Weight Lifting Exercise

Kalani

May 15, 2015

I built a model to classify how well a dumbell bicep curl is being done using the Weight Lifting Exercises dataset. The classifications were specified as: exactly according to the specification (Class A), throwing the elbows to the front (Class B), lifting the dumbbell only halfway (Class C), lowering the dumbbell only halfway (Class D) and throwing the hips to the front (Class E).

Preprocessing

The following r code was used to load the caret package and the datasets.

```r
# Load Packages
library(caret)    #Predictions

#Load the training and test data
train = read.csv("training.csv")
test = read.csv("testing.csv")
```

I split the 'train' dataset into 'training' and 'validation' datasets.

```r
#create training and validation datasets
set.seed(123123)
inTrain = createDataPartition(y=train$classe,p=0.1,list=FALSE)
training = train[inTrain,]
validation = train[-inTrain,]
```

I used the test dataset to determine which columns we would have access to when doing the prediction. I also wanted to filter out the first 7 columns.

```r
# For Filtering Columns
numericColumns = sapply(test,is.numeric)
numericColumns[c(1,2,3,4,5,6,7,160)] = FALSE
```

I created a preprocessing object to center and scale the data. Then I defined a function for determining the accuracy.

```r
#create a preprocessing object
proc = preProcess(training[,numericColumns],
                  method=c("center","scale"))

#create a function for determining the accuracy
getaccuracy = function(df,process,model){
  prediction = predict(model,predict(process,df[,numericColumns]))
  confusionMatrix(prediction,df$classe)$overall[1]
}
```

Model Buidling

I built a linear discrimination model.

```
ldaMod = train(training$classe~.,method="lda",
               data=predict(proc,training[,numericColumns]))
```

Estimating Error

I used the training dataset to generate the in-sample error and the validation dataset to estimate the out-of-sample error.

```
#Generate in-somple error
getaccuracy(training,proc,ldaMod)
```

```
##   Accuracy
## 0.7240326
```

```
#Estimate out-of-somple error
getaccuracy(validation,proc,ldaMod)
```

```
##   Accuracy
## 0.6938498
```

I then used the model to predict the test dataset.

```
prediction=predict(ldaMod,predict(proc,test[,numericColumns]))
```