Predicting Curl Quality

Donald Hescht

11/28/2016

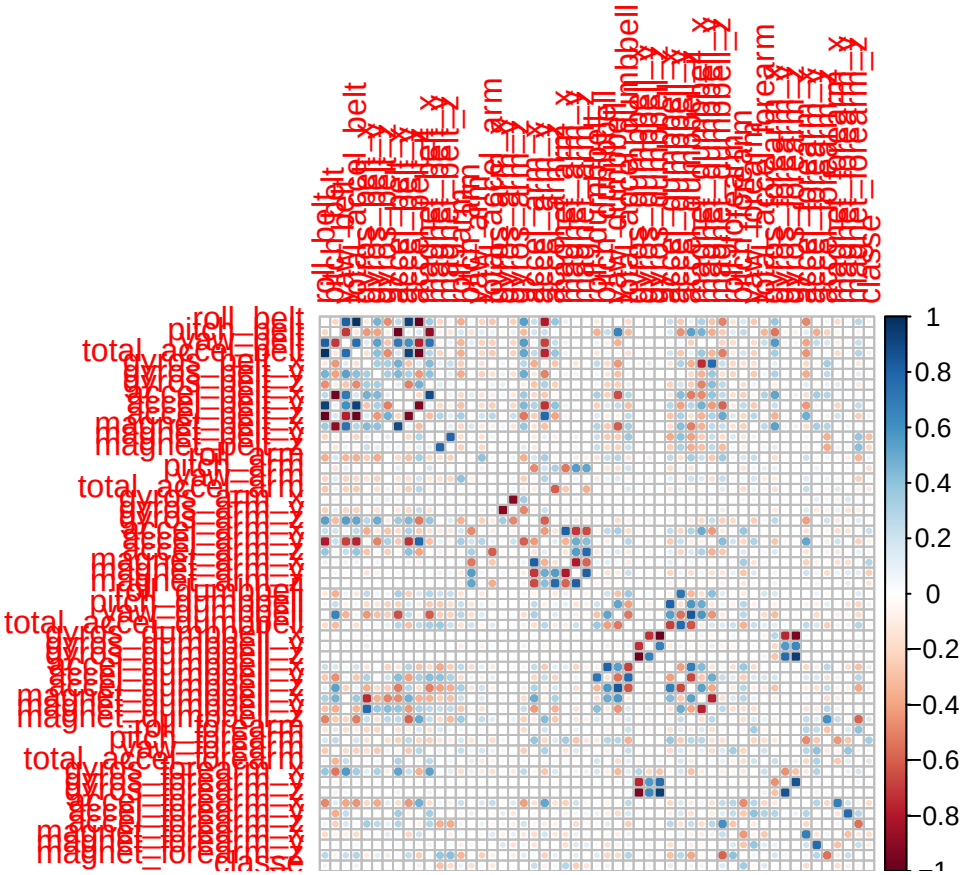
Summary

This paper decribes the author's approach to designing a 99% accurate model for predicting curl exercise type as defined in the paper "Qualitative Activity Recognition of Weight Lifting Exercises" [1]. This activity's data was captured by sampling 6 lifters with 4 attached sensors. These sensors were attached to the lifter's forearm, arm, belt and dumbell. They were them monitored while performing 5 distinct curl exercises: one that was "correct" and four others with distinct errors. The aggregate of these sensor data and post calcuations resulted in 19622 observations of 160 variables.

Cleaning and Analysis

The lift data contains N/As and divide by zero errors. These are by removing feature data that has incomplete values. These values were post calculated as summary type data (various angles, max/mins and other stats calucated in 2.5 second windows) from the orginal sensors data. Therefore, actual sensor information is preserved by the cleaning. The cleaning resulted in a clean Training (14718, 54) and Testing (4904, 54) data set.

The following diagram shows the correlation plot of the 53 features to the "classe"" output (changed to numeric to make correlation work).



Model Generation

The author chose the Random Forest alorthim for the prediction model due to its acurracy performance and reduction in bias due to the way it keeps out a porition of the training data to ensure good out of sample accuracy. The features were ordered by their absolute correlation value. The author manually picked the highest correlated features, off-line, thus reducing features from 53 to 37while keeping the model 99% accurate. The final Random Forest *features* =

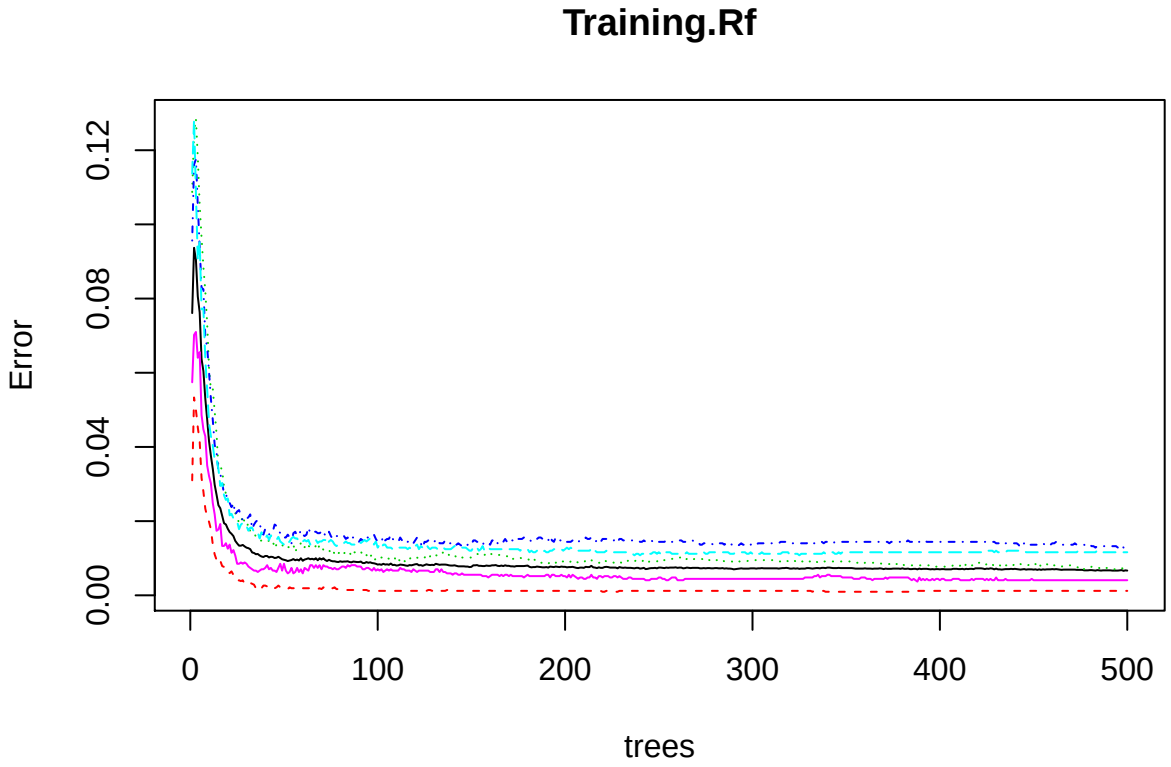
{ pitch_forearm, magnet_belt_y, magnet_arm_x, magnet_arm_y, accel_arm_x, magnet_belt_z, accel_forearm_x, magnet_forearm_x, pitch_arm, total_accel_forearm, magnet_dumbbell_z, magnet_arm_z, total_accel_arm, accel_dumbbell_x, magnet_forearm_y, roll_arm, accel_belt_z, accel_arm_y, total_accel_belt, pitch_dumbbell, accel_dumbbell_z, magnet_dumbbell_x, roll_belt, accel_arm_z, total_accel_dumbbell, yaw_forearm, yaw_arm, roll_dumbbell, magnet_forearm_z, gyros_dumbbell_y, accel_forearm_y, roll_forearm, magnet_belt_x, gyros_arm_y, gyros_belt_y, accel_dumbbell_y, yaw_belt}

```
set.seed(19652)
Training.Rf <- randomForest(Formula, data=Training, mtry=as.integer(sqrt(length(Features))), importance=
Training.Rf
```

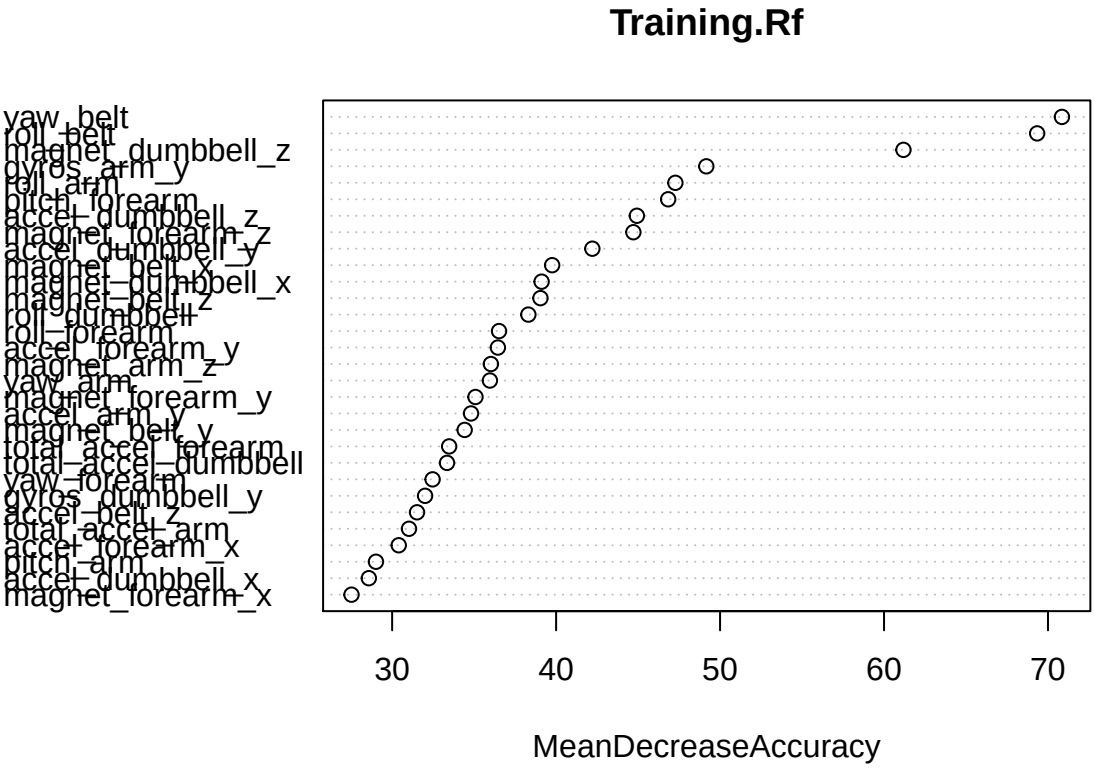
```
##
## Call:
##  randomForest(formula = Formula, data = Training, mtry = as.integer(sqrt(length(Features))),      imp
##                Type of random forest: classification
##                      Number of trees: 500
```

```
## No. of variables tried at each split: 6
##
##         OOB estimate of  error rate: 0.67%
## Confusion matrix:
##      A    B    C    D    E class.error
## A 4180    4    1    0    0 0.001194743
## B   13 2829    6    0    0 0.006671348
## C    0   24 2532   11    0 0.013634593
## D    1    0   24 2384    3 0.011608624
## E    0    0    2    9 2695 0.004065041
```

The final model error is shown in this plot indicating that all classe (A-E) features are running around an 99% accuracy.



While the importance of each Random Forest model feature is shown below. Removing any of the features reduced the accuracy below 99%. Interestingly, the calculation of the Euler angles [1] (roll/pitch/yaw) created important features for classification accuracy.



Validation

```
# Function to report accuracy of model
set.seed(19653)
Training.Test.Pred <- predict(Training.Rf, Testing)
ctable <- table(Training.Test.Pred, Testing$classe)
Testing.Acc <-  sum(diag(ctable)) / sum(ctable)
Training.OOB.Acc <- sum(diag(Training.Rf$confusion[,1:5])) / sum(Training.Rf$confusion[,1:5])
```

The Random Forest model was validated by using a 25% hold out of the "pml-training.csv" training data. Using this hold out data gives an out of sample accuracy of **0.9918434**. This closely matches the accuracy giving by the Random Forest Training Confusion Matrix **(0.9933415)** which is ~(1 - OOB).

Citations

[1] Velloso, E.; Bulling, A.; Gellersen, H.; Ugulino, W.; Fuks, H. Qualitative Activity Recognition of Weight Lifting Exercises. Proceedings of 4th International Conference in Cooperation with SIGCHI (Augmented Human '13) . Stuttgart, Germany: ACM SIGCHI, 2013. Read more: http://groupware.les.inf.puc-rio.br/har#ixzz4QYU9sHrZ

[2] https://d396qusza40orc.cloudfront.net/predmachlearn/pml-training.csv