**Coursera machine learning exercise**

**Title:** Can accelerometer measures predict how well individuals exercise?

The goal of this exercise was to use data from accelerometers on the belt, forearm, arm and dumbbell of 6 participants to predict their performance on barbell lifting. The participants were instructed to lift barbells correctly and incorrectly in 5 different ways. The variable “classe” is the categorical variable we want to predict and includes information on how well individuals exercised.

I downloaded the training and testing datasets as instructed and then partitioned the training dataset into two sets: a training (mytraining; 60% of the initial dataset) and a testing (mytesting) sets. I excluded variables with more than 60% of missing values, and zero variance variables, e.g. variables containing “0” observations and/or useless predictors. I removed the ID column from mytraining as it was of no use for further prediction purposes. I then made sure that both mytraining, mytesting, and testing would contain the same set of variables and adjusted the three datasets (coerce) so that they are of the same data type. I used Recursive Partitioning (rpart) for building a classification tree and generated a decision tree from the mytraining dataset and generated a forest tree. I generated a probability matrix for each of the observations of the mytesting dataset. I then used an alternative prediction model based on randomForest. This model uses multiple models for better performance and adjusts to a number of settings (e.g. large number of variables, can provide a measure of variable importance). In mytesting dataset random forests gave an accuracy of 99.87% , which is greater than that observed with recursive partitioning (86.26%). The expected out-of-sample error is 100-99.89 = 0.11%.