Submission 13

Shannon Rush July 23, 2014

ADA

Load and Source

```
source("../helpers/predictions.R")

## Loading required package: lattice
## Loading required package: ggplot2

library(ada)

## Loading required package: rpart

train <- read.csv("../../data/original/training.csv")
test <- read.csv("../../data/original/test.csv")</pre>
```

Data Processing

```
set.seed(1111)
training.indices <- createDataPartition(train$Label, p=0.6, list=F)
training <- train[training.indices,]
validation <- train[-training.indices,]</pre>
```

Models

```
set.seed(646)
predictors <- training[,setdiff(names(training),c("EventId","Label","Weight"))]
ada.fit <- ada(x=predictors, y=training$Label, verbose=T)

load("RData/adafit.RData")

pred.val <- predict(ada.fit, validation[,setdiff(names(training),c("EventId","Weight"))])
confusionMatrix(pred.val, validation$Label)</pre>
```

Warning: NAs produced by integer overflow

```
## Confusion Matrix and Statistics
##
##
             Reference
## Prediction
                h
            b 58819 10573
##
            s 6914 23693
##
                  Accuracy: 0.825
##
##
                    95% CI: (0.823, 0.827)
##
       No Information Rate: 0.657
##
       P-Value [Acc > NIR] : <2e-16
##
##
                     Kappa: NA
##
    Mcnemar's Test P-Value : <2e-16
##
##
               Sensitivity: 0.895
##
               Specificity: 0.691
##
            Pos Pred Value: 0.848
##
            Neg Pred Value: 0.774
##
                Prevalence: 0.657
##
            Detection Rate: 0.588
##
      Detection Prevalence: 0.694
##
         Balanced Accuracy: 0.793
##
##
          'Positive' Class : b
##
    set.seed(464)
    predictors <- train[,setdiff(names(train),c("EventId","Label","Weight"))]</pre>
    adafull.fit <- ada(x=predictors, y=train$Label, verbose=T)</pre>
## Warning: NAs produced by integer overflow
```

```
## Warning: NAs produced by integer overflow
```

adafull.fit

```
## Call:
## ada(predictors, y = train$Label, verbose = T)
## Loss: exponential Method: discrete
                                        Iteration: 50
## Final Confusion Matrix for Data:
            Final Prediction
## True value
                   b
           b 147393 16940
            s 26445 59222
##
## Train Error: 0.174
## Out-Of-Bag Error: 0.177 iteration= 50
## Additional Estimates of number of iterations:
## $train.err1
## [1] 50
##
## $train.kap1
## integer(0)
```

```
sub.test <- test[,setdiff(names(test), c("EventId", "Weight"))]</pre>
    pred <- predict(adafull.fit, sub.test, type=c("probs"))</pre>
    head(pred)
##
           [,1]
                   [,2]
## [1,] 0.9242 0.07581
## [2,] 0.8430 0.15705
## [3,] 0.4574 0.54263
## [4,] 0.1393 0.86067
## [5,] 0.9094 0.09058
## [6,] 0.8792 0.12076
    pred <- as.data.frame(pred)</pre>
    names(pred) <- c("b","s")</pre>
    pred.final <- PrepPrediction(pred, test)</pre>
    WriteSubmission(pred.final, 13)
```

Results

2.66179