

Machine Learning Project

Contents

tidying 4

load data

```
training<-read.csv('./data/pml-training.csv',stringsAsFactors=TRUE,na.strings = c("#DIV/0!","NA"))
dim(training)
```

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

```
str(training,list.len=ncol(training))
```

```
## 'data.frame': 19622 obs. of 160 variables:
## $ X : int 1 2 3 4 5 6 7 8 9 10 ...
## $ user_name : Factor w/ 6 levels "adelmo","carlitos",...: 2 2 2 2 2 2 2 2 2 2 ...
## $ raw_timestamp_part_1 : int 1323084231 1323084231 1323084231 1323084232 1323084232 1323084232 ...
## $ raw_timestamp_part_2 : int 788290 808298 820366 120339 196328 304277 368296 440390 484323 484...
## $ cvtd_timestamp : Factor w/ 20 levels "02/12/2011 13:32",...: 9 9 9 9 9 9 9 9 9 9 ...
## $ new_window : Factor w/ 2 levels "no","yes": 1 1 1 1 1 1 1 1 1 1 ...
## $ num_window : int 11 11 11 12 12 12 12 12 12 12 ...
## $ 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 3 ...
## $ kurtosis_roll_belt : num NA NA NA NA NA NA NA NA NA NA ...
## $ kurtosis_pitch_belt : num NA NA NA NA NA NA NA NA NA NA ...
## $ kurtosis_yaw_belt : logi NA NA NA NA NA NA ...
## $ skewness_roll_belt : num NA NA NA NA NA NA NA NA NA NA ...
## $ skewness_roll_belt.1 : num NA NA NA NA NA NA NA NA NA NA ...
## $ skewness_yaw_belt : logi NA NA NA NA NA NA ...
## $ max_roll_belt : num NA NA NA NA NA NA NA NA NA NA ...
## $ max_pitch_belt : int NA NA NA NA NA NA NA NA NA NA ...
## $ max_yaw_belt : num NA NA NA NA NA NA NA NA NA NA ...
## $ min_roll_belt : num NA NA NA NA NA NA NA NA NA NA ...
## $ min_pitch_belt : int NA NA NA NA NA NA NA NA NA NA ...
## $ min_yaw_belt : num NA NA NA NA NA NA NA NA NA NA ...
## $ amplitude_roll_belt : num NA NA NA NA NA NA NA NA NA NA ...
## $ amplitude_pitch_belt : int NA NA NA NA NA NA NA NA NA NA ...
## $ amplitude_yaw_belt : num NA NA NA NA NA NA NA NA NA NA ...
## $ var_total_accel_belt : num NA NA NA NA NA NA NA NA NA NA ...
## $ avg_roll_belt : num NA NA NA NA NA NA NA NA NA NA ...
## $ stddev_roll_belt : num NA NA NA NA NA NA NA NA NA NA ...
## $ var_roll_belt : num NA NA NA NA NA NA NA NA NA NA ...
## $ avg_pitch_belt : num NA NA NA NA NA NA NA NA NA NA ...
## $ stddev_pitch_belt : num NA NA NA NA NA NA NA NA NA NA ...
## $ var_pitch_belt : num NA NA NA NA NA NA NA NA NA NA ...
## $ avg_yaw_belt : num NA NA NA NA NA NA NA NA NA NA ...
## $ stddev_yaw_belt : num NA NA NA NA NA NA NA NA NA NA ...
```

```

## $ var_yaw_belt : num NA NA NA NA NA NA NA NA NA NA ...
## $ 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 ...
## $ var_accel_arm : num NA NA NA NA NA NA NA NA NA NA ...
## $ avg_roll_arm : num NA NA NA NA NA NA NA NA NA NA ...
## $ stddev_roll_arm : num NA NA NA NA NA NA NA NA NA NA ...
## $ var_roll_arm : num NA NA NA NA NA NA NA NA NA NA ...
## $ avg_pitch_arm : num NA NA NA NA NA NA NA NA NA NA ...
## $ stddev_pitch_arm : num NA NA NA NA NA NA NA NA NA NA ...
## $ var_pitch_arm : num NA NA NA NA NA NA NA NA NA NA ...
## $ avg_yaw_arm : num NA NA NA NA NA NA NA NA NA NA ...
## $ stddev_yaw_arm : num NA NA NA NA NA NA NA NA NA NA ...
## $ var_yaw_arm : num NA NA NA NA NA NA NA NA NA NA ...
## $ 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 ...
## $ kurtosis_roll_arm : num NA NA NA NA NA NA NA NA NA NA ...
## $ kurtosis_pitch_arm : num NA NA NA NA NA NA NA NA NA NA ...
## $ kurtosis_yaw_arm : num NA NA NA NA NA NA NA NA NA NA ...
## $ skewness_roll_arm : num NA NA NA NA NA NA NA NA NA NA ...
## $ skewness_pitch_arm : num NA NA NA NA NA NA NA NA NA NA ...
## $ skewness_yaw_arm : num NA NA NA NA NA NA NA NA NA NA ...
## $ max_roll_arm : num NA NA NA NA NA NA NA NA NA NA ...
## $ max_pitch_arm : num NA NA NA NA NA NA NA NA NA NA ...
## $ max_yaw_arm : int NA NA NA NA NA NA NA NA NA NA ...
## $ min_roll_arm : num NA NA NA NA NA NA NA NA NA NA ...
## $ min_pitch_arm : num NA NA NA NA NA NA NA NA NA NA ...
## $ min_yaw_arm : int NA NA NA NA NA NA NA NA NA NA ...
## $ amplitude_roll_arm : num NA NA NA NA NA NA NA NA NA NA ...
## $ amplitude_pitch_arm : num NA NA NA NA NA NA NA NA NA NA ...
## $ amplitude_yaw_arm : int NA NA NA NA NA NA NA NA NA NA ...
## $ 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 ...
## $ kurtosis_roll_dumbbell : num NA NA NA NA NA NA NA NA NA NA ...
## $ kurtosis_pitch_dumbbell : num NA NA NA NA NA NA NA NA NA NA ...
## $ kurtosis_yaw_dumbbell : logi NA NA NA NA NA NA ...

```

```

## $ skewness_roll_dumbbell : num NA NA NA NA NA NA NA NA NA NA ...
## $ skewness_pitch_dumbbell : num NA NA NA NA NA NA NA NA NA NA ...
## $ skewness_yaw_dumbbell : logi NA NA NA NA NA NA ...
## $ max_roll_dumbbell : num NA NA NA NA NA NA NA NA NA NA ...
## $ max_pitch_dumbbell : num NA NA NA NA NA NA NA NA NA NA ...
## $ max_yaw_dumbbell : num NA NA NA NA NA NA NA NA NA NA ...
## $ min_roll_dumbbell : num NA NA NA NA NA NA NA NA NA NA ...
## $ min_pitch_dumbbell : num NA NA NA NA NA NA NA NA NA NA ...
## $ min_yaw_dumbbell : num NA NA NA NA NA NA NA NA NA NA ...
## $ amplitude_roll_dumbbell : num NA NA NA NA NA NA NA NA NA NA ...
## $ amplitude_pitch_dumbbell : num NA NA NA NA NA NA NA NA NA NA ...
## $ amplitude_yaw_dumbbell : num NA NA NA NA NA NA NA NA NA NA ...
## $ total_accel_dumbbell : int 37 37 37 37 37 37 37 37 37 37 ...
## $ var_accel_dumbbell : num NA NA NA NA NA NA NA NA NA NA ...
## $ avg_roll_dumbbell : num NA NA NA NA NA NA NA NA NA NA ...
## $ stddev_roll_dumbbell : num NA NA NA NA NA NA NA NA NA NA ...
## $ var_roll_dumbbell : num NA NA NA NA NA NA NA NA NA NA ...
## $ avg_pitch_dumbbell : num NA NA NA NA NA NA NA NA NA NA ...
## $ stddev_pitch_dumbbell : num NA NA NA NA NA NA NA NA NA NA ...
## $ var_pitch_dumbbell : num NA NA NA NA NA NA NA NA NA NA ...
## $ avg_yaw_dumbbell : num NA NA NA NA NA NA NA NA NA NA ...
## $ stddev_yaw_dumbbell : num NA NA NA NA NA NA NA NA NA NA ...
## $ var_yaw_dumbbell : num NA NA NA NA NA NA NA NA NA NA ...
## $ 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 ...
## $ 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 ...
## $ kurtosis_roll_forearm : num NA NA NA NA NA NA NA NA NA NA ...
## $ kurtosis_pitch_forearm : num NA NA NA NA NA NA NA NA NA NA ...
## $ kurtosis_yaw_forearm : logi NA NA NA NA NA NA ...
## $ skewness_roll_forearm : num NA NA NA NA NA NA NA NA NA NA ...
## $ skewness_pitch_forearm : num NA NA NA NA NA NA NA NA NA NA ...
## $ skewness_yaw_forearm : logi NA NA NA NA NA NA ...
## $ max_roll_forearm : num NA NA NA NA NA NA NA NA NA NA ...
## $ max_pitch_forearm : num NA NA NA NA NA NA NA NA NA NA ...
## $ max_yaw_forearm : num NA NA NA NA NA NA NA NA NA NA ...
## $ min_roll_forearm : num NA NA NA NA NA NA NA NA NA NA ...
## $ min_pitch_forearm : num NA NA NA NA NA NA NA NA NA NA ...
## $ min_yaw_forearm : num NA NA NA NA NA NA NA NA NA NA ...
## $ amplitude_roll_forearm : num NA NA NA NA NA NA NA NA NA NA ...
## $ amplitude_pitch_forearm : num NA NA NA NA NA NA NA NA NA NA ...
## $ amplitude_yaw_forearm : num NA NA NA NA NA NA NA NA NA NA ...
## $ total_accel_forearm : int 36 36 36 36 36 36 36 36 36 36 ...
## $ var_accel_forearm : num NA NA NA NA NA NA NA NA NA NA ...
## $ avg_roll_forearm : num NA NA NA NA NA NA NA NA NA NA ...
## $ stddev_roll_forearm : num NA NA NA NA NA NA NA NA NA NA ...

```

```
## $ var_roll_forearm      : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ avg_pitch_forearm     : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ stddev_pitch_forearm  : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ var_pitch_forearm     : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ avg_yaw_forearm       : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ stddev_yaw_forearm    : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ var_yaw_forearm       : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ 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 ...
```

```
training$cvtd_time<-strptime(training$cvtd_timestamp,format = "%d/%m/%Y %H:%M")
```

tidying

write a function to do this, as we will need to tidy the test data also cvtd timestamp -> to date \$ max_roll_belt

max_pitch_belt min_roll_belt

min_pitch_belt amplitude_roll_belt

A bunch of others

set things to missing, impute the missing values

Then procede