PML_HAR

Ricardo Portilla

Sunday, April 26, 2015

```r
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.0.3
```
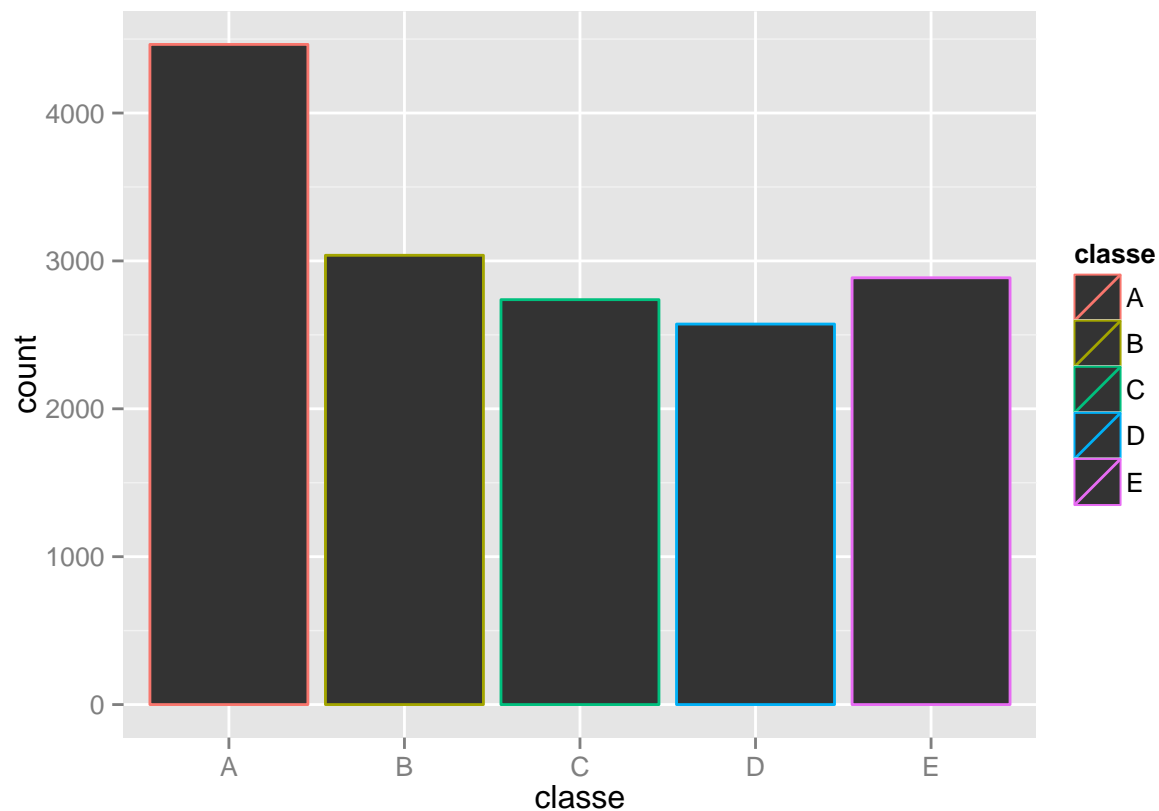
```r
library(caret)
```

```
## Warning: package 'caret' was built under R version 3.0.3
```

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

```
## Warning: package 'lattice' was built under R version 3.0.3
```

```r
setwd("C:/Users/Ricardo/Desktop")
training<-read.csv("pml-training.csv")
inTrain<-createDataPartition(training$classe, p=0.8, list=FALSE)
train<-training[inTrain,]
qplot(classe, data=train, colour=classe)
```

We used the ggplot2 library in order to get a visualization of the classes of exercise in the HAR data. It looked evenly distributed and no variables were missing. In order to get a good prediction, we decieded to build a model using a random forest and a subset of variables corresponding to the positions for each exercise for each subject in the data set.

Note that we use the caret library to split the data into a training and test subset so we can compute an out-of-sample error. We use accuracy as our metric for a good prediction. The out-of-sample on the cross validation dataset is computed below.

```
library(caret)
library(randomForest)
```

```
## Warning: package 'randomForest' was built under R version 3.0.3
```

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

```
set.seed(1000)
setwd("C:/Users/Ricardo/Desktop")
training<-read.csv("pml-training.csv")
inTrain<-createDataPartition(training$classe, p=0.8, list=FALSE)
train<-training[inTrain,]
cv<-training[-inTrain,]
train_modRf<-randomForest(classe ~ gyros_belt_x + gyros_belt_y + gyros_belt_z
                    + accel_belt_x + accel_belt_y + accel_belt_z
                    + magnet_belt_x + magnet_belt_y + magnet_belt_z
                    + gyros_arm_x + gyros_arm_y + gyros_arm_z
                    + accel_arm_x + accel_arm_y + accel_arm_z
                    + magnet_arm_x + magnet_arm_y + magnet_arm_z
                    + gyros_dumbbell_x + gyros_dumbbell_y + gyros_dumbbell_z
                    + accel_dumbbell_x + accel_dumbbell_y + accel_dumbbell_z
                    + magnet_dumbbell_x + magnet_dumbbell_y + magnet_dumbbell_z
                    + accel_forearm_x + accel_forearm_y + accel_forearm_z
                    + magnet_forearm_x + magnet_forearm_y + magnet_forearm_z
                    + gyros_forearm_x + gyros_forearm_y + gyros_forearm_z,train)
train_rf_pred<-predict(train_modRf, train)
train_cmbnd<-cbind(train, train_rf_pred)
sum(train_cmbnd$classe == train_rf_pred)/length(train_rf_pred)*100
```

```
## [1] 100
```

```
cv_rf_pred<-predict(train_modRf, cv)
cv_cmbnd<-cbind(cv, cv_rf_pred)
```

Out of sample error:

```
sum(cv_cmbnd$classe == cv_rf_pred)/length(cv_rf_pred)*100
```

```
## [1] 99.08233
```