PMLProject

TFH

15 4 2020

Read Data and cleaning

```
library(caret)
```

```
## Warning: package 'caret' was built under R version 3.5.2
```

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

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

```
## Warning: package 'ggplot2' was built under R version 3.5.2
```

```
train <- read.csv("https://d396qusza40orc.cloudfront.net/predmachlearn/pml-training.csv")
fintest <- read.csv("https://d396qusza40orc.cloudfront.net/predmachlearn/pml-testing.csv")
names <- names(fintest[, colSums(is.na(fintest)) == 0])
train <- train[, c(names[-length(names)], "classe")]
train <- train[, -1:-7]
```

Split the trainingset for cross validation

```
inTrain <- createDataPartition(train$classe, p = 0.7, list = F)
train_1 <- train[inTrain, ]
test <- train[-inTrain, ]
```

Know we have three dataframes, train_1, cross_vali and test.

After I testes some models, I decides, that the randomForest makes the best out of it. Use the caret library. But it was so long, that I did a classification for the final pdf.

```
set.seed(1)
model <- train(classe ~., data = train_1, method="rpart")
```

Check the Accurancy

```
table(predict(model, test), test$classe)
```

```
##
##        A   B   C   D   E
##   A 983 168  32  34  17
##   B 186 659  42 155 272
##   C 363 268 812 485 235
##   D 135  44 140 290  72
##   E   7   0   0   0 486
```

Conclusion

Train the RF model is really slow, maybe their are some faster, but else it's okey. Clasification tree is fast, but teh accurancy is not good.