Practical Machine Learning

Yefu Wang

November 24, 2017

Introduction

This report is to solve the final project in the coursera course "practical machine learning". Two .csv documents are used in this report: pml-testing.csv, and pml-training.csv. The training document contains all the information from six user's movement, which includes five classes (A, B, C, D, and E). Then, we need to build and train a model from all the training information, and finally evaluate the model in the test data set.

Data clean

First of all, let's import all the data into R. And then, let's take a look what they look like.

```
train.data <- read.csv("pml-training.csv", header = TRUE, stringsAsFactors = TRUE)
test.data <- read.csv("pml-testing.csv", header = TRUE, stringsAsFactors = TRUE)
cat(noquote("Here is the summary of training data: "))
```

```
## Here is the summary of training data:
```

```
summary(train.data)
```

```
##        X            user_name    raw_timestamp_part_1 raw_timestamp_part_2
##  Min.   :    1   adelmo  :3892   Min.   :1.322e+09    Min.   :   294
##  1st Qu.: 4906   carlitos:3112   1st Qu.:1.323e+09    1st Qu.:252912
##  Median : 9812   charles :3536   Median :1.323e+09    Median :496380
##  Mean   : 9812   eurico  :3070   Mean   :1.323e+09    Mean   :500656
##  3rd Qu.:14717   jeremy  :3402   3rd Qu.:1.323e+09    3rd Qu.:751891
##  Max.   :19622   pedro   :2610   Max.   :1.323e+09    Max.   :998801
##
##           cvtd_timestamp  new_window    num_window      roll_belt
##  28/11/2011 14:14: 1498   no :19216   Min.   :  1.0   Min.   :-28.90
##  05/12/2011 11:24: 1497   yes:  406   1st Qu.:222.0   1st Qu.:  1.10
##  30/11/2011 17:11: 1440               Median :424.0   Median :113.00
##  05/12/2011 11:25: 1425               Mean   :430.6   Mean   : 64.41
##  02/12/2011 14:57: 1380               3rd Qu.:644.0   3rd Qu.:123.00
##  02/12/2011 13:34: 1375               Max.   :864.0   Max.   :162.00
##  (Other)         :11007
##    pitch_belt          yaw_belt       total_accel_belt kurtosis_roll_belt
##  Min.   :-55.8000   Min.   :-180.00   Min.   : 0.00             :19216
##  1st Qu.:  1.7600   1st Qu.: -88.30   1st Qu.: 3.00    #DIV/0!  :   10
##  Median :  5.2800   Median : -13.00   Median :17.00    -1.908453:    2
##  Mean   :  0.3053   Mean   : -11.21   Mean   :11.31    -0.016850:    1
##  3rd Qu.: 14.9000   3rd Qu.:  12.90   3rd Qu.:18.00    -0.021024:    1
##  Max.   : 60.3000   Max.   : 179.00   Max.   :29.00    -0.025513:    1
##                                                        (Other)  :  391
##  kurtosis_picth_belt kurtosis_yaw_belt skewness_roll_belt
##            :19216            :19216             :19216
##  #DIV/0!   :   32     #DIV/0!:  406     #DIV/0!  :    9
##  47.000000:    4                       0.000000 :    4
```

```
##  -0.150950:    3                      0.422463 :    2
##  -0.684748:    3                      -0.003095:    1
##  -1.750749:    3                      -0.010002:    1
##  (Other)  :  361                      (Other)  :  389
##  skewness_roll_belt.1 skewness_yaw_belt max_roll_belt     max_picth_belt
##           :19216            :19216     Min.   :-94.300   Min.   : 3.00
##  #DIV/0!  :   32      #DIV/0!:  406     1st Qu.:-88.000   1st Qu.: 5.00
##  0.000000 :    4                       Median : -5.100   Median :18.00
##  -2.156553:    3                       Mean   : -6.667   Mean   :12.92
##  -3.072669:    3                       3rd Qu.: 18.500   3rd Qu.:19.00
##  -6.324555:    3                       Max.   :180.000   Max.   :30.00
##  (Other)  :  361                       NA's   :19216     NA's   :19216
##   max_yaw_belt   min_roll_belt     min_pitch_belt   min_yaw_belt
##          :19216   Min.   :-180.00   Min.   : 0.00          :19216
##  -1.1    :   30   1st Qu.: -88.40   1st Qu.: 3.00   -1.1   :   30
##  -1.4    :   29   Median :  -7.85   Median :16.00   -1.4   :   29
##  -1.2    :   26   Mean   : -10.44   Mean   :10.76   -1.2   :   26
##  -0.9    :   24   3rd Qu.:   9.05   3rd Qu.:17.00   -0.9   :   24
##  -1.3    :   22   Max.   : 173.00   Max.   :23.00   -1.3   :   22
##  (Other):  275   NA's   :19216     NA's   :19216   (Other):  275
##  amplitude_roll_belt amplitude_pitch_belt amplitude_yaw_belt
##  Min.   :  0.000     Min.   : 0.000              :19216
##  1st Qu.:  0.300     1st Qu.: 1.000       #DIV/0!:   10
##  Median :  1.000     Median : 1.000       0.00   :   12
##  Mean   :  3.769     Mean   : 2.167       0.0000 :  384
##  3rd Qu.:  2.083     3rd Qu.: 2.000
##  Max.   :360.000     Max.   :12.000
##  NA's   :19216       NA's   :19216
##  var_total_accel_belt avg_roll_belt    stddev_roll_belt var_roll_belt
##  Min.   : 0.000       Min.   :-27.40   Min.   : 0.000   Min.   :  0.000
##  1st Qu.: 0.100       1st Qu.:  1.10   1st Qu.: 0.200   1st Qu.:  0.000
##  Median : 0.200       Median :116.35   Median : 0.400   Median :  0.100
##  Mean   : 0.926       Mean   : 68.06   Mean   : 1.337   Mean   :  7.699
##  3rd Qu.: 0.300       3rd Qu.:123.38   3rd Qu.: 0.700   3rd Qu.:  0.500
##  Max.   :16.500       Max.   :157.40   Max.   :14.200   Max.   :200.700
##  NA's   :19216        NA's   :19216    NA's   :19216    NA's   :19216
##  avg_pitch_belt    stddev_pitch_belt var_pitch_belt    avg_yaw_belt
##  Min.   :-51.400   Min.   :0.000     Min.   : 0.000   Min.   :-138.300
##  1st Qu.:  2.025   1st Qu.:0.200     1st Qu.: 0.000   1st Qu.: -88.175
##  Median :  5.200   Median :0.400     Median : 0.100   Median :  -6.550
##  Mean   :  0.520   Mean   :0.603     Mean   : 0.766   Mean   :  -8.831
##  3rd Qu.: 15.775   3rd Qu.:0.700     3rd Qu.: 0.500   3rd Qu.:  14.125
##  Max.   : 59.700   Max.   :4.000     Max.   :16.200   Max.   : 173.500
##  NA's   :19216     NA's   :19216     NA's   :19216    NA's   :19216
##  stddev_yaw_belt    var_yaw_belt        gyros_belt_x
##  Min.   :  0.000   Min.   :    0.000   Min.   :-1.040000
##  1st Qu.:  0.100   1st Qu.:    0.010   1st Qu.:-0.030000
##  Median :  0.300   Median :    0.090   Median : 0.030000
##  Mean   :  1.341   Mean   :  107.487   Mean   :-0.005592
##  3rd Qu.:  0.700   3rd Qu.:    0.475   3rd Qu.: 0.110000
##  Max.   :176.600   Max.   :31183.240   Max.   : 2.220000
##  NA's   :19216     NA's   :19216
##   gyros_belt_y       gyros_belt_z      accel_belt_x       accel_belt_y
##  Min.   :-0.64000   Min.   :-1.4600   Min.   :-120.000   Min.   :-69.00
```

```
##  1st Qu.: 0.00000   1st Qu.:-0.2000   1st Qu.: -21.000   1st Qu.:  3.00
##  Median : 0.02000   Median :-0.1000   Median : -15.000   Median : 35.00
##  Mean   : 0.03959   Mean   :-0.1305   Mean   :  -5.595   Mean   : 30.15
##  3rd Qu.: 0.11000   3rd Qu.:-0.0200   3rd Qu.:  -5.000   3rd Qu.: 61.00
##  Max.   : 0.64000   Max.   : 1.6200   Max.   :  85.000   Max.   :164.00
##
##   accel_belt_z     magnet_belt_x   magnet_belt_y   magnet_belt_z
##  Min.   :-275.00   Min.   :-52.0   Min.   :354.0   Min.   :-623.0
##  1st Qu.:-162.00   1st Qu.:  9.0   1st Qu.:581.0   1st Qu.:-375.0
##  Median :-152.00   Median : 35.0   Median :601.0   Median :-320.0
##  Mean   : -72.59   Mean   : 55.6   Mean   :593.7   Mean   :-345.5
##  3rd Qu.:  27.00   3rd Qu.: 59.0   3rd Qu.:610.0   3rd Qu.:-306.0
##  Max.   : 105.00   Max.   :485.0   Max.   :673.0   Max.   : 293.0
##
##     roll_arm         pitch_arm          yaw_arm          total_accel_arm
##  Min.   :-180.00   Min.   :-88.800   Min.   :-180.0000   Min.   : 1.00
##  1st Qu.: -31.77   1st Qu.:-25.900   1st Qu.: -43.1000   1st Qu.:17.00
##  Median :   0.00   Median :  0.000   Median :   0.0000   Median :27.00
##  Mean   :  17.83   Mean   : -4.612   Mean   :  -0.6188   Mean   :25.51
##  3rd Qu.:  77.30   3rd Qu.: 11.200   3rd Qu.:  45.8750   3rd Qu.:33.00
##  Max.   : 180.00   Max.   : 88.500   Max.   : 180.0000   Max.   :66.00
##
##  var_accel_arm     avg_roll_arm     stddev_roll_arm    var_roll_arm
##  Min.   :  0.00   Min.   :-166.67   Min.   :  0.000   Min.   :    0.000
##  1st Qu.:  9.03   1st Qu.: -38.37   1st Qu.:  1.376   1st Qu.:    1.898
##  Median : 40.61   Median :   0.00   Median :  5.702   Median :   32.517
##  Mean   : 53.23   Mean   :  12.68   Mean   : 11.201   Mean   :  417.264
##  3rd Qu.: 75.62   3rd Qu.:  76.33   3rd Qu.: 14.921   3rd Qu.:  222.647
##  Max.   :331.70   Max.   : 163.33   Max.   :161.964   Max.   :26232.208
##  NA's   :19216    NA's   :19216     NA's   :19216     NA's   :19216
##  avg_pitch_arm     stddev_pitch_arm var_pitch_arm       avg_yaw_arm
##  Min.   :-81.773   Min.   : 0.000   Min.   :   0.000   Min.   :-173.440
##  1st Qu.:-22.770   1st Qu.: 1.642   1st Qu.:   2.697   1st Qu.: -29.198
##  Median :  0.000   Median : 8.133   Median :  66.146   Median :   0.000
##  Mean   : -4.901   Mean   :10.383   Mean   : 195.864   Mean   :   2.359
##  3rd Qu.:  8.277   3rd Qu.:16.327   3rd Qu.: 266.576   3rd Qu.:  38.185
##  Max.   : 75.659   Max.   :43.412   Max.   :1884.565   Max.   : 152.000
##  NA's   :19216     NA's   :19216    NA's   :19216      NA's   :19216
##  stddev_yaw_arm     var_yaw_arm         gyros_arm_x
##  Min.   :  0.000   Min.   :    0.000   Min.   :-6.37000
##  1st Qu.:  2.577   1st Qu.:    6.642   1st Qu.:-1.33000
##  Median : 16.682   Median :  278.309   Median : 0.08000
##  Mean   : 22.270   Mean   : 1055.933   Mean   : 0.04277
##  3rd Qu.: 35.984   3rd Qu.: 1294.850   3rd Qu.: 1.57000
##  Max.   :177.044   Max.   :31344.568   Max.   : 4.87000
##  NA's   :19216     NA's   :19216
##   gyros_arm_y       gyros_arm_z       accel_arm_x       accel_arm_y
##  Min.   :-3.4400   Min.   :-2.3300   Min.   :-404.00   Min.   :-318.0
##  1st Qu.:-0.8000   1st Qu.:-0.0700   1st Qu.:-242.00   1st Qu.: -54.0
##  Median :-0.2400   Median : 0.2300   Median : -44.00   Median :  14.0
##  Mean   :-0.2571   Mean   : 0.2695   Mean   : -60.24   Mean   :  32.6
##  3rd Qu.: 0.1400   3rd Qu.: 0.7200   3rd Qu.:  84.00   3rd Qu.: 139.0
##  Max.   : 2.8400   Max.   : 3.0200   Max.   : 437.00   Max.   : 308.0
##
```

```
##   accel_arm_z       magnet_arm_x     magnet_arm_y     magnet_arm_z
##  Min.   :-636.00   Min.   :-584.0   Min.   :-392.0   Min.   :-597.0
##  1st Qu.:-143.00   1st Qu.:-300.0   1st Qu.:  -9.0   1st Qu.: 131.2
##  Median : -47.00   Median : 289.0   Median : 202.0   Median : 444.0
##  Mean   : -71.25   Mean   : 191.7   Mean   : 156.6   Mean   : 306.5
##  3rd Qu.:  23.00   3rd Qu.: 637.0   3rd Qu.: 323.0   3rd Qu.: 545.0
##  Max.   : 292.00   Max.   : 782.0   Max.   : 583.0   Max.   : 694.0
##
##  kurtosis_roll_arm kurtosis_picth_arm kurtosis_yaw_arm skewness_roll_arm
##          :19216            :19216             :19216           :19216
##  #DIV/0! :   78    #DIV/0! :   80     #DIV/0! :   11   #DIV/0! :   77
##  -0.02438:    1    -0.00484:    1     0.55844 :    2   -0.00051:    1
##  -0.04190:    1    -0.01311:    1     0.65132 :    2   -0.00696:    1
##  -0.05051:    1    -0.02967:    1     -0.01548:    1   -0.01884:    1
##  -0.05695:    1    -0.07394:    1     -0.01749:    1   -0.03359:    1
##  (Other) :  324    (Other) :  322     (Other) :  389   (Other) :  325
##  skewness_pitch_arm skewness_yaw_arm  max_roll_arm     max_picth_arm
##          :19216             :19216   Min.   :-73.100   Min.   :-173.000
##  #DIV/0! :   80     #DIV/0! :   11   1st Qu.: -0.175   1st Qu.:  -1.975
##  -0.00184:    1     -1.62032:    2   Median :  4.950   Median :  23.250
##  -0.01185:    1     0.55053 :    2   Mean   : 11.236   Mean   :  35.751
##  -0.01247:    1     -0.00311:    1   3rd Qu.: 26.775   3rd Qu.:  95.975
##  -0.02063:    1     -0.00562:    1   Max.   : 85.500   Max.   : 180.000
##  (Other) :  322     (Other) :  389   NA's   :19216     NA's   :19216
##   max_yaw_arm     min_roll_arm    min_pitch_arm      min_yaw_arm
##  Min.   : 4.00   Min.   :-89.10   Min.   :-180.00   Min.   : 1.00
##  1st Qu.:29.00   1st Qu.:-41.98   1st Qu.: -72.62   1st Qu.: 8.00
##  Median :34.00   Median :-22.45   Median : -33.85   Median :13.00
##  Mean   :35.46   Mean   :-21.22   Mean   : -33.92   Mean   :14.66
##  3rd Qu.:41.00   3rd Qu.:  0.00   3rd Qu.:   0.00   3rd Qu.:19.00
##  Max.   :65.00   Max.   : 66.40   Max.   : 152.00   Max.   :38.00
##  NA's   :19216   NA's   :19216    NA's   :19216     NA's   :19216
##  amplitude_roll_arm amplitude_pitch_arm amplitude_yaw_arm
##  Min.   :  0.000    Min.   :  0.000     Min.   : 0.00
##  1st Qu.:  5.425    1st Qu.:  9.925     1st Qu.:13.00
##  Median : 28.450    Median : 54.900     Median :22.00
##  Mean   : 32.452    Mean   : 69.677     Mean   :20.79
##  3rd Qu.: 50.960    3rd Qu.:115.175     3rd Qu.:28.75
##  Max.   :119.500    Max.   :360.000     Max.   :52.00
##  NA's   :19216      NA's   :19216       NA's   :19216
##  roll_dumbbell     pitch_dumbbell     yaw_dumbbell
##  Min.   :-153.71   Min.   :-149.59   Min.   :-150.871
##  1st Qu.: -18.49   1st Qu.: -40.89   1st Qu.: -77.644
##  Median :  48.17   Median : -20.96   Median :  -3.324
##  Mean   :  23.84   Mean   : -10.78   Mean   :   1.674
##  3rd Qu.:  67.61   3rd Qu.:  17.50   3rd Qu.:  79.643
##  Max.   : 153.55   Max.   : 149.40   Max.   : 154.952
##
##  kurtosis_roll_dumbbell kurtosis_picth_dumbbell kurtosis_yaw_dumbbell
##         :19216                 :19216                  :19216
##  #DIV/0!:    5          -0.5464:    2           #DIV/0!:  406
##  -0.2583:    2          -0.9334:    2
##  -0.3705:    2          -2.0833:    2
##  -0.5855:    2          -2.0851:    2
```

```
##  -2.0851:    2          -2.0889:    2
##  (Other):  393          (Other):  396
##  skewness_roll_dumbbell skewness_pitch_dumbbell skewness_yaw_dumbbell
##         :19216                 :19216                 :19216
##  #DIV/0!:    4          -0.2328:    2          #DIV/0!:  406
##  -0.9324:    2          -0.3521:    2
##  0.1110 :    2          -0.7036:    2
##  1.0312 :    2          0.1090 :    2
##  -0.0082:    1          1.0326 :    2
##  (Other):  395          (Other):  396
##  max_roll_dumbbell max_picth_dumbbell max_yaw_dumbbell min_roll_dumbbell
##  Min.   :-70.10    Min.   :-112.90          :19216    Min.   :-149.60
##  1st Qu.:-27.15    1st Qu.: -66.70    -0.6   :   20    1st Qu.: -59.67
##  Median : 14.85    Median :  40.05    0.2    :   19    Median : -43.55
##  Mean   : 13.76    Mean   :  32.75    -0.8   :   18    Mean   : -41.24
##  3rd Qu.: 50.58    3rd Qu.: 133.22    -0.3   :   16    3rd Qu.: -25.20
##  Max.   :137.00    Max.   : 155.00    -0.2   :   15    Max.   :  73.20
##  NA's   :19216     NA's   :19216      (Other):  318    NA's   :19216
##  min_pitch_dumbbell min_yaw_dumbbell amplitude_roll_dumbbell
##  Min.   :-147.00          :19216    Min.   :  0.00
##  1st Qu.: -91.80    -0.6   :   20    1st Qu.: 14.97
##  Median : -66.15    0.2    :   19    Median : 35.05
##  Mean   : -33.18    -0.8   :   18    Mean   : 55.00
##  3rd Qu.:  21.20    -0.3   :   16    3rd Qu.: 81.04
##  Max.   : 120.90    -0.2   :   15    Max.   :256.48
##  NA's   :19216      (Other):  318    NA's   :19216
##  amplitude_pitch_dumbbell amplitude_yaw_dumbbell total_accel_dumbbell
##  Min.   :  0.00                  :19216          Min.   : 0.00
##  1st Qu.: 17.06           #DIV/0!:    5          1st Qu.: 4.00
##  Median : 41.73           0.00   :  401          Median :10.00
##  Mean   : 65.93                                  Mean   :13.72
##  3rd Qu.: 99.55                                  3rd Qu.:19.00
##  Max.   :273.59                                  Max.   :58.00
##  NA's   :19216
##  var_accel_dumbbell avg_roll_dumbbell stddev_roll_dumbbell
##  Min.   :  0.000    Min.   :-128.96   Min.   :  0.000
##  1st Qu.:  0.378    1st Qu.: -12.33   1st Qu.:  4.639
##  Median :  1.000    Median :  48.23   Median : 12.204
##  Mean   :  4.388    Mean   :  23.86   Mean   : 20.761
##  3rd Qu.:  3.434    3rd Qu.:  64.37   3rd Qu.: 26.356
##  Max.   :230.428    Max.   : 125.99   Max.   :123.778
##  NA's   :19216      NA's   :19216     NA's   :19216
##  var_roll_dumbbell  avg_pitch_dumbbell stddev_pitch_dumbbell
##  Min.   :    0.00   Min.   :-70.73     Min.   : 0.000
##  1st Qu.:   21.52   1st Qu.:-42.00     1st Qu.: 3.482
##  Median :  148.95   Median :-19.91     Median : 8.089
##  Mean   : 1020.27   Mean   :-12.33     Mean   :13.147
##  3rd Qu.:  694.65   3rd Qu.: 13.21     3rd Qu.:19.238
##  Max.   :15321.01   Max.   : 94.28     Max.   :82.680
##  NA's   :19216      NA's   :19216      NA's   :19216
##  var_pitch_dumbbell avg_yaw_dumbbell   stddev_yaw_dumbbell
##  Min.   :   0.00    Min.   :-117.950   Min.   :  0.000
##  1st Qu.:  12.12    1st Qu.: -76.696   1st Qu.:  3.885
##  Median :  65.44    Median :  -4.505   Median : 10.264
```

```
##  Mean   : 350.31    Mean   :   0.202   Mean   : 16.647
##  3rd Qu.: 370.11    3rd Qu.:  71.234   3rd Qu.: 24.674
##  Max.   :6836.02    Max.   : 134.905   Max.   :107.088
##  NA's   :19216      NA's   :19216      NA's   :19216
##  var_yaw_dumbbell   gyros_dumbbell_x    gyros_dumbbell_y
##  Min.   :    0.00   Min.   :-204.0000   Min.   :-2.10000
##  1st Qu.:   15.09   1st Qu.:  -0.0300   1st Qu.:-0.14000
##  Median :  105.35   Median :   0.1300   Median : 0.03000
##  Mean   :  589.84   Mean   :   0.1611   Mean   : 0.04606
##  3rd Qu.:  608.79   3rd Qu.:   0.3500   3rd Qu.: 0.21000
##  Max.   :11467.91   Max.   :   2.2200   Max.   :52.00000
##  NA's   :19216
##  gyros_dumbbell_z  accel_dumbbell_x  accel_dumbbell_y  accel_dumbbell_z
##  Min.   : -2.380   Min.   :-419.00   Min.   :-189.00   Min.   :-334.00
##  1st Qu.: -0.310   1st Qu.: -50.00   1st Qu.:  -8.00   1st Qu.:-142.00
##  Median : -0.130   Median :  -8.00   Median :  41.50   Median :  -1.00
##  Mean   : -0.129   Mean   : -28.62   Mean   :  52.63   Mean   : -38.32
##  3rd Qu.:  0.030   3rd Qu.:  11.00   3rd Qu.: 111.00   3rd Qu.:  38.00
##  Max.   :317.000   Max.   : 235.00   Max.   : 315.00   Max.   : 318.00
##
##  magnet_dumbbell_x magnet_dumbbell_y magnet_dumbbell_z  roll_forearm
##  Min.   :-643.0    Min.   :-3600     Min.   :-262.00   Min.   :-180.0000
##  1st Qu.:-535.0    1st Qu.:  231     1st Qu.: -45.00   1st Qu.:  -0.7375
##  Median :-479.0    Median :  311     Median :  13.00   Median :  21.7000
##  Mean   :-328.5    Mean   :  221     Mean   :  46.05   Mean   :  33.8265
##  3rd Qu.:-304.0    3rd Qu.:  390     3rd Qu.:  95.00   3rd Qu.: 140.0000
##  Max.   : 592.0    Max.   :  633     Max.   : 452.00   Max.   : 180.0000
##
##  pitch_forearm     yaw_forearm      kurtosis_roll_forearm
##  Min.   :-72.50   Min.   :-180.00          :19216
##  1st Qu.:  0.00   1st Qu.: -68.60   #DIV/0!:   84
##  Median :  9.24   Median :   0.00   -0.8079:    2
##  Mean   : 10.71   Mean   :  19.21   -0.9169:    2
##  3rd Qu.: 28.40   3rd Qu.: 110.00   -0.0227:    1
##  Max.   : 89.80   Max.   : 180.00   -0.0359:    1
##                                     (Other):  316
##  kurtosis_picth_forearm kurtosis_yaw_forearm skewness_roll_forearm
##         :19216                  :19216               :19216
##  #DIV/0!:   85          #DIV/0!:  406        #DIV/0!:   83
##  -0.0073:    1                               -0.1912:    2
##  -0.0442:    1                               -0.4126:    2
##  -0.0489:    1                               -0.0004:    1
##  -0.0523:    1                               -0.0013:    1
##  (Other):  317                               (Other):  317
##  skewness_pitch_forearm skewness_yaw_forearm max_roll_forearm
##         :19216                  :19216        Min.   :-66.60
##  #DIV/0!:   85          #DIV/0!:  406        1st Qu.:  0.00
##  0.0000 :    4                               Median : 26.80
##  -0.6992:    2                               Mean   : 24.49
##  -0.0113:    1                               3rd Qu.: 45.95
##  -0.0131:    1                               Max.   : 89.80
##  (Other):  313                               NA's   :19216
##  max_picth_forearm max_yaw_forearm min_roll_forearm  min_pitch_forearm
##  Min.   :-151.00          :19216   Min.   :-72.500   Min.   :-180.00
```

```
##  1st Qu.:   0.00   #DIV/0!:   84   1st Qu.: -6.075   1st Qu.:-175.00
##  Median : 113.00   -1.2   :   32   Median :  0.000   Median : -61.00
##  Mean   :  81.49   -1.3   :   31   Mean   : -0.167   Mean   : -57.57
##  3rd Qu.: 174.75   -1.4   :   24   3rd Qu.: 12.075   3rd Qu.:   0.00
##  Max.   : 180.00   -1.5   :   24   Max.   : 62.100   Max.   : 167.00
##  NA's   :19216     (Other):  211   NA's   :19216     NA's   :19216
##  min_yaw_forearm amplitude_roll_forearm amplitude_pitch_forearm
##         :19216   Min.   :  0.000        Min.   :  0.0
##  #DIV/0!:   84   1st Qu.:  1.125        1st Qu.:  2.0
##  -1.2   :   32   Median : 17.770        Median : 83.7
##  -1.3   :   31   Mean   : 24.653        Mean   :139.1
##  -1.4   :   24   3rd Qu.: 39.875        3rd Qu.:350.0
##  -1.5   :   24   Max.   :126.000        Max.   :360.0
##  (Other):  211   NA's   :19216          NA's   :19216
##  amplitude_yaw_forearm total_accel_forearm var_accel_forearm
##         :19216         Min.   :  0.00      Min.   :  0.000
##  #DIV/0!:   84         1st Qu.: 29.00      1st Qu.:  6.759
##  0.00   :  322         Median : 36.00      Median : 21.165
##                        Mean   : 34.72      Mean   : 33.502
##                        3rd Qu.: 41.00      3rd Qu.: 51.240
##                        Max.   :108.00      Max.   :172.606
##                                            NA's   :19216
##  avg_roll_forearm   stddev_roll_forearm var_roll_forearm
##  Min.   :-177.234   Min.   :  0.000     Min.   :    0.00
##  1st Qu.:  -0.909   1st Qu.:  0.428     1st Qu.:    0.18
##  Median :  11.172   Median :  8.030     Median :   64.48
##  Mean   :  33.165   Mean   : 41.986     Mean   : 5274.10
##  3rd Qu.: 107.132   3rd Qu.: 85.373     3rd Qu.: 7289.08
##  Max.   : 177.256   Max.   :179.171     Max.   :32102.24
##  NA's   :19216      NA's   :19216       NA's   :19216
##  avg_pitch_forearm stddev_pitch_forearm var_pitch_forearm
##  Min.   :-68.17    Min.   : 0.000       Min.   :   0.000
##  1st Qu.:  0.00    1st Qu.: 0.336       1st Qu.:   0.113
##  Median : 12.02    Median : 5.516       Median :  30.425
##  Mean   : 11.79    Mean   : 7.977       Mean   : 139.593
##  3rd Qu.: 28.48    3rd Qu.:12.866       3rd Qu.: 165.532
##  Max.   : 72.09    Max.   :47.745       Max.   :2279.617
##  NA's   :19216     NA's   :19216        NA's   :19216
##  avg_yaw_forearm   stddev_yaw_forearm var_yaw_forearm    gyros_forearm_x
##  Min.   :-155.06   Min.   :  0.000    Min.   :    0.00   Min.   :-22.000
##  1st Qu.: -26.26   1st Qu.:  0.524    1st Qu.:    0.27   1st Qu.: -0.220
##  Median :   0.00   Median : 24.743    Median :  612.21   Median :  0.050
##  Mean   :  18.00   Mean   : 44.854    Mean   : 4639.85   Mean   :  0.158
##  3rd Qu.:  85.79   3rd Qu.: 85.817    3rd Qu.: 7368.41   3rd Qu.:  0.560
##  Max.   : 169.24   Max.   :197.508    Max.   :39009.33   Max.   :  3.970
##  NA's   :19216     NA's   :19216      NA's   :19216
##  gyros_forearm_y     gyros_forearm_z    accel_forearm_x   accel_forearm_y
##  Min.   : -7.02000   Min.   : -8.0900   Min.   :-498.00   Min.   :-632.0
##  1st Qu.: -1.46000   1st Qu.: -0.1800   1st Qu.:-178.00   1st Qu.:  57.0
##  Median :  0.03000   Median :  0.0800   Median : -57.00   Median : 201.0
##  Mean   :  0.07517   Mean   :  0.1512   Mean   : -61.65   Mean   : 163.7
##  3rd Qu.:  1.62000   3rd Qu.:  0.4900   3rd Qu.:  76.00   3rd Qu.: 312.0
##  Max.   :311.00000   Max.   :231.0000   Max.   : 477.00   Max.   : 923.0
##
```

```
##  accel_forearm_z   magnet_forearm_x  magnet_forearm_y magnet_forearm_z
##  Min.   :-446.00   Min.   :-1280.0   Min.   :-896.0   Min.   :-973.0
##  1st Qu.:-182.00   1st Qu.: -616.0   1st Qu.:   2.0   1st Qu.: 191.0
##  Median : -39.00   Median : -378.0   Median : 591.0   Median : 511.0
##  Mean   : -55.29   Mean   : -312.6   Mean   : 380.1   Mean   : 393.6
##  3rd Qu.:  26.00   3rd Qu.:  -73.0   3rd Qu.: 737.0   3rd Qu.: 653.0
##  Max.   : 291.00   Max.   :  672.0   Max.   :1480.0   Max.   :1090.0
##
##  classe
##  A:5580
##  B:3797
##  C:3422
##  D:3216
##  E:3607
##
##
```

```
cat(noquote("Here is the summary of test data: "))
```

```
## Here is the summary of test data:
```

```
summary(test.data)
```

```
##        X            user_name raw_timestamp_part_1 raw_timestamp_part_2
##  Min.   : 1.00   adelmo  :1   Min.   :1.322e+09    Min.   : 36553
##  1st Qu.: 5.75   carlitos:3   1st Qu.:1.323e+09    1st Qu.:268655
##  Median :10.50   charles :1   Median :1.323e+09    Median :530706
##  Mean   :10.50   eurico  :4   Mean   :1.323e+09    Mean   :512167
##  3rd Qu.:15.25   jeremy  :8   3rd Qu.:1.323e+09    3rd Qu.:787738
##  Max.   :20.00   pedro   :3   Max.   :1.323e+09    Max.   :920315
##
##           cvtd_timestamp new_window   num_window      roll_belt
##  30/11/2011 17:11:4      no:20      Min.   : 48.0   Min.   : -5.9200
##  05/12/2011 11:24:3                 1st Qu.:250.0   1st Qu.:  0.9075
##  30/11/2011 17:12:3                 Median :384.5   Median :  1.1100
##  05/12/2011 14:23:2                 Mean   :379.6   Mean   : 31.3055
##  28/11/2011 14:14:2                 3rd Qu.:467.0   3rd Qu.: 32.5050
##  02/12/2011 13:33:1                 Max.   :859.0   Max.   :129.0000
##  (Other)         :5
##    pitch_belt         yaw_belt      total_accel_belt kurtosis_roll_belt
##  Min.   :-41.600   Min.   :-93.70   Min.   : 2.00    Mode:logical
##  1st Qu.:  3.013   1st Qu.:-88.62   1st Qu.: 3.00    NA's:20
##  Median :  4.655   Median :-87.85   Median : 4.00
##  Mean   :  5.824   Mean   :-59.30   Mean   : 7.55
##  3rd Qu.:  6.135   3rd Qu.:-63.50   3rd Qu.: 8.00
##  Max.   : 27.800   Max.   :162.00   Max.   :21.00
##
##  kurtosis_picth_belt kurtosis_yaw_belt skewness_roll_belt
##  Mode:logical        Mode:logical      Mode:logical
##  NA's:20             NA's:20           NA's:20
##
##
##
##
##
```

```
##  skewness_roll_belt.1 skewness_yaw_belt max_roll_belt  max_picth_belt
##  Mode:logical         Mode:logical      Mode:logical   Mode:logical
##  NA's:20              NA's:20           NA's:20        NA's:20
##
##
##
##
##
##  max_yaw_belt   min_roll_belt  min_pitch_belt min_yaw_belt
##  Mode:logical   Mode:logical   Mode:logical   Mode:logical
##  NA's:20        NA's:20        NA's:20        NA's:20
##
##
##
##
##
##  amplitude_roll_belt amplitude_pitch_belt amplitude_yaw_belt
##  Mode:logical        Mode:logical         Mode:logical
##  NA's:20             NA's:20              NA's:20
##
##
##
##
##
##  var_total_accel_belt avg_roll_belt  stddev_roll_belt var_roll_belt
##  Mode:logical         Mode:logical   Mode:logical     Mode:logical
##  NA's:20              NA's:20        NA's:20          NA's:20
##
##
##
##
##
##  avg_pitch_belt stddev_pitch_belt var_pitch_belt avg_yaw_belt
##  Mode:logical   Mode:logical      Mode:logical   Mode:logical
##  NA's:20        NA's:20           NA's:20        NA's:20
##
##
##
##
##
##  stddev_yaw_belt var_yaw_belt    gyros_belt_x     gyros_belt_y
##  Mode:logical    Mode:logical   Min.   :-0.500   Min.   :-0.050
##  NA's:20         NA's:20        1st Qu.:-0.070   1st Qu.:-0.005
##                                 Median : 0.020   Median : 0.000
##                                 Mean   :-0.045   Mean   : 0.010
##                                 3rd Qu.: 0.070   3rd Qu.: 0.020
##                                 Max.   : 0.240   Max.   : 0.110
##
##   gyros_belt_z      accel_belt_x     accel_belt_y     accel_belt_z
##  Min.   :-0.4800   Min.   :-48.00   Min.   :-16.00   Min.   :-187.00
##  1st Qu.:-0.1375   1st Qu.:-19.00   1st Qu.:  2.00   1st Qu.: -24.00
##  Median :-0.0250   Median :-13.00   Median :  4.50   Median :  27.00
##  Mean   :-0.1005   Mean   :-13.50   Mean   : 18.35   Mean   : -17.60
##  3rd Qu.: 0.0000   3rd Qu.: -8.75   3rd Qu.: 25.50   3rd Qu.:  38.25
```

```
##  Max.   : 0.0500   Max.   : 46.00   Max.   : 72.00   Max.   :  49.00
##
##  magnet_belt_x    magnet_belt_y   magnet_belt_z       roll_arm
##  Min.   :-13.00   Min.   :566.0   Min.   :-426.0   Min.   :-137.00
##  1st Qu.:  5.50   1st Qu.:578.5   1st Qu.:-398.5   1st Qu.:   0.00
##  Median : 33.50   Median :600.5   Median :-313.5   Median :   0.00
##  Mean   : 35.15   Mean   :601.5   Mean   :-346.9   Mean   :  16.42
##  3rd Qu.: 46.25   3rd Qu.:631.2   3rd Qu.:-305.0   3rd Qu.:  71.53
##  Max.   :169.00   Max.   :638.0   Max.   :-291.0   Max.   : 152.00
##
##    pitch_arm          yaw_arm        total_accel_arm var_accel_arm
##  Min.   :-63.800   Min.   :-167.00   Min.   : 3.00   Mode:logical
##  1st Qu.: -9.188   1st Qu.: -60.15   1st Qu.:20.25   NA's:20
##  Median :  0.000   Median :   0.00   Median :29.50
##  Mean   : -3.950   Mean   :  -2.80   Mean   :26.40
##  3rd Qu.:  3.465   3rd Qu.:  25.50   3rd Qu.:33.25
##  Max.   : 55.000   Max.   : 178.00   Max.   :44.00
##
##  avg_roll_arm   stddev_roll_arm var_roll_arm   avg_pitch_arm
##  Mode:logical   Mode:logical    Mode:logical   Mode:logical
##  NA's:20        NA's:20         NA's:20        NA's:20
##
##
##
##
##
##  stddev_pitch_arm var_pitch_arm  avg_yaw_arm    stddev_yaw_arm
##  Mode:logical     Mode:logical   Mode:logical   Mode:logical
##  NA's:20          NA's:20        NA's:20        NA's:20
##
##
##
##
##
##  var_yaw_arm     gyros_arm_x      gyros_arm_y       gyros_arm_z
##  Mode:logical   Min.   :-3.710   Min.   :-2.0900   Min.   :-0.6900
##  NA's:20        1st Qu.:-0.645   1st Qu.:-0.6350   1st Qu.:-0.1800
##                 Median : 0.020   Median :-0.0400   Median :-0.0250
##                 Mean   : 0.077   Mean   :-0.1595   Mean   : 0.1205
##                 3rd Qu.: 1.248   3rd Qu.: 0.2175   3rd Qu.: 0.5650
##                 Max.   : 3.660   Max.   : 1.8500   Max.   : 1.1300
##
##   accel_arm_x      accel_arm_y      accel_arm_z       magnet_arm_x
##  Min.   :-341.0   Min.   :-65.00   Min.   :-404.00   Min.   :-428.00
##  1st Qu.:-277.0   1st Qu.: 52.25   1st Qu.:-128.50   1st Qu.:-373.75
##  Median :-194.5   Median :112.00   Median : -83.50   Median :-265.00
##  Mean   :-134.6   Mean   :103.10   Mean   : -87.85   Mean   : -38.95
##  3rd Qu.:   5.5   3rd Qu.:168.25   3rd Qu.: -27.25   3rd Qu.: 250.50
##  Max.   : 106.0   Max.   :245.00   Max.   :  93.00   Max.   : 750.00
##
##   magnet_arm_y     magnet_arm_z    kurtosis_roll_arm kurtosis_picth_arm
##  Min.   :-307.0   Min.   :-499.0   Mode:logical      Mode:logical
##  1st Qu.: 205.2   1st Qu.: 403.0   NA's:20           NA's:20
##  Median : 291.0   Median : 476.5
```

```
##  Mean   : 239.4   Mean   : 369.8
##  3rd Qu.: 358.8   3rd Qu.: 517.0
##  Max.   : 474.0   Max.   : 633.0
##
##  kurtosis_yaw_arm skewness_roll_arm skewness_pitch_arm skewness_yaw_arm
##  Mode:logical     Mode:logical      Mode:logical       Mode:logical
##  NA's:20          NA's:20           NA's:20            NA's:20
##
##
##
##
##
##  max_roll_arm   max_picth_arm  max_yaw_arm    min_roll_arm
##  Mode:logical   Mode:logical   Mode:logical   Mode:logical
##  NA's:20        NA's:20        NA's:20        NA's:20
##
##
##
##
##
##  min_pitch_arm  min_yaw_arm    amplitude_roll_arm amplitude_pitch_arm
##  Mode:logical   Mode:logical   Mode:logical       Mode:logical
##  NA's:20        NA's:20        NA's:20            NA's:20
##
##
##
##
##
##  amplitude_yaw_arm roll_dumbbell      pitch_dumbbell    yaw_dumbbell
##  Mode:logical      Min.   :-111.118   Min.   :-54.97   Min.   :-103.3200
##  NA's:20           1st Qu.:   7.494   1st Qu.:-51.89   1st Qu.: -75.2809
##                    Median :  50.403   Median :-40.81   Median :  -8.2863
##                    Mean   :  33.760   Mean   :-19.47   Mean   :  -0.9385
##                    3rd Qu.:  58.129   3rd Qu.: 16.12   3rd Qu.:  55.8335
##                    Max.   : 123.984   Max.   : 96.87   Max.   : 132.2337
##
##  kurtosis_roll_dumbbell kurtosis_picth_dumbbell kurtosis_yaw_dumbbell
##  Mode:logical           Mode:logical            Mode:logical
##  NA's:20                NA's:20                 NA's:20
##
##
##
##
##
##  skewness_roll_dumbbell skewness_pitch_dumbbell skewness_yaw_dumbbell
##  Mode:logical           Mode:logical            Mode:logical
##  NA's:20                NA's:20                 NA's:20
##
##
##
##
##
##  max_roll_dumbbell max_picth_dumbbell max_yaw_dumbbell min_roll_dumbbell
##  Mode:logical      Mode:logical       Mode:logical     Mode:logical
```

```
##  NA's:20           NA's:20            NA's:20          NA's:20
##
##
##
##
##
##  min_pitch_dumbbell min_yaw_dumbbell amplitude_roll_dumbbell
##  Mode:logical       Mode:logical     Mode:logical
##  NA's:20            NA's:20          NA's:20
##
##
##
##
##
##  amplitude_pitch_dumbbell amplitude_yaw_dumbbell total_accel_dumbbell
##  Mode:logical             Mode:logical           Min.   : 1.0
##  NA's:20                  NA's:20                1st Qu.: 7.0
##                                                  Median :15.5
##                                                  Mean   :17.2
##                                                  3rd Qu.:29.0
##                                                  Max.   :31.0
##
##  var_accel_dumbbell avg_roll_dumbbell stddev_roll_dumbbell
##  Mode:logical       Mode:logical      Mode:logical
##  NA's:20            NA's:20           NA's:20
##
##
##
##
##
##  var_roll_dumbbell avg_pitch_dumbbell stddev_pitch_dumbbell
##  Mode:logical      Mode:logical       Mode:logical
##  NA's:20           NA's:20            NA's:20
##
##
##
##
##
##  var_pitch_dumbbell avg_yaw_dumbbell stddev_yaw_dumbbell var_yaw_dumbbell
##  Mode:logical       Mode:logical     Mode:logical        Mode:logical
##  NA's:20            NA's:20          NA's:20             NA's:20
##
##
##
##
##
##  gyros_dumbbell_x  gyros_dumbbell_y  gyros_dumbbell_z accel_dumbbell_x
##  Min.   :-1.0300   Min.   :-1.1100   Min.   :-1.180   Min.   :-159.00
##  1st Qu.: 0.1600   1st Qu.:-0.2100   1st Qu.:-0.485   1st Qu.:-140.25
##  Median : 0.3600   Median : 0.0150   Median :-0.280   Median : -19.00
##  Mean   : 0.2690   Mean   : 0.0605   Mean   :-0.266   Mean   : -47.60
##  3rd Qu.: 0.4625   3rd Qu.: 0.1450   3rd Qu.:-0.165   3rd Qu.:  15.75
##  Max.   : 1.0600   Max.   : 1.9100   Max.   : 1.100   Max.   : 185.00
##
```

```
##  accel_dumbbell_y accel_dumbbell_z magnet_dumbbell_x magnet_dumbbell_y
##  Min.   :-30.00   Min.   :-221.0   Min.   :-576.0    Min.   :-558.0
##  1st Qu.:  5.75   1st Qu.:-192.2   1st Qu.:-528.0    1st Qu.: 259.5
##  Median : 71.50   Median :  -3.0   Median :-508.5    Median : 316.0
##  Mean   : 70.55   Mean   : -60.0   Mean   :-304.2    Mean   : 189.3
##  3rd Qu.:151.25   3rd Qu.:  76.5   3rd Qu.:-317.0    3rd Qu.: 348.2
##  Max.   :166.00   Max.   : 100.0   Max.   : 523.0    Max.   : 403.0
##
##  magnet_dumbbell_z  roll_forearm     pitch_forearm      yaw_forearm
##  Min.   :-164.00   Min.   :-176.00   Min.   :-63.500   Min.   :-168.000
##  1st Qu.: -33.00   1st Qu.: -40.25   1st Qu.:-11.457   1st Qu.: -93.375
##  Median :  49.50   Median :  94.20   Median :  8.830   Median : -19.250
##  Mean   :  71.40   Mean   :  38.66   Mean   :  7.099   Mean   :   2.195
##  3rd Qu.:  96.25   3rd Qu.: 143.25   3rd Qu.: 28.500   3rd Qu.: 104.500
##  Max.   : 368.00   Max.   : 176.00   Max.   : 59.300   Max.   : 159.000
##
##  kurtosis_roll_forearm kurtosis_picth_forearm kurtosis_yaw_forearm
##  Mode:logical          Mode:logical           Mode:logical
##  NA's:20               NA's:20                NA's:20
##
##
##
##
##
##  skewness_roll_forearm skewness_pitch_forearm skewness_yaw_forearm
##  Mode:logical          Mode:logical           Mode:logical
##  NA's:20               NA's:20                NA's:20
##
##
##
##
##
##  max_roll_forearm max_picth_forearm max_yaw_forearm min_roll_forearm
##  Mode:logical     Mode:logical      Mode:logical    Mode:logical
##  NA's:20          NA's:20           NA's:20         NA's:20
##
##
##
##
##
##  min_pitch_forearm min_yaw_forearm amplitude_roll_forearm
##  Mode:logical      Mode:logical    Mode:logical
##  NA's:20           NA's:20         NA's:20
##
##
##
##
##
##  amplitude_pitch_forearm amplitude_yaw_forearm total_accel_forearm
##  Mode:logical            Mode:logical          Min.   :21.00
##  NA's:20                 NA's:20               1st Qu.:24.00
##                                                Median :32.50
##                                                Mean   :32.05
##                                                3rd Qu.:36.75
```

```
##                                                Max.   :47.00
##
##  var_accel_forearm avg_roll_forearm stddev_roll_forearm var_roll_forearm
##  Mode:logical      Mode:logical     Mode:logical        Mode:logical
##  NA's:20           NA's:20          NA's:20             NA's:20
##
##
##
##
##
##  avg_pitch_forearm stddev_pitch_forearm var_pitch_forearm avg_yaw_forearm
##  Mode:logical      Mode:logical         Mode:logical      Mode:logical
##  NA's:20           NA's:20              NA's:20           NA's:20
##
##
##
##
##
##  stddev_yaw_forearm var_yaw_forearm gyros_forearm_x   gyros_forearm_y
##  Mode:logical       Mode:logical    Min.   :-1.0600   Min.   :-5.9700
##  NA's:20            NA's:20         1st Qu.:-0.5850   1st Qu.:-1.2875
##                                     Median : 0.0200   Median : 0.0350
##                                     Mean   :-0.0200   Mean   :-0.0415
##                                     3rd Qu.: 0.2925   3rd Qu.: 2.0475
##                                     Max.   : 1.3800   Max.   : 4.2600
##
##  gyros_forearm_z   accel_forearm_x  accel_forearm_y  accel_forearm_z
##  Min.   :-1.2600   Min.   :-212.0   Min.   :-331.0   Min.   :-282.0
##  1st Qu.:-0.0975   1st Qu.:-114.8   1st Qu.:   8.5   1st Qu.:-199.0
##  Median : 0.2300   Median :  86.0   Median : 138.0   Median :-148.5
##  Mean   : 0.2610   Mean   :  38.8   Mean   : 125.3   Mean   : -93.7
##  3rd Qu.: 0.7625   3rd Qu.: 166.2   3rd Qu.: 268.0   3rd Qu.: -31.0
##  Max.   : 1.8000   Max.   : 232.0   Max.   : 406.0   Max.   : 179.0
##
##  magnet_forearm_x magnet_forearm_y magnet_forearm_z   problem_id
##  Min.   :-714.0   Min.   :-787.0   Min.   :-32.0    Min.   : 1.00
##  1st Qu.:-427.2   1st Qu.:-328.8   1st Qu.:275.2    1st Qu.: 5.75
##  Median :-189.5   Median : 487.0   Median :491.5    Median :10.50
##  Mean   :-159.2   Mean   : 191.8   Mean   :460.2    Mean   :10.50
##  3rd Qu.:  41.5   3rd Qu.: 720.8   3rd Qu.:661.5    3rd Qu.:15.25
##  Max.   : 532.0   Max.   : 800.0   Max.   :884.0    Max.   :20.00
##
```

It seems that both data sets have 160 columns, and there are many columns in the test data set containing only NA values. So, we will have to select the columns containing no missing values.

```
test.data.select <- test.data[, colSums(is.na(test.data)) < nrow(test.data)]
cat(noquote("Here is the list of all columns with NO nas in test data set: "))
```

```
## Here is the list of all columns with NO nas in test data set:
```

```
colnames(test.data.select)
```

```
##  [1] "X"                    "user_name"            "raw_timestamp_part_1"
##  [4] "raw_timestamp_part_2" "cvtd_timestamp"       "new_window"
##  [7] "num_window"           "roll_belt"            "pitch_belt"
```

```
## [10] "yaw_belt"             "total_accel_belt"     "gyros_belt_x"
## [13] "gyros_belt_y"         "gyros_belt_z"         "accel_belt_x"
## [16] "accel_belt_y"         "accel_belt_z"         "magnet_belt_x"
## [19] "magnet_belt_y"        "magnet_belt_z"        "roll_arm"
## [22] "pitch_arm"            "yaw_arm"              "total_accel_arm"
## [25] "gyros_arm_x"          "gyros_arm_y"          "gyros_arm_z"
## [28] "accel_arm_x"          "accel_arm_y"          "accel_arm_z"
## [31] "magnet_arm_x"         "magnet_arm_y"         "magnet_arm_z"
## [34] "roll_dumbbell"        "pitch_dumbbell"       "yaw_dumbbell"
## [37] "total_accel_dumbbell" "gyros_dumbbell_x"     "gyros_dumbbell_y"
## [40] "gyros_dumbbell_z"     "accel_dumbbell_x"     "accel_dumbbell_y"
## [43] "accel_dumbbell_z"     "magnet_dumbbell_x"    "magnet_dumbbell_y"
## [46] "magnet_dumbbell_z"    "roll_forearm"         "pitch_forearm"
## [49] "yaw_forearm"          "total_accel_forearm"  "gyros_forearm_x"
## [52] "gyros_forearm_y"      "gyros_forearm_z"      "accel_forearm_x"
## [55] "accel_forearm_y"      "accel_forearm_z"      "magnet_forearm_x"
## [58] "magnet_forearm_y"     "magnet_forearm_z"     "problem_id"
```

```
cat(noquote("Check whether the names of all these features are the same as the names in train.data:"))
```

```
## Check whether the names of all these features are the same as the names in train.data:
```

```
cat("\n")
```

```
colnames(test.data.select) %in% colnames(train.data)
```

```
##  [1]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
## [12]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
## [23]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
## [34]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
## [45]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
## [56]  TRUE  TRUE  TRUE  TRUE FALSE
```

We can see that these names are common in the train data set, except the last feature "problem_id", so we can select the all other features to build the model. Of course, we also need the predicted output "classe" in the model. Also, the first feature "X" is not that important in the model at all, so we will just abandon it. In addition, I don't think the features "raw_timestamp_part_1", "raw_timestamp_part_2", "cvtd_timestamp", "new_window", "num_window" are really related to the "classe", so these features are dropped also.

```
features.select <- colnames(test.data.select)
features.delete <- c("X", "raw_timestamp_part_1", "raw_timestamp_part_2", "cvtd_timestamp", "new_window
features.input <- features.select[!(features.select %in% features.delete)]
train.data <- train.data[, c(features.input, "classe")]
```

So, we will use only two data sets in the following analysis: "train.data" and "test.data".

Let's start with simple decision tree first.

Decision tree

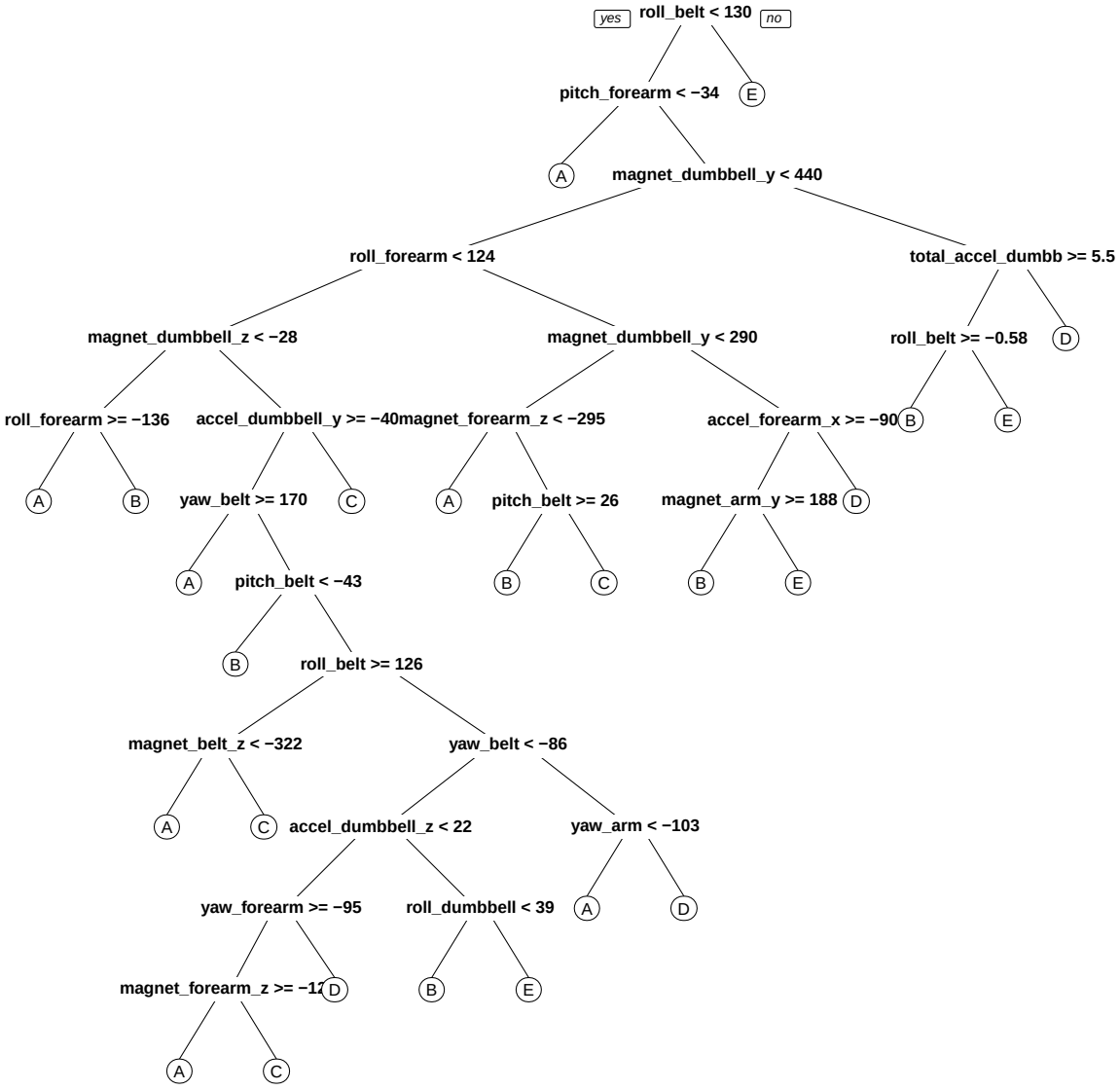
In this basic decision tree model, we can use "classe" as the output, and all other variables as "inputs".

```
set.seed(1217)
# Get the random 30% of the rows in the train.data as crossing validation test, and 70% as train.train
index.train <- sample(c(1:nrow(train.data)), 0.7 * nrow(train.data))
train.train <- train.data[index.train, ]
```

```
train.cv <- train.data[-index.train, ]
tree <- rpart(classe ~ ., train.train, method = "class")
cat(noquote("Here is the decision tree built with all features from train data:"))
```

```
## Here is the decision tree built with all features from train data:
```

```
prp(tree, faclen = 0, cex = 0.8)
```



Then, let's take a look at how the model performs.

```
prediction.tree <- predict(tree, train.cv, type = "class")
tree.performance <- confusionMatrix(prediction.tree, train.cv$classe)
cat(noquote("Here is the performance of the decision tree model in the cross validation data set:"))
```

```
## Here is the performance of the decision tree model in the cross validation data set:
```

```
cat("\n")
```

```
tree.performance
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction    A    B    C    D    E
##          A 1557  174   19   40   17
##          B   56  677   94   83  100
##          C   53  109  829  151  121
##          D   19   87   63  608   62
##          E   33   46   28   90  771
##
## Overall Statistics
##
##                Accuracy : 0.7545
##                  95% CI : (0.7433, 0.7655)
##     No Information Rate : 0.2918
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.6883
##  Mcnemar's Test P-Value : < 2.2e-16
##
## Statistics by Class:
##
##                      Class: A Class: B Class: C Class: D Class: E
## Sensitivity            0.9063   0.6194   0.8025   0.6255   0.7199
## Specificity            0.9400   0.9305   0.9106   0.9530   0.9591
## Pos Pred Value         0.8616   0.6703   0.6564   0.7247   0.7965
## Neg Pred Value         0.9605   0.9147   0.9559   0.9279   0.9390
## Prevalence             0.2918   0.1857   0.1755   0.1651   0.1819
## Detection Rate         0.2645   0.1150   0.1408   0.1033   0.1310
## Detection Prevalence   0.3069   0.1716   0.2145   0.1425   0.1644
## Balanced Accuracy      0.9232   0.7750   0.8566   0.7893   0.8395
```

So, we can see that the overall accuracy is 0.7545.

Random forest

In this section, the random forest model is used in the train.train data set.

```
rf <- randomForest(classe ~ ., data = train.train, type = "class")
prediction.rf <- predict(rf, train.cv, type = "class")
rf.performance <- confusionMatrix(prediction.rf, train.cv$classe)
cat(noquote("Here is the performance of the random forest model in the cross validation data set:"))
```

```
## Here is the performance of the random forest model in the cross validation data set:
```

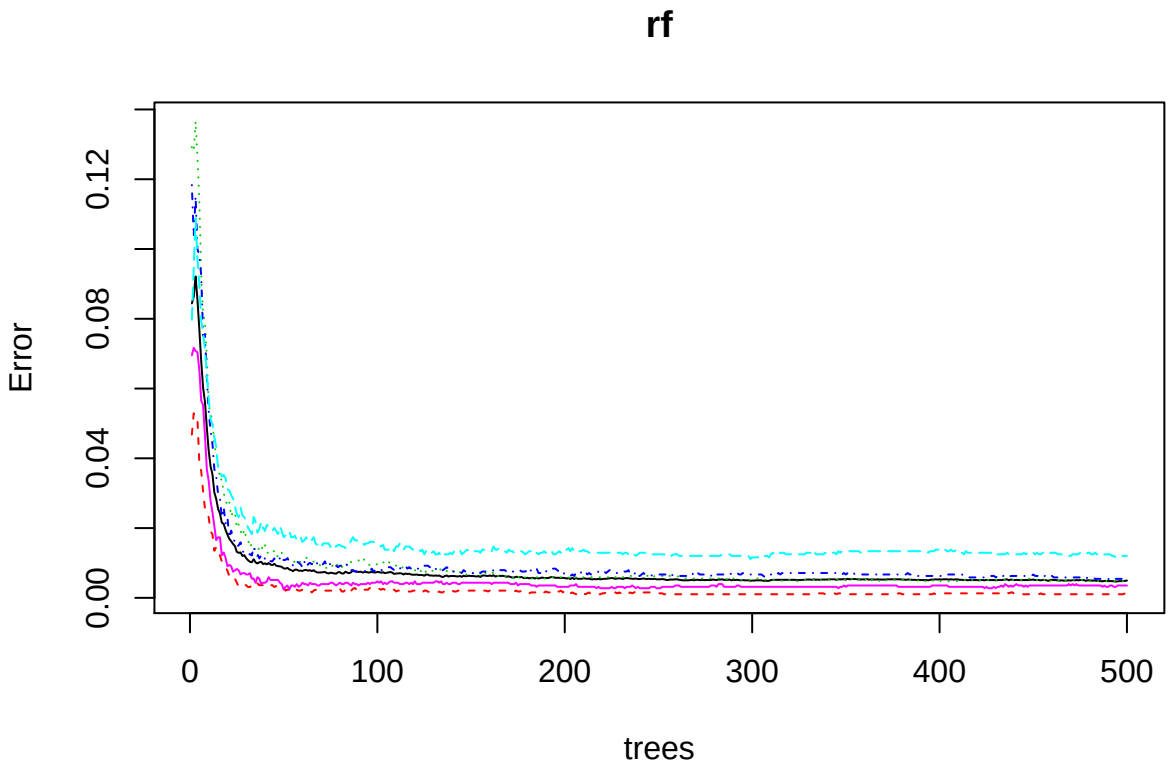
```
cat("\n")
```

```
rf.performance
```

```
## Confusion Matrix and Statistics
##
##           Reference
```

```
## Prediction    A    B    C    D    E
##          A 1718    4    0    0    0
##          B    0 1087    7    0    0
##          C    0    2 1025   13    0
##          D    0    0    1  957    1
##          E    0    0    0    2 1070
##
## Overall Statistics
##
##                Accuracy : 0.9949
##                  95% CI : (0.9927, 0.9966)
##     No Information Rate : 0.2918
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.9935
##  Mcnemar's Test P-Value : NA
##
## Statistics by Class:
##
##                      Class: A Class: B Class: C Class: D Class: E
## Sensitivity            1.0000   0.9945   0.9923   0.9846   0.9991
## Specificity            0.9990   0.9985   0.9969   0.9996   0.9996
## Pos Pred Value         0.9977   0.9936   0.9856   0.9979   0.9981
## Neg Pred Value         1.0000   0.9987   0.9983   0.9970   0.9998
## Prevalence             0.2918   0.1857   0.1755   0.1651   0.1819
## Detection Rate         0.2918   0.1846   0.1741   0.1626   0.1818
## Detection Prevalence   0.2925   0.1858   0.1767   0.1629   0.1821
## Balanced Accuracy      0.9995   0.9965   0.9946   0.9921   0.9993
```

```
plot(rf)
```



It seems that the random forest has a very good estimation already, the overall accuracy is 0.9949 in the cross validation data set. So, it should perform quite well in the test data also.

In this model, the number of trees is set as default (500). However, it might not be necessary to have those many trees in a forest, the plot shows that the error can be almost no-decreasing once the number of trees is smaller than 50.

However, let's take a look at the boosting method also.

Boosting method

```
fitControl <- trainControl(method = "repeatedcv", number = 2, repeats = 1)

boosting <- train(classe ~ ., data = train.train, method = "gbm",
                  trControl = fitControl,
                  verbose = FALSE)
```

```
## Warning: package 'gbm' was built under R version 3.4.2
```

```
## Loading required package: survival
```

```
##
## Attaching package: 'survival'
```

```
## The following object is masked from 'package:caret':
##
##     cluster
```

```
## Loading required package: splines

## Loading required package: parallel

## Loaded gbm 2.1.3

## Warning: package 'plyr' was built under R version 3.4.2
```

```
boosting.model <- boosting$finalModel
prediction.boosting <- predict(boosting, train.cv)
boosting.performance <- confusionMatrix(prediction.boosting, train.cv$classe)
cat(noquote("The performance for the general boosting model in cross validation data set is:"))
```

```
## The performance for the general boosting model in cross validation data set is:
```

```
cat("\n")
```

```
boosting.performance
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction    A    B    C    D    E
##          A 1690   21    0    0    1
##          B   21 1038   29    3    9
##          C    5   31  985   33   11
##          D    1    2   17  924   12
##          E    1    1    2   12 1038
##
## Overall Statistics
##
##                Accuracy : 0.964
##                  95% CI : (0.9589, 0.9686)
##     No Information Rate : 0.2918
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.9544
##  Mcnemar's Test P-Value : 0.007554
##
## Statistics by Class:
##
##                      Class: A Class: B Class: C Class: D Class: E
## Sensitivity            0.9837   0.9497   0.9535   0.9506   0.9692
## Specificity            0.9947   0.9871   0.9835   0.9935   0.9967
## Pos Pred Value         0.9871   0.9436   0.9249   0.9665   0.9848
## Neg Pred Value         0.9933   0.9885   0.9900   0.9903   0.9932
## Prevalence             0.2918   0.1857   0.1755   0.1651   0.1819
## Detection Rate         0.2871   0.1763   0.1673   0.1570   0.1763
## Detection Prevalence   0.2908   0.1869   0.1809   0.1624   0.1790
## Balanced Accuracy      0.9892   0.9684   0.9685   0.9721   0.9829
```

The performance of the boosting method is 0.9616 overall accuracy. I think the accuracy would be much better if we can increase the number in the "trainControl"; however, it would take too much CPU time on my machine. So, I am unable to do that.

Now, let's take a prediction of all values in test data set.

Test data set

```
test.tree <- predict(tree, test.data, type = "class")
test.rf <- predict(rf, test.data)
test.boosting <- predict(boosting, test.data)

cat(noquote("Here are the summary of all predicted results in test data set:"))
```

```
## Here are the summary of all predicted results in test data set:
```

```
test.tree
```

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

```
test.rf
```

```
##  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20
##  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
```

```
test.boosting
```

```
##  [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
```

So, let's use the results from random forest to estimate the classe in test data set.