit$finalModel)
```

```
##
## Call:
##  randomForest(x = x, y = y, mtry = param$mtry)
##                Type of random forest: classification
##                      Number of trees: 500
## No. of variables tried at each split: 27
##
##         OOB estimate of  error rate: 0.65%
## Confusion matrix:
##      A    B    C    D    E class.error
## A 3898    7    0    0    1 0.002048131
## B   16 2632    8    1    1 0.009781791
## C    0    8 2377   11    0 0.007929883
## D    0    0   24 2226    2 0.011545293
## E    0    1    2    7 2515 0.003960396
```

Look at the variable importance to the model

```
varImp(modFit, useModel=TRUE)
```

```
## rf variable importance
##
##   only 20 most important variables shown (out of 52)
##
##                      Overall
## roll_belt             100.00
## pitch_forearm          61.89
## yaw_belt               57.45
## pitch_belt             46.02
## magnet_dumbbell_z      44.62
## magnet_dumbbell_y      42.50
## roll_forearm           42.13
## accel_dumbbell_y       20.36
## magnet_dumbbell_x      19.97
## roll_dumbbell          19.59
## accel_forearm_x        17.59
## accel_dumbbell_z       15.00
## magnet_belt_z          14.66
## accel_belt_z           14.34
## total_accel_dumbbell   13.61
## magnet_belt_y          13.21
```

```
## magnet_forearm_z       13.05
## gyros_belt_z           11.44
## magnet_belt_x          11.22
## yaw_arm                10.95
```

Predict values on testing data set

```
predictions <- predict(modFit, newdata = testing)
pred <- data.frame(predictions, classe=testing$classe)

confusionMatrix(predictions, testing$classe)
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction    A    B    C    D    E
##          A 1669    6    0    0    0
##          B    1 1130    0    0    0
##          C    4    3 1023   13    2
##          D    0    0    3  951    3
##          E    0    0    0    0 1077
##
## Overall Statistics
##
##                Accuracy : 0.9941
##                  95% CI : (0.9917, 0.9959)
##     No Information Rate : 0.2845
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.9925
##  Mcnemar's Test P-Value : NA
##
## Statistics by Class:
##
##                      Class: A Class: B Class: C Class: D Class: E
## Sensitivity            0.9970   0.9921   0.9971   0.9865   0.9954
## Specificity            0.9986   0.9998   0.9955   0.9988   1.0000
## Pos Pred Value         0.9964   0.9991   0.9789   0.9937   1.0000
## Neg Pred Value         0.9988   0.9981   0.9994   0.9974   0.9990
## Prevalence             0.2845   0.1935   0.1743   0.1638   0.1839
## Detection Rate         0.2836   0.1920   0.1738   0.1616   0.1830
## Detection Prevalence   0.2846   0.1922   0.1776   0.1626   0.1830
## Balanced Accuracy      0.9978   0.9959   0.9963   0.9926   0.9977
```

Use random forest model to predict the outcome of the 20 test cases for submission

```
submission_outcomes <- predict(modFit, newdata = Rawtesting)
submission_outcomes
```

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