

SYNOPSIS

We are provided data for six participants who participated in a dumbbell lifting exercise five different ways. The five ways are Class A-Exactly according to the specification, Class B-Throwing the elbows to the front, Class C-Lifting the dumbbell only halfway, Class D-lowering the dumbbell only halfway and Class E-throwing the hips to the front. Class A corresponds to the specified execution of the exercise, while the other 4 classes correspond to common mistakes.

By processing data gathered from accelerometers on the belt, forearm, arm, and dumbbell of the participants in a machine learning algorithm we will try to answer the question - what was the quality of exercise which is a variable “classe” in training data set, class A to E will be predicted? We will utilize other variables in data to make the prediction.

Exploratory Data Analysis

Load required libraries.

```
#Load library
```

```
library(caret)
```

```
library(knitr)
#Set seed
set.seed(12895)
```

Load Training and Test data downloaded.

```
train <- read.csv("C:/Users/home/Documents/pml-training.csv", stringsAsFactors=FALSE)
test <- read.csv("C:/Users/home/Documents/pml-testing.csv", stringsAsFactors=FALSE)
```

To prepare data eliminated both NA columns and other extraneous columns.

```
filterData <- function(df) {
  # Since we have lots of variables, remove any with NA's
  # or have empty strings
  dx.keep <- !sapply(df, function(x) any(is.na(x)))
  df <- df[, dx.keep]
  dx.keep <- !sapply(df, function(x) any(x==""))
  df <- df[, dx.keep]

  # Remove the columns that aren't the predictor variables
  col.rm <- c("X", "user_name", "raw_timestamp_part_1", "raw_timestamp_part_2",
             "cvtd_timestamp", "new_window", "num_window")
  dx.rm <- which(colnames(df) %in% col.rm)
  df <- df[, -dx.rm]

  return(df)
}
```

Applying filter function built for preparing data.

```
train <- filterData(train)
train$classe <- factor(train$classe)

test <- filterData(test)
```

Cross Validation and Modelling

Strategy for making model is to make two models Random Forest and SVM-radial kernel and then compare the two using voting mechanism for the classe predictions

There is no variable variance so there's no need to eliminate any covariates due to lack of variability.

```
# Creating prediction models on the training data
# Utilizing cross validation with trainControl to help optimize
# the model parameters we use 5-fold cross validation
cvCtrl <- trainControl(method = "cv", number = 5, verboseIter = TRUE)
```

```
#Making the two model.
modelRF <- train(classe ~ ., data = train, method = "rf", trControl = cvCtrl)
modelSVM <- train(classe ~ ., data = train, method = "svmRadial", trControl = cvCtrl)
```

Investigating the cross-validation performance accuracy for the two models.

```
# Make a data frame with the maximum accuracy values from the models obtained
# via the cross validation on the training data
accuracy <- data.frame(Model=c("Random Forest", "SVM (radial)"),
  Accuracy=c(round(max(head(modelRF$results)$Accuracy), 3),
    round(max(head(modelSVM$results)$Accuracy), 3)))
kable(accuracy)
```

```
##
##
## |Model          | Accuracy|
## |:-----:|:-----:|
## |Random Forest |    0.994|
## |SVM (radial)  |    0.933|
```

From the table above the Random Forest model appears to have the higher cross-validation accuracy compare to SVM.

Prediction

Here we do predictions on test data using the 2 models and look for concordance in the classifications.

```
# Predictions with test data
testPredictionRF <- predict(modelRF, test)
testPredictionSVM <- predict(modelSVM, test)

# compare prediction of two models
pred <- data.frame(rfPred = testPredictionRF, svmPred = testPredictionSVM)
pred$agree <- with(pred, rfPred == svmPred)
bothAgree <- all(pred$agree)
```

Comparing results of two models.

```
colnames(pred) <- c("Random Forest", "SVM", "Agreement?")
kable(pred)
```

```
##
##
## |Random Forest |SVM |Agreement? |
## |:-----|:---|:-----|
## |B           |B   |TRUE      |
## |A           |A   |TRUE      |
## |B           |B   |TRUE      |
## |A           |A   |TRUE      |
## |A           |A   |TRUE      |
## |E           |E   |TRUE      |
## |D           |D   |TRUE      |
## |B           |B   |TRUE      |
## |A           |A   |TRUE      |
## |A           |A   |TRUE      |
## |B           |B   |TRUE      |
## |C           |C   |TRUE      |
## |B           |B   |TRUE      |
## |A           |A   |TRUE      |
## |E           |E   |TRUE      |
## |E           |E   |TRUE      |
## |A           |A   |TRUE      |
## |B           |B   |TRUE      |
## |B           |B   |TRUE      |
## |B           |B   |TRUE      |
```

Conclusion

It appears that we achieved high accuracy from the cross-validation procedure and predictions from the two models. Results from the model are submitted and are all correct. The results shows that we have good prediction results from both models.