Exercise Quality

Gail Muldoon

May 4, 2016

Summary

Predictions of exercise quality are determined from various sources of accelerometer data. Two prediction algorithms are compared: a random forest and gradient boosting machine. The random forest is found to be most accurate and have the correspondingly lowest out-of-sample error on a validation dataset.

Background

Using devices such as Jawbone Up, Nike FuelBand, and Fitbit it is now possible to collect a large amount of data about personal activity relatively inexpensively. These type of devices are part of the quantified self movement – a group of enthusiasts who take measurements about themselves regularly to improve their health, to find patterns in their behavior, or because they are tech geeks. 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, your goal will be 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. More information is available from the website here: http://groupware.les.inf.puc-rio.br/har (see the section on the Weight Lifting Exercise Dataset).

Data

The training data for this project are available here:

https://d396qusza40orc.cloudfront.net/predmachlearn/pml-training.csv

The test data are available here:

https://d396qusza40orc.cloudfront.net/predmachlearn/pml-testing.csv

The data for this project come from this source: http://groupware.les.inf.puc-rio.br/har. If you use the document you create for this class for any purpose please cite them as they have been very generous in allowing their data to be used for this kind of assignment.

Load and Clean Data

```
library(caret)
library(randomForest)

#Read in data, putting NAs in empty strings
training = read.csv("pml-training.csv", na.strings = c("NA", ""))
testing = read.csv("pml-testing.csv", na.strings = c("NA", ""))

#Create a validation set from 20% of the training set.
set.seed(100)
inTrain = createDataPartition(training$classe, p = 0.5, list = FALSE)
training = training[inTrain, ]
validating = training[-inTrain, ]

#Get rid of predictors that have NAs because that won't be useful
training = training[, colSums(is.na(training)) == 0]
validatiing = validating[, colSums(is.na(validating)) == 0]
testing = testing[, colSums(is.na(testing)) == 0]

#Get rid of non-acceleromator columns like name, timestamps, etc. because they're not predictor
s
trainClean =  training[, -c(1:7)]
```

```
validClean = validating[, -c(1:7)]
testClean = testing[, -c(1:7)]
```

Prediction Models

I test two prediction models, random forests and gradient boosting machine, for comparison to see which has the highest accuracy and lowest out-of-sample error rate.

Random Forest

I construct a random forest using all the non-NA accelerometer data in the training set. I then predict the classe variable, the manner in which the subect did the exercise (correctly or incorrectly) using the validation data set. I then output the accuracy and out-of-sample error rate of this method.

```
#use k=10 folds for cross-validating
control = trainControl(method = "cv", number = 5)
rfMod = train(classe ~ ., data = trainClean, method = "rf")
rfPredict = predict(rfMod,validClean)
rfConfusion = confusionMatrix(validClean$classe, rfPredict)

#Accuracy
rfAccuracy = rfConfusion$overall[1]

#Out of sample error
rfError = 1 - rfConfusion$overall[1]
```

The accuracy of the random forest method is an incredible 99.9% leaving the error rate at less than 0.1%. This is likely to most accurate method already.

Gradient Boosting Machine

As an alternative method, I use GBM to predict classe.

```
gbmMod = train(classe ~ ., data=trainClean, method = "gbm", trControl = control)
```

```
## Iter   TrainDeviance   ValidDeviance   StepSize   Improve
##      1        1.6094             nan     0.1000    0.1279
##      2        1.5230             nan     0.1000    0.0899
##      3        1.4636             nan     0.1000    0.0682
##      4        1.4188             nan     0.1000    0.0516
##      5        1.3829             nan     0.1000    0.0452
##      6        1.3526             nan     0.1000    0.0450
##      7        1.3232             nan     0.1000    0.0405
##      8        1.2974             nan     0.1000    0.0333
##      9        1.2752             nan     0.1000    0.0346
##     10        1.2527             nan     0.1000    0.0280
##     20        1.0937             nan     0.1000    0.0140
##     40        0.9181             nan     0.1000    0.0096
##     60        0.8116             nan     0.1000    0.0072
##     80        0.7324             nan     0.1000    0.0055
##    100        0.6655             nan     0.1000    0.0024
##    120        0.6133             nan     0.1000    0.0034
##    140        0.5691             nan     0.1000    0.0025
##    150        0.5496             nan     0.1000    0.0017
##
## Iter   TrainDeviance   ValidDeviance   StepSize   Improve
##      1        1.6094             nan     0.1000    0.1879
##      2        1.4884             nan     0.1000    0.1337
##      3        1.4028             nan     0.1000    0.1065
##      4        1.3337             nan     0.1000    0.0828
##      5        1.2802             nan     0.1000    0.0710
##      6        1.2336             nan     0.1000    0.0658
```

```
##      7        1.1908             nan     0.1000    0.0618
##      8        1.1516             nan     0.1000    0.0543
##      9        1.1163             nan     0.1000    0.0437
##     10        1.0873             nan     0.1000    0.0402
##     20        0.8801             nan     0.1000    0.0195
##     40        0.6633             nan     0.1000    0.0109
##     60        0.5401             nan     0.1000    0.0073
##     80        0.4560             nan     0.1000    0.0053
##    100        0.3881             nan     0.1000    0.0024
##    120        0.3370             nan     0.1000    0.0026
##    140        0.2914             nan     0.1000    0.0017
##    150        0.2732             nan     0.1000    0.0010
##
## Iter   TrainDeviance   ValidDeviance   StepSize   Improve
##      1        1.6094             nan     0.1000    0.2404
##      2        1.4575             nan     0.1000    0.1572
##      3        1.3565             nan     0.1000    0.1291
##      4        1.2737             nan     0.1000    0.1109
##      5        1.2027             nan     0.1000    0.0835
##      6        1.1489             nan     0.1000    0.0697
##      7        1.1027             nan     0.1000    0.0677
##      8        1.0577             nan     0.1000    0.0746
##      9        1.0111             nan     0.1000    0.0588
##     10        0.9738             nan     0.1000    0.0449
##     20        0.7370             nan     0.1000    0.0254
##     40        0.5189             nan     0.1000    0.0087
##     60        0.3939             nan     0.1000    0.0057
##     80        0.3143             nan     0.1000    0.0056
##    100        0.2565             nan     0.1000    0.0025
##    120        0.2150             nan     0.1000    0.0020
##    140        0.1810             nan     0.1000    0.0015
```

```
##    150        0.1657             nan     0.1000    0.0014
##
## Iter   TrainDeviance   ValidDeviance   StepSize   Improve
##      1        1.6094             nan     0.1000    0.1307
##      2        1.5225             nan     0.1000    0.0876
##      3        1.4651             nan     0.1000    0.0644
##      4        1.4204             nan     0.1000    0.0528
##      5        1.3847             nan     0.1000    0.0441
##      6        1.3558             nan     0.1000    0.0440
##      7        1.3269             nan     0.1000    0.0355
##      8        1.3024             nan     0.1000    0.0365
##      9        1.2791             nan     0.1000    0.0295
##     10        1.2589             nan     0.1000    0.0314
##     20        1.1001             nan     0.1000    0.0184
##     40        0.9270             nan     0.1000    0.0088
##     60        0.8176             nan     0.1000    0.0072
##     80        0.7382             nan     0.1000    0.0048
##    100        0.6737             nan     0.1000    0.0037
##    120        0.6226             nan     0.1000    0.0019
##    140        0.5799             nan     0.1000    0.0018
##    150        0.5617             nan     0.1000    0.0020
##
## Iter   TrainDeviance   ValidDeviance   StepSize   Improve
##      1        1.6094             nan     0.1000    0.1803
##      2        1.4896             nan     0.1000    0.1256
##      3        1.4053             nan     0.1000    0.1119
##      4        1.3357             nan     0.1000    0.0861
##      5        1.2804             nan     0.1000    0.0721
##      6        1.2340             nan     0.1000    0.0597
##      7        1.1947             nan     0.1000    0.0649
##      8        1.1550             nan     0.1000    0.0531
```

```
##      9        1.1209             nan     0.1000    0.0499
##     10        1.0895             nan     0.1000    0.0369
##     20        0.8874             nan     0.1000    0.0241
##     40        0.6762             nan     0.1000    0.0109
##     60        0.5566             nan     0.1000    0.0082
##     80        0.4658             nan     0.1000    0.0048
##    100        0.3995             nan     0.1000    0.0044
##    120        0.3480             nan     0.1000    0.0029
##    140        0.3046             nan     0.1000    0.0025
##    150        0.2873             nan     0.1000    0.0009
##
## Iter   TrainDeviance   ValidDeviance   StepSize   Improve
##      1        1.6094             nan     0.1000    0.2305
##      2        1.4599             nan     0.1000    0.1615
##      3        1.3569             nan     0.1000    0.1211
##      4        1.2801             nan     0.1000    0.1143
##      5        1.2089             nan     0.1000    0.0835
##      6        1.1544             nan     0.1000    0.0698
##      7        1.1088             nan     0.1000    0.0813
##      8        1.0587             nan     0.1000    0.0569
##      9        1.0216             nan     0.1000    0.0493
##     10        0.9893             nan     0.1000    0.0460
##     20        0.7532             nan     0.1000    0.0233
##     40        0.5248             nan     0.1000    0.0115
##     60        0.4014             nan     0.1000    0.0059
##     80        0.3194             nan     0.1000    0.0046
##    100        0.2611             nan     0.1000    0.0014
##    120        0.2171             nan     0.1000    0.0033
##    140        0.1842             nan     0.1000    0.0014
##    150        0.1717             nan     0.1000    0.0010
##
```

```
## Iter   TrainDeviance   ValidDeviance   StepSize   Improve
##      1        1.6094             nan     0.1000    0.1292
##      2        1.5229             nan     0.1000    0.0862
##      3        1.4646             nan     0.1000    0.0677
##      4        1.4216             nan     0.1000    0.0492
##      5        1.3876             nan     0.1000    0.0456
##      6        1.3587             nan     0.1000    0.0434
##      7        1.3301             nan     0.1000    0.0408
##      8        1.3044             nan     0.1000    0.0323
##      9        1.2827             nan     0.1000    0.0323
##     10        1.2602             nan     0.1000    0.0340
##     20        1.0982             nan     0.1000    0.0174
##     40        0.9252             nan     0.1000    0.0095
##     60        0.8152             nan     0.1000    0.0052
##     80        0.7373             nan     0.1000    0.0050
##    100        0.6741             nan     0.1000    0.0015
##    120        0.6209             nan     0.1000    0.0025
##    140        0.5780             nan     0.1000    0.0022
##    150        0.5588             nan     0.1000    0.0020
##
## Iter   TrainDeviance   ValidDeviance   StepSize   Improve
##      1        1.6094             nan     0.1000    0.1849
##      2        1.4888             nan     0.1000    0.1387
##      3        1.4000             nan     0.1000    0.1072
##      4        1.3315             nan     0.1000    0.0873
##      5        1.2741             nan     0.1000    0.0667
##      6        1.2303             nan     0.1000    0.0737
##      7        1.1835             nan     0.1000    0.0564
##      8        1.1459             nan     0.1000    0.0534
##      9        1.1119             nan     0.1000    0.0404
##     10        1.0843             nan     0.1000    0.0443
```

```
##     20        0.8874             nan     0.1000    0.0195
##     40        0.6759             nan     0.1000    0.0091
##     60        0.5437             nan     0.1000    0.0121
##     80        0.4564             nan     0.1000    0.0057
##    100        0.3912             nan     0.1000    0.0039
##    120        0.3402             nan     0.1000    0.0016
##    140        0.2980             nan     0.1000    0.0018
##    150        0.2825             nan     0.1000    0.0014
##
## Iter   TrainDeviance   ValidDeviance   StepSize   Improve
##      1        1.6094             nan     0.1000    0.2342
##      2        1.4595             nan     0.1000    0.1651
##      3        1.3559             nan     0.1000    0.1309
##      4        1.2741             nan     0.1000    0.1022
##      5        1.2079             nan     0.1000    0.0919
##      6        1.1494             nan     0.1000    0.0782
##      7        1.0987             nan     0.1000    0.0653
##      8        1.0564             nan     0.1000    0.0537
##      9        1.0210             nan     0.1000    0.0606
##     10        0.9809             nan     0.1000    0.0494
##     20        0.7467             nan     0.1000    0.0267
##     40        0.5243             nan     0.1000    0.0071
##     60        0.3994             nan     0.1000    0.0082
##     80        0.3172             nan     0.1000    0.0048
##    100        0.2602             nan     0.1000    0.0021
##    120        0.2160             nan     0.1000    0.0023
##    140        0.1812             nan     0.1000    0.0008
##    150        0.1689             nan     0.1000    0.0019
##
## Iter   TrainDeviance   ValidDeviance   StepSize   Improve
##      1        1.6094             nan     0.1000    0.1230
```

```
##      2        1.5247             nan     0.1000    0.0903
##      3        1.4656             nan     0.1000    0.0670
##      4        1.4215             nan     0.1000    0.0523
##      5        1.3866             nan     0.1000    0.0517
##      6        1.3530             nan     0.1000    0.0357
##      7        1.3292             nan     0.1000    0.0407
##      8        1.3037             nan     0.1000    0.0317
##      9        1.2822             nan     0.1000    0.0325
##     10        1.2606             nan     0.1000    0.0276
##     20        1.1008             nan     0.1000    0.0152
##     40        0.9276             nan     0.1000    0.0097
##     60        0.8193             nan     0.1000    0.0045
##     80        0.7382             nan     0.1000    0.0055
##    100        0.6749             nan     0.1000    0.0024
##    120        0.6231             nan     0.1000    0.0023
##    140        0.5775             nan     0.1000    0.0028
##    150        0.5566             nan     0.1000    0.0027
##
## Iter   TrainDeviance   ValidDeviance   StepSize   Improve
##      1        1.6094             nan     0.1000    0.1841
##      2        1.4891             nan     0.1000    0.1299
##      3        1.4032             nan     0.1000    0.1045
##      4        1.3371             nan     0.1000    0.0847
##      5        1.2817             nan     0.1000    0.0770
##      6        1.2326             nan     0.1000    0.0620
##      7        1.1915             nan     0.1000    0.0529
##      8        1.1563             nan     0.1000    0.0534
##      9        1.1221             nan     0.1000    0.0382
##     10        1.0961             nan     0.1000    0.0437
##     20        0.8890             nan     0.1000    0.0222
##     40        0.6800             nan     0.1000    0.0125
```

```
##     60         0.5496             nan     0.1000    0.0114
##     80         0.4584             nan     0.1000    0.0043
##    100         0.3904             nan     0.1000    0.0031
##    120         0.3395             nan     0.1000    0.0020
##    140         0.2985             nan     0.1000    0.0018
##    150         0.2812             nan     0.1000    0.0014
##
## Iter   TrainDeviance   ValidDeviance   StepSize   Improve
##      1         1.6094             nan     0.1000    0.2304
##      2         1.4604             nan     0.1000    0.1613
##      3         1.3571             nan     0.1000    0.1262
##      4         1.2765             nan     0.1000    0.0979
##      5         1.2141             nan     0.1000    0.0860
##      6         1.1579             nan     0.1000    0.0736
##      7         1.1109             nan     0.1000    0.0759
##      8         1.0620             nan     0.1000    0.0699
##      9         1.0169             nan     0.1000    0.0592
##     10         0.9787             nan     0.1000    0.0425
##     20         0.7498             nan     0.1000    0.0208
##     40         0.5232             nan     0.1000    0.0108
##     60         0.3968             nan     0.1000    0.0067
##     80         0.3160             nan     0.1000    0.0050
##    100         0.2580             nan     0.1000    0.0026
##    120         0.2122             nan     0.1000    0.0022
##    140         0.1795             nan     0.1000    0.0013
##    150         0.1658             nan     0.1000    0.0012
##
## Iter   TrainDeviance   ValidDeviance   StepSize   Improve
##      1         1.6094             nan     0.1000    0.1281
##      2         1.5246             nan     0.1000    0.0849
##      3         1.4661             nan     0.1000    0.0659
```

```
##      4        1.4230             nan     0.1000    0.0536
##      5        1.3880             nan     0.1000    0.0493
##      6        1.3558             nan     0.1000    0.0383
##      7        1.3288             nan     0.1000    0.0397
##      8        1.3032             nan     0.1000    0.0360
##      9        1.2803             nan     0.1000    0.0338
##     10        1.2571             nan     0.1000    0.0277
##     20        1.0999             nan     0.1000    0.0170
##     40        0.9242             nan     0.1000    0.0102
##     60        0.8188             nan     0.1000    0.0064
##     80        0.7398             nan     0.1000    0.0036
##    100        0.6776             nan     0.1000    0.0043
##    120        0.6244             nan     0.1000    0.0025
##    140        0.5806             nan     0.1000    0.0026
##    150        0.5613             nan     0.1000    0.0027
##
## Iter   TrainDeviance   ValidDeviance   StepSize   Improve
##      1        1.6094             nan     0.1000    0.1824
##      2        1.4891             nan     0.1000    0.1361
##      3        1.4018             nan     0.1000    0.0996
##      4        1.3372             nan     0.1000    0.0872
##      5        1.2822             nan     0.1000    0.0650
##      6        1.2378             nan     0.1000    0.0717
##      7        1.1932             nan     0.1000    0.0608
##      8        1.1542             nan     0.1000    0.0444
##      9        1.1245             nan     0.1000    0.0530
##     10        1.0907             nan     0.1000    0.0397
##     20        0.8920             nan     0.1000    0.0228
##     40        0.6783             nan     0.1000    0.0084
##     60        0.5495             nan     0.1000    0.0057
##     80        0.4593             nan     0.1000    0.0067
```

```
##    100        0.3949             nan     0.1000    0.0028
##    120        0.3405             nan     0.1000    0.0015
##    140        0.3008             nan     0.1000    0.0015
##    150        0.2836             nan     0.1000    0.0028
##
## Iter   TrainDeviance   ValidDeviance   StepSize   Improve
##      1        1.6094             nan     0.1000    0.2299
##      2        1.4601             nan     0.1000    0.1621
##      3        1.3566             nan     0.1000    0.1257
##      4        1.2767             nan     0.1000    0.1023
##      5        1.2105             nan     0.1000    0.0998
##      6        1.1493             nan     0.1000    0.0786
##      7        1.1003             nan     0.1000    0.0617
##      8        1.0607             nan     0.1000    0.0723
##      9        1.0142             nan     0.1000    0.0535
##     10        0.9801             nan     0.1000    0.0444
##     20        0.7494             nan     0.1000    0.0249
##     40        0.5260             nan     0.1000    0.0101
##     60        0.4024             nan     0.1000    0.0057
##     80        0.3162             nan     0.1000    0.0044
##    100        0.2588             nan     0.1000    0.0043
##    120        0.2166             nan     0.1000    0.0010
##    140        0.1840             nan     0.1000    0.0026
##    150        0.1690             nan     0.1000    0.0011
##
## Iter   TrainDeviance   ValidDeviance   StepSize   Improve
##      1        1.6094             nan     0.1000    0.2302
##      2        1.4640             nan     0.1000    0.1694
##      3        1.3579             nan     0.1000    0.1241
##      4        1.2799             nan     0.1000    0.1089
##      5        1.2108             nan     0.1000    0.0963
```

```
##      6         1.1514             nan     0.1000    0.0857
##      7         1.0991             nan     0.1000    0.0576
##      8         1.0612             nan     0.1000    0.0566
##      9         1.0252             nan     0.1000    0.0636
##     10         0.9840             nan     0.1000    0.0459
##     20         0.7558             nan     0.1000    0.0214
##     40         0.5309             nan     0.1000    0.0100
##     60         0.4073             nan     0.1000    0.0069
##     80         0.3235             nan     0.1000    0.0062
##    100         0.2631             nan     0.1000    0.0033
##    120         0.2188             nan     0.1000    0.0029
##    140         0.1863             nan     0.1000    0.0019
##    150         0.1712             nan     0.1000    0.0014
```

```
gbmPredict = predict(gbmMod,validClean)
gbmConfusion = confusionMatrix(validClean$classe, gbmPredict)

#Accuracy
gbmAccuracy = gbmConfusion$overall[1]

#Out of sample error
gbmError = 1 - gbmConfusion$overall[1]
```

The accuracy and corresponding expected out-of-sample error rate are 97% and 3%, respectively. While still highly accurate, the random forest is a more accurate algorithm for this case so I will use it to make predictions of the test set.

Prediction of classe in test set

The most accurate algorithm is random forest, with an accuracy of ~99%. I therefore use it to do the final prediction of classe on the test set.

```
(predict(rfMod, testClean))
```

```
##  [1] 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
```

This predictive method is 100% accurate on this test set.