Prediction Model For Movement Execution

mml

Sunday, January 25, 2015

The dumbell lift, are you doing it right?

Introduction

In this project, your goal is 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. The five different ways are :

– exactly according to the specification (Class A)
– throwing the elbows to the front (Class B)
– lifting the dumbbell only halfway (Class C)
– lowering the dumbbell only halfway (Class D)
– throwing the hips to the front (Class E)

The goal of the project is to predict the manner in which they did the exercise according to their respective letter *A B C D* or *E*. This is the “classe” variable in the training set. We can use any of the other variables to predict with. We should create a report describing how we built our model, how we used cross validation, what we think the expected out of sample error is, and why we made the choices we did. We will therefore use our prediction model to predict 20 different test cases.

training data set for this project is `https://d396qusza40orc.cloudfront.net/predmachlearn/pml-training.csv`
test data set for this project is `https://d396qusza40orc.cloudfront.net/predmachlearn/pml-testing.csv`
more info is avalaible at `http://groupware.les.inf.puc-rio.br/har`

Pre-predictive work

Cross-validation:

1. Use the training set
2. Split it into training/test sets
3. Build a model on the training set
4. Evaluate on the test set
5. Repeat and average estimated errors

subtraining data: 75% of the training data set
subtest data: 25% of the training data
The models will fit on the subtraining data and tested on subset.
We will test on the real test set once the most/best predictive model is found.

The expected out of sample error will be 1-accuracy found from the cross-validation data set because `classe` is unordered (although we could have put an order to it) and is a factor variable.

Code and Predictive model

Libraries, seed and data

```
library(caret)
library(randomForest)
library(rpart)
library(rpart.plot)
library(e1071)
set.seed(1210)
train_set <-read.csv("malearntrain.csv", na.strings=c("NA","#DIV/0!",
                                                                    ""))
test_set <- read.csv("machlearntest.csv", na.strings=c("NA","#DIV/0!", "
"))
```

Cleaning data and exploration

```
dim(train_set)
dim(test_set)
head(train_set)
head(test_set)
##delete col with missing values
train_set<-train_set[,colSums(is.na(train_set)) == 0]
test_set <-test_set[,colSums(is.na(test_set)) == 0]
train_set <-train_set[,-c(1:7)]
test_set <-test_set[,-c(1:7)]
##check the new data set
dim(train_set)
dim(test_set)
head(train_set)
head(test_set)
```

Partitioning

```
subsamples <- createDataPartition(y=train_set$classe, p=0.75, list=FALSE)
subTraining <- train_set[subsamples, ]
subTesting <- train_set[-subsamples, ]
dim(subTraining)
```

```
## [1] 14718    53
```

```
dim(subTesting)
```

```
## [1] 4904   53
```

```
head(subTraining)
```

```
##    roll_belt pitch_belt yaw_belt total_accel_belt gyros_belt_x
## 1       1.41       8.07    -94.4                3         0.00
## 3       1.42       8.07    -94.4                3         0.00
## 7       1.42       8.09    -94.4                3         0.02
## 9       1.43       8.16    -94.4                3         0.02
## 11      1.45       8.18    -94.4                3         0.03
## 13      1.42       8.20    -94.4                3         0.02
##    gyros_belt_y gyros_belt_z accel_belt_x accel_belt_y accel_belt_z
## 1             0        -0.02          -21            4           22
## 3             0        -0.02          -20            5           23
## 7             0        -0.02          -22            3           21
## 9             0        -0.02          -20            2           24
## 11            0        -0.02          -21            2           23
## 13            0         0.00          -22            4           21
##    magnet_belt_x magnet_belt_y magnet_belt_z roll_arm pitch_arm yaw_arm
## 1             -3           599          -313     -128      22.5    -161
## 3             -2           600          -305     -128      22.5    -161
## 7             -4           599          -311     -128      21.9    -161
## 9              1           602          -312     -128      21.7    -161
## 11            -5           596          -317     -128      21.5    -161
## 13            -3           606          -309     -128      21.4    -161
##    total_accel_arm gyros_arm_x gyros_arm_y gyros_arm_z accel_arm_x
## 1               34        0.00        0.00       -0.02        -288
## 3               34        0.02       -0.02       -0.02        -289
## 7               34        0.00       -0.03        0.00        -289
## 9               34        0.02       -0.03       -0.02        -288
## 11              34        0.02       -0.03        0.00        -290
## 13              34        0.02       -0.02       -0.02        -287
##    accel_arm_y accel_arm_z magnet_arm_x magnet_arm_y magnet_arm_z
## 1          109        -123         -368          337          516
## 3          110        -126         -368          344          513
## 7          111        -125         -373          336          509
## 9          109        -122         -369          341          518
## 11         110        -123         -366          339          509
## 13         111        -124         -372          338          509
##    roll_dumbbell pitch_dumbbell yaw_dumbbell total_accel_dumbbell
## 1       13.05217      -70.49400    -84.87394                   37
## 3       12.85075      -70.27812    -85.14078                   37
## 7       13.12695      -70.24757    -85.09961                   37
```

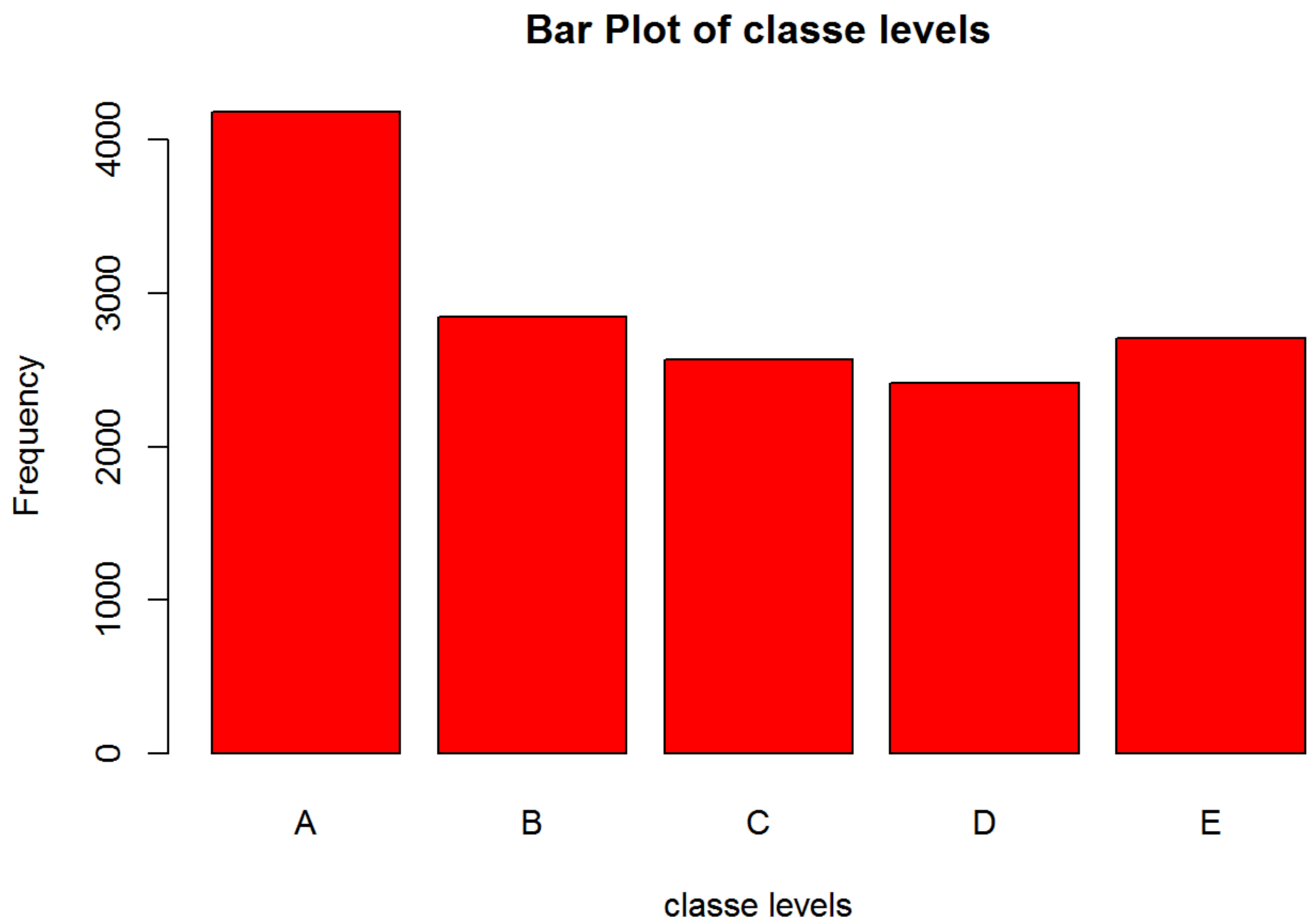
```
## 9      13.15463     -70.42520    -84.91563                  37
## 11     13.13074     -70.63751    -84.71065                  37
## 13     13.38246     -70.81759    -84.46500                  37
##    gyros_dumbbell_x gyros_dumbbell_y gyros_dumbbell_z accel_dumbbell_x
## 1                 0            -0.02             0.00             -234
## 3                 0            -0.02             0.00             -232
## 7                 0            -0.02             0.00             -232
## 9                 0            -0.02             0.00             -232
## 11                0            -0.02             0.00             -233
## 13                0            -0.02            -0.02             -234
##    accel_dumbbell_y accel_dumbbell_z magnet_dumbbell_x magnet_dumbbell_y
## 1                47             -271              -559               293
## 3                46             -270              -561               298
## 7                47             -270              -551               295
## 9                47             -269              -549               292
## 11               47             -269              -564               299
## 13               48             -269              -552               302
##    magnet_dumbbell_z roll_forearm pitch_forearm yaw_forearm
## 1                -65         28.4         -63.9        -153
## 3                -63         28.3         -63.9        -152
## 7                -70         27.9         -63.9        -152
## 9                -65         27.7         -63.8        -152
## 11               -64         27.6         -63.8        -152
## 13               -69         27.2         -63.9        -151
##    total_accel_forearm gyros_forearm_x gyros_forearm_y gyros_forearm_z
## 1                   36            0.03            0.00           -0.02
## 3                   36            0.03           -0.02            0.00
## 7                   36            0.02            0.00           -0.02
## 9                   36            0.03            0.00           -0.02
## 11                  36            0.02           -0.02           -0.02
## 13                  36            0.00            0.00           -0.03
##    accel_forearm_x accel_forearm_y accel_forearm_z magnet_forearm_x
## 1              192             203            -215              -17
## 3              196             204            -213              -18
## 7              195             205            -215              -18
## 9              193             204            -214              -16
## 11             193             205            -214              -17
## 13             193             205            -215              -15
##    magnet_forearm_y magnet_forearm_z classe
## 1               654              476      A
## 3               658              469      A
## 7               659              470      A
## 9               653              476      A
## 11              657              465      A
## 13              655              472      A
```

```
head(subTesting)
```

```
##    roll_belt pitch_belt yaw_belt total_accel_belt gyros_belt_x
## 2       1.41       8.07    -94.4                3         0.02
## 4       1.48       8.05    -94.4                3         0.02
## 5       1.48       8.07    -94.4                3         0.02
## 6       1.45       8.06    -94.4                3         0.02
## 8       1.42       8.13    -94.4                3         0.02
## 10      1.45       8.17    -94.4                3         0.03
##    gyros_belt_y gyros_belt_z accel_belt_x accel_belt_y accel_belt_z
## 2          0.00        -0.02          -22            4           22
## 4          0.00        -0.03          -22            3           21
## 5          0.02        -0.02          -21            2           24
## 6          0.00        -0.02          -21            4           21
## 8          0.00        -0.02          -22            4           21
## 10         0.00         0.00          -21            4           22
##    magnet_belt_x magnet_belt_y magnet_belt_z roll_arm pitch_arm yaw_arm
## 2             -7           608          -311     -128      22.5    -161
## 4             -6           604          -310     -128      22.1    -161
## 5             -6           600          -302     -128      22.1    -161
## 6              0           603          -312     -128      22.0    -161
## 8             -2           603          -313     -128      21.8    -161
## 10            -3           609          -308     -128      21.6    -161
##    total_accel_arm gyros_arm_x gyros_arm_y gyros_arm_z accel_arm_x
## 2               34        0.02       -0.02       -0.02        -290
## 4               34        0.02       -0.03        0.02        -289
## 5               34        0.00       -0.03        0.00        -289
## 6               34        0.02       -0.03        0.00        -289
## 8               34        0.02       -0.02        0.00        -289
## 10              34        0.02       -0.03       -0.02        -288
##    accel_arm_y accel_arm_z magnet_arm_x magnet_arm_y magnet_arm_z
## 2          110        -125         -369          337          513
## 4          111        -123         -372          344          512
## 5          111        -123         -374          337          506
## 6          111        -122         -369          342          513
## 8          111        -124         -372          338          510
## 10         110        -124         -376          334          516
##    roll_dumbbell pitch_dumbbell yaw_dumbbell total_accel_dumbbell
## 2       13.13074      -70.63751    -84.71065                   37
## 4       13.43120      -70.39379    -84.87363                   37
## 5       13.37872      -70.42856    -84.85306                   37
## 6       13.38246      -70.81759    -84.46500                   37
## 8       12.75083      -70.34768    -85.09708                   37
## 10      13.33034      -70.85059    -84.44602                   37
##    gyros_dumbbell_x gyros_dumbbell_y gyros_dumbbell_z accel_dumbbell_x
## 2                 0            -0.02             0.00             -233
```

```
## 4                0           -0.02            -0.02            -232
## 5                0           -0.02             0.00            -233
## 6                0           -0.02             0.00            -234
## 8                0           -0.02             0.00            -234
## 10               0           -0.02             0.00            -235
##    accel_dumbbell_y accel_dumbbell_z magnet_dumbbell_x magnet_dumbbell_y
## 2                47             -269              -555               296
## 4                48             -269              -552               303
## 5                48             -270              -554               292
## 6                48             -269              -558               294
## 8                46             -272              -555               300
## 10               48             -270              -558               291
##    magnet_dumbbell_z roll_forearm pitch_forearm yaw_forearm
## 2                -64         28.3         -63.9        -153
## 4                -60         28.1         -63.9        -152
## 5                -68         28.0         -63.9        -152
## 6                -66         27.9         -63.9        -152
## 8                -74         27.8         -63.8        -152
## 10               -69         27.7         -63.8        -152
##    total_accel_forearm gyros_forearm_x gyros_forearm_y gyros_forearm_z
## 2                   36            0.02            0.00           -0.02
## 4                   36            0.02           -0.02            0.00
## 5                   36            0.02            0.00           -0.02
## 6                   36            0.02           -0.02           -0.03
## 8                   36            0.02           -0.02            0.00
## 10                  36            0.02            0.00           -0.02
##    accel_forearm_x accel_forearm_y accel_forearm_z magnet_forearm_x
## 2              192             203            -216              -18
## 4              189             206            -214              -16
## 5              189             206            -214              -17
## 6              193             203            -215               -9
## 8              193             205            -213               -9
## 10             190             205            -215              -22
##    magnet_forearm_y magnet_forearm_z classe
## 2               661              473      A
## 4               658              469      A
## 5               655              473      A
## 6               660              478      A
## 8               660              474      A
## 10              656              473      A
```

```
#overview
plot(subTraining$classe, col="red", main="Bar Plot of classe levels", xlab="classe levels", ylab="Fre
quency")
```

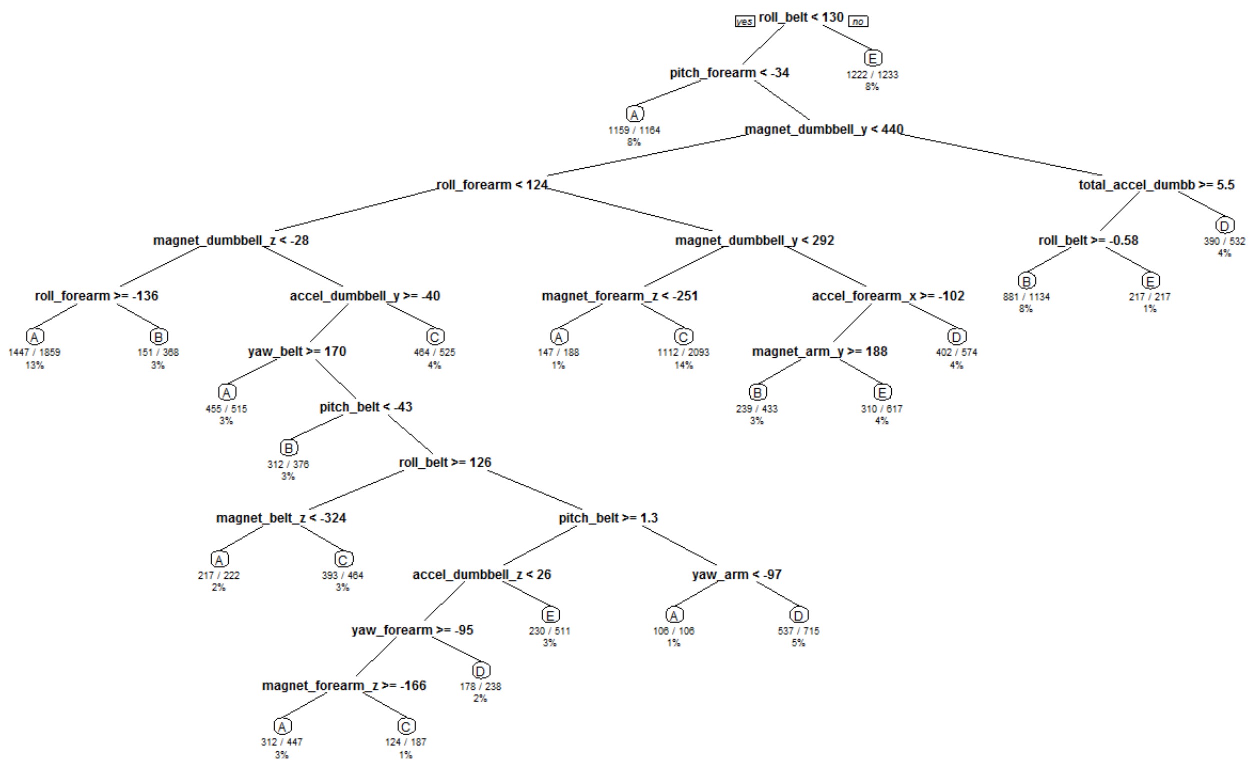


Modeling

First Model

```
mod1 <- rpart(classe ~ ., data=subTraining, method="class")
# Predicting:
pred1 <- predict(mod1, subTesting, type = "class")
# Plot of the Decision Tree
rpart.plot(mod1, main="classe Tree", extra=102, under=TRUE, faclen=0)
```

classe Tree




```
##test results on subTesting data
confusionMatrix(pred1, subTesting$classe)
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction    A    B    C    D    E
##          A 1261  130   20   47   16
##          B   47  505   77   63   62
##          C   46  148  689  106  127
##          D   15   69   47  533   51
##          E   26   97   22   55  645
##
## Overall Statistics
##
##                Accuracy : 0.7408
##                  95% CI : (0.7283, 0.753)
##     No Information Rate : 0.2845
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.6717
##  Mcnemar's Test P-Value : < 2.2e-16
##
## Statistics by Class:
##
##                      Class: A Class: B Class: C Class: D Class: E
## Sensitivity            0.9039   0.5321   0.8058   0.6629   0.7159
## Specificity            0.9393   0.9370   0.8945   0.9556   0.9500
## Pos Pred Value         0.8555   0.6698   0.6174   0.7455   0.7633
## Neg Pred Value         0.9609   0.8930   0.9562   0.9353   0.9369
## Prevalence             0.2845   0.1935   0.1743   0.1639   0.1837
## Detection Rate         0.2571   0.1030   0.1405   0.1087   0.1315
## Detection Prevalence   0.3006   0.1538   0.2276   0.1458   0.1723
## Balanced Accuracy      0.9216   0.7346   0.8502   0.8093   0.8330
```

Second Model

```
mod2 <- randomForest(classe ~. , data=subTraining, method="class")
# Predicting:
pred2 <- predict(mod2, subTesting, type = "class")
# Test results on subTesting data set:
confusionMatrix(pred2, subTesting$classe)
```

```
## Confusion Matrix and Statistics
## 
##           Reference
## Prediction    A    B    C    D    E
##          A 1395    0    0    0    0
##          B    0  948    5    0    0
##          C    0    1  850    7    1
##          D    0    0    0  795    4
##          E    0    0    0    2  896
## 
## Overall Statistics
##                                           
##                Accuracy : 0.9959          
##                  95% CI : (0.9937, 0.9975)
##     No Information Rate : 0.2845          
##     P-Value [Acc > NIR] : < 2.2e-16       
##                                           
##                   Kappa : 0.9948          
##  Mcnemar's Test P-Value : NA              
## 
## Statistics by Class:
## 
##                      Class: A Class: B Class: C Class: D Class: E
## Sensitivity            1.0000   0.9989   0.9942   0.9888   0.9945
## Specificity            1.0000   0.9987   0.9978   0.9990   0.9995
## Pos Pred Value         1.0000   0.9948   0.9895   0.9950   0.9978
## Neg Pred Value         1.0000   0.9997   0.9988   0.9978   0.9988
## Prevalence             0.2845   0.1935   0.1743   0.1639   0.1837
## Detection Rate         0.2845   0.1933   0.1733   0.1621   0.1827
## Detection Prevalence   0.2845   0.1943   0.1752   0.1629   0.1831
## Balanced Accuracy      1.0000   0.9988   0.9960   0.9939   0.9970
```

Conclusion:

Random Forest is performing better than model 1.

Submission

```
pred_subm <- predict(mod2, test_set, type="class")
pred_subm
```

```
##  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
```

```
pml_write_files = function(x){
  n = length(x)
  for(i in 1:n){
    filename = paste0("problem_id_",i,".txt")
    write.table(x[i],file=filename,quote=FALSE,row.names=FALSE,col.names=FALSE)
  }
}
pml_write_files(pred_subm)
```

References

– 1. http://groupware.les.inf.puc-rio.br/har (http://groupware.les.inf.puc-rio.br/har)
– 2.Ugulino, W.; Cardador, D.; Vega, K.; Velloso, E.; Milidiu, R.; Fuks, H. Wearable Computing: Accelerometers' Data Classification of Body Postures and Movements. Proceedings of 21st Brazilian Symposium on Artificial Intelligence. Advances in Artificial Intelligence - SBIA 2012. In: Lecture Notes in Computer Science. , pp. 52-61. Curitiba, PR: Springer Berlin / Heidelberg, 2012. ISBN 978-3-642-34458-9. DOI: 10.1007/978-3-642-34459-6_6.
– 3. 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.