Predicting the manner in which peole did the exercise

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. This work use data from accelerometers on the belt, forearm, arm, and dumbell of 6 participants to to predict the manner in which people did the exercise.

Reading data

We download and read the data that we will use to train and predict the model (training) and the data that we will use to to predict 20 different test cases (testing)

```
# Training data
url_train <- "https://d396qusza40orc.cloudfront.net/predmachlearn/pml-training.csv"
download.file(url_train, paste0(tempdir(),"/pml-training.csv"))
training <- read.csv(paste0(tempdir(),"/pml-training.csv"))

# Testing data
url_test <- "https://d396qusza40orc.cloudfront.net/predmachlearn/pml-testing.csv"
download.file(url_test, paste0(tempdir(),"/pml-testing.csv"))
testing <- read.csv(paste0(tempdir(),"/pml-testing.csv"))
```

Cleaning data

The non predictors variables, near zero variance predictor and NA variables are excluded from the data bases.

```
# Remove non predictor variables
training <- training %>% select(-X, -user_name)

# Remove near zero variance predictor
nzv <- training %>% nearZeroVar()
training <- training[, -nzv]

# Remove columns with NA
col <- !sapply(training, function(x) any(is.na(x)))
training <- training[, col]

# Cleaning 'testing' data
clean <- names(training)
clean <- clean[1:56] #remove 'classe' variable
testing <- testing %>% select(clean)
```

Partition

The training data is split in newTraining (70%) and newTesting (30%)

```
inTrain <- createDataPartition(training$classe, p = 0.7, list = FALSE)
newTraining <- training[ inTrain,]
newTesting <- training[-inTrain,]
dim(newTraining)
```

```
## [1] 13737    57
```

```
dim(newTesting)
```

```
## [1] 5885   57
```

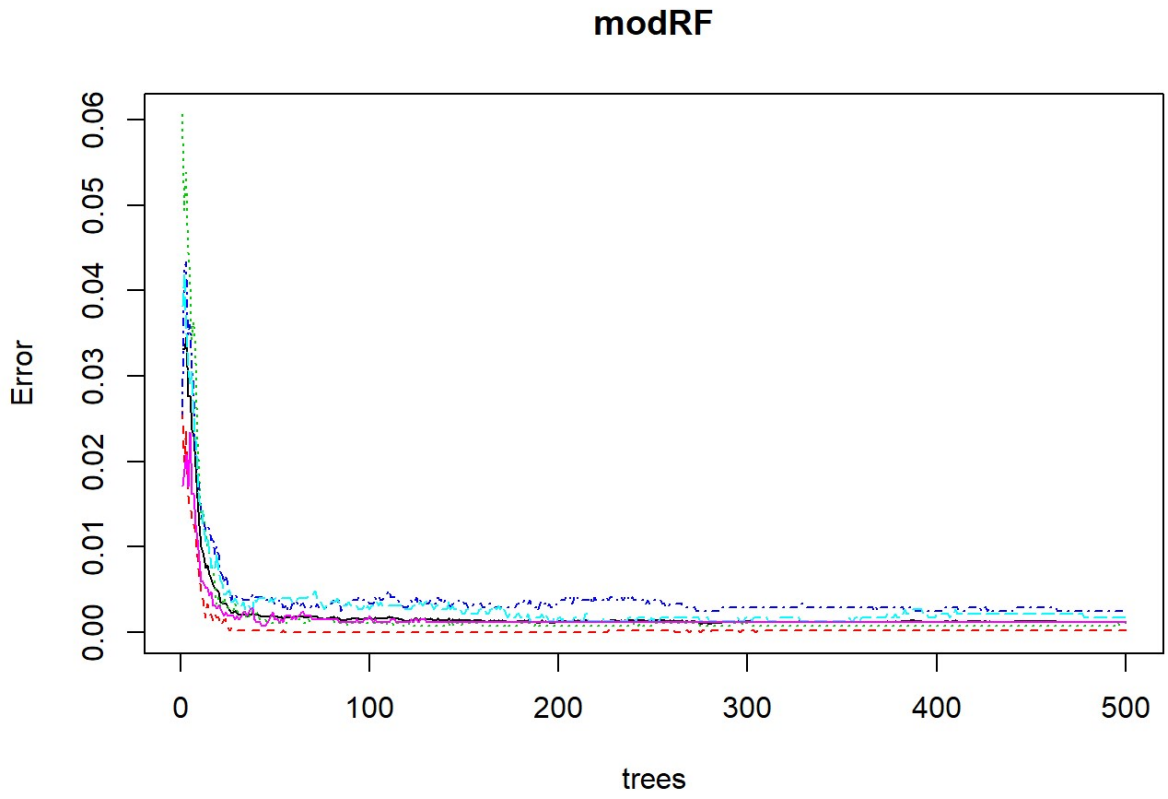
Training and predicting

We use several methods and specifications (random forest, boosting, linear discriminant analysis).

```
modGBM <- train(classe~., method = "gbm", data = newTraining)
modLDA <- train(classe~., method = "lda", data = newTraining)

predGBM <- predict(modGBM, newdata = newTesting)
predLDA <- predict(modLDA, newdata = newTesting)
```

However, the random forest model, with 5-Fold cross validation, presented the highest accuracy.

```
set.seed(325)

# Cross validation
ctr <- trainControl(method="cv",number=5)

# Training
modRF <- randomForest(classe~., data = newTraining, trControl= ctr)
plot(modRF)
```




```
# Predicting
predRF <- predict(modRF, newdata = newTesting)
confusionMatrix(predRF, newTesting$classe)
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction    A    B    C    D    E
##          A 1674    0    0    0    0
##          B    0 1139    2    0    0
##          C    0    0 1024    2    0
##          D    0    0    0  962    1
##          E    0    0    0    0 1081
##
## Overall Statistics
##
##                Accuracy : 0.9992
##                  95% CI : (0.998, 0.9997)
##     No Information Rate : 0.2845
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.9989
##  Mcnemar's Test P-Value : NA
##
## Statistics by Class:
##
##                      Class: A Class: B Class: C Class: D Class: E
## Sensitivity            1.0000   1.0000   0.9981   0.9979   0.9991
## Specificity            1.0000   0.9996   0.9996   0.9998   1.0000
## Pos Pred Value         1.0000   0.9982   0.9981   0.9990   1.0000
## Neg Pred Value         1.0000   1.0000   0.9996   0.9996   0.9998
## Prevalence             0.2845   0.1935   0.1743   0.1638   0.1839
## Detection Rate         0.2845   0.1935   0.1740   0.1635   0.1837
## Detection Prevalence   0.2845   0.1939   0.1743   0.1636   0.1837
## Balanced Accuracy      1.0000   0.9998   0.9988   0.9989   0.9995
```

This is the importance of each variable

```
imp <- varImp(modRF)
head(order(imp, decreasing = T), 10)
```

```
##  [1]  3  1  4  5  7 42  6 43 45 44
```

Test cases

The final model is applied to predict 20 different test cases.

```
pred <- predict(modRF, testing, type = "class")
pred
```

Reference

Velloso, E.; Bulling, A.; Gellersen, H.; Ugulino, W.; Fuks, H. Qualitative Activity Recognition of Weight Lifting Exercises. Proceedings of 4th International Conference in Cooperation with SIGCHI (Augmented Human ’13) . Stuttgart, Germany: ACM SIGCHI, 2013.