random_forest_gbm_tuning.R

vikaskamath

2021-03-22

```r
#####################################################################
###    Compare random forest, gradient boosting & linear models    #
#####################################################################

# Using the default data
original_train    = read.csv("train.csv", header = T)
original_train$Id = NULL

head(original_train)
```

```
##   y        x1 x2 x3         x4    x5 x6 x7 x8 x9 x10
## 1 y 0.7661266 45  2 0.80298213  9120 13  0  6  0   2
## 2 n 0.9571510 40  0 0.12187620  2600  4  0  0  0   1
## 3 n 0.6581801 38  1 0.08511338  3042  2  1  0  0   0
## 4 n 0.2338098 30  0 0.03604968  3300  5  0  0  0   0
## 5 n 0.9072394 49  1 0.02492570 63588  7  0  1  0   0
## 6 n 0.2131787 74  0 0.37560697  3500  3  0  1  0   1
```

```r
set.seed(321)
ind = sample(1:3, size = nrow(original_train), replace = TRUE)

test_index = which(ind == 1)
train_data = original_train[ -test_index, ]
test_data  = original_train[  test_index, ]

library(randomForest)
```

```
## randomForest 4.6-14

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

```r
library(gbm)
```

```
## Loaded gbm 2.1.8
```

```r
library(pROC)
```

```
## Type 'citation("pROC")' for a citation.
```

```
##
## Attaching package: 'pROC'

## The following objects are masked from 'package:stats':
##
##     cov, smooth, var
```

```
set.seed(770077)

system.time({
  rf_model  = randomForest(as.factor(y) ~ . ,  nodesize = 250, data = train_data)
})
```

```
##    user  system elapsed
##  22.011   1.173  23.437
```

```
y_for_gbm = ifelse(train_data$y == "y", 1, 0)
system.time({
  gb_model  = gbm(y_for_gbm ~ ., data = train_data[,-1], distribution = "bernoulli", cv.folds = 2,
                  interaction.depth = 6, n.tree = 2000, shrinkage = 0.05 )
})
```

```
##    user  system elapsed
##  62.888   0.295 103.449
```

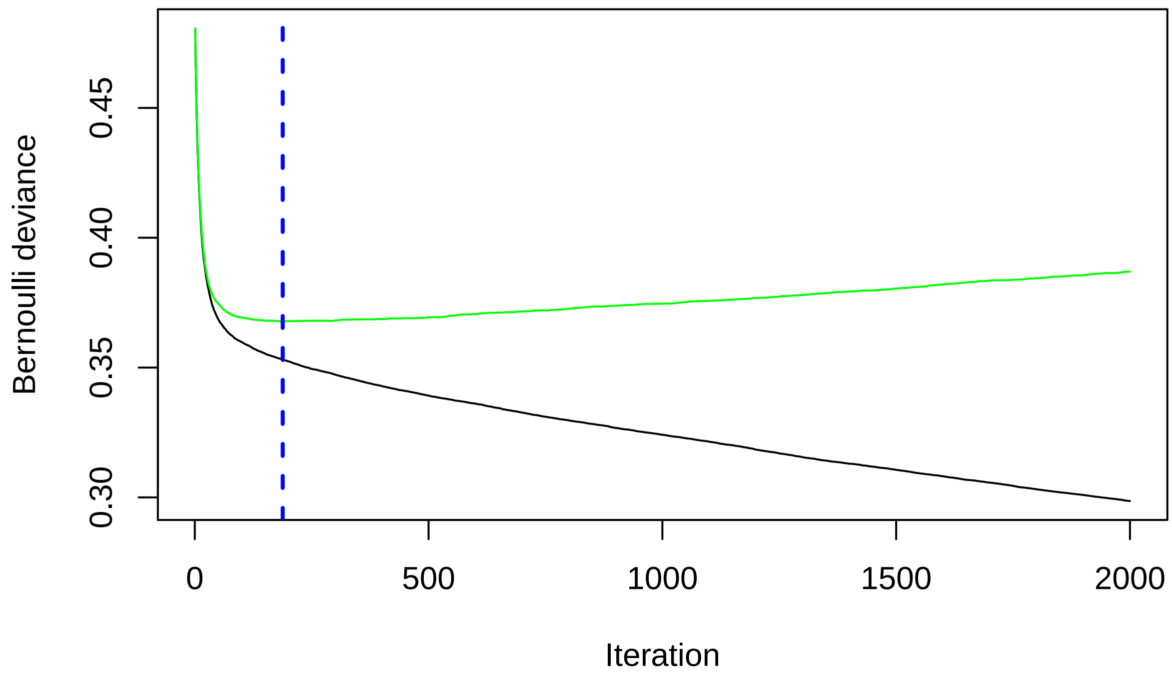
```
rf_model
```

```
##
## Call:
##  randomForest(formula = as.factor(y) ~ ., data = train_data, nodesize = 250)
##                Type of random forest: classification
##                      Number of trees: 500
## No. of variables tried at each split: 3
##
##         OOB estimate of  error rate: 6.5%
## Confusion matrix:
##       n   y class.error
## n 46377 253 0.005425692
## y  2998 407 0.880469897
```

```
gb_model
```

```
## gbm(formula = y_for_gbm ~ ., distribution = "bernoulli", data = train_data[,
##     -1], n.trees = 2000, interaction.depth = 6, shrinkage = 0.05,
##     cv.folds = 2)
## A gradient boosted model with bernoulli loss function.
## 2000 iterations were performed.
## The best cross-validation iteration was 188.
## There were 10 predictors of which 10 had non-zero influence.
```

```r
optimal_gbm_tree = gbm.perf(gb_model, method = "cv")
```




```r
optimal_gbm_tree
```

```
## [1] 188
```

```r
rf_y = predict(rf_model, newdata = train_data, type = "prob")[,2]
gb_y = predict(gb_model, newdata = train_data, n.tree = optimal_gbm_tree, type="response")

roc_rf = roc(response = train_data$y, predictor = rf_y , plot=T, col = "red")
```

```
## Setting levels: control = n, case = y
```

```
## Setting direction: controls < cases
```

```r
roc_rf$auc
```

```
## Area under the curve: 0.9418
```

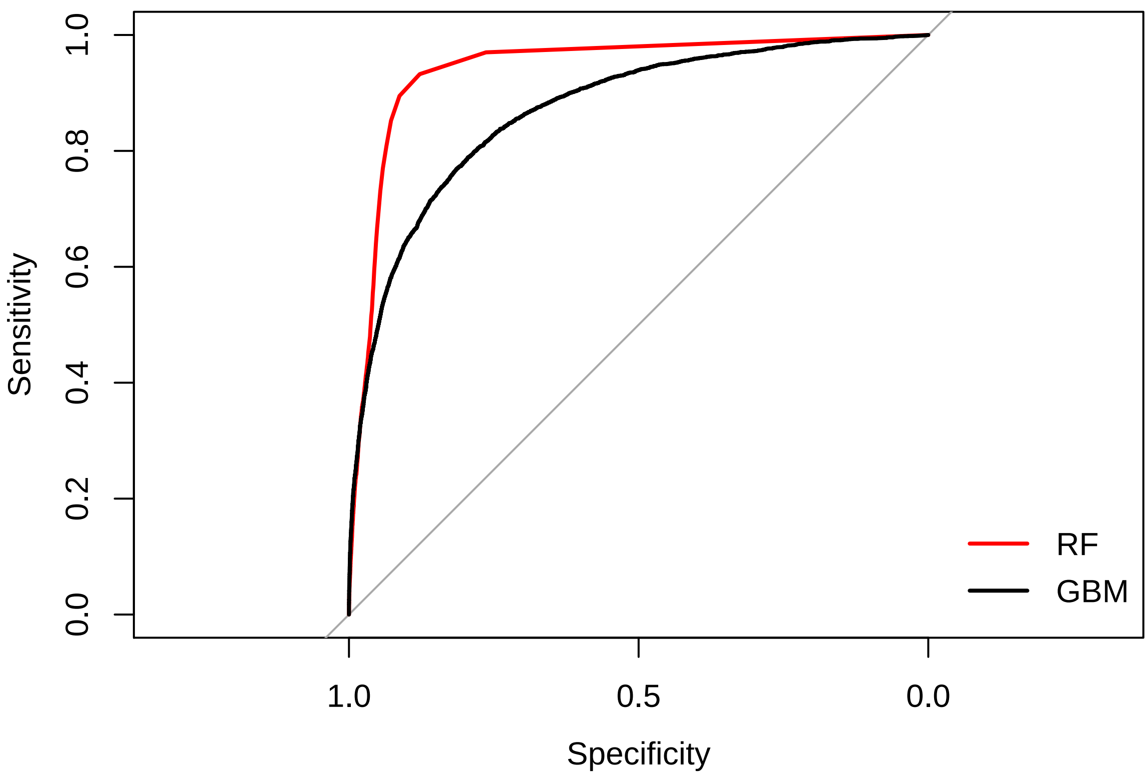
```r
roc_gb = roc(response = train_data$y, predictor = gb_y, plot=T, add = T, col = "black")
```

```
## Setting levels: control = n, case = y
## Setting direction: controls < cases
```

```r
roc_gb$auc
```

```
## Area under the curve: 0.87
```

```r
legend("bottomright", c("RF","GBM"), lwd = "2", col = c("red","black"), bty = "n")
```




```r
rf_y_test = predict(rf_model, newdata = test_data, type = "prob")[,2]
gb_y_test = predict(gb_model, newdata = test_data, n.tree = optimal_gbm_tree, type="response")

roc_rf_test = roc(response = test_data$y, predictor = rf_y_test , plot=T, col = "red")
```

```
## Setting levels: control = n, case = y
## Setting direction: controls < cases
```

```r
roc_rf_test$auc
```

```
## Area under the curve: 0.7814
```

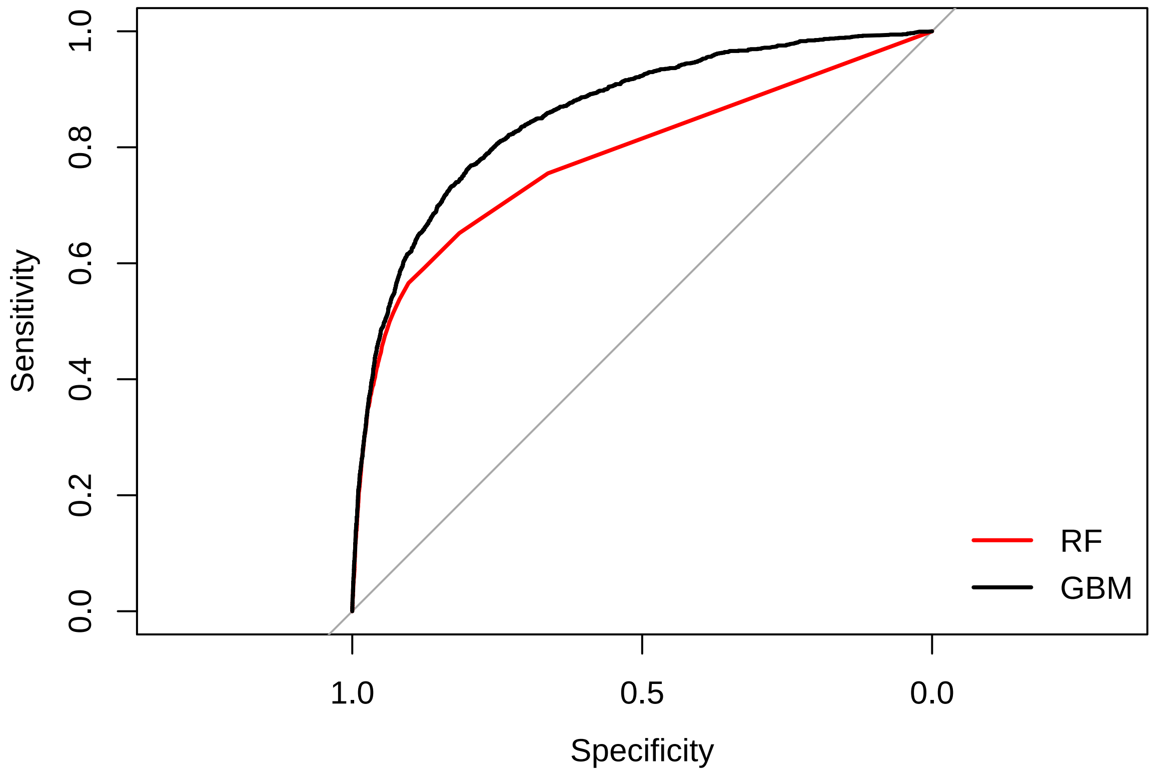
```r
roc_gb_test = roc(response = test_data$y, predictor = gb_y_test, plot=T, add = T, col = "black")
```

```
## Setting levels: control = n, case = y
## Setting direction: controls < cases
```

```r
roc_gb_test$auc
```

```
## Area under the curve: 0.8568
```

```r
legend("bottomright", c("RF","GBM"), lwd = "2", col = c("red","black"), bty = "n")
```




```r
## Looks like the gbm model won with a higher AUC

################################################################################
###   Lets try grid search to tune the model
################################################################################

original_train    = read.csv("train.csv", header = T)
original_train$Id = NULL

original_train    = original_train[1:3000,]

set.seed(321)
ind = sample(1:3, size = nrow(original_train), replace = TRUE)

test_index = which(ind == 1)
train_data = original_train[ -test_index, ]
test_data  = original_train[  test_index, ]
```

```r
library(gbm)
library(pROC)

train_data$y = ifelse(train_data$y == "y", 1, 0)
y_test       = test_data$y
test_data$y  = NULL

gbm_grid = expand.grid(interaction.depth = seq(from = 6, to = 8, by = 1),
                       n.trees = seq(from = 200, to = 2000, by = 400),
                       shrinkage = c(0.001, 0.05, 1), bag.fraction = c(0.5,1))
gbm_grid
```

```
##    interaction.depth n.trees shrinkage bag.fraction
## 1                  6     200     0.001          0.5
## 2                  7     200     0.001          0.5
## 3                  8     200     0.001          0.5
## 4                  6     600     0.001          0.5
## 5                  7     600     0.001          0.5
## 6                  8     600     0.001          0.5
## 7                  6    1000     0.001          0.5
## 8                  7    1000     0.001          0.5
## 9                  8    1000     0.001          0.5
## 10                 6    1400     0.001          0.5
## 11                 7    1400     0.001          0.5
## 12                 8    1400     0.001          0.5
## 13                 6    1800     0.001          0.5
## 14                 7    1800     0.001          0.5
## 15                 8    1800     0.001          0.5
## 16                 6     200     0.050          0.5
## 17                 7     200     0.050          0.5
## 18                 8     200     0.050          0.5
## 19                 6     600     0.050          0.5
## 20                 7     600     0.050          0.5
## 21                 8     600     0.050          0.5
## 22                 6    1000     0.050          0.5
## 23                 7    1000     0.050          0.5
## 24                 8    1000     0.050          0.5
## 25                 6    1400     0.050          0.5
## 26                 7    1400     0.050          0.5
## 27                 8    1400     0.050          0.5
## 28                 6    1800     0.050          0.5
## 29                 7    1800     0.050          0.5
## 30                 8    1800     0.050          0.5
## 31                 6     200     1.000          0.5
## 32                 7     200     1.000          0.5
## 33                 8     200     1.000          0.5
## 34                 6     600     1.000          0.5
## 35                 7     600     1.000          0.5
## 36                 8     600     1.000          0.5
## 37                 6    1000     1.000          0.5
## 38                 7    1000     1.000          0.5
## 39                 8    1000     1.000          0.5
## 40                 6    1400     1.000          0.5
```

```
## 41                   7     1400     1.000          0.5
## 42                   8     1400     1.000          0.5
## 43                   6     1800     1.000          0.5
## 44                   7     1800     1.000          0.5
## 45                   8     1800     1.000          0.5
## 46                   6      200     0.001          1.0
## 47                   7      200     0.001          1.0
## 48                   8      200     0.001          1.0
## 49                   6      600     0.001          1.0
## 50                   7      600     0.001          1.0
## 51                   8      600     0.001          1.0
## 52                   6     1000     0.001          1.0
## 53                   7     1000     0.001          1.0
## 54                   8     1000     0.001          1.0
## 55                   6     1400     0.001          1.0
## 56                   7     1400     0.001          1.0
## 57                   8     1400     0.001          1.0
## 58                   6     1800     0.001          1.0
## 59                   7     1800     0.001          1.0
## 60                   8     1800     0.001          1.0
## 61                   6      200     0.050          1.0
## 62                   7      200     0.050          1.0
## 63                   8      200     0.050          1.0
## 64                   6      600     0.050          1.0
## 65                   7      600     0.050          1.0
## 66                   8      600     0.050          1.0
## 67                   6     1000     0.050          1.0
## 68                   7     1000     0.050          1.0
## 69                   8     1000     0.050          1.0
## 70                   6     1400     0.050          1.0
## 71                   7     1400     0.050          1.0
## 72                   8     1400     0.050          1.0
## 73                   6     1800     0.050          1.0
## 74                   7     1800     0.050          1.0
## 75                   8     1800     0.050          1.0
## 76                   6      200     1.000          1.0
## 77                   7      200     1.000          1.0
## 78                   8      200     1.000          1.0
## 79                   6      600     1.000          1.0
## 80                   7      600     1.000          1.0
## 81                   8      600     1.000          1.0
## 82                   6     1000     1.000          1.0
## 83                   7     1000     1.000          1.0
## 84                   8     1000     1.000          1.0
## 85                   6     1400     1.000          1.0
## 86                   7     1400     1.000          1.0
## 87                   8     1400     1.000          1.0
## 88                   6     1800     1.000          1.0
## 89                   7     1800     1.000          1.0
## 90                   8     1800     1.000          1.0
```

```
m = dim(gbm_grid)[1]
m
```

```
## [1] 90
```

```r
gbm_auc  = rep(0, m)
gbm_auc
```

```
##  [1] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## [39] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## [77] 0 0 0 0 0 0 0 0 0 0 0 0 0 0
```

```r
no_of_folds = 2
set.seed(2000)
index_values = sample(1:no_of_folds, size = dim(train_data)[1], replace = TRUE)

system.time({

  for (i in 1:m)
  {

    tmp_auc     = rep(0, no_of_folds)

    for (j in 1:no_of_folds)
    {
      index_out     = which(index_values == j)
      left_out_data = train_data[  index_out, ]
      left_in_data  = train_data[ -index_out, ]

      tmp_model     = gbm( y ~ ., data = left_in_data, dist = "bernoulli",
                           interaction.depth = gbm_grid$interaction.depth[i],
                           shrinkage         = gbm_grid$shrinkage[i],
                           n.trees           = gbm_grid$n.trees[i],
                           bag.fraction      = gbm_grid$bag.fraction[i])

      tmp_pred     = predict(tmp_model, newdata = left_out_data, type="response",
                             n.trees = gbm_grid$n.trees[i])

      tmp_auc[j]   = roc(response = left_out_data$y, predictor = tmp_pred ,
                         plot=F)$auc[1]

    }

    gbm_auc[i]     = mean(tmp_auc)

  }

})
```

```
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases
```

```
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
```

```
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
```

```
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
```

```
## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases
```

```
## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls > cases
```

```
## Setting levels: control = 0, case = 1

## Setting direction: controls > cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls > cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls > cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases
```

```
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls > cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
```

```
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
```

```
## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases
```

```
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
```

```
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
```

```
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
```

```
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls > cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
```

```
## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls > cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls > cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases
```

```
## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls > cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

## Setting levels: control = 0, case = 1

## Setting direction: controls > cases

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

##    user  system elapsed
## 113.528   0.230 113.761
```

```
results     = cbind(gbm_grid, gbm_auc)
results
```

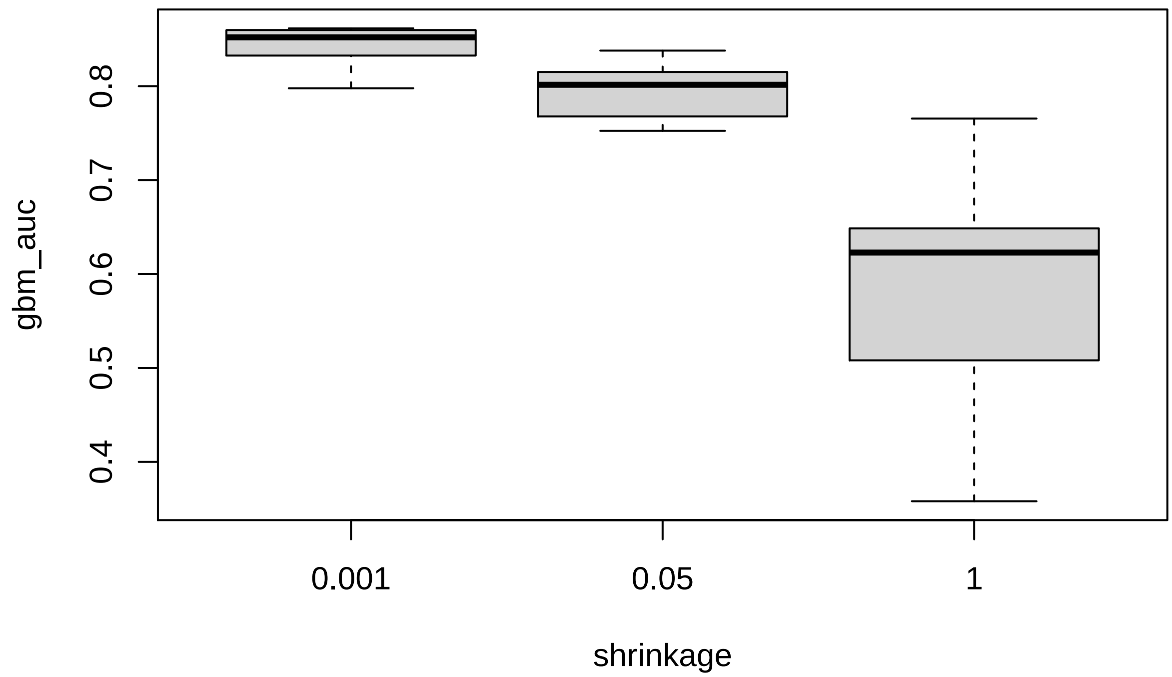
```
##    interaction.depth n.trees shrinkage bag.fraction   gbm_auc
## 1                  6     200     0.001          0.5 0.8526093
## 2                  7     200     0.001          0.5 0.8547386
## 3                  8     200     0.001          0.5 0.8523970
## 4                  6     600     0.001          0.5 0.8611400
## 5                  7     600     0.001          0.5 0.8597702
## 6                  8     600     0.001          0.5 0.8545992
## 7                  6    1000     0.001          0.5 0.8611926
## 8                  7    1000     0.001          0.5 0.8596881
## 9                  8    1000     0.001          0.5 0.8569756
## 10                 6    1400     0.001          0.5 0.8615255
## 11                 7    1400     0.001          0.5 0.8603313
## 12                 8    1400     0.001          0.5 0.8564688
## 13                 6    1800     0.001          0.5 0.8615581
## 14                 7    1800     0.001          0.5 0.8602522
## 15                 8    1800     0.001          0.5 0.8574719
## 16                 6     200     0.050          0.5 0.8332664
## 17                 7     200     0.050          0.5 0.8350017
## 18                 8     200     0.050          0.5 0.8378801
## 19                 6     600     0.050          0.5 0.8196220
## 20                 7     600     0.050          0.5 0.8160940
## 21                 8     600     0.050          0.5 0.8181727
## 22                 6    1000     0.050          0.5 0.8111935
## 23                 7    1000     0.050          0.5 0.8062048
## 24                 8    1000     0.050          0.5 0.8016698
## 25                 6    1400     0.050          0.5 0.8012629
## 26                 7    1400     0.050          0.5 0.8150131
## 27                 8    1400     0.050          0.5 0.8104474
## 28                 6    1800     0.050          0.5 0.7987020
## 29                 7    1800     0.050          0.5 0.8061494
## 30                 8    1800     0.050          0.5 0.7981257
## 31                 6     200     1.000          0.5 0.6174450
## 32                 7     200     1.000          0.5 0.5801106
## 33                 8     200     1.000          0.5 0.5679759
## 34                 6     600     1.000          0.5 0.5168504
## 35                 7     600     1.000          0.5 0.3733013
## 36                 8     600     1.000          0.5 0.5247367
## 37                 6    1000     1.000          0.5 0.4905207
## 38                 7    1000     1.000          0.5 0.4716068
## 39                 8    1000     1.000          0.5 0.3580825
## 40                 6    1400     1.000          0.5 0.4189320
## 41                 7    1400     1.000          0.5 0.5617828
## 42                 8    1400     1.000          0.5 0.4990526
## 43                 6    1800     1.000          0.5 0.5683858
## 44                 7    1800     1.000          0.5 0.3862428
## 45                 8    1800     1.000          0.5 0.5081039
## 46                 6     200     0.001          1.0 0.7977753
## 47                 7     200     0.001          1.0 0.8026006
## 48                 8     200     0.001          1.0 0.7984577
## 49                 6     600     0.001          1.0 0.8325608
## 50                 7     600     0.001          1.0 0.8279926
## 51                 8     600     0.001          1.0 0.8269859
## 52                 6    1000     0.001          1.0 0.8331427
## 53                 7    1000     0.001          1.0 0.8335406
```

```
## 54                 8    1000     0.001          1.0 0.8326023
## 55                 6    1400     0.001          1.0 0.8361088
## 56                 7    1400     0.001          1.0 0.8309918
## 57                 8    1400     0.001          1.0 0.8355960
## 58                 6    1800     0.001          1.0 0.8514355
## 59                 7    1800     0.001          1.0 0.8516643
## 60                 8    1800     0.001          1.0 0.8433010
## 61                 6     200     0.050          1.0 0.8204419
## 62                 7     200     0.050          1.0 0.8125805
## 63                 8     200     0.050          1.0 0.8037757
## 64                 6     600     0.050          1.0 0.7866060
## 65                 7     600     0.050          1.0 0.7852306
## 66                 8     600     0.050          1.0 0.7823792
## 67                 6    1000     0.050          1.0 0.7781384
## 68                 7    1000     0.050          1.0 0.7662574
## 69                 8    1000     0.050          1.0 0.7678180
## 70                 6    1400     0.050          1.0 0.7659670
## 71                 7    1400     0.050          1.0 0.7586969
## 72                 8    1400     0.050          1.0 0.7631792
## 73                 6    1800     0.050          1.0 0.7566323
## 74                 7    1800     0.050          1.0 0.7524287
## 75                 8    1800     0.050          1.0 0.7614500
## 76                 6     200     1.000          1.0 0.7466947
## 77                 7     200     1.000          1.0 0.7655161
## 78                 8     200     1.000          1.0 0.6282487
## 79                 6     600     1.000          1.0 0.7508422
## 80                 7     600     1.000          1.0 0.7228175
## 81                 8     600     1.000          1.0 0.6310778
## 82                 6    1000     1.000          1.0 0.7109823
## 83                 7    1000     1.000          1.0 0.6929100
## 84                 8    1000     1.000          1.0 0.6417185
## 85                 6    1400     1.000          1.0 0.6842151
## 86                 7    1400     1.000          1.0 0.6463390
## 87                 8    1400     1.000          1.0 0.6486354
## 88                 6    1800     1.000          1.0 0.6464135
## 89                 7    1800     1.000          1.0 0.6367831
## 90                 8    1800     1.000          1.0 0.6473517
```

```
best_result = results[which.max(gbm_auc),]
best_result
```

```
##    interaction.depth n.trees shrinkage bag.fraction   gbm_auc
## 13                 6    1800     0.001          0.5 0.8615581
```

```
boxplot( gbm_auc ~ shrinkage, data = results)
```

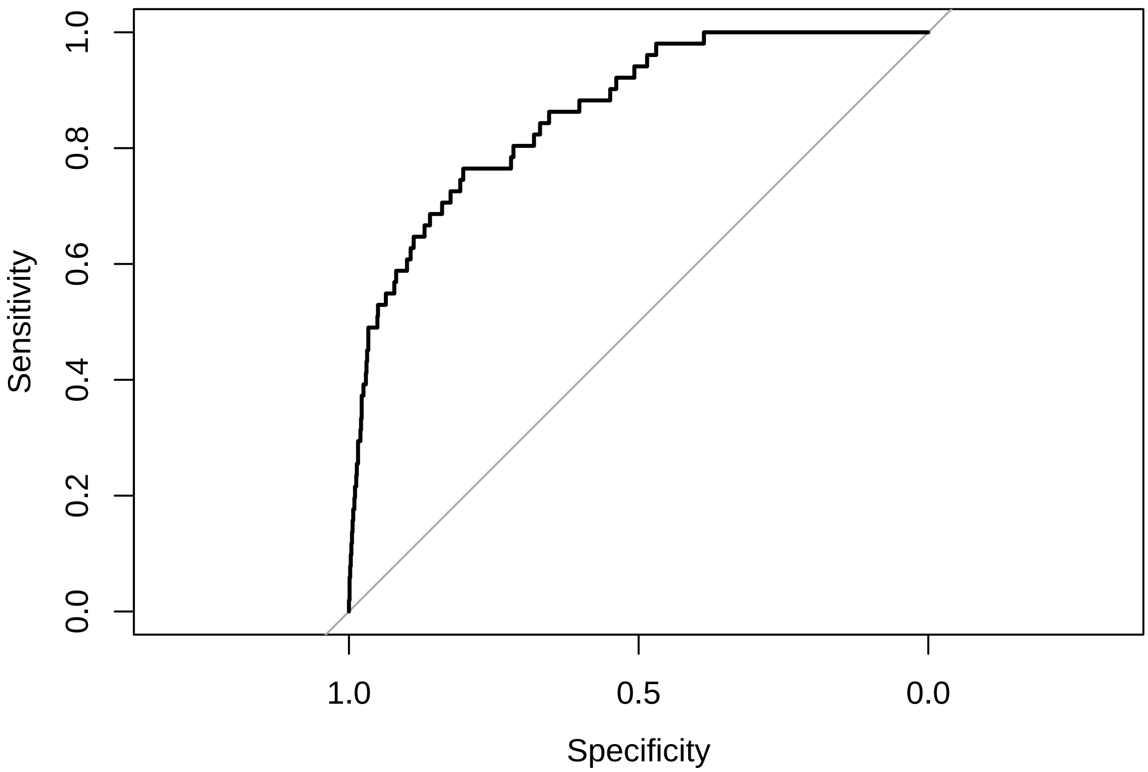



```
final_model = gbm(y ~ ., data = train_data, dist = "bernoulli",
                  interaction.depth = best_result$interaction.depth,
                  shrinkage         = best_result$shrinkage,
                  bag.fraction      = best_result$bag.fraction,
                  n.trees           = best_result$n.trees)

gb_y = predict(final_model, newdata = test_data,
               n.tree = best_result$n.trees, type="response")

roc_gb = roc(response = y_test, predictor = gb_y, plot=T, col = "black")
```

```
## Setting levels: control = n, case = y
## Setting direction: controls < cases
```




```
roc_gb$auc
```

```
## Area under the curve: 0.8622
```