Weight Lifting Exercises

Predicts how well the weight lifting exercise is performed using human activity recognition data. See http://groupware.les.inf.puc-rio.br/har for more information. Load Package, set the working directory and the seed

```
library(caret)
setwd("/Your Path/")
set.seed(123)
```

Load the data, exclude columns containing NA's and delete the first seven columns since they do not contain useful predictors.

```
data = read.csv("pml-training.csv", na.strings = c("NA", "","<NA>"))
data<-data[,colSums(is.na(data))==0]
data<-data[,-(1:7)]
```

Delete highly correlated Predictors (column 53 is excluded since it contains the performance classe which should be predicted)

```
correlationMatrix=cor(data[,-53])
highlyCorr<-findCorrelation(correlationMatrix,cutoff=0.9)
data<-data[,-highlyCorr]
```

Divide data into a training and a test set

```
inTrain<-createDataPartition(y=data$classe,p=0.6,list=FALSE)
training<-data[inTrain,]
testing<-data[-inTrain,]
```

Fit the model to the training set using stochastic gradient boosting ("gbm"). The data is centered and scaled taking care of zero and near zero variance predictors.

```
modelFit<-train(classe~.,data=training,preProcess=c("center","scale"),method="gbm")
```

Predict the outcome for the test data

```
predFit<-predict(modelFit,newdata=testing)
```

Compute the confusion matrix and determine the accuracy (true classification divided by all outcomes) on the test set poviding a measure for the out of sample error.

```
confusionMatrix<-table(testing$classe,predFit)
sum(diag(confusionMatrix))/sum(confusionMatrix)
```

```
## [1] 0.96
```

An out of sample accuracy of about 96% seems okay for me. All 20 test cases were predicted correctly.