

Exercises Predict

xiaoyi_Li

2016 年 9 月 8 日

Question

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, our goal will be to use data from accelerometers on the belt, forearm, arm, and dumbbell of 6 participants. #Input Data

```
setwd("d:/test data/homework")
training<-read.csv("pml-training.csv",header = T,stringsAsFactors = F)
testing<-read.csv("pml-testing.csv",header = T,stringsAsFactors = F)
names(training)
## [1] "X" "user_name"
## [3] "raw_timestamp_part_1" "raw_timestamp_part_2"
## [5] "cvtd_timestamp" "new_window"
## [7] "num_window" "roll_belt"
## [9] "pitch_belt" "yaw_belt"
## [11] "total_accel_belt" "kurtosis_roll_belt"
## [13] "kurtosis_pitch_belt" "kurtosis_yaw_belt"
## [15] "skewness_roll_belt" "skewness_roll_belt.1"
## [17] "skewness_yaw_belt" "max_roll_belt"
## [19] "max_pitch_belt" "max_yaw_belt"
## [21] "min_roll_belt" "min_pitch_belt"
## [23] "min_yaw_belt" "amplitude_roll_belt"
## [25] "amplitude_pitch_belt" "amplitude_yaw_belt"
## [27] "var_total_accel_belt" "avg_roll_belt"
## [29] "stddev_roll_belt" "var_roll_belt"
## [31] "avg_pitch_belt" "stddev_pitch_belt"
## [33] "var_pitch_belt" "avg_yaw_belt"
## [35] "stddev_yaw_belt" "var_yaw_belt"
## [37] "gyros_belt_x" "gyros_belt_y"
## [39] "gyros_belt_z" "accel_belt_x"
## [41] "accel_belt_y" "accel_belt_z"
## [43] "magnet_belt_x" "magnet_belt_y"
## [45] "magnet_belt_z" "roll_arm"
## [47] "pitch_arm" "yaw_arm"
## [49] "total_accel_arm" "var_accel_arm"
## [51] "avg_roll_arm" "stddev_roll_arm"
## [53] "var_roll_arm" "avg_pitch_arm"
## [55] "stddev_pitch_arm" "var_pitch_arm"
## [57] "avg_yaw_arm" "stddev_yaw_arm"
## [59] "var_yaw_arm" "gyros_arm_x"
```

## [61] "gyros_arm_y"	"gyros_arm_z"
## [63] "accel_arm_x"	"accel_arm_y"
## [65] "accel_arm_z"	"magnet_arm_x"
## [67] "magnet_arm_y"	"magnet_arm_z"
## [69] "kurtosis_roll_arm"	"kurtosis_pitch_arm"
## [71] "kurtosis_yaw_arm"	"skewness_roll_arm"
## [73] "skewness_pitch_arm"	"skewness_yaw_arm"
## [75] "max_roll_arm"	"max_pitch_arm"
## [77] "max_yaw_arm"	"min_roll_arm"
## [79] "min_pitch_arm"	"min_yaw_arm"
## [81] "amplitude_roll_arm"	"amplitude_pitch_arm"
## [83] "amplitude_yaw_arm"	"roll_dumbbell"
## [85] "pitch_dumbbell"	"yaw_dumbbell"
## [87] "kurtosis_roll_dumbbell"	"kurtosis_pitch_dumbbell"
## [89] "kurtosis_yaw_dumbbell"	"skewness_roll_dumbbell"
## [91] "skewness_pitch_dumbbell"	"skewness_yaw_dumbbell"
## [93] "max_roll_dumbbell"	"max_pitch_dumbbell"
## [95] "max_yaw_dumbbell"	"min_roll_dumbbell"
## [97] "min_pitch_dumbbell"	"min_yaw_dumbbell"
## [99] "amplitude_roll_dumbbell"	"amplitude_pitch_dumbbell"
## [101] "amplitude_yaw_dumbbell"	"total_accel_dumbbell"
## [103] "var_accel_dumbbell"	"avg_roll_dumbbell"
## [105] "stddev_roll_dumbbell"	"var_roll_dumbbell"
## [107] "avg_pitch_dumbbell"	"stddev_pitch_dumbbell"
## [109] "var_pitch_dumbbell"	"avg_yaw_dumbbell"
## [111] "stddev_yaw_dumbbell"	"var_yaw_dumbbell"
## [113] "gyros_dumbbell_x"	"gyros_dumbbell_y"
## [115] "gyros_dumbbell_z"	"accel_dumbbell_x"
## [117] "accel_dumbbell_y"	"accel_dumbbell_z"
## [119] "magnet_dumbbell_x"	"magnet_dumbbell_y"
## [121] "magnet_dumbbell_z"	"roll_forearm"
## [123] "pitch_forearm"	"yaw_forearm"
## [125] "kurtosis_roll_forearm"	"kurtosis_pitch_forearm"
## [127] "kurtosis_yaw_forearm"	"skewness_roll_forearm"
## [129] "skewness_pitch_forearm"	"skewness_yaw_forearm"
## [131] "max_roll_forearm"	"max_pitch_forearm"
## [133] "max_yaw_forearm"	"min_roll_forearm"
## [135] "min_pitch_forearm"	"min_yaw_forearm"
## [137] "amplitude_roll_forearm"	"amplitude_pitch_forearm"
## [139] "amplitude_yaw_forearm"	"total_accel_forearm"
## [141] "var_accel_forearm"	"avg_roll_forearm"
## [143] "stddev_roll_forearm"	"var_roll_forearm"
## [145] "avg_pitch_forearm"	"stddev_pitch_forearm"
## [147] "var_pitch_forearm"	"avg_yaw_forearm"

```

## [149] "stddev_yaw_forearm"      "var_yaw_forearm"
## [151] "gyros_forearm_x"         "gyros_forearm_y"
## [153] "gyros_forearm_z"         "accel_forearm_x"
## [155] "accel_forearm_y"         "accel_forearm_z"
## [157] "magnet_forearm_x"        "magnet_forearm_y"
## [159] "magnet_forearm_z"        "classe"
str(training)
## 'data.frame':    19622 obs. of  160 variables:
##  $ X                      : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ user_name               : chr  "carlitos" "carlitos" "carlitos" "carlitos" ...
##  $ raw_timestamp_part_1    : int  1323084231 1323084231 1323084231 1323084232
1323084232 1323084232 1323084232 1323084232 1323084232 ...
##  $ raw_timestamp_part_2    : int  788290 808298 820366 120339 196328 304277 368296
440390 484323 484434 ...
##  $ cvtd_timestamp         : chr  "05/12/2011 11:23" "05/12/2011 11:23" "05/12/2011
11:23" "05/12/2011 11:23" ...
##  $ new_window             : chr  "no" "no" "no" "no" ...
##  $ 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 ...
##  $ kurtosis_roll_belt     : chr  "" "" "" "" ...
##  $ kurtosis_pitch_belt    : chr  "" "" "" "" ...
##  $ kurtosis_yaw_belt      : chr  "" "" "" "" ...
##  $ skewness_roll_belt     : chr  "" "" "" "" ...
##  $ skewness_roll_belt.1   : chr  "" "" "" "" ...
##  $ skewness_yaw_belt      : chr  "" "" "" "" ...
##  $ 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           : chr  "" "" "" "" ...
##  $ 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           : chr  "" "" "" "" ...
##  $ 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     : chr  "" "" "" "" ...
##  $ 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
-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   : chr   "" "" "" "" "" ...
## $ kurtosis_pitch_arm  : chr   "" "" "" "" "" ...
## $ kurtosis_yaw_arm    : chr   "" "" "" "" "" ...
## $ skewness_roll_arm   : chr   "" "" "" "" "" ...
## $ skewness_pitch_arm  : chr   "" "" "" "" "" ...
## $ skewness_yaw_arm    : chr   "" "" "" "" "" ...

```

```
## $ max_roll_arm      : num  NA NA NA NA NA NA NA NA NA NA ...
## $ max_picth_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 : chr  "" "" "" "" "" ...
## $ kurtosis_picth_dumbbell : chr  "" "" "" "" "" ...
## $ kurtosis_yaw_dumbbell  : chr  "" "" "" "" "" ...
## $ skewness_roll_dumbbell : chr  "" "" "" "" "" ...
## $ skewness_pitch_dumbbell : chr  "" "" "" "" "" ...
## $ skewness_yaw_dumbbell  : chr  "" "" "" "" "" ...
## $ max_roll_dumbbell     : num  NA NA NA NA NA NA NA NA NA NA ...
## $ max_picth_dumbbell    : num  NA NA NA NA NA NA NA NA NA NA ...
## $ max_yaw_dumbbell      : chr   "" "" "" "" "" ...
## $ 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      : chr   "" "" "" "" "" ...
## $ amplitude_roll_dumbbell : num  NA NA NA NA NA NA NA NA NA NA ...
## [list output truncated]
```

Data Preprocess

```
library(caret,ggplot2)
```

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

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

```
training1<-training[,-c(12:36,50:59,69:83,87:100)]
```

```
training1<-training1[,-c(37,39:48,61:75,77:86)]
```

```
set.seed(1111)
```

```
ktraining<-createFolds(training1$X,k=10,list=TRUE,returnTrain=TRUE)
```

```
sapply(ktraining,length)
```

```
## Fold01 Fold02 Fold03 Fold04 Fold05 Fold06 Fold07 Fold08 Fold09 Fold10
```

```
## 17661 17661 17662 17660 17658 17658 17661 17659 17659 17659
```

train the model

in this project,we choose randomforest to fit the model;use the cross validation such as fold2 to test model1(fold1 fitted),fold4 to test model2(fold3 fitted),fold6 to test model3(fold5 fitted),fold8 to test model4(fold7 fitted),fold10 to test model5(fold9 fitted);use the confusion Matrix to compare the model expected out of sample error, then choose the min error which mode predict as the final model.

```

modelFit1<-train(classe~.,data=training1[ktraining[[1]],-1],preProcess=c("center","scale"),metho
d="rf")
## Loading required package: randomForest
## 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
pred1<-predict(modelFit1,training1[ktraining[[2]],-1])
modelFit2<-train(classe~.,data=training1[ktraining[[3]],-1],preProcess=c("center","scale"),metho
d="rf")
pred2<-predict(modelFit2,training1[ktraining[[4]],-1])
modelFit3<-train(classe~.,data=training1[ktraining[[5]],-1],preProcess=c("center","scale"),metho
d="rf")
pred3<-predict(modelFit3,training1[ktraining[[6]],-1])
modelFit4<-train(classe~.,data=training1[ktraining[[7]],-1],preProcess=c("center","scale"),metho
d="rf")
pred4<-predict(modelFit4,training1[ktraining[[8]],-1])
modelFit5<-train(classe~.,data=training1[ktraining[[9]],-1],preProcess=c("center","scale"),metho
d="rf")
pred5<-predict(modelFit5,training1[ktraining[[10]],-1])
confuse1<-confusionMatrix(pred1,training1[ktraining[[2]],60])
confuse2<-confusionMatrix(pred2,training1[ktraining[[4]],60])
confuse3<-confusionMatrix(pred3,training1[ktraining[[6]],60])
confuse4<-confusionMatrix(pred4,training1[ktraining[[8]],60])
confuse5<-confusionMatrix(pred5,training1[ktraining[[10]],60])
pred<-predict(pred5,testing)
Predict Result
B A B A A E D B A A B C B A E E A B B B

```