Weight Lifting Exercise Prediction

Hathairat

18 May 2019

Executive Summary

This analysis uses machine learning to predict the manner of 6 participants in their exercise. The manner of exercise is represented in 5 classes (A:E). The author created a predicting model using random forest method from the training dataset. The model is later used to predict 20 cases in the test dataset.

Data Preparation

The training and testing dataset is available in http://groupware.les.inf.puc-rio.br/har. I downloaded the dataset into my working directory and read the data which I called TrainData and TestData.

```
library(dplyr); library(ggplot2); library(caret)
TrainData <- read.csv("~/pml-training.csv")
TestData <- read.csv("~/pml-testing.csv")
```

The Train data has 160 variables and 19,622 observations. I remove any rows that have NAs or blank cells and also the columns with descriptive texts. Eventually the Traindata has 53 columns. Classe which is one of 53 columns is a predictor.

```
TrainData[TrainData == ""] <- NA
TrainData <- TrainData[ , colSums(is.na(TrainData)) == 0]
TrainData <- subset(TrainData, select = -c(1:7))
```

A a quick glance and statistical summary of 53 variables which will be used in the analysis are included in **Appendix - Exploratory Data Analysis**.

Before I started prediction models, I split the Train data into testing and training dataset. I also set seed for reproducible purposes.

```
inTrain <- createDataPartition(y=TrainData$classe,p=0.7,list=FALSE)
training <- TrainData[inTrain,]
testing <- TrainData[-inTrain,]
set.seed(123)
```

Creating model with random forests

Since the data has labels, I choose random forest which is one of the supervised predicting methods to build a machine learning model. This is mainly because random forest method can handle classification prediction and returns high accuracy.

```
mod1 <- train(classe~., method = "rf", data =training)
pred1 <- predict(mod1, testing)
confusionMatrix(pred1, testing$classe)
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction    A    B    C    D    E
##          A 1674    8    0    1    0
##          B    0 1129    2    0    0
##          C    0    2 1021    5    3
```

```
##          D    0    0    3  957    8
##          E    0    0    0    1 1071
##
## Overall Statistics
##
##                Accuracy : 0.9944
##                  95% CI : (0.9921, 0.9961)
##     No Information Rate : 0.2845
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.9929
##
##  Mcnemar's Test P-Value : NA
##
## Statistics by Class:
##
##                      Class: A Class: B Class: C Class: D Class: E
## Sensitivity            1.0000   0.9912   0.9951   0.9927   0.9898
## Specificity            0.9979   0.9996   0.9979   0.9978   0.9998
## Pos Pred Value         0.9947   0.9982   0.9903   0.9886   0.9991
## Neg Pred Value         1.0000   0.9979   0.9990   0.9986   0.9977
## Prevalence             0.2845   0.1935   0.1743   0.1638   0.1839
## Detection Rate         0.2845   0.1918   0.1735   0.1626   0.1820
## Detection Prevalence   0.2860   0.1922   0.1752   0.1645   0.1822
## Balanced Accuracy      0.9989   0.9954   0.9965   0.9953   0.9948
```

Here is the model variables ranked by its importance.

```
varImp(mod1)
```

```
## rf variable importance
##
##   only 20 most important variables shown (out of 52)
##
##                      Overall
## roll_belt             100.00
## pitch_forearm          58.48
## yaw_belt               54.09
## pitch_belt             45.04
## magnet_dumbbell_z      44.60
## magnet_dumbbell_y      42.61
## roll_forearm           42.28
## accel_dumbbell_y       23.10
## magnet_dumbbell_x      20.41
## accel_forearm_x        18.60
## roll_dumbbell          18.14
## accel_dumbbell_z       14.69
## magnet_belt_z          14.17
## total_accel_dumbbell   13.50
## accel_belt_z           12.79
## magnet_forearm_z       12.69
## magnet_belt_y          12.60
## gyros_belt_z           12.23
## yaw_arm                10.99
## magnet_belt_x          10.64
```

Predicting test data

The model returns high accuracy in the Train data and is expected to give good prediction to the Test dataset. Here is the classe prediction of 20 cases.

```
pred1Test <- predict(mod1, TestData)
table(TestData$problem_id,pred1Test)
```

```
##     pred1Test
##      A B C D E
##   1  0 1 0 0 0
##   2  1 0 0 0 0
##   3  0 1 0 0 0
##   4  1 0 0 0 0
##   5  1 0 0 0 0
##   6  0 0 0 0 1
##   7  0 0 0 1 0
##   8  0 1 0 0 0
##   9  1 0 0 0 0
##   10 1 0 0 0 0
##   11 0 1 0 0 0
##   12 0 0 1 0 0
##   13 0 1 0 0 0
##   14 1 0 0 0 0
##   15 0 0 0 0 1
##   16 0 0 0 0 1
##   17 1 0 0 0 0
##   18 0 1 0 0 0
##   19 0 1 0 0 0
##   20 0 1 0 0 0
```

Appendix

Exploratory data analysis

Tables below exhibit characteristics and statistic summary of all 53 variables including the prediction - classe from Train dataset.

```
str(TrainData)
```

```
## 'data.frame':    19622 obs. of  53 variables:
##  $ roll_belt            : num  1.41 1.41 1.42 1.48 1.48 1.45 1.42 1.42 1.43 1.45 ...
##  $ pitch_belt           : num  8.07 8.07 8.07 8.05 8.07 8.06 8.09 8.13 8.16 8.17 ...
##  $ yaw_belt             : num  -94.4 -94.4 -94.4 -94.4 -94.4 -94.4 -94.4 -94.4 -94.4 -94.4 ...
##  $ total_accel_belt     : int  3 3 3 3 3 3 3 3 3 ...
##  $ gyros_belt_x         : num  0 0.02 0 0.02 0.02 0.02 0.02 0.02 0.02 0.03 ...
##  $ gyros_belt_y         : num  0 0 0 0 0.02 0 0 0 0 0 ...
##  $ gyros_belt_z         : num  -0.02 -0.02 -0.02 -0.03 -0.02 -0.02 -0.02 -0.02 -0.02 0 ...
##  $ accel_belt_x         : int  -21 -22 -20 -22 -21 -21 -22 -22 -20 -21 ...
##  $ accel_belt_y         : int  4 4 5 3 2 4 3 4 2 4 ...
##  $ accel_belt_z         : int  22 22 23 21 24 21 21 21 24 22 ...
##  $ magnet_belt_x        : int  -3 -7 -2 -6 -6 0 -4 -2 1 -3 ...
##  $ magnet_belt_y        : int  599 608 600 604 600 603 599 603 602 609 ...
##  $ magnet_belt_z        : int  -313 -311 -305 -310 -302 -312 -311 -313 -312 -308 ...
##  $ roll_arm             : num  -128 -128 -128 -128 -128 -128 -128 -128 -128 -128 ...
##  $ pitch_arm            : num  22.5 22.5 22.5 22.1 22.1 22 21.9 21.8 21.7 21.6 ...
```

```
##  $ yaw_arm             : num  -161 -161 -161 -161 -161 -161 -161 -161 -161 -161 ...
##  $ total_accel_arm     : int  34 34 34 34 34 34 34 34 34 34 ...
##  $ gyros_arm_x         : num  0 0.02 0.02 0.02 0 0.02 0 0.02 0.02 0.02 ...
##  $ gyros_arm_y         : num  0 -0.02 -0.02 -0.03 -0.03 -0.03 -0.03 -0.02 -0.03 -0.03 ...
##  $ gyros_arm_z         : num  -0.02 -0.02 -0.02 0.02 0 0 0 0 -0.02 -0.02 ...
##  $ accel_arm_x         : int  -288 -290 -289 -289 -289 -289 -289 -289 -288 -288 ...
##  $ accel_arm_y         : int  109 110 110 111 111 111 111 111 109 110 ...
##  $ accel_arm_z         : int  -123 -125 -126 -123 -123 -122 -125 -124 -122 -124 ...
##  $ magnet_arm_x        : int  -368 -369 -368 -372 -374 -369 -373 -372 -369 -376 ...
##  $ magnet_arm_y        : int  337 337 344 344 337 342 336 338 341 334 ...
##  $ magnet_arm_z        : int  516 513 513 512 506 513 509 510 518 516 ...
##  $ roll_dumbbell       : num  13.1 13.1 12.9 13.4 13.4 ...
##  $ pitch_dumbbell      : num  -70.5 -70.6 -70.3 -70.4 -70.4 ...
##  $ yaw_dumbbell        : num  -84.9 -84.7 -85.1 -84.9 -84.9 ...
##  $ total_accel_dumbbell: int  37 37 37 37 37 37 37 37 37 37 ...
##  $ gyros_dumbbell_x    : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ gyros_dumbbell_y    : num  -0.02 -0.02 -0.02 -0.02 -0.02 -0.02 -0.02 -0.02 -0.02 -0.02 ...
##  $ gyros_dumbbell_z    : num  0 0 0 -0.02 0 0 0 0 0 0 ...
##  $ accel_dumbbell_x    : int  -234 -233 -232 -232 -233 -234 -232 -234 -232 -235 ...
##  $ accel_dumbbell_y    : int  47 47 46 48 48 48 47 46 47 48 ...
##  $ accel_dumbbell_z    : int  -271 -269 -270 -269 -270 -269 -270 -272 -269 -270 ...
##  $ magnet_dumbbell_x   : int  -559 -555 -561 -552 -554 -558 -551 -555 -549 -558 ...
##  $ magnet_dumbbell_y   : int  293 296 298 303 292 294 295 300 292 291 ...
##  $ magnet_dumbbell_z   : num  -65 -64 -63 -60 -68 -66 -70 -74 -65 -69 ...
##  $ roll_forearm        : num  28.4 28.3 28.3 28.1 28 27.9 27.9 27.8 27.7 27.7 ...
##  $ pitch_forearm       : num  -63.9 -63.9 -63.9 -63.9 -63.9 -63.9 -63.9 -63.8 -63.8 -63.8 ...
##  $ yaw_forearm         : num  -153 -153 -152 -152 -152 -152 -152 -152 -152 -152 ...
##  $ total_accel_forearm : int  36 36 36 36 36 36 36 36 36 36 ...
##  $ gyros_forearm_x     : num  0.03 0.02 0.03 0.02 0.02 0.02 0.02 0.02 0.03 0.02 ...
##  $ gyros_forearm_y     : num  0 0 -0.02 -0.02 0 -0.02 0 -0.02 0 0 ...
##  $ gyros_forearm_z     : num  -0.02 -0.02 0 0 -0.02 -0.03 -0.02 0 -0.02 -0.02 ...
##  $ accel_forearm_x     : int  192 192 196 189 189 193 195 193 193 190 ...
##  $ accel_forearm_y     : int  203 203 204 206 206 203 205 205 204 205 ...
##  $ accel_forearm_z     : int  -215 -216 -213 -214 -214 -215 -215 -213 -214 -215 ...
##  $ magnet_forearm_x    : int  -17 -18 -18 -16 -17 -9 -18 -9 -16 -22 ...
##  $ magnet_forearm_y    : num  654 661 658 658 655 660 659 660 653 656 ...
##  $ magnet_forearm_z    : num  476 473 469 469 473 478 470 474 476 473 ...
##  $ classe              : Factor w/ 5 levels "A","B","C","D",..: 1 1 1 1 1 1 1 1 1 1 ...
```

```
library(psych)
psych::describe(TrainData)
```

```
##                     vars     n    mean     sd  median trimmed    mad
## roll_belt              1 19622   64.41  62.75  113.00   63.37  62.27
## pitch_belt             2 19622    0.31  22.35    5.28    2.44  10.35
## yaw_belt               3 19622  -11.21  95.19  -13.00  -23.19 111.19
## total_accel_belt       4 19622   11.31   7.74   17.00   11.12  11.86
## gyros_belt_x           5 19622   -0.01   0.21    0.03    0.02   0.12
## gyros_belt_y           6 19622    0.04   0.08    0.02    0.04   0.07
## gyros_belt_z           7 19622   -0.13   0.24   -0.10   -0.12   0.15
## accel_belt_x           8 19622   -5.59  29.64  -15.00   -8.08  11.86
## accel_belt_y           9 19622   30.15  28.58   35.00   29.22  45.96
## accel_belt_z          10 19622  -72.59 100.45 -152.00  -71.74 121.57
## magnet_belt_x         11 19622   55.60  64.18   35.00   47.24  38.55
## magnet_belt_y         12 19622  593.68  35.68  601.00  598.21  22.24
```

```
## magnet_belt_z          13 19622 -345.48  65.21 -320.00 -338.48  29.65
## roll_arm               14 19622   17.83  72.74    0.00   19.66  81.91
## pitch_arm              15 19622   -4.61  30.68    0.00   -5.34  28.02
## yaw_arm                16 19622   -0.62  71.36    0.00    0.35  65.53
## total_accel_arm        17 19622   25.51  10.52   27.00   25.49  10.38
## gyros_arm_x            18 19622    0.04   1.99    0.08    0.11   2.16
## gyros_arm_y            19 19622   -0.26   0.85   -0.24   -0.27   0.71
## gyros_arm_z            20 19622    0.27   0.55    0.23    0.28   0.58
## accel_arm_x            21 19622  -60.24 182.04  -44.00  -71.09 244.63
## accel_arm_y            22 19622   32.60 109.87   14.00   32.34 134.92
## accel_arm_z            23 19622  -71.25 134.65  -47.00  -59.64 117.13
## magnet_arm_x           24 19622  191.72 443.64  289.00  200.29 622.69
## magnet_arm_y           25 19622  156.61 201.91  202.00  167.96 210.53
## magnet_arm_z           26 19622  306.49 326.62  444.00  356.24 195.70
## roll_dumbbell          27 19622   23.84  69.93   48.17   29.40  54.55
## pitch_dumbbell         28 19622  -10.78  36.99  -20.96  -13.18  42.16
## yaw_dumbbell           29 19622    1.67  82.52   -3.32   -1.14 112.15
## total_accel_dumbbell   30 19622   13.72  10.23   10.00   12.91  10.38
## gyros_dumbbell_x       31 19622    0.16   1.51    0.13    0.16   0.28
## gyros_dumbbell_y       32 19622    0.05   0.61    0.03    0.04   0.27
## gyros_dumbbell_z       33 19622   -0.13   2.29   -0.13   -0.14   0.25
## accel_dumbbell_x       34 19622  -28.62  67.32   -8.00  -24.88  47.44
## accel_dumbbell_y       35 19622   52.63  80.75   41.50   49.34  82.28
## accel_dumbbell_z       36 19622  -38.32 109.47   -1.00  -35.97 121.57
## magnet_dumbbell_x      37 19622 -328.48 339.72 -479.00 -402.07 108.23
## magnet_dumbbell_y      38 19622  220.97 326.87  311.00  280.65 117.13
## magnet_dumbbell_z      39 19622   46.05 139.96   13.00   32.75 100.82
## roll_forearm           40 19622   33.83 108.04   21.70   42.25 142.77
## pitch_forearm          41 19622   10.71  28.15    9.24   12.39  17.72
## yaw_forearm            42 19622   19.21 103.22    0.00   23.04 145.00
## total_accel_forearm    43 19622   34.72  10.06   36.00   35.23   8.90
## gyros_forearm_x        44 19622    0.16   0.65    0.05    0.15   0.52
## gyros_forearm_y        45 19622    0.08   3.10    0.03    0.10   2.28
## gyros_forearm_z        46 19622    0.15   1.75    0.08    0.13   0.49
## accel_forearm_x        47 19622  -61.65 180.59  -57.00  -55.32 189.77
## accel_forearm_y        48 19622  163.66 200.13  201.00  179.20 180.88
## accel_forearm_z        49 19622  -55.29 138.40  -39.00  -65.30 186.81
## magnet_forearm_x       50 19622 -312.58 346.96 -378.00 -345.19 383.99
## magnet_forearm_y       51 19622  380.12 509.37  591.00  427.03 283.18
## magnet_forearm_z       52 19622  393.61 369.27  511.00  446.24 246.11
## classe*                53 19622    2.77   1.48    3.00    2.71   1.48
##                           min     max   range    skew kurtosis   se
## roll_belt              -28.90  162.00  190.90   -0.02    -1.93 0.45
## pitch_belt             -55.80   60.30  116.10   -1.00    -0.09 0.16
## yaw_belt              -180.00  179.00  359.00    0.91    -0.58 0.68
## total_accel_belt         0.00   29.00   29.00    0.03    -1.75 0.06
## gyros_belt_x            -1.04    2.22    3.26   -0.58     6.56 0.00
## gyros_belt_y            -0.64    0.64    1.28   -0.06     4.72 0.00
## gyros_belt_z            -1.46    1.62    3.08    0.21     6.29 0.00
## accel_belt_x          -120.00   85.00  205.00    0.97    -0.09 0.21
## accel_belt_y           -69.00  164.00  233.00    0.16    -1.57 0.20
## accel_belt_z          -275.00  105.00  380.00    0.02    -1.83 0.72
## magnet_belt_x          -52.00  485.00  537.00    1.43     2.10 0.46
## magnet_belt_y          354.00  673.00  319.00   -2.24     7.97 0.25
```

```
## magnet_belt_z         -623.00  293.00  916.00    0.17    12.35 0.47
## roll_arm              -180.00  180.00  360.00   -0.18    -0.41 0.52
## pitch_arm              -88.80   88.50  177.30    0.20     0.28 0.22
## yaw_arm               -180.00  180.00  360.00   -0.09    -0.13 0.51
## total_accel_arm          1.00   66.00   65.00    0.06    -0.29 0.08
## gyros_arm_x             -6.37    4.87   11.24   -0.29    -0.43 0.01
## gyros_arm_y             -3.44    2.84    6.28    0.14     0.65 0.01
## gyros_arm_z             -2.33    3.02    5.35   -0.17     0.04 0.00
## accel_arm_x           -404.00  437.00  841.00    0.34    -0.78 1.30
## accel_arm_y           -318.00  308.00  626.00    0.10    -1.04 0.78
## accel_arm_z           -636.00  292.00  928.00   -0.85     0.88 0.96
## magnet_arm_x          -584.00  782.00 1366.00   -0.15    -1.60 3.17
## magnet_arm_y          -392.00  583.00  975.00   -0.47    -0.83 1.44
## magnet_arm_z          -597.00  694.00 1291.00   -1.14     0.10 2.33
## roll_dumbbell         -153.71  153.55  307.26   -0.74    -0.44 0.50
## pitch_dumbbell        -149.59  149.40  299.00    0.52    -0.13 0.26
## yaw_dumbbell          -150.87  154.95  305.82    0.22    -1.43 0.59
## total_accel_dumbbell     0.00   58.00   58.00    0.60    -0.83 0.07
## gyros_dumbbell_x      -204.00    2.22  206.22 -126.32 17093.48 0.01
## gyros_dumbbell_y        -2.10   52.00   54.10   31.65  2681.48 0.00
## gyros_dumbbell_z        -2.38  317.00  319.38  135.95 18852.61 0.02
## accel_dumbbell_x      -419.00  235.00  654.00   -0.46     1.07 0.48
## accel_dumbbell_y      -189.00  315.00  504.00    0.35    -0.43 0.58
## accel_dumbbell_z      -334.00  318.00  652.00   -0.10    -0.54 0.78
## magnet_dumbbell_x     -643.00  592.00 1235.00    1.69     1.44 2.43
## magnet_dumbbell_y    -3600.00  633.00 4233.00   -1.81     2.78 2.33
## magnet_dumbbell_z     -262.00  452.00  714.00    0.88     0.27 1.00
## roll_forearm          -180.00  180.00  360.00   -0.46    -0.76 0.77
## pitch_forearm          -72.50   89.80  162.30   -0.51     0.88 0.20
## yaw_forearm           -180.00  180.00  360.00   -0.27    -1.14 0.74
## total_accel_forearm      0.00  108.00  108.00   -0.58     1.04 0.07
## gyros_forearm_x        -22.00    3.97   25.97   -1.92    69.88 0.00
## gyros_forearm_y         -7.02  311.00  318.02   51.32  5150.22 0.02
## gyros_forearm_z         -8.09  231.00  239.09  116.08 15271.76 0.01
## accel_forearm_x       -498.00  477.00  975.00   -0.23    -0.69 1.29
## accel_forearm_y       -632.00  923.00 1555.00   -0.65    -0.12 1.43
## accel_forearm_z       -446.00  291.00  737.00    0.44    -0.91 0.99
## magnet_forearm_x     -1280.00  672.00 1952.00    0.64    -0.08 2.48
## magnet_forearm_y      -896.00 1480.00 2376.00   -0.74    -0.46 3.64
## magnet_forearm_z      -973.00 1090.00 2063.00   -1.22     0.90 2.64
## classe*                  1.00    5.00    4.00    0.21    -1.36 0.01
```