Practical Machine Learning - Quantify Exercises Levels

Jay Yanamandala

November 03, 2021

Executive Summary

Goal of this project is to use dataset from accelerometers on the belt, forearm, arm, and dumbell of 6 participants quantify how much of a particular activity they do correctly and incorrectly, using the 'classe' variable.

Create a report describing how you built your model, and how you used cross validation.

Questions to answer:
1. what you think the expected out of sample error is?
2. Why you made the choices you did.
3. Prediction model to predict 20 different test cases - test set.

The five different 'classe' factors in this dataset are: * 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) * Throwing the hips to the front (Class E)

For more details, please read the section on Weight Lifting Exercise Dataset

Data Preprocessing

1. Load Required libraries
2. Read CSV files
3. Remove first 7 columns since they do not add value to this analysis
4. Convert columns with all NAs to 0s
5. Convert all 'integer' columns to 'numeric'
6. Split training data set (70%) and a validation data set (30%).
7. Use the validation data set to conduct cross validation in future steps.

Data Modeling

We fit a predictive model for activity recognition using **Random Forest** algorithm because it automatically selects important variables and is robust to correlated covariates & outliers in general. We will use **5-fold cross validation** when applying the algorithm.

Confusion Matrix Table:

```
##           Reference
## Prediction    A    B    C    D    E
##          A 1672    2    0    0    0
##          B   14 1121    3    1    0
##          C    0    3 1020    3    0
##          D    0    0   11  951    2
##          E    0    1    1    1 1079
```

Accuracy and Prediction

The estimated accuracy of the model is **99.29**
The estimated out-of-sample error is **0.71**

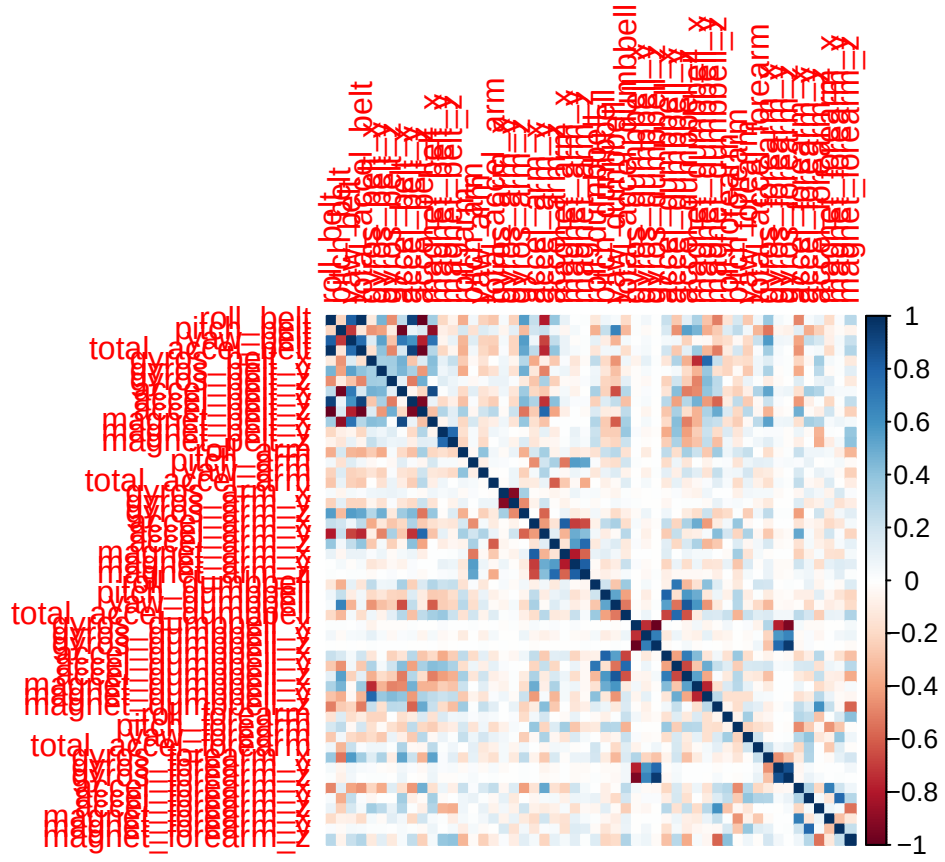
Predicted Result on Test Data:
B, A, B, A, A, E, D, B, A, A, B, C, B, A, E, E, A, B, B, B

Citation

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.

Appendix: Figures

Correlation Matrix Visualization



Decision Tree Visualization