Machine Learning

LK

21 June 2015

Load the caret library and read the trainign data file

```
library(caret)
```

```
## Warning: package 'caret' was built under R version 3.2.1
```

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

```
set.seed(1234)

        wd<-"C:/Users/lnkkota.Home-PC/datasciencecoursera/MachineLearning"
        if (!is.null(wd)) setwd(wd)

rawdata <- read.csv("pml-training.csv", na.strings=c("NA",""),strip.white = T)
dim(rawdata)
```

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

```
isNA <- apply(rawdata, 2, function(x) { sum(is.na(x)) })
validData <- subset(rawdata[, which(isNA == 0)],
                    select=-c(X, user_name, new_window, num_window, raw_timestamp_part_1, r
aw_timestamp_part_2, cvtd_timestamp))

dim(validData)
```

```
## [1] 19622    53
```

```
inTrain <- createDataPartition(validData$classe, p=0.7, list=F)
training <- validData[inTrain,]
testing <- validData[-inTrain,]
```

Train the model - Random Forest Model

```
## Loading required package: randomForest
```

```
## Warning: package 'randomForest' was built under R version 3.2.1
```

```
## randomForest 4.6-10
## Type rfNews() to see new features/changes/bug fixes.
```

Check the prediction

```
sum(pred == testing$classe) / length(pred)
```

```
## [1] 0.9938828
```

```
confusionMatrix(testing$classe, pred)$table
```

```
##           Reference
## Prediction    A    B    C    D    E
##          A 1674    0    0    0    0
##          B   13 1125    1    0    0
##          C    0    4 1018    4    0
##          D    0    1    6  956    1
##          E    0    0    2    4 1076
```

```
rawTestData <- read.csv("pml-testing.csv", na.strings=c("NA",""), strip.white=T)
validTestData <- subset(rawTestData[, which(isNA == 0)],
                        select=-c(X, user_name, new_window, num_window, raw_timestamp_par
t_1, raw_timestamp_part_2, cvtd_timestamp))
predict(model, newdata=validTestData)
```

```
##  [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
```