Human Activity Recognition Based on Three Classification Methods

Introduction

Many applications and devices such as *Nike FuelBand* are possible to collect personal activity data inexpensively. Thus many enthusiasts try to improve their health by taking measurements about themselves regularly. People like to measure how long they work out, but they rarely quantify how well they do it. Data in this project comes from 6 participants, who use accelerometers on the belt, forearm, arm, and dumbell and are asked to perform barbell lifts correctly and incorrectly in 5 different ways. We will employ three classification methods to recognize how well people do in activity based on this data.

Methods

Prepossessing

The goal of this project is to predict the manner in which they did the exercise. This is the "classe" variable in the training set. “classe” variable contains five level identified by capital letter: 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).

And the other 159 variables are either user’s personal information or the activity data from devices involving features on the Euler angles (roll, pitch and yaw), as well as the raw accelerometer, gyroscope, and magnetometer readings and 96 derived feature sets based on Euler angles.

In this study, 19622 observations are collected from 6 participants in this dataset. In addition, only 406 observations in which new window are used have more available data, but for other observations, many variables are missing.

As we can see, Euler angles are the most important index in this measurement, since these indexes can represent the changes of all body activities and no missing value on these variables. Although raw readings are more comprehensive, too many variables could affect the performance of our classification methods. And the useful information in this problem could be all obtained from Euler angles

Therefore in our model, we use 12 variables involving raw Euler angles(roll, pitch and yaw) from each device as predictors.

Tree Based Prediction

Boosting

For boosting model, we use bootstrap resampling method with 25 resamples, the result are followed.

interaction.depth	n.trees	Accuracy	Kappa	AccuracySD	KappaSD
1	50	0.6232	0.5154	0.0098	0.0130
1	100	0.6667	0.5751	0.0063	0.0080
1	150	0.6915	0.6076	0.0067	0.0087
2	50	0.7252	0.6509	0.0071	0.0091
2	100	0.7894	0.7330	0.0065	0.0082
2	150	0.8226	0.7752	0.0064	0.0081
3	50	0.7908	0.7345	0.0055	0.0070
3	100	0.8506	0.8106	0.0045	0.0058
3	150	0.8766	0.8436	0.0039	0.0049

Based on the result in “Course Project: Submission”. The Accuracy using 20 test observations= 17/20=85%

Random Forest

For boosting model, we use 10 fold cross validation, the result are followed.

mtry	Accuracy	Kappa	AccuracySD	KappaSD
2	0.9551	0.9432	0.0027	0.0034
5	0.9585	0.9474	0.0027	0.0034
9	0.9502	0.9370	0.0030	0.0038
2	0.9551	0.9432	0.0027	0.0034
5	0.9585	0.9474	0.0027	0.0034

The Accuracy using 20 test observations= 20/20=100%

Model Based Prediction

Naïve Bayes

For Naïve Bayes model, we use bootstrap resampling method with 25 resamples, the result are followed.

mtry	Accuracy	Kappa	AccuracySD	KappaSD
FALSE	0.4301	0.2827	0.0102	0.0124
TRUE	0.6317	0.5358	0.0122	0.0152

The Accuracy using 20 test observations= 9/20=45%

LDA(Linear Discriminant Analysis)

For LDA model, we use 10 fold cross validation, the result are followed.

Accuracy	Kappa	AccuracySD	KappaSD
0.4347	0.2791	0.0108	0.0139

The Accuracy using 20 test observations= 9/20=45%

Conclusions

Therefore, tree based prediction performed much better than model based prediction, since there is no obvious functional relationship between “classe” and its predictors. Meanwhile, random forest model perform best