Practical Machine Learning

Yading Song

21 December 2014

1. Set path and load the correct working directory and necessary packages.

```
setwd("~/Google Drive/JobApplication/DataScicentist/courses/08_PracticalMachineLearning/writeUp")
library(caret)
```

```
## Warning: package 'caret' was built under R version 3.1.2
```

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

2. Load the data for training and testing

```
# "classe" is the target variable
data <- read.csv("pml-training.csv")
```

3. Clean the data

```
factor_names <- names(data)
observations <- nrow(data)
noise <- vector()
count <- 1

# remove NA data points
for (i in seq(1:160)){
    na_number <- sum(is.na(data[factor_names[i]]))
    if (na_number > observations/2 ){
        noise[count] <- factor_names[i]
        count <- count + 1
        print(factor_names[i])
    }
}
```

```
## [1] "max_roll_belt"
## [1] "max_picth_belt"
## [1] "min_roll_belt"
## [1] "min_pitch_belt"
## [1] "amplitude_roll_belt"
## [1] "amplitude_pitch_belt"
## [1] "var_total_accel_belt"
## [1] "avg_roll_belt"
## [1] "stddev_roll_belt"
## [1] "var_roll_belt"
## [1] "avg_pitch_belt"
## [1] "stddev_pitch_belt"
## [1] "var_pitch_belt"
## [1] "avg_yaw_belt"
```

```
## [1] "stddev_yaw_belt"
## [1] "var_yaw_belt"
## [1] "var_accel_arm"
## [1] "avg_roll_arm"
## [1] "stddev_roll_arm"
## [1] "var_roll_arm"
## [1] "avg_pitch_arm"
## [1] "stddev_pitch_arm"
## [1] "var_pitch_arm"
## [1] "avg_yaw_arm"
## [1] "stddev_yaw_arm"
## [1] "var_yaw_arm"
## [1] "max_roll_arm"
## [1] "max_picth_arm"
## [1] "max_yaw_arm"
## [1] "min_roll_arm"
## [1] "min_pitch_arm"
## [1] "min_yaw_arm"
## [1] "amplitude_roll_arm"
## [1] "amplitude_pitch_arm"
## [1] "amplitude_yaw_arm"
## [1] "max_roll_dumbbell"
## [1] "max_picth_dumbbell"
## [1] "min_roll_dumbbell"
## [1] "min_pitch_dumbbell"
## [1] "amplitude_roll_dumbbell"
## [1] "amplitude_pitch_dumbbell"
## [1] "var_accel_dumbbell"
## [1] "avg_roll_dumbbell"
## [1] "stddev_roll_dumbbell"
## [1] "var_roll_dumbbell"
## [1] "avg_pitch_dumbbell"
## [1] "stddev_pitch_dumbbell"
## [1] "var_pitch_dumbbell"
## [1] "avg_yaw_dumbbell"
## [1] "stddev_yaw_dumbbell"
## [1] "var_yaw_dumbbell"
## [1] "max_roll_forearm"
## [1] "max_picth_forearm"
## [1] "min_roll_forearm"
## [1] "min_pitch_forearm"
## [1] "amplitude_roll_forearm"
## [1] "amplitude_pitch_forearm"
## [1] "var_accel_forearm"
## [1] "avg_roll_forearm"
## [1] "stddev_roll_forearm"
## [1] "var_roll_forearm"
## [1] "avg_pitch_forearm"
## [1] "stddev_pitch_forearm"
## [1] "var_pitch_forearm"
## [1] "avg_yaw_forearm"
## [1] "stddev_yaw_forearm"
## [1] "var_yaw_forearm"
```

```
redundent <- c("raw_timestamp_part_1", "raw_timestamp_part_2", "cvtd_timestamp", "new_window", "num_win
null_part1 <- c("kurtosis_roll_belt", "kurtosis_picth_belt", "kurtosis_yaw_belt",
                "skewness_roll_belt", "skewness_roll_belt.1", "skewness_yaw_belt",
                "max_yaw_belt", "min_yaw_belt", "amplitude_yaw_belt")
null_part2 <- c("kurtosis_roll_arm", "kurtosis_picth_arm", "kurtosis_yaw_arm",
                "skewness_roll_arm", "skewness_pitch_arm", "skewness_yaw_arm")
null_part3 <- c("kurtosis_roll_dumbbell", "kurtosis_picth_dumbbell",
                "kurtosis_yaw_dumbbell", "skewness_roll_dumbbell", "skewness_pitch_dumbbell",
                "skewness_yaw_dumbbell", "max_yaw_dumbbell", "min_yaw_dumbbell", "amplitude_yaw_dumbbel
null_part4 <- c("kurtosis_roll_forearm", "kurtosis_picth_forearm", "kurtosis_yaw_forearm",
                "skewness_roll_forearm", "skewness_pitch_forearm", "skewness_yaw_forearm",
                "max_yaw_forearm", "min_yaw_forearm", "amplitude_yaw_forearm")
noises <- c(noise, redundent, null_part1, null_part2, null_part3, null_part4)

my_vars <- factor_names %in% noises
clean_data <- data[!my_vars]
clean_data <- clean_data[, 3:55]
```