Machine Learning Project

Xu Zhang

Friday, January 23, 2015

Data Clearning

The columns with "NA" are removed, the row index in that column are also removed. For the simplicity of this report, I skip the data clearning part, and load the data for the later work.

```
setwd("C:/Xu's work/coursera/Data-Science/R-programing")
load("myTrainD.RData")
load("myTestD.RData")
```

Exploratory Data Analysis

```
library(randomForest)
```

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

```
library(lattice)
library(ggplot2)
library(caret)

set.seed(2000)
# Define the cross-validation experiment
fitControl=trainControl(method="cv",number=2)
# Apply the cross validation
Result1=train(myTrain1$X.classe~.,data=myTrain1,method="rf",trControl=fitControl)
#Result1$bestTune$mtry
```

Build random forest model with full training model

```
RandomForest=randomForest(myTrain1$X.classe~.,data=myTrain1,mtry=Result1$bestTune$mtry)

PredictForTrain=predict(RandomForest)
PredictTable=table(PredictForTrain,myTrain1$X.classe)
```

```
Predict1=predict(RandomForest,myTest1)
```

Get the Prediction for the Submission part

```r
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(Predict1)
```