Practical Machine Learning Project

Tianyuan Xie

December 27, 2016

1. Backgroud

Using devices such as Jawbone Up, Nike FuelBand, and Fitbit it is now possible to collect a large amount of data about personal activity relatively inexpensively. These type of devices are part of the quantified self movement - a group of enthusiasts who take measurements about themselves regularly to improve their health, to find patterns in their behavior, or because they are tech geeks. 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. They were asked to perform barbell lifts correctly and incorrectly in 5 different ways. More information is available from the website here: http://groupware.les.inf.puc-rio.br/har (http://groupware.les.inf.puc-rio.br/har) (see the section on the Weight Lifting Exercise Dataset).

2. Load data

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.2.5
```

```
library(caret)
```

```
## Warning: package 'caret' was built under R version 3.2.5
```

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

```
library(e1071)
```

```
## Warning: package 'e1071' was built under R version 3.2.5
```

```
library(randomForest)
```

```
## Warning: package 'randomForest' was built under R version 3.2.5
```

```
## 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
```

```
setwd('C:/Courses/coursera/08 Machine learning/project')
source('multiplot.R')

url_train <- 'https://d396qusza40orc.cloudfront.net/predmachlearn/pml-training.csv'
url_test <- 'https://d396qusza40orc.cloudfront.net/predmachlearn/pml-testing.csv'
filename_train <- 'train.csv'
filename_test <- 'test.csv'
if(!file.exists(filename_train) & !file.exists(filename_test)) {
        download.file(url_train, filename_train)
        download.file(url_test, filename_test)
}
training <- read.csv(filename_train, na.strings = c('NA', ''))
testing <- read.csv(filename_test, na.strings = c('NA', ''))
```

3. explore data

```
dim(training)
```

```
## [1] 19622   160
```

```
head(training)
```

```
##   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               <NA>                <NA>              <NA>
## 2               <NA>                <NA>              <NA>
## 3               <NA>                <NA>              <NA>
## 4               <NA>                <NA>              <NA>
## 5               <NA>                <NA>              <NA>
## 6               <NA>                <NA>              <NA>
##   skewness_roll_belt skewness_roll_belt.1 skewness_yaw_belt max_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
##   max_picth_belt max_yaw_belt min_roll_belt min_pitch_belt min_yaw_belt
## 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>
##   amplitude_roll_belt amplitude_pitch_belt amplitude_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>
##   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              <NA>               <NA>
## 2          337          513              <NA>               <NA>
## 3          344          513              <NA>               <NA>
## 4          344          512              <NA>               <NA>
## 5          337          506              <NA>               <NA>
## 6          342          513              <NA>               <NA>
##   kurtosis_yaw_arm skewness_roll_arm skewness_pitch_arm skewness_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>
##   max_roll_arm max_picth_arm max_yaw_arm min_roll_arm min_pitch_arm
## 1           NA            NA          NA           NA            NA
## 2           NA            NA          NA           NA            NA
## 3           NA            NA          NA           NA            NA
## 4           NA            NA          NA           NA            NA
## 5           NA            NA          NA           NA            NA
## 6           NA            NA          NA           NA            NA
##   min_yaw_arm amplitude_roll_arm amplitude_pitch_arm amplitude_yaw_arm
## 1          NA                 NA                  NA                NA
## 2          NA                 NA                  NA                NA
## 3          NA                 NA                  NA                NA
## 4          NA                 NA                  NA                NA
## 5          NA                 NA                  NA                NA
## 6          NA                 NA                  NA                NA
##   roll_dumbbell pitch_dumbbell yaw_dumbbell kurtosis_roll_dumbbell
## 1      13.05217      -70.49400    -84.87394                   <NA>
## 2      13.13074      -70.63751    -84.71065                   <NA>
## 3      12.85075      -70.27812    -85.14078                   <NA>
## 4      13.43120      -70.39379    -84.87363                   <NA>
## 5      13.37872      -70.42856    -84.85306                   <NA>
## 6      13.38246      -70.81759    -84.46500                   <NA>
##   kurtosis_picth_dumbbell kurtosis_yaw_dumbbell skewness_roll_dumbbell
## 1                    <NA>                  <NA>                   <NA>
## 2                    <NA>                  <NA>                   <NA>
## 3                    <NA>                  <NA>                   <NA>
## 4                    <NA>                  <NA>                   <NA>
## 5                    <NA>                  <NA>                   <NA>
## 6                    <NA>                  <NA>                   <NA>
##   skewness_pitch_dumbbell skewness_yaw_dumbbell max_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
##   max_picth_dumbbell max_yaw_dumbbell min_roll_dumbbell min_pitch_dumbbell
## 1                 NA             <NA>                NA                 NA
## 2                 NA             <NA>                NA                 NA
## 3                 NA             <NA>                NA                 NA
## 4                 NA             <NA>                NA                 NA
## 5                 NA             <NA>                NA                 NA
## 6                 NA             <NA>                NA                 NA
##   min_yaw_dumbbell amplitude_roll_dumbbell amplitude_pitch_dumbbell
## 1             <NA>                      NA                       NA
## 2             <NA>                      NA                       NA
## 3             <NA>                      NA                       NA
## 4             <NA>                      NA                       NA
## 5             <NA>                      NA                       NA
## 6             <NA>                      NA                       NA
##   amplitude_yaw_dumbbell total_accel_dumbbell var_accel_dumbbell
## 1                   <NA>                   37                 NA
## 2                   <NA>                   37                 NA
## 3                   <NA>                   37                 NA
## 4                   <NA>                   37                 NA
## 5                   <NA>                   37                 NA
## 6                   <NA>                   37                 NA
##   avg_roll_dumbbell stddev_roll_dumbbell var_roll_dumbbell
## 1                NA                   NA                NA
## 2                NA                   NA                NA
## 3                NA                   NA                NA
## 4                NA                   NA                NA
## 5                NA                   NA                NA
## 6                NA                   NA                NA
##   avg_pitch_dumbbell stddev_pitch_dumbbell var_pitch_dumbbell
## 1                 NA                    NA                 NA
## 2                 NA                    NA                 NA
## 3                 NA                    NA                 NA
## 4                 NA                    NA                 NA
## 5                 NA                    NA                 NA
## 6                 NA                    NA                 NA
##   avg_yaw_dumbbell stddev_yaw_dumbbell var_yaw_dumbbell gyros_dumbbell_x
## 1               NA                  NA               NA                0
## 2               NA                  NA               NA                0
## 3               NA                  NA               NA                0
## 4               NA                  NA               NA                0
## 5               NA                  NA               NA                0
## 6               NA                  NA               NA                0
##   gyros_dumbbell_y gyros_dumbbell_z accel_dumbbell_x accel_dumbbell_y
## 1            -0.02             0.00             -234               47
## 2            -0.02             0.00             -233               47
## 3            -0.02             0.00             -232               46
## 4            -0.02            -0.02             -232               48
## 5            -0.02             0.00             -233               48
## 6            -0.02             0.00             -234               48
##   accel_dumbbell_z magnet_dumbbell_x magnet_dumbbell_y magnet_dumbbell_z
## 1             -271              -559               293               -65
## 2             -269              -555               296               -64
## 3             -270              -561               298               -63
## 4             -269              -552               303               -60
## 5             -270              -554               292               -68
## 6             -269              -558               294               -66
##   roll_forearm pitch_forearm yaw_forearm kurtosis_roll_forearm
## 1         28.4         -63.9        -153                  <NA>
## 2         28.3         -63.9        -153                  <NA>
## 3         28.3         -63.9        -152                  <NA>
## 4         28.1         -63.9        -152                  <NA>
## 5         28.0         -63.9        -152                  <NA>
## 6         27.9         -63.9        -152                  <NA>
##   kurtosis_picth_forearm kurtosis_yaw_forearm skewness_roll_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>
##   skewness_pitch_forearm skewness_yaw_forearm max_roll_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
##   max_picth_forearm max_yaw_forearm min_roll_forearm min_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
##   min_yaw_forearm amplitude_roll_forearm amplitude_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
##   amplitude_yaw_forearm total_accel_forearm var_accel_forearm
## 1                  <NA>                  36                NA
## 2                  <NA>                  36                NA
## 3                  <NA>                  36                NA
## 4                  <NA>                  36                NA
## 5                  <NA>                  36                NA
## 6                  <NA>                  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
```

4. data preprocessing remove missing value

```
value.column <- apply(training, 2, function(x) sum(is.na(x))/dim(training)[1] < 0.8)
# 100 columns are almost empty so remove them
value.training <- training[, unname(value.column)]
```

The first seven columns are meaningless to model, so remove them as well. do the same thing to testing dataset

```
value.training <- value.training[,-(1:7)]
value.testing <- testing[, unname(value.column)]
value.testing <- value.testing[, -(1:7)]
```

check if the variables in the training and testing dataset are the same

```
all.equal(names(value.training)[-53], names(value.testing)[-53])
```

```
## [1] TRUE
```

check zero covariates

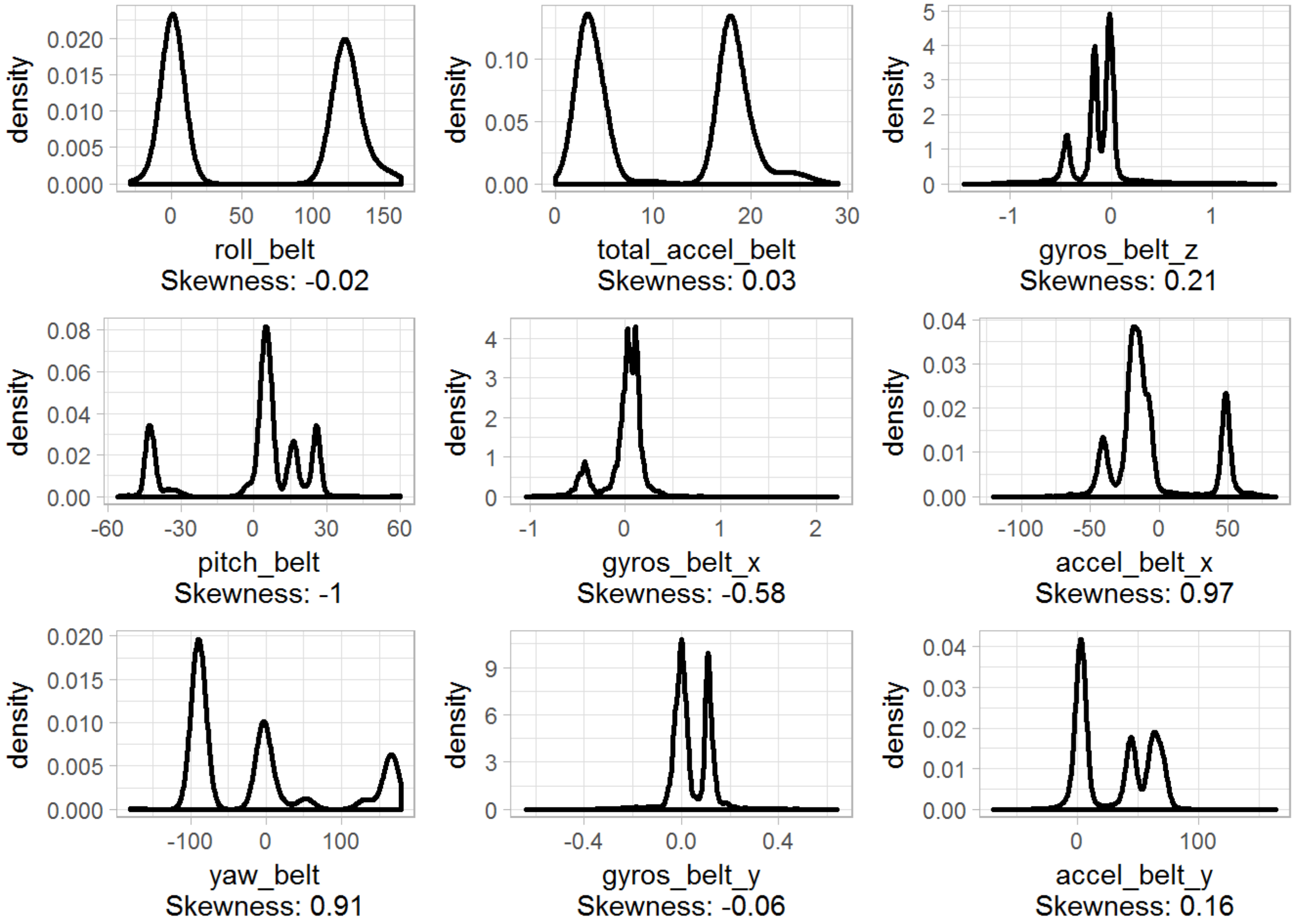
```
nsv <- nearZeroVar(value.training, saveMetrics = T)
nsv
```

```
##                       freqRatio percentUnique zeroVar   nzv
## roll_belt              1.101904     6.7781062   FALSE FALSE
## pitch_belt             1.036082     9.3772296   FALSE FALSE
## yaw_belt               1.058480     9.9734991   FALSE FALSE
## total_accel_belt       1.063160     0.1477933   FALSE FALSE
## gyros_belt_x           1.058651     0.7134849   FALSE FALSE
## gyros_belt_y           1.144000     0.3516461   FALSE FALSE
## gyros_belt_z           1.066214     0.8612782   FALSE FALSE
## accel_belt_x           1.055412     0.8357966   FALSE FALSE
## accel_belt_y           1.113725     0.7287738   FALSE FALSE
## accel_belt_z           1.078767     1.5237998   FALSE FALSE
## magnet_belt_x          1.090141     1.6664968   FALSE FALSE
## magnet_belt_y          1.099688     1.5187035   FALSE FALSE
## magnet_belt_z          1.006369     2.3290184   FALSE FALSE
## roll_arm              52.338462    13.5256345   FALSE FALSE
## pitch_arm             87.256410    15.7323412   FALSE FALSE
## yaw_arm               33.029126    14.6570176   FALSE FALSE
## total_accel_arm        1.024526     0.3363572   FALSE FALSE
## gyros_arm_x            1.015504     3.2769341   FALSE FALSE
## gyros_arm_y            1.454369     1.9162165   FALSE FALSE
## gyros_arm_z            1.110687     1.2638875   FALSE FALSE
## accel_arm_x            1.017341     3.9598410   FALSE FALSE
## accel_arm_y            1.140187     2.7367241   FALSE FALSE
## accel_arm_z            1.128000     4.0362858   FALSE FALSE
## magnet_arm_x           1.000000     6.8239731   FALSE FALSE
## magnet_arm_y           1.056818     4.4439914   FALSE FALSE
## magnet_arm_z           1.036364     6.4468454   FALSE FALSE
## roll_dumbbell          1.022388    84.2065029   FALSE FALSE
## pitch_dumbbell         2.277372    81.7449801   FALSE FALSE
## yaw_dumbbell           1.132231    83.4828254   FALSE FALSE
## total_accel_dumbbell   1.072634     0.2191418   FALSE FALSE
## gyros_dumbbell_x       1.003268     1.2282132   FALSE FALSE
## gyros_dumbbell_y       1.264957     1.4167771   FALSE FALSE
## gyros_dumbbell_z       1.060100     1.0498420   FALSE FALSE
## accel_dumbbell_x       1.018018     2.1659362   FALSE FALSE
## accel_dumbbell_y       1.053061     2.3748853   FALSE FALSE
## accel_dumbbell_z       1.133333     2.0894914   FALSE FALSE
## magnet_dumbbell_x      1.098266     5.7486495   FALSE FALSE
## magnet_dumbbell_y      1.197740     4.3012945   FALSE FALSE
## magnet_dumbbell_z      1.020833     3.4451126   FALSE FALSE
## roll_forearm          11.589286    11.0895933   FALSE FALSE
## pitch_forearm         65.983051    14.8557741   FALSE FALSE
## yaw_forearm           15.322835    10.1467740   FALSE FALSE
## total_accel_forearm    1.128928     0.3567424   FALSE FALSE
## gyros_forearm_x        1.059273     1.5187035   FALSE FALSE
## gyros_forearm_y        1.036554     3.7763735   FALSE FALSE
## gyros_forearm_z        1.122917     1.5645704   FALSE FALSE
## accel_forearm_x        1.126437     4.0464784   FALSE FALSE
## accel_forearm_y        1.059406     5.1116094   FALSE FALSE
## accel_forearm_z        1.006250     2.9558659   FALSE FALSE
## magnet_forearm_x       1.012346     7.7667924   FALSE FALSE
## magnet_forearm_y       1.246914     9.5403119   FALSE FALSE
## magnet_forearm_z      1.000000     8.5771073   FALSE FALSE
## classe                1.469581     0.0254816   FALSE FALSE
```

see the density distribution

```
plotDen <- function(data_in, i, lab){
        data <- data.frame(x=data_in[[i]], y=lab)
        p <- ggplot(data= data) + geom_density(aes(x = x), size = 1,alpha = 1.0) +
                xlab(paste0((colnames(data_in)[i]), '\n', 'Skewness:
',round(skewness(data_in[[i]], na.rm = TRUE), 2))) +
                theme_light()
        return(p)
}

pb = list()
for (i in 1:9) {
        pb[[i]] <- plotDen(value.training, i, 'den')
}
# Plot the first nine variables' distribution
multiplot(plotlist = pb, cols = 3)
```



All the skewness are calculated, the skewness correction is done. However, since the nonlinear model will be implemented. skewness correction does help to improve the model accuracy. So this step is ignored.

```
# skewness calculation
#skew <- apply(value.training[,-53], 2, skewness, na.rm = T)
# some variables have very large skewness
#high_skew_name <- names(value.training[,abs(unname(skew)) > 10])
#value.training[, high_skew_name] <- apply(value.training[, high_skew_name], 2, function(x) log1
0(x-min(x)+1))
```

Split the training data to training and testing dataset

```
set.seed(1234)
inTrain <- createDataPartition(y = value.training$classe, p = 0.75, list = F)
s.train <- value.training[inTrain,]
s.test <- value.training[-inTrain,]
```

5. model implementation CART

```
set.seed(1234)
modelfit1 <- train(classe ~ ., method = 'rpart', data = s.train)
```

```
## Loading required package: rpart
```

```
pred1 <- predict(modelfit1, s.test)
confusionMatrix(s.test$classe, pred1)
```

```
## Confusion Matrix and Statistics
## 
##           Reference
## Prediction    A    B    C    D    E
##          A 1275   28   88    0    4
##          B  390  340  219    0    0
##          C  416   31  408    0    0
##          D  356  141  307    0    0
##          E  134  131  230    0  406
## 
## Overall Statistics
## 
##                Accuracy : 0.4953          
##                  95% CI : (0.4812, 0.5094)
##     No Information Rate : 0.5243          
##     P-Value [Acc > NIR] : 1               
## 
##                   Kappa : 0.3399          
##  Mcnemar's Test P-Value : NA              
## 
## Statistics by Class:
## 
##                      Class: A Class: B Class: C Class: D Class: E
## Sensitivity            0.4959  0.50671   0.3259       NA  0.99024
## Specificity            0.9486  0.85613   0.8776   0.8361  0.88985
## Pos Pred Value         0.9140  0.35827   0.4772       NA  0.45061
## Neg Pred Value         0.6307  0.91631   0.7916       NA  0.99900
## Prevalence             0.5243  0.13683   0.2553   0.0000  0.08361
## Detection Rate         0.2600  0.06933   0.0832   0.0000  0.08279
## Detection Prevalence   0.2845  0.19352   0.1743   0.1639  0.18373
## Balanced Accuracy      0.7222  0.68142   0.6017       NA  0.94005
```

The model accuracy is around 50%, which is not acceptable.

Random Forest

```
modelfit2 <- randomForest(classe ~ .,data = s.train, trControl=trainControl(method = "cv", numbe
r = 4))
pred2 <- predict(modelfit2, s.test)
confusionMatrix(s.test$classe, pred2)
```

```
## Confusion Matrix and Statistics
## 
##           Reference
## Prediction    A    B    C    D    E
##          A 1395    0    0    0    0
##          B    3  944    2    0    0
##          C    0    9  845    1    0
##          D    0    0    6  798    0
##          E    0    0    0    0  901
## 
## Overall Statistics
## 
##                Accuracy : 0.9957
##                  95% CI : (0.9935, 0.9973)
##     No Information Rate : 0.2851
##     P-Value [Acc > NIR] : < 2.2e-16
## 
##                   Kappa : 0.9946
##  Mcnemar's Test P-Value : NA
## 
## Statistics by Class:
## 
##                      Class: A Class: B Class: C Class: D Class: E
## Sensitivity            0.9979   0.9906   0.9906   0.9987   1.0000
## Specificity            1.0000   0.9987   0.9975   0.9985   1.0000
## Pos Pred Value         1.0000   0.9947   0.9883   0.9925   1.0000
## Neg Pred Value         0.9991   0.9977   0.9980   0.9998   1.0000
## Prevalence             0.2851   0.1943   0.1739   0.1629   0.1837
## Detection Rate         0.2845   0.1925   0.1723   0.1627   0.1837
## Detection Prevalence   0.2845   0.1935   0.1743   0.1639   0.1837
## Balanced Accuracy      0.9989   0.9946   0.9941   0.9986   1.0000
```

The random forest model shows 99% accurcy in data classification. So this algorithm is chosen to predict the class in the testing dataset.

6. Testing dataset prediction

```
predict(modelfit2, value.testing)
```

```
##  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
```