RandomForest_&_Boosting

JaimeGoB

11/27/2020

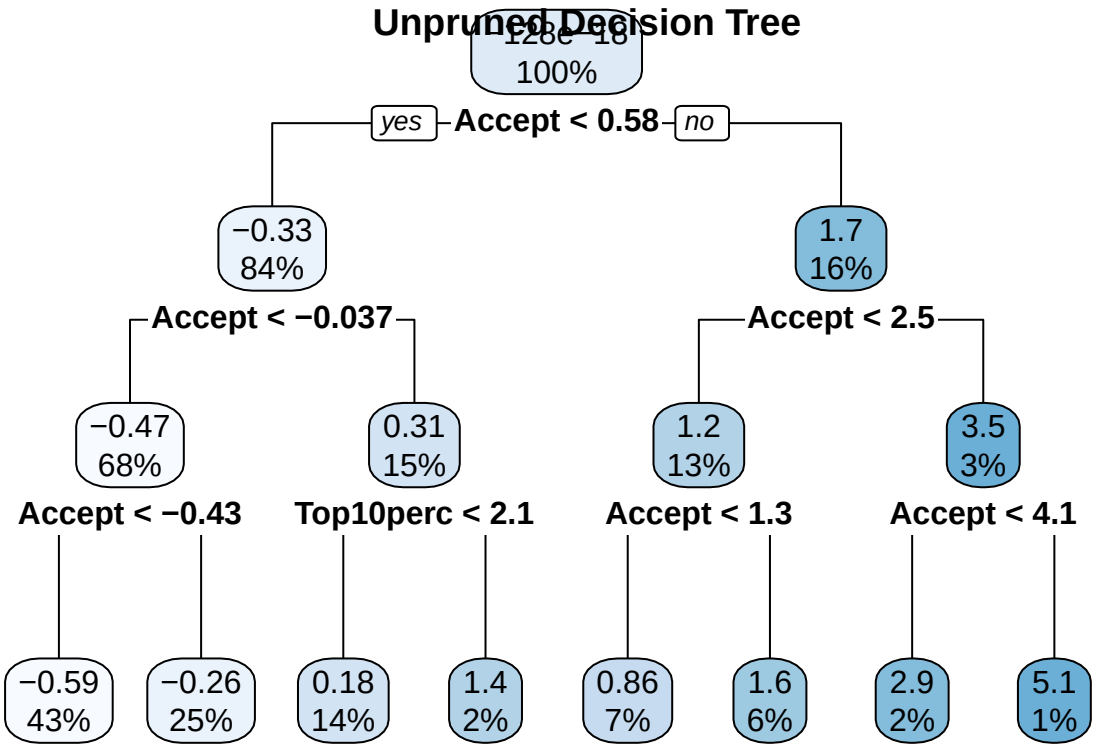
(a) Fit a tree to the data. i) Summarize the results. Unless the number of terminal nodes is large, ii)display the tree graphically and explicitly iii)describe the regions corresponding to the terminal nodes that provide a partition of the predictor space (i.e., provide expressions for the regions R1,. . . ,RJ). iiii)Report its MSE.

i) Decision Tree - Summary

```
##
## Regression tree:
## tree(formula = Apps ~ ., data = college)
## Variables actually used in tree construction:
## [1] "Accept"    "Top10perc"
## Number of terminal nodes:  8
## Residual mean deviance:  0.1356 = 104.3 / 769
## Distribution of residuals:
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## -2.16300 -0.11090 -0.02694  0.00000  0.07409  5.97800
```

ii) Visualizing Tree

```
##   Apps
##  -0.59 when Accept <  -0.434
##  -0.26 when Accept is -0.434 to -0.037
##   0.18 when Accept is -0.037 to  0.580 & Top10perc <  2.1
##   0.86 when Accept is  0.580 to  1.283
##   1.41 when Accept is -0.037 to  0.580 & Top10perc >= 2.1
##   1.64 when Accept is  1.283 to  2.465
##   2.94 when Accept is  2.465 to  4.099
##   5.07 when Accept >=            4.099
```

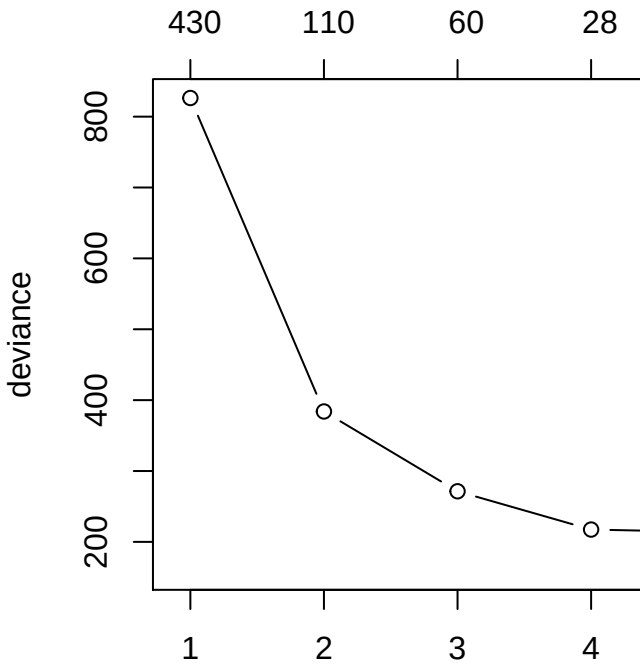


iii) Describe Regions

iiii) Report MSE - Decision Tree

```
## [1] 0.1342
```

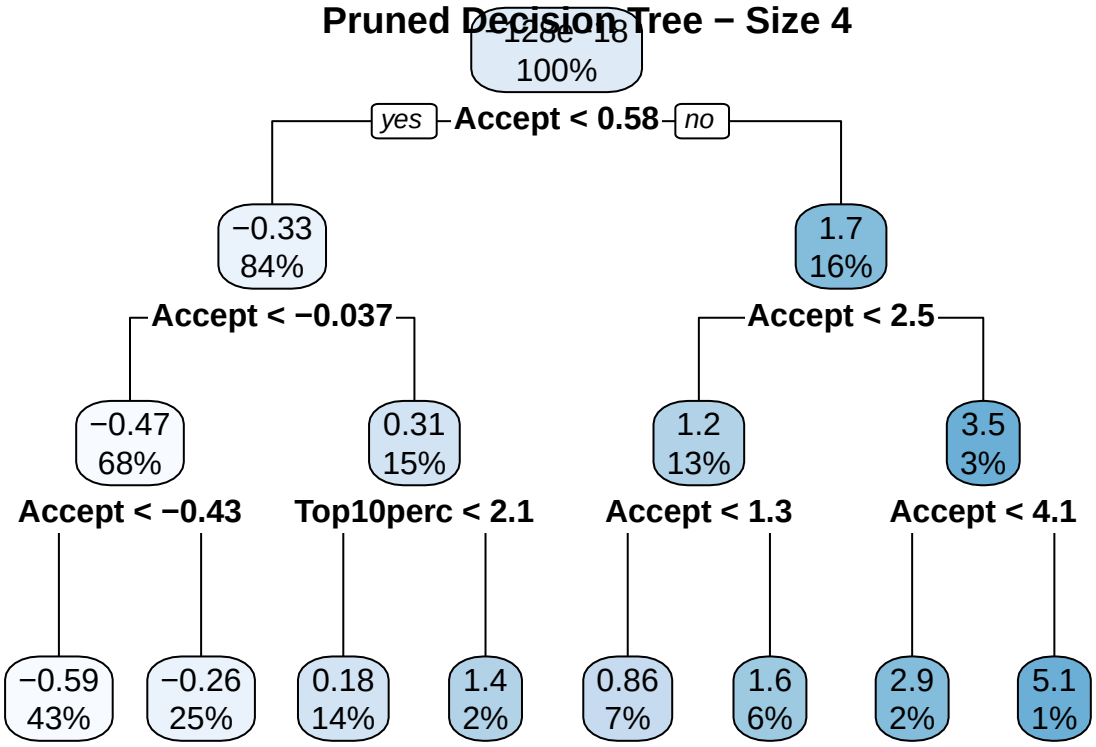
(b) i) Use LOOCV to determine whether pruning is helpful and determine the optimal size for the pruned tree. ii)Compare the pruned and un-pruned trees. iii) Report MSE for the pruned tree. iiii) Which predictors seem to be the most important?



i) Using LOOCV to Determine Optimal Size of Pruned Tree

ii)Compare the pruned and un-pruned trees.

```
##   Apps
##  -0.59 when Accept <  -0.434
##  -0.26 when Accept is -0.434 to -0.037
##   0.18 when Accept is -0.037 to  0.580 & Top10perc <  2.1
##   0.86 when Accept is  0.580 to  1.283
##   1.41 when Accept is -0.037 to  0.580 & Top10perc >= 2.1
##   1.64 when Accept is  1.283 to  2.465
##   2.94 when Accept is  2.465 to  4.099
##   5.07 when Accept >=            4.099
```



iii) Report MSE for the pruned tree.

```
## [1] 0.2322
```

iiii) Which predictors seem to be the most important?

(c) i) Use a bagging approach to analyze the data with B = 1000. ii) Compute the MSE. iii) Which predictors seem to be the most important?

i) Use a bagging approach to analyze the data with B = 1000

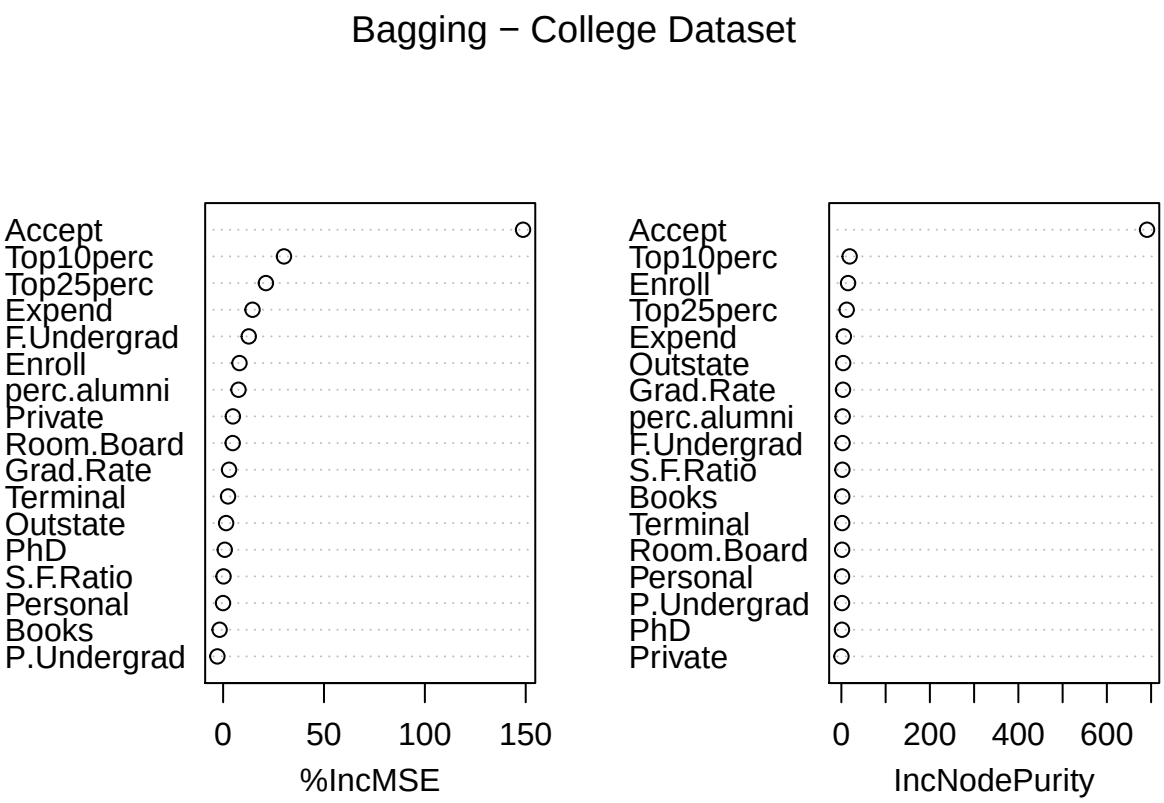
ii) Compute the MSE

```
## [1] 0.0219
```

iii) Which predictors seem to be the most important?

```
##                  %IncMSE IncNodePurity
## Private       4.83118027      0.108938
## Accept      148.66907566    690.797592
## Enroll        8.10698599     15.094219
## Top10perc    30.16960702     18.684319
## Top25perc    21.21159943     12.010591
## F.Undergrad  12.70584773      2.667243
## P.Undergrad  -2.84818427      1.465628
## Outstate      1.47851517      4.080293
## Room.Board    4.75165971      1.717679
## Books        -1.74304307      1.968720
## Personal     -0.05034011      1.469174
## PhD           0.80769147      1.305400
```

```
## Terminal      2.43831328      1.858596
## S.F.Ratio     0.19743642      2.311502
## perc.alumni   7.62388170      2.924923
## Expend       14.57450042      5.594058
## Grad.Rate     2.93474712      3.689698
```



(d) Repeat (c) with a random forest approach with B = 1000 and m = p/3

i) Use random forest approach with B = 1000 and m = p/3

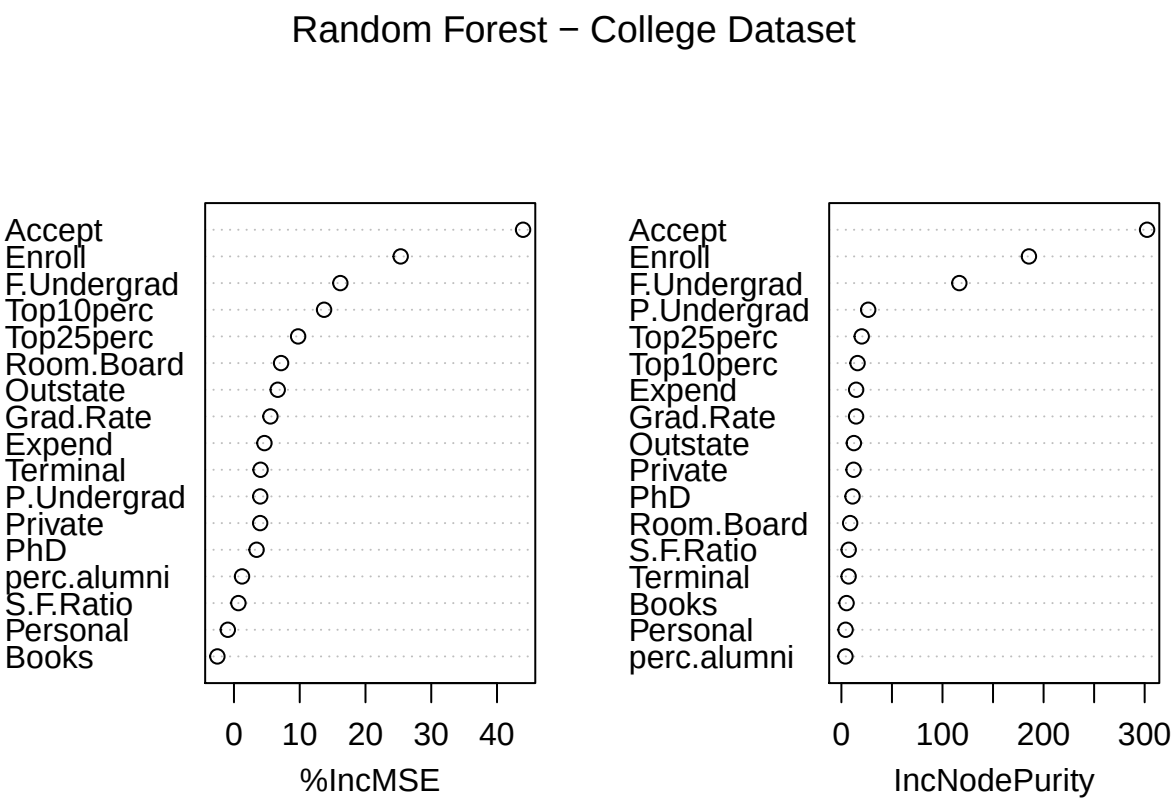
ii) Compute the MSE

```
## [1] 0.0303
```

iii) Which predictors seem to be the most important?

```
##                %IncMSE IncNodePurity
## Private      3.9801409     12.046810
## Accept      43.9617854    302.374555
## Enroll      25.3494828    185.498229
## Top10perc   13.7034103     15.972794
## Top25perc    9.7541225     20.087012
## F.Undergrad 16.1767345    116.642220
## P.Undergrad  3.9921228     26.496841
## Outstate     6.6549077     12.328685
## Room.Board   7.1638670      8.640198
## Books       -2.5230514      5.084508
## Personal    -0.9391046      4.127072
## PhD          3.4391029     11.005382
```

```
## Terminal     4.0426396      7.151158
## S.F.Ratio    0.6722498      7.222853
## perc.alumni  1.2235795      3.968985
## Expend       4.6306753     14.583698
## Grad.Rate    5.5432328     14.537672
```

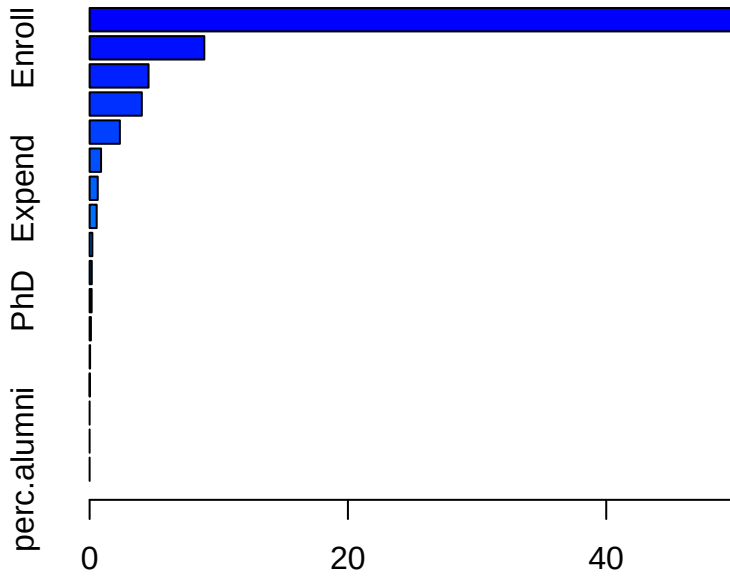


(e) Repeat (c) with a boosting approach with B = 1000, d = 1, and lambda = 0.01

i) Use boosting approach with B = 1000, d = 1, and lambda = 0.01

ii) Compute the MSE

```
## [1] 0.1229
```



iii) Which predictors seem to be the most important?

```
##                     var     rel.inf
## Accept           Accept 77.42304552
## Enroll           Enroll  8.88043821
## F.Undergrad F.Undergrad  4.56276803
## Top10perc     Top10perc  4.04347092
## Top25perc     Top25perc  2.34322672
## P.Undergrad P.Undergrad  0.88538917
## Expend           Expend  0.62927580
## Grad.Rate     Grad.Rate  0.54635854
## Books             Books  0.21184201
## Outstate       Outstate  0.16634838
## PhD                 PhD  0.15744648
## Terminal       Terminal  0.10031255
## S.F.Ratio     S.F.Ratio  0.03710821
## Room.Board   Room.Board  0.01296945
## Private         Private  0.00000000
## Personal       Personal  0.00000000
## perc.alumni perc.alumni  0.00000000
```

(f) Compare the results from the various methods. Which method would you recommend?

```
##      mse_tree mse_pruned_tree mse_bag mse_rf mse_boosting
## [1,]   0.1342          0.2322  0.0219 0.0303       0.1229
```