Practical machine learning course project

Anton Protopopov

We will consider data about human activity recognition research. You could take a look to short description of dataset or to full article

The goal of our project is to predict the manner in which our participants did the exercise.

If we will take a look to our dataset via `str` command we will able to see that some of our variables have NA values. Let's decrease number of vartiables in our prediction model, we will take into account variables containing gyros, accel and magnet, first 7 (username, X, times), and classe and exclude variables containing var:

```
trainMod <- train[,1:7]
trainMod <- cbind(trainMod, train[, grepl("accel|gyros|magnet", names(train))])
trainMod$classe <- train$classe
trainMod <- trainMod[,!grepl("var", names(trainMod))]
```

Now we have dataset with size:

```
dim(trainMod)
```

```
## [1] 19622    48
```

We could see that dataset doesn't contain any NA values:

```
table(sapply(trainMod, is.na))
```

```
##
##  FALSE
## 941856
```

First approach, with random forest:

```
fitMod <- train(factor(classe) ~ ., data=trainMod, method="rf",
  trControl = trainControl(method="cv"), number=3)
```

To be continued...

Unfortunately my computer is slow (1 core 1.6GHz) and I didn't have enough time to do all project to the deadline..

Appendix

Codes.