STAT 447 project

Ken Mawer

2023-04-05

Calcification_mode <> 3.5

① False 448 obs

Subtlety_mean <> 4.125

Lobulation_entropy <> 0.405639   Spiculation_mean <> 2.125

Spiculation_mean <> 1.125   ④   Margin_entropy <> 0.405639   ⑦

② ③   No_consensus 1383 obs   ⑤ ⑥   True 552 obs

False 825 obs   Ambiguous 279 obs   Ambiguous 235 obs   No_consensus 732 obs
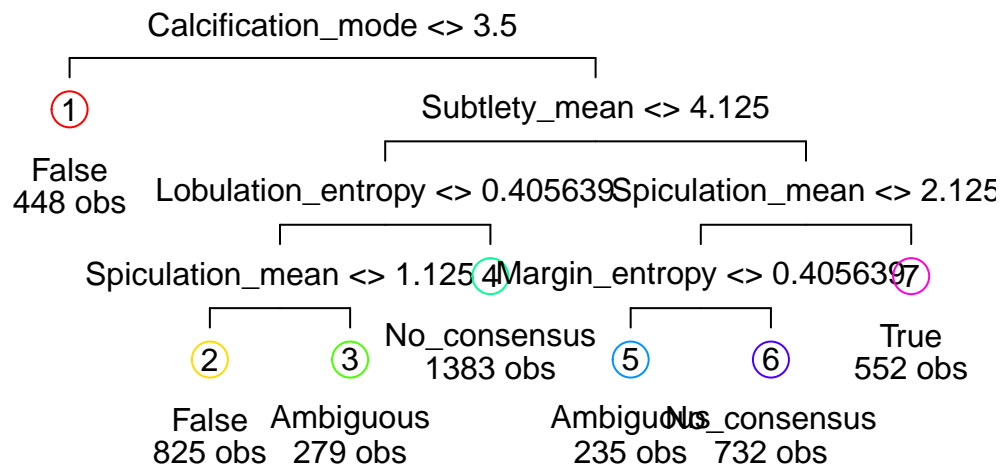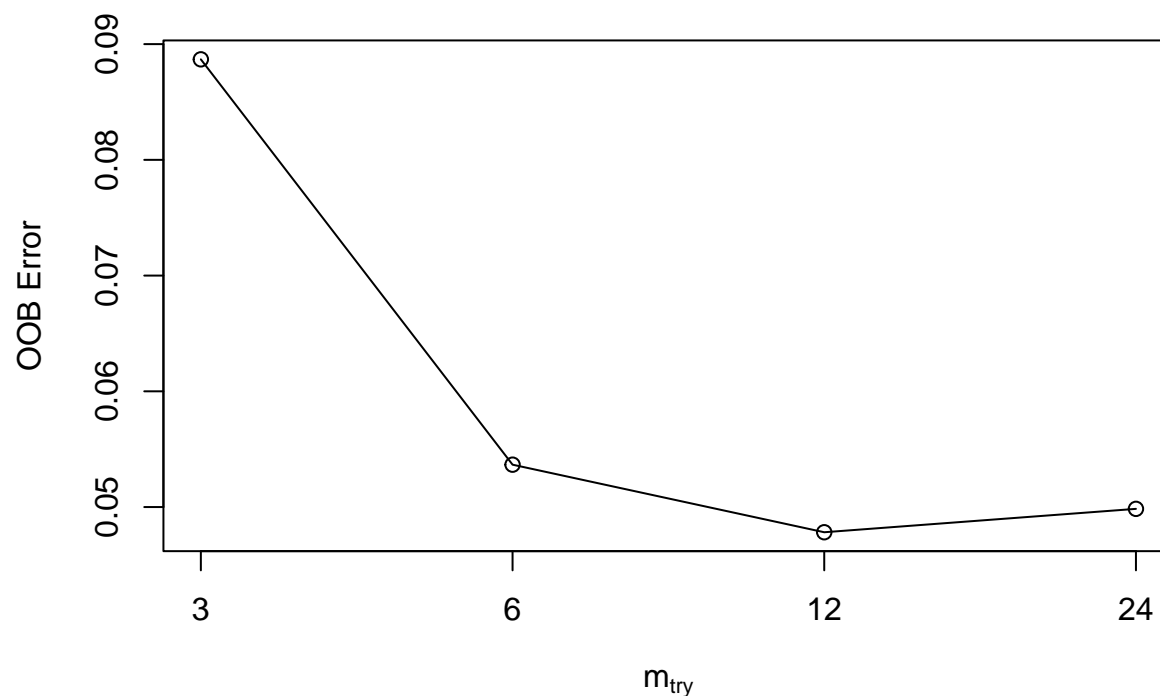
```
## mtry = 6   OOB error = 5.37%
## Searching left ...
## mtry = 3     OOB error = 8.87%
## -0.6527197 0.05
## Searching right ...
## mtry = 12     OOB error = 4.78%
## 0.1087866 0.05
## mtry = 24     OOB error = 4.98%
## -0.04225352 0.05
```

```
##         mtry    OOBError
## 3.OOB      3 0.08868433
## 6.OOB      6 0.05365963
## 12.OOB    12 0.04782218
## 24.OOB    24 0.04984284
```

```
## Warning in confusionMatrix.default(predict(randomForest(Is_cancer ~ ., data =
## t), : Levels are not in the same order for reference and data. Refactoring data
## to match.
```

```
## Confusion Matrix and Statistics
##
##               Reference
## Prediction    True Ambiguous False No_consensus
##   True         173        32    10           69
##   Ambiguous     51       295   125          138
##   False         30       159   558           63
##   No_consensus  86       135   103          308
##
## Overall Statistics
##
##                Accuracy : 0.5713
##                  95% CI : (0.5509, 0.5915)
##     No Information Rate : 0.3409
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.4109
```

```
##
##  Mcnemar's Test P-Value : 4.008e-05
##
## Statistics by Class:
##
##                      Class: True Class: Ambiguous Class: False
## Sensitivity              0.50882            0.4750       0.7010
## Specificity              0.94436            0.8168       0.8363
## Pos Pred Value           0.60915            0.4844       0.6889
## Neg Pred Value           0.91858            0.8111       0.8439
## Prevalence               0.14561            0.2660       0.3409
## Detection Rate           0.07409            0.1263       0.2390
## Detection Prevalence     0.12163            0.2608       0.3469
## Balanced Accuracy        0.72659            0.6459       0.7686
##                      Class: No_consensus
## Sensitivity                       0.5329
## Specificity                       0.8156
## Pos Pred Value                    0.4873
## Neg Pred Value                    0.8415
## Prevalence                        0.2475
## Detection Rate                    0.1319
## Detection Prevalence              0.2707
## Balanced Accuracy                 0.6742

## Warning in confusionMatrix.default(predict(randomForest(Is_cancer ~ ., data =
## t, : Levels are not in the same order for reference and data. Refactoring data
## to match.

## Confusion Matrix and Statistics
##
##             Reference
## Prediction    True Ambiguous False No_consensus
##   True         173        27    13           76
##   Ambiguous     45       302   139          137
##   False         36       163   550           63
##   No_consensus  86       129    94          302
##
## Overall Statistics
##
##                Accuracy : 0.5683
##                  95% CI : (0.5479, 0.5885)
##     No Information Rate : 0.3409
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.4068
##
##  Mcnemar's Test P-Value : 0.0004835
##
## Statistics by Class:
##
##                      Class: True Class: Ambiguous Class: False
## Sensitivity              0.50882            0.4863       0.6910
## Specificity              0.94185            0.8127       0.8298
## Pos Pred Value           0.59862            0.4848       0.6773
## Neg Pred Value           0.91838            0.8137       0.8385
```

```
## Prevalence                    0.14561          0.2660      0.3409
## Detection Rate                 0.07409          0.1293      0.2355
## Detection Prevalence           0.12377          0.2668      0.3478
## Balanced Accuracy              0.72534          0.6495      0.7604
##                       Class: No_consensus
## Sensitivity                       0.5225
## Specificity                       0.8241
## Pos Pred Value                    0.4943
## Neg Pred Value                    0.8399
## Prevalence                        0.2475
## Detection Rate                    0.1293
## Detection Prevalence              0.2617
## Balanced Accuracy                 0.6733
```

```
plot(my_cv$n.var, my_cv$error.cv)
```