Project #2

J.Bonacci, Stefan Jonsson, James Adler, Nikki Barden

5/15/2018

```
remove(list=ls())
```

R Markdown

Nikki's method consistently performed the best so we chose to go with hers. We tried to cut our predictors by making the non zero variance function stricter but it actually hurt our prediction out of sample. We chose not to validate the data because it did not improve the random forest model much or at all. Nikki_finalpred is the final output for the out of sample test set.

Libraries

```
library('ddalpha')
```

```
## Warning: package 'ddalpha' was built under R version 3.4.4
```

```
## Loading required package: MASS
```

```
## Loading required package: class
```

```
## Loading required package: robustbase
```

```
## Loading required package: sfsmisc
```

```
## Warning: package 'sfsmisc' was built under R version 3.4.4
```

```
## Loading required package: geometry
```

```
## Loading required package: magic
```

```
## Loading required package: abind
```

```
library('kernlab')
library('caret')
```

```
## Warning: package 'caret' was built under R version 3.4.4
```

```
## Loading required package: lattice
```

```
## Loading required package: ggplot2
```

```
##
## Attaching package: 'ggplot2'

## The following object is masked from 'package:kernlab':
##
##     alpha

library('MASS')
library('randomForest')

## Warning: package 'randomForest' was built under R version 3.4.4

## randomForest 4.6-14

## Type rfNews() to see new features/changes/bug fixes.

##
## Attaching package: 'randomForest'

## The following object is masked from 'package:ggplot2':
##
##     margin

library("dplyr")

##
## Attaching package: 'dplyr'

## The following object is masked from 'package:randomForest':
##
##     combine

## The following object is masked from 'package:sfsmisc':
##
##     last

## The following object is masked from 'package:MASS':
##
##     select

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

library("e1071")
library('rpart')
library('rpart.plot')
library('neuralnet')
```

```
##
## Attaching package: 'neuralnet'

## The following object is masked from 'package:dplyr':
##
##     compute

library('nnet')
library('ggvis')

##
## Attaching package: 'ggvis'

## The following object is masked from 'package:ggplot2':
##
##     resolution

library('class')
library('gmodels')
```

Reading in the datasets

```
train_data <-
read.csv('https://raw.githubusercontent.com/slevkoff/ECON386REPO/master/Prediction%20Project/training.csv')
test_data <-
read.csv('https://raw.githubusercontent.com/slevkoff/ECON386REPO/master/Prediction%20Project/testing.csv')
```

Count NA values

```
sapply(train_data, function(x) sum(is.na(x)))

##                        X                user_name     raw_timestamp_part_1
##                        0                        0                        0
##     raw_timestamp_part_2           cvtd_timestamp               new_window
##                        0                        0                        0
##               num_window                roll_belt               pitch_belt
##                        0                        0                        0
##                yaw_belt         total_accel_belt       kurtosis_roll_belt
##                        0                        0                        0
##      kurtosis_picth_belt        kurtosis_yaw_belt       skewness_roll_belt
##                        0                        0                        0
##     skewness_roll_belt.1        skewness_yaw_belt            max_roll_belt
##                        0                        0                    19216
##           max_picth_belt             max_yaw_belt            min_roll_belt
##                    19216                        0                    19216
##           min_pitch_belt             min_yaw_belt      amplitude_roll_belt
##                    19216                        0                    19216
##     amplitude_pitch_belt       amplitude_yaw_belt     var_total_accel_belt
##                    19216                        0                    19216
```

```
##            avg_roll_belt         stddev_roll_belt            var_roll_belt
##                    19216                    19216                    19216
##           avg_pitch_belt        stddev_pitch_belt           var_pitch_belt
##                    19216                    19216                    19216
##             avg_yaw_belt          stddev_yaw_belt             var_yaw_belt
##                    19216                    19216                    19216
##             gyros_belt_x             gyros_belt_y             gyros_belt_z
##                        0                        0                        0
##             accel_belt_x             accel_belt_y             accel_belt_z
##                        0                        0                        0
##            magnet_belt_x            magnet_belt_y            magnet_belt_z
##                        0                        0                        0
##                 roll_arm                 pitch_arm                   yaw_arm
##                        0                        0                        0
##          total_accel_arm            var_accel_arm            avg_roll_arm
##                        0                    19216                    19216
##          stddev_roll_arm             var_roll_arm            avg_pitch_arm
##                    19216                    19216                    19216
##         stddev_pitch_arm            var_pitch_arm              avg_yaw_arm
##                    19216                    19216                    19216
##           stddev_yaw_arm             var_yaw_arm              gyros_arm_x
##                    19216                    19216                        0
##              gyros_arm_y              gyros_arm_z              accel_arm_x
##                        0                        0                        0
##              accel_arm_y              accel_arm_z             magnet_arm_x
##                        0                        0                        0
##             magnet_arm_y             magnet_arm_z        kurtosis_roll_arm
##                        0                        0                        0
##        kurtosis_picth_arm         kurtosis_yaw_arm        skewness_roll_arm
##                        0                        0                        0
##        skewness_pitch_arm         skewness_yaw_arm             max_roll_arm
##                        0                        0                    19216
##             max_picth_arm              max_yaw_arm             min_roll_arm
##                    19216                    19216                    19216
##             min_pitch_arm              min_yaw_arm       amplitude_roll_arm
##                    19216                    19216                    19216
##       amplitude_pitch_arm        amplitude_yaw_arm            roll_dumbbell
##                    19216                    19216                        0
##            pitch_dumbbell             yaw_dumbbell   kurtosis_roll_dumbbell
##                        0                        0                        0
##  kurtosis_picth_dumbbell    kurtosis_yaw_dumbbell   skewness_roll_dumbbell
##                        0                        0                        0
##  skewness_pitch_dumbbell    skewness_yaw_dumbbell        max_roll_dumbbell
##                        0                        0                    19216
##        max_picth_dumbbell         max_yaw_dumbbell        min_roll_dumbbell
##                    19216                        0                    19216
##        min_pitch_dumbbell         min_yaw_dumbbell  amplitude_roll_dumbbell
##                    19216                        0                    19216
## amplitude_pitch_dumbbell   amplitude_yaw_dumbbell     total_accel_dumbbell
##                    19216                        0                        0
```

```
##        var_accel_dumbbell         avg_roll_dumbbell     stddev_roll_dumbbell
##                     19216                     19216                     19216
##         var_roll_dumbbell        avg_pitch_dumbbell    stddev_pitch_dumbbell
##                     19216                     19216                     19216
##        var_pitch_dumbbell          avg_yaw_dumbbell      stddev_yaw_dumbbell
##                     19216                     19216                     19216
##          var_yaw_dumbbell         gyros_dumbbell_x         gyros_dumbbell_y
##                     19216                         0                         0
##          gyros_dumbbell_z         accel_dumbbell_x         accel_dumbbell_y
##                         0                         0                         0
##          accel_dumbbell_z        magnet_dumbbell_x        magnet_dumbbell_y
##                         0                         0                         0
##         magnet_dumbbell_z             roll_forearm            pitch_forearm
##                         0                         0                         0
##               yaw_forearm    kurtosis_roll_forearm   kurtosis_picth_forearm
##                         0                         0                         0
##     kurtosis_yaw_forearm    skewness_roll_forearm   skewness_pitch_forearm
##                         0                         0                         0
##     skewness_yaw_forearm        max_roll_forearm        max_picth_forearm
##                         0                     19216                     19216
##          max_yaw_forearm        min_roll_forearm        min_pitch_forearm
##                         0                     19216                     19216
##          min_yaw_forearm   amplitude_roll_forearm  amplitude_pitch_forearm
##                         0                     19216                     19216
##    amplitude_yaw_forearm      total_accel_forearm        var_accel_forearm
##                         0                         0                     19216
##         avg_roll_forearm      stddev_roll_forearm         var_roll_forearm
##                     19216                     19216                     19216
##        avg_pitch_forearm     stddev_pitch_forearm        var_pitch_forearm
##                     19216                     19216                     19216
##          avg_yaw_forearm       stddev_yaw_forearm          var_yaw_forearm
##                     19216                     19216                     19216
##          gyros_forearm_x         gyros_forearm_y         gyros_forearm_z
##                         0                         0                         0
##          accel_forearm_x         accel_forearm_y         accel_forearm_z
##                         0                         0                         0
##         magnet_forearm_x        magnet_forearm_y        magnet_forearm_z
##                         0                         0                         0
##                   classe
##                         0
```

```
sapply(test_data, function(x) sum(is.na(x)))
```

```
##                        X                user_name     raw_timestamp_part_1
##                        0                        0                        0
##     raw_timestamp_part_2           cvtd_timestamp               new_window
##                        0                        0                        0
##               num_window                roll_belt               pitch_belt
##                        0                        0                        0
##                yaw_belt         total_accel_belt       kurtosis_roll_belt
```

```
##                        0                        0                       20
##      kurtosis_picth_belt        kurtosis_yaw_belt       skewness_roll_belt
##                       20                       20                       20
##     skewness_roll_belt.1        skewness_yaw_belt            max_roll_belt
##                       20                       20                       20
##           max_picth_belt             max_yaw_belt            min_roll_belt
##                       20                       20                       20
##           min_pitch_belt             min_yaw_belt      amplitude_roll_belt
##                       20                       20                       20
##     amplitude_pitch_belt       amplitude_yaw_belt     var_total_accel_belt
##                       20                       20                       20
##            avg_roll_belt         stddev_roll_belt            var_roll_belt
##                       20                       20                       20
##           avg_pitch_belt        stddev_pitch_belt           var_pitch_belt
##                       20                       20                       20
##            avg_yaw_belt          stddev_yaw_belt            var_yaw_belt
##                       20                       20                       20
##            gyros_belt_x             gyros_belt_y            gyros_belt_z
##                       0                        0                       0
##            accel_belt_x             accel_belt_y            accel_belt_z
##                       0                        0                       0
##           magnet_belt_x            magnet_belt_y           magnet_belt_z
##                       0                        0                       0
##                roll_arm                pitch_arm                 yaw_arm
##                       0                        0                       0
##          total_accel_arm            var_accel_arm            avg_roll_arm
##                       0                       20                       20
##          stddev_roll_arm             var_roll_arm            avg_pitch_arm
##                       20                       20                       20
##         stddev_pitch_arm            var_pitch_arm              avg_yaw_arm
##                       20                       20                       20
##           stddev_yaw_arm             var_yaw_arm              gyros_arm_x
##                       20                       20                       0
##              gyros_arm_y              gyros_arm_z              accel_arm_x
##                       0                        0                       0
##              accel_arm_y              accel_arm_z             magnet_arm_x
##                       0                        0                       0
##             magnet_arm_y             magnet_arm_z        kurtosis_roll_arm
##                       0                        0                       20
##        kurtosis_picth_arm         kurtosis_yaw_arm        skewness_roll_arm
##                       20                       20                       20
##        skewness_pitch_arm         skewness_yaw_arm             max_roll_arm
##                       20                       20                       20
##             max_picth_arm              max_yaw_arm             min_roll_arm
##                       20                       20                       20
##             min_pitch_arm              min_yaw_arm       amplitude_roll_arm
##                       20                       20                       20
##       amplitude_pitch_arm        amplitude_yaw_arm             roll_dumbbell
##                       20                       20                       0
##            pitch_dumbbell             yaw_dumbbell   kurtosis_roll_dumbbell
```

```
##                         0                         0                        20
##  kurtosis_picth_dumbbell    kurtosis_yaw_dumbbell   skewness_roll_dumbbell
##                        20                        20                        20
##  skewness_pitch_dumbbell    skewness_yaw_dumbbell        max_roll_dumbbell
##                        20                        20                        20
##       max_picth_dumbbell         max_yaw_dumbbell        min_roll_dumbbell
##                        20                        20                        20
##       min_pitch_dumbbell         min_yaw_dumbbell  amplitude_roll_dumbbell
##                        20                        20                        20
## amplitude_pitch_dumbbell   amplitude_yaw_dumbbell     total_accel_dumbbell
##                        20                        20                         0
##       var_accel_dumbbell        avg_roll_dumbbell     stddev_roll_dumbbell
##                        20                        20                        20
##        var_roll_dumbbell       avg_pitch_dumbbell    stddev_pitch_dumbbell
##                        20                        20                        20
##       var_pitch_dumbbell         avg_yaw_dumbbell      stddev_yaw_dumbbell
##                        20                        20                        20
##         var_yaw_dumbbell         gyros_dumbbell_x         gyros_dumbbell_y
##                        20                         0                         0
##         gyros_dumbbell_z         accel_dumbbell_x         accel_dumbbell_y
##                         0                         0                         0
##         accel_dumbbell_z        magnet_dumbbell_x        magnet_dumbbell_y
##                         0                         0                         0
##        magnet_dumbbell_z             roll_forearm            pitch_forearm
##                         0                         0                         0
##              yaw_forearm    kurtosis_roll_forearm   kurtosis_picth_forearm
##                         0                        20                        20
##     kurtosis_yaw_forearm    skewness_roll_forearm   skewness_pitch_forearm
##                        20                        20                        20
##     skewness_yaw_forearm         max_roll_forearm        max_picth_forearm
##                        20                        20                        20
##          max_yaw_forearm         min_roll_forearm        min_pitch_forearm
##                        20                        20                        20
##          min_yaw_forearm   amplitude_roll_forearm  amplitude_pitch_forearm
##                        20                        20                        20
##    amplitude_yaw_forearm      total_accel_forearm        var_accel_forearm
##                        20                         0                        20
##         avg_roll_forearm      stddev_roll_forearm         var_roll_forearm
##                        20                        20                        20
##        avg_pitch_forearm     stddev_pitch_forearm        var_pitch_forearm
##                        20                        20                        20
##          avg_yaw_forearm       stddev_yaw_forearm          var_yaw_forearm
##                        20                        20                        20
##          gyros_forearm_x         gyros_forearm_y          gyros_forearm_z
##                         0                         0                         0
##          accel_forearm_x         accel_forearm_y          accel_forearm_z
##                         0                         0                         0
##         magnet_forearm_x        magnet_forearm_y         magnet_forearm_z
##                         0                         0                         0
```

```
##               problem_id
##                        0
```

Remove all columns containing at least one NA

```
train_data2 <- train_data[ , apply(train_data, 2, function(x)
!any(is.na(x)))]
test_data2 <- test_data[ , apply(test_data, 2, function(x) !any(is.na(x)))]
```

input NAs into all blank observations

```
train_data2[train_data2==""] <- NA
test_data2[test_data2==""] <- NA
```

Count NA values again to check

```
sapply(train_data2, function(x) sum(is.na(x)))
```

```
##                        X                user_name     raw_timestamp_part_1
##                        0                        0                        0
##     raw_timestamp_part_2           cvtd_timestamp               new_window
##                        0                        0                        0
##               num_window                roll_belt               pitch_belt
##                        0                        0                        0
##                yaw_belt         total_accel_belt       kurtosis_roll_belt
##                        0                        0                    19216
##      kurtosis_picth_belt        kurtosis_yaw_belt       skewness_roll_belt
##                    19216                    19216                    19216
##     skewness_roll_belt.1        skewness_yaw_belt            max_yaw_belt
##                    19216                    19216                    19216
##            min_yaw_belt       amplitude_yaw_belt            gyros_belt_x
##                    19216                    19216                        0
##            gyros_belt_y            gyros_belt_z            accel_belt_x
##                        0                        0                        0
##            accel_belt_y            accel_belt_z           magnet_belt_x
##                        0                        0                        0
##           magnet_belt_y           magnet_belt_z                roll_arm
##                        0                        0                        0
##               pitch_arm                 yaw_arm         total_accel_arm
##                        0                        0                        0
##             gyros_arm_x              gyros_arm_y              gyros_arm_z
##                        0                        0                        0
##             accel_arm_x              accel_arm_y              accel_arm_z
##                        0                        0                        0
##            magnet_arm_x             magnet_arm_y             magnet_arm_z
##                        0                        0                        0
##        kurtosis_roll_arm       kurtosis_picth_arm         kurtosis_yaw_arm
##                    19216                    19216                    19216
##        skewness_roll_arm       skewness_pitch_arm         skewness_yaw_arm
```

```
##                   19216                   19216                   19216
##            roll_dumbbell          pitch_dumbbell             yaw_dumbbell
##                       0                       0                       0
##  kurtosis_roll_dumbbell kurtosis_picth_dumbbell   kurtosis_yaw_dumbbell
##                   19216                   19216                   19216
##  skewness_roll_dumbbell skewness_pitch_dumbbell   skewness_yaw_dumbbell
##                   19216                   19216                   19216
##        max_yaw_dumbbell        min_yaw_dumbbell  amplitude_yaw_dumbbell
##                   19216                   19216                   19216
##    total_accel_dumbbell        gyros_dumbbell_x        gyros_dumbbell_y
##                       0                       0                       0
##        gyros_dumbbell_z        accel_dumbbell_x        accel_dumbbell_y
##                       0                       0                       0
##        accel_dumbbell_z       magnet_dumbbell_x       magnet_dumbbell_y
##                       0                       0                       0
##       magnet_dumbbell_z            roll_forearm           pitch_forearm
##                       0                       0                       0
##             yaw_forearm   kurtosis_roll_forearm  kurtosis_picth_forearm
##                       0                   19216                   19216
##    kurtosis_yaw_forearm   skewness_roll_forearm  skewness_pitch_forearm
##                   19216                   19216                   19216
##    skewness_yaw_forearm         max_yaw_forearm         min_yaw_forearm
##                   19216                   19216                   19216
##   amplitude_yaw_forearm     total_accel_forearm         gyros_forearm_x
##                   19216                       0                       0
##         gyros_forearm_y         gyros_forearm_z         accel_forearm_x
##                       0                       0                       0
##         accel_forearm_y         accel_forearm_z        magnet_forearm_x
##                       0                       0                       0
##        magnet_forearm_y        magnet_forearm_z                  classe
##                       0                       0                       0
```

```
sapply(test_data2, function(x) sum(is.na(x)))
```

```
##                    X            user_name raw_timestamp_part_1
##                    0                    0                    0
## raw_timestamp_part_2       cvtd_timestamp           new_window
##                    0                    0                    0
##           num_window            roll_belt           pitch_belt
##                    0                    0                    0
##             yaw_belt     total_accel_belt         gyros_belt_x
##                    0                    0                    0
##         gyros_belt_y         gyros_belt_z         accel_belt_x
##                    0                    0                    0
##         accel_belt_y         accel_belt_z        magnet_belt_x
##                    0                    0                    0
##        magnet_belt_y        magnet_belt_z             roll_arm
##                    0                    0                    0
##            pitch_arm              yaw_arm      total_accel_arm
##                    0                    0                    0
```

```
##          gyros_arm_x          gyros_arm_y          gyros_arm_z
##                    0                    0                    0
##          accel_arm_x          accel_arm_y          accel_arm_z
##                    0                    0                    0
##         magnet_arm_x         magnet_arm_y         magnet_arm_z
##                    0                    0                    0
##        roll_dumbbell       pitch_dumbbell         yaw_dumbbell
##                    0                    0                    0
## total_accel_dumbbell     gyros_dumbbell_x     gyros_dumbbell_y
##                    0                    0                    0
##     gyros_dumbbell_z     accel_dumbbell_x     accel_dumbbell_y
##                    0                    0                    0
##     accel_dumbbell_z    magnet_dumbbell_x    magnet_dumbbell_y
##                    0                    0                    0
##    magnet_dumbbell_z         roll_forearm        pitch_forearm
##                    0                    0                    0
##          yaw_forearm  total_accel_forearm      gyros_forearm_x
##                    0                    0                    0
##      gyros_forearm_y      gyros_forearm_z      accel_forearm_x
##                    0                    0                    0
##      accel_forearm_y      accel_forearm_z     magnet_forearm_x
##                    0                    0                    0
##     magnet_forearm_y     magnet_forearm_z           problem_id
##                    0                    0                    0
```

input NAs into all blank observations

```
train_data3<- train_data2[ , apply(train_data2, 2, function(x)
!any(is.na(x)))]
test_data3<- test_data2[ , apply(test_data2, 2, function(x) !any(is.na(x)))]
```

omits zero variance predictors

```
##freq cut and unique cut arguments can be ommitted if it fits better with
out them (leaving arguments in cuts more predictors)

remove_cols <- nearZeroVar(train_data3,names=TRUE)
all_cols<-names(train_data3)
train_data4<-train_data3[ , setdiff(all_cols,remove_cols)]



remove_cols2<-nearZeroVar(test_data3,names=TRUE)
all_cols2<-names(test_data3)
test_data4<-test_data3[ , setdiff(all_cols2,remove_cols2)]
```

rename datasets

```
train <- train_data4
test <- test_data4
```

removing timestamps and factor variables

```
train_final<- train[c(7:59)]
test_final<- test[c(7:59)]
```

partitioning data

```
#partitions 70% of data into training set
trainingRowIndex<-sample(1:nrow(train_final), size = .7*nrow(train_final))
part_training<-train_final[trainingRowIndex, ]

#leaves 30% for testing and validating
part_test <-train_final[-trainingRowIndex, ]
```

removing old datasets

```
remove(train_data, train_data2, train_data3,train_data4, train, test_data,
test_data2, test_data3, test,test_data4,train_final)
```

random forest with default number of variables at each node (Jack Bonacci)

```
set.seed(1234)
#creates the random forest
part_training$classe<- as.factor(part_training$classe)
jack_rf<- randomForest(classe~., part_training)
jack_pred<- predict(jack_rf, part_test)
confusionMatrix(jack_pred, part_test$classe,  dnn = c("Prediction",
"Reference"))

## Confusion Matrix and Statistics
##
##           Reference
## Prediction    A    B    C    D    E
##          A 1687    4    0    0    0
##          B    0 1115    9    0    0
##          C    0    1 1009   13    2
##          D    0    0    2  952    0
##          E    0    0    0    2 1091
##
## Overall Statistics
##
##                Accuracy : 0.9944
```

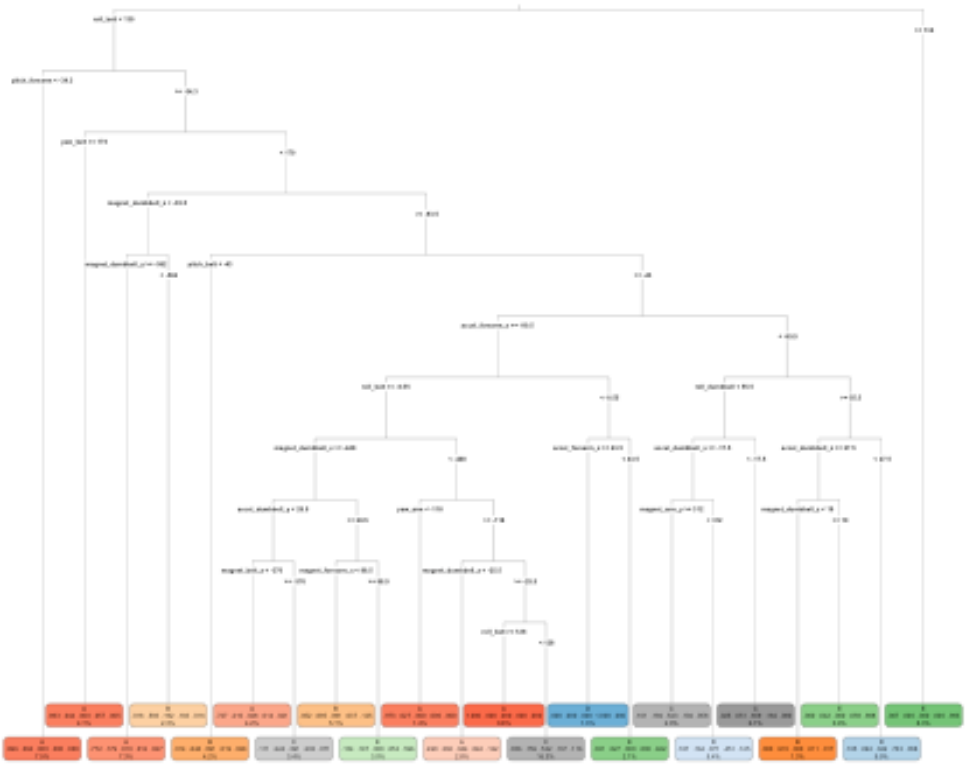
```
##                  95% CI : (0.9921, 0.9961)
##     No Information Rate : 0.2866
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.9929
##  Mcnemar's Test P-Value : NA
##
## Statistics by Class:
##
##                      Class: A Class: B Class: C Class: D Class: E
## Sensitivity            1.0000   0.9955   0.9892   0.9845   0.9982
## Specificity            0.9990   0.9981   0.9967   0.9996   0.9996
## Pos Pred Value         0.9976   0.9920   0.9844   0.9979   0.9982
## Neg Pred Value         1.0000   0.9990   0.9977   0.9970   0.9996
## Prevalence             0.2866   0.1902   0.1733   0.1643   0.1857
## Detection Rate         0.2866   0.1894   0.1714   0.1617   0.1853
## Detection Prevalence   0.2872   0.1909   0.1741   0.1621   0.1857
## Balanced Accuracy      0.9995   0.9968   0.9930   0.9920   0.9989
```

```
#Running algorithm on test data
jack_final_prediction<- predict(jack_rf, test_final)
jack_final_prediction
```

```
##  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20
##  B  A  B  A  A  E  D  B  A  A  B  C  B  A  E  E  A  B  B  B
## Levels: A B C D E
```

cart (James Adler)

```
#creating cart model
set.seed(117)
james_tree <- rpart(classe ~. ,data = part_training, method = "class")
#gets the optimal cp (cp with minimum erro)
optimal_cp<-james_tree$cptable[which.min(james_tree$cptable[,"xerror"]),"CP"]
#uses optimal cp to create new tree
james_ptree<-prune(james_tree,optimal_cp)
#viewing new tree
rpart.plot(james_ptree,type = 3,digits = 3, fallen.leaves = TRUE)
```



```r
#runs new pruned tree model with partitioned test data
james_pred<-predict(james_ptree, part_test, type="class")
confusionMatrix(james_pred,part_test$classe)
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction    A    B    C    D    E
##          A 1418  201   24   15   32
##          B   20  554   60   66   62
##          C  104  222  877  241  160
##          D  111  108   59  636   64
##          E   34   35    0    9  775
##
## Overall Statistics
##
##                Accuracy : 0.7236
##                  95% CI : (0.712, 0.735)
##     No Information Rate : 0.2866
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.6509
##  Mcnemar's Test P-Value : < 2.2e-16
```

```
##
## Statistics by Class:
##
##                      Class: A Class: B Class: C Class: D Class: E
## Sensitivity            0.8405  0.49464   0.8598   0.6577   0.7091
## Specificity            0.9352  0.95637   0.8506   0.9305   0.9837
## Pos Pred Value         0.8391  0.72703   0.5468   0.6503   0.9086
## Neg Pred Value         0.9359  0.88956   0.9666   0.9326   0.9368
## Prevalence             0.2866  0.19025   0.1733   0.1643   0.1857
## Detection Rate         0.2409  0.09411   0.1490   0.1080   0.1316
## Detection Prevalence   0.2871  0.12944   0.2725   0.1661   0.1449
## Balanced Accuracy      0.8879  0.72550   0.8552   0.7941   0.8464
```

```
#tests our algorithm on the 20 outputless observations
james_final_prediction<-predict(james_ptree,test_final)
james_final_prediction
```

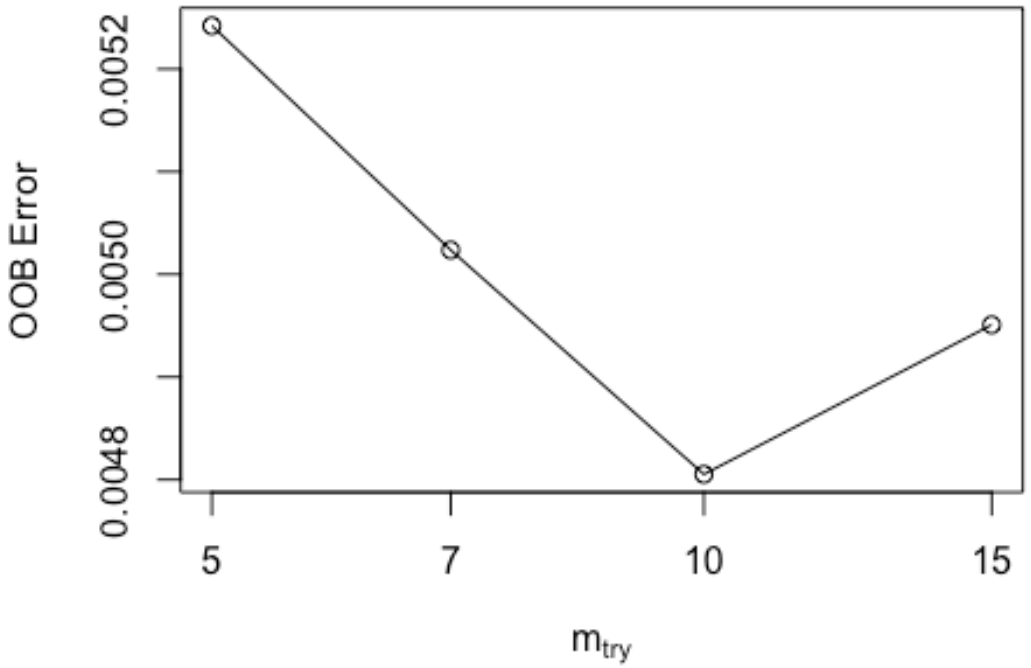
```
##             A           B          C          D          E
## 1  0.10071942 0.194244604 0.54316547 0.10431655 0.05755396
## 2  0.48958333 0.302083333 0.04427083 0.04166667 0.12239583
## 3  0.05553068 0.158978952 0.53246753 0.13703538 0.11598746
## 4  0.97311828 0.026881720 0.00000000 0.00000000 0.00000000
## 5  0.72743056 0.210069444 0.02777778 0.01388889 0.02083333
## 6  0.03050847 0.027118644 0.00000000 0.00000000 0.94237288
## 7  0.10530191 0.092047128 0.04197349 0.70250368 0.05817378
## 8  0.10530191 0.092047128 0.04197349 0.70250368 0.05817378
## 9  0.99631336 0.003686636 0.00000000 0.00000000 0.00000000
## 10 0.77222777 0.178821179 0.00999001 0.01198801 0.02697303
## 11 0.05553068 0.158978952 0.53246753 0.13703538 0.11598746
## 12 0.05553068 0.158978952 0.53246753 0.13703538 0.11598746
## 13 0.72743056 0.210069444 0.02777778 0.01388889 0.02083333
## 14 0.99631336 0.003686636 0.00000000 0.00000000 0.00000000
## 15 0.05553068 0.158978952 0.53246753 0.13703538 0.11598746
## 16 0.15384615 0.197115385 0.00000000 0.05288462 0.59615385
## 17 0.72743056 0.210069444 0.02777778 0.01388889 0.02083333
## 18 0.10530191 0.092047128 0.04197349 0.70250368 0.05817378
## 19 0.10071942 0.194244604 0.54316547 0.10431655 0.05755396
## 20 0.05187320 0.694524496 0.09077810 0.03746398 0.12536023
```

Random forest variation with regularization of data (Nikki)

```
set.seed(1234)
x <- subset(part_training, select=-classe)
y <- part_training$classe
bestmtry <- tuneRF(x, y, stepFactor = 1.5, improve=1e-5, ntree=500)
```

```
## mtry = 7  OOB error = 0.5%
## Searching left ...
## mtry = 5     OOB error = 0.52%
## -0.04347826 1e-05
```

```
## Searching right ...
## mtry = 10    OOB error = 0.48%
## 0.04347826 1e-05
## mtry = 15    OOB error = 0.5%
## -0.03030303 1e-05
```




```
print(bestmtry)

##        mtry    OOBError
## 5.OOB     5 0.005242082
## 7.OOB     7 0.005023662
## 10.OOB   10 0.004805242
## 15.OOB   15 0.004950855

nikki_rf <- randomForest(classe~., data=part_training, ntree=300, mtry=10,
importance=TRUE)
nikki_pred <- predict(nikki_rf, part_test)
confusionMatrix(nikki_pred, part_test$classe,
dnn=c("prediction","reference"))

## Confusion Matrix and Statistics
##
##           reference
## prediction    A    B    C    D    E
```

```
##          A 1685    3    0    0    0
##          B    2 1116    7    0    0
##          C    0    1 1011   12    2
##          D    0    0    2  954    0
##          E    0    0    0    1 1091
##
## Overall Statistics
##
##                Accuracy : 0.9949
##                  95% CI : (0.9927, 0.9966)
##     No Information Rate : 0.2866
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.9936
##  Mcnemar's Test P-Value : NA
##
## Statistics by Class:
##
##                      Class: A Class: B Class: C Class: D Class: E
## Sensitivity            0.9988   0.9964   0.9912   0.9866   0.9982
## Specificity            0.9993   0.9981   0.9969   0.9996   0.9998
## Pos Pred Value         0.9982   0.9920   0.9854   0.9979   0.9991
## Neg Pred Value         0.9995   0.9992   0.9981   0.9974   0.9996
## Prevalence             0.2866   0.1902   0.1733   0.1643   0.1857
## Detection Rate         0.2862   0.1896   0.1717   0.1621   0.1853
## Detection Prevalence   0.2867   0.1911   0.1743   0.1624   0.1855
## Balanced Accuracy      0.9991   0.9973   0.9940   0.9931   0.9990
```

```r
# Run final prediction
nikki_finalpred <- predict(nikki_rf, test_final)
nikki_finalpred
```

```
##  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20
##  B  A  B  A  A  E  D  B  A  A  B  C  B  A  E  E  A  B  B  B
## Levels: A B C D E
```

Support Vector Machine by Stefan (we attempted to tune it but the accuracy only minorly improved and it was quite time consuming)

```r
stefan_svm <- svm(classe ~. ,data = part_training)
stefan_pred<- predict(stefan_svm, part_test)
confusionMatrix(stefan_pred,part_test$classe,
dnn=c("prediction","reference"))
```

```
## Confusion Matrix and Statistics
##
##           reference
## prediction    A    B    C    D    E
##          A 1678   65    0    3    0
##          B    0 1033   23    0    4
```

```
##          C    7   22  986   82   30
##          D    0    0    8  878   17
##          E    2    0    3    4 1042
##
## Overall Statistics
##
##                Accuracy : 0.9541
##                  95% CI : (0.9485, 0.9593)
##     No Information Rate : 0.2866
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.9419
##  Mcnemar's Test P-Value : NA
##
## Statistics by Class:
##
##                      Class: A Class: B Class: C Class: D Class: E
## Sensitivity            0.9947   0.9223   0.9667   0.9080   0.9533
## Specificity            0.9838   0.9943   0.9710   0.9949   0.9981
## Pos Pred Value         0.9611   0.9745   0.8749   0.9723   0.9914
## Neg Pred Value         0.9978   0.9820   0.9929   0.9821   0.9895
## Prevalence             0.2866   0.1902   0.1733   0.1643   0.1857
## Detection Rate         0.2850   0.1755   0.1675   0.1491   0.1770
## Detection Prevalence   0.2966   0.1801   0.1914   0.1534   0.1785
## Balanced Accuracy      0.9892   0.9583   0.9688   0.9514   0.9757
```

This is Stefan's neural network, it wouldn't run so we left it commented out

scaledata<-scale(train_final[,1:52]) #scale normalization

normalize <- function(x) { return ((x - min(x)) / (max(x) - min(x))) }#max-min normalization

maxmindf<-as.data.frame(lapply(scaledata, normalize))#normilization takes too much computing power

test_final_2<-cbind(maxmindf,class.ind(train_final$classe))

partitions 70% of data into training set

trainingRowIndex<-sample(1:nrow(train_final2), size = .7*nrow(train_final2)) trainset<-train_final2[trainingRowIndex,]

leaves 30% for testing and validating

testset<-train_final2[-trainingRowIndex,]

library(neuralnet) #Neural Network

nn <- neuralnet(A + B + C + D + E ~
roll_belt+pitch_belt+yaw_belt+total_accel_belt+gyros_belt_x+gyros_belt_y
+gyros_belt_z+accel_belt_x+accel_belt_y+accel_belt_z+magnet_belt_x+magnet_belt_y
+magnet_belt_z+roll_arm+pitch_arm+yaw_arm+total_accel_arm+gyros_arm_x+gyros_arm_y
+gyros_arm_z+accel_arm_x+accel_arm_y+accel_arm_z+magnet_arm_x+magnet_arm_y+mag
net_arm_z
+roll_dumbbell+pitch_dumbbell+yaw_dumbbell+total_accel_dumbbell+gyros_dumbbell_x
+gyros_dumbbell_y+gyros_dumbbell_z+accel_dumbbell_x+accel_dumbbell_y+accel_dumbbe
ll_z
+magnet_dumbbell_x+magnet_dumbbell_y+magnet_dumbbell_z+roll_forearm+pitch_forear
m +yaw_forearm+total_accel_forearm+gyros_forearm_x+gyros_forearm_y+gyros_forearm_z
+accel_forearm_x+accel_forearm_y+accel_forearm_z+magnet_forearm_x+magnet_forearm_
y +magnet_forearm_z, data=test_final_2, hidden=c(10,5),linear.output=FALSE,
threshold=0.01) nn$result.matrix plot(nn)

Test the resulting output

temp_test <- subset(testset, select = c("roll_belt","pitch_belt",
"yaw_belt","total_accel_belt","gyros_belt_x","gyros_belt_y","gyros_belt_z","accel_belt_x","acce
l_belt_y","accel_belt_z","magnet_belt_x","magnet_belt_y","magnet_belt_z","roll_arm","pitch_a
rm","yaw_arm","total_accel_arm","gyros_arm_x","gyros_arm_y","gyros_arm_z","accel_arm_x"
,"accel_arm_y","accel_arm_z","magnet_arm_x","magnet_arm_y","magnet_arm_z","roll_dumbb
ell","pitch_dumbbell","yaw_dumbbell","total_accel_dumbbell","gyros_dumbbell_x","gyros_d
umbbell_y","gyros_dumbbell_z","accel_dumbbell_x","accel_dumbbell_y","accel_dumbbell_z",
"magnet_dumbbell_x","magnet_dumbbell_y","magnet_dumbbell_z","roll_forearm","pitch_for
earm","yaw_forearm","total_accel_forearm","gyros_forearm_x","gyros_forearm_y","gyros_fo
rearm_z","accel_forearm_x","accel_forearm_y","accel_forearm_z","magnet_forearm_x","magn
et_forearm_y","magnet_forearm_z"))

head(temp_test)

nn.results <- compute(nn, temp_test)

Accuracy

results <- data.frame(actual = testset("A","B","C","D","E"), prediction =
nn.results$net.result)

```
results roundedresults<-sapply(results,round,digits=0)
roundedresultsdf=data.frame(roundedresults) attach(roundedresultsdf)
table(actual,prediction)
```