```
[ ] system("gdown --id 1puITTv_Y_2g9rKg7jSaT4XN41HuoFywr")
    system("gdown --id 1JCy0LLNnrauv8RVi0oZLfjRugEDe_Qc3")
```

```
[ ] hmeqtrain <- read.csv("hmeqN_train.csv", header = T)
    str(hmeqtrain)
```

```
'data.frame':   2000 obs. of  7 variables:
 $ ID     : int  4952 5546 938 277 5204 5762 2354 2896 76 4700 ...
 $ BAD    : int  1 1 1 1 1 1 1 1 1 1 ...
 $ LOAN   : int  26100 35000 9100 5700 28100 44000 14200 16000 3900 24800 ...
 $ MORTDUE: num  73525 391000 17218 58400 61000 ...
 $ VALUE  : num  89870 505000 36721 75000 99000 ...
 $ REASON : chr  "DebtCon" "DebtCon" "DebtCon" "HomeImp" ...
 $ JOB    : chr  "Office" "ProfExe" "Other" "ProfExe" ...
```

```
[ ] hmeqtest <- read.csv("hmeqN_test.csv", header=T)
    str(hmeqtest)
```

```
'data.frame':   378 obs. of  7 variables:
 $ ID     : int  5632 675 3234 3537 3804 926 462 2229 4770 184 ...
 $ BAD    : int  1 1 1 1 1 1 1 1 1 1 ...
 $ LOAN   : int  38700 8000 17300 18600 20000 9000 6900 13700 25000 5000 ...
 $ MORTDUE: num  119847 37871 73000 64248 60336 ...
 $ VALUE  : num  162365 89870 95000 82690 132430 ...
 $ REASON : chr  "HomeImp" "HomeImp" "DebtCon" "DebtCon" ...
 $ JOB    : chr  "ProfExe" "ProfExe" "Other" "Mgr" ...
```

```
[ ] install.packages("reshape")
    library(reshape)
```

```
Installing package into '/usr/local/lib/R/site-library'
(as 'lib' is unspecified)

also installing the dependencies 'Rcpp', 'plyr'
```

```
[ ] hmeqtrain <- as.data.frame(hmeqtrain)
    BAD <- as.factor(hmeqtrain$BAD)
    LOAN <- as.numeric(hmeqtrain$LOAN)
    hmeqtrain <- cbind(BAD,LOAN,hmeqtrain[,-2:-3])
    str(hmeqtrain)
```

```
'data.frame':   2000 obs. of  7 variables:
 $ BAD    : Factor w/ 2 levels "1","2": 1 1 1 1 1 1 1 1 1 1 ...
 $ LOAN   : num  26100 35000 9100 5700 28100 44000 14200 16000 3900 24800 ...
 $ ID     : int  4952 5546 938 277 5204 5762 2354 2896 76 4700 ...
 $ MORTDUE: num  73525 391000 17218 58400 61000 ...
 $ VALUE  : num  89870 505000 36721 75000 99000 ...
 $ REASON : chr  "DebtCon" "DebtCon" "DebtCon" "HomeImp" ...
 $ JOB    : chr  "Office" "ProfExe" "Other" "ProfExe" ...
```

```
[ ] hmeqtest <- as.data.frame(hmeqtest)
    BAD <- as.factor(hmeqtest$BAD)
    LOAN <- as.numeric(hmeqtest$LOAN)
    hmeqtest <- cbind(BAD,LOAN,hmeqtest[,-2:-3])
    str(hmeqtest)
```

```
'data.frame':   378 obs. of  7 variables:
 $ BAD    : Factor w/ 2 levels "1","2": 1 1 1 1 1 1 1 1 1 1 ...
 $ LOAN   : num  38700 8000 17300 18600 20000 9000 6900 13700 25000 5000 ...
 $ ID     : int  5632 675 3234 3537 3804 926 462 2229 4770 184 ...
 $ MORTDUE: num  119847 37871 73000 64248 60336 ...
 $ VALUE  : num  162365 89870 95000 82690 132430 ...
 $ REASON : chr  "HomeImp" "HomeImp" "DebtCon" "DebtCon" ...
 $ JOB    : chr  "ProfExe" "ProfExe" "Other" "Mgr" ...
```

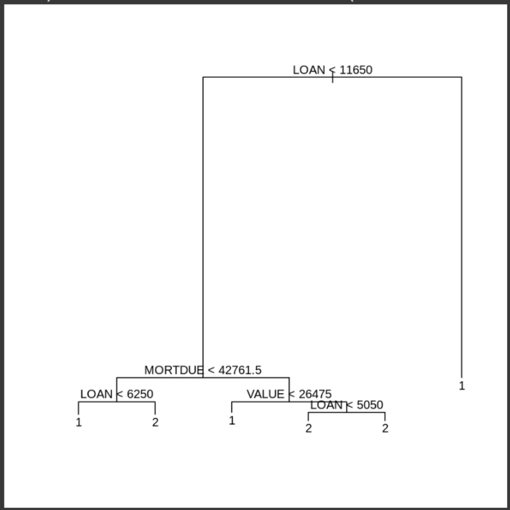
```
[ ] install.packages("tree")
    library(tree)
```

```
Installing package into '/usr/local/lib/R/site-library'
(as 'lib' is unspecified)
```

```
[ ] tree1 <- tree(BAD~.-ID, data=hmeqtrain)
    plot(tree1);text(tree1)
    summary(tree1)
    print(tree1)
```

```
Warning message in tree(BAD ~ . - ID, data = hmeqtrain):
```

```
"NAs introduced by coercion"

Classification tree:
tree(formula = BAD ~ . - ID, data = hmeqtrain)
Variables actually used in tree construction:
[1] "LOAN"    "MORTDUE" "VALUE"
Number of terminal nodes:  6
Residual mean deviance:  0.6881 = 1372 / 1994
Misclassification error rate: 0.15 = 300 / 2000
node), split, n, deviance, yval, (yprob)
      * denotes terminal node

 1) root 2000 2773.000 1 ( 0.5000 0.5000 )
   2) LOAN < 11650 1367 1590.000 2 ( 0.2685 0.7315 )
     4) MORTDUE < 42761.5 278  385.300 1 ( 0.5108 0.4892 )
       8) LOAN < 6250 81   74.580 1 ( 0.8272 0.1728 ) *
       9) LOAN > 6250 197  261.800 2 ( 0.3807 0.6193 ) *
     5) MORTDUE > 42761.5 1089 1110.000 2 ( 0.2066 0.7934 )
      10) VALUE < 26475 16    7.481 1 ( 0.9375 0.0625 ) *
      11) VALUE > 26475 1073 1061.000 2 ( 0.1957 0.8043 )
        22) LOAN < 5050 106  144.500 2 ( 0.4245 0.5755 ) *
        23) LOAN > 5050 967  883.600 2 ( 0.1706 0.8294 ) *
   3) LOAN > 11650 633    0.000 1 ( 1.0000 0.0000 ) *
```

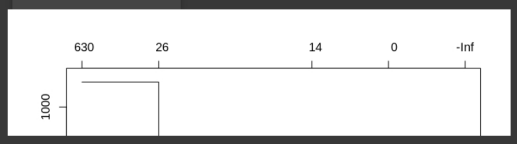


[ ]
```
cv.tree1 <- cv.tree(tree1, FUN=prune.misclass, K=10)
plot(cv.tree1);text(cv.tree1)
```

```
Warning message in tree(model = m[rand != i, , drop = FALSE]):
"NAs introduced by coercion"
Warning message in pred1.tree(tree, tree.matrix(nd)):
"NAs introduced by coercion"
Warning message in tree(model = m[rand != i, , drop = FALSE]):
"NAs introduced by coercion"
Warning message in pred1.tree(tree, tree.matrix(nd)):
"NAs introduced by coercion"
Warning message in tree(model = m[rand != i, , drop = FALSE]):
"NAs introduced by coercion"
Warning message in pred1.tree(tree, tree.matrix(nd)):
"NAs introduced by coercion"
Warning message in tree(model = m[rand != i, , drop = FALSE]):
"NAs introduced by coercion"
Warning message in pred1.tree(tree, tree.matrix(nd)):
"NAs introduced by coercion"
Warning message in tree(model = m[rand != i, , drop = FALSE]):
"NAs introduced by coercion"
Warning message in pred1.tree(tree, tree.matrix(nd)):
"NAs introduced by coercion"
Warning message in tree(model = m[rand != i, , drop = FALSE]):
"NAs introduced by coercion"
Warning message in pred1.tree(tree, tree.matrix(nd)):
"NAs introduced by coercion"
Warning message in tree(model = m[rand != i, , drop = FALSE]):
"NAs introduced by coercion"
Warning message in pred1.tree(tree, tree.matrix(nd)):
"NAs introduced by coercion"
Warning message in tree(model = m[rand != i, , drop = FALSE]):
"NAs introduced by coercion"
Warning message in pred1.tree(tree, tree.matrix(nd)):
"NAs introduced by coercion"
Warning message in tree(model = m[rand != i, , drop = FALSE]):
"NAs introduced by coercion"
Warning message in pred1.tree(tree, tree.matrix(nd)):
"NAs introduced by coercion"
Warning message in tree(model = m[rand != i, , drop = FALSE]):
"NAs introduced by coercion"
Warning message in pred1.tree(tree, tree.matrix(nd)):
"NAs introduced by coercion"
Error in xy.coords(x, y, recycle = TRUE, setLab = FALSE): 'x' is a list, but does not have components 'x' and 'y'
Traceback:
```

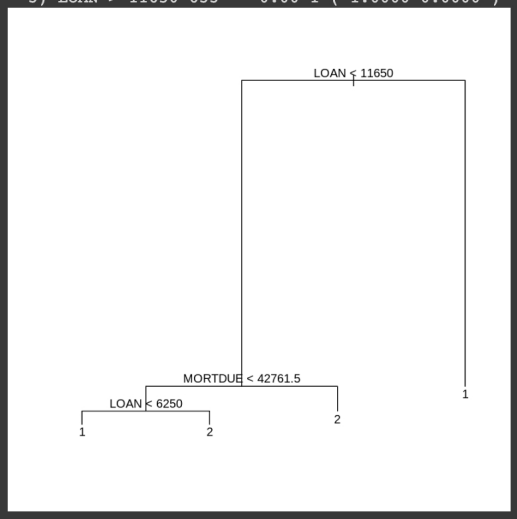
```
1. text(cv.tree1)
2. text.default(cv.tree1)
3. xy.coords(x, y, recycle = TRUE, setLab = FALSE)
4. stop("'x' is a list, but does not have components 'x' and 'y'")
```

SEARCH STACK OVERFLOW




```
[ ] tree1 <- prune.misclass(tree1, best=4)
    plot(tree1);text(tree1)
    summary(tree1)
    print(tree1)
```

```
Classification tree:
snip.tree(tree = tree1, nodes = 5L)
Variables actually used in tree construction:
[1] "LOAN"    "MORTDUE"
Number of terminal nodes:  4
Residual mean deviance:  0.7244 = 1446 / 1996
Misclassification error rate: 0.157 = 314 / 2000
node), split, n, deviance, yval, (yprob)
      * denotes terminal node

1) root 2000 2773.00 1 ( 0.5000 0.5000 )
  2) LOAN < 11650 1367 1590.00 2 ( 0.2685 0.7315 )
    4) MORTDUE < 42761.5 278  385.30 1 ( 0.5108 0.4892 )
      8) LOAN < 6250 81   74.58 1 ( 0.8272 0.1728 ) *
      9) LOAN > 6250 197  261.80 2 ( 0.3807 0.6193 ) *
    5) MORTDUE > 42761.5 1089 1110.00 2 ( 0.2066 0.7934 ) *
  3) LOAN > 11650 633    0.00 1 ( 1.0000 0.0000 ) *
```




```
[ ] tree2 <- predict(tree1, hmeqtest, type="vector")
    print(tree2)
```

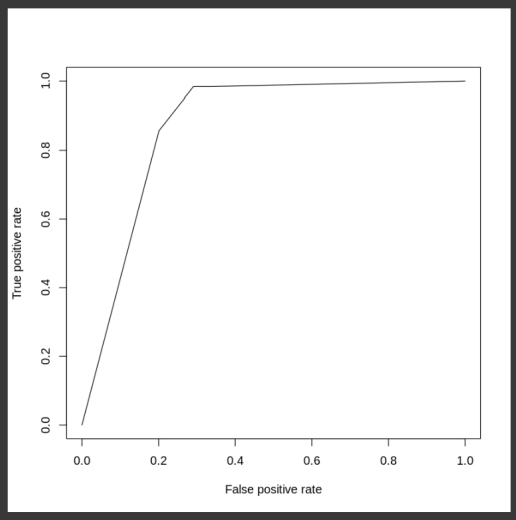
```
321 0.3807107 0.6192893
322 0.2066116 0.7933884
323 0.2066116 0.7933884
324 0.2066116 0.7933884
325 0.2066116 0.7933884
326 0.2066116 0.7933884
327 0.2066116 0.7933884
328 0.3807107 0.6192893
329 0.2066116 0.7933884
330 0.2066116 0.7933884
331 0.2066116 0.7933884
332 0.2066116 0.7933884
333 0.3807107 0.6192893
334 0.2066116 0.7933884
335 0.2066116 0.7933884
336 0.2066116 0.7933884
337 0.3807107 0.6192893
338 0.2066116 0.7933884
339 0.2066116 0.7933884
340 0.2066116 0.7933884
341 0.2066116 0.7933884
342 0.2066116 0.7933884
343 0.2066116 0.7933884
344 0.2066116 0.7933884
345 0.2066116 0.7933884
346 0.2066116 0.7933884
347 0.2066116 0.7933884
348 0.2066116 0.7933884
349 0.2066116 0.7933884
```

```
350 0.3807107 0.6192893
351 0.2066116 0.7933884
352 0.2066116 0.7933884
353 0.2066116 0.7933884
354 0.2066116 0.7933884
355 0.2066116 0.7933884
356 0.2066116 0.7933884
357 0.2066116 0.7933884
358 0.2066116 0.7933884
359 0.2066116 0.7933884
360 0.2066116 0.7933884
361 0.2066116 0.7933884
362 0.2066116 0.7933884
363 0.2066116 0.7933884
364 0.2066116 0.7933884
365 0.2066116 0.7933884
366 0.2066116 0.7933884
367 0.2066116 0.7933884
368 0.3807107 0.6192893
369 0.2066116 0.7933884
370 0.3807107 0.6192893
371 0.2066116 0.7933884
372 0.2066116 0.7933884
373 0.2066116 0.7933884
374 0.2066116 0.7933884
375 0.2066116 0.7933884
376 0.2066116 0.7933884
377 0.2066116 0.7933884
378 0.2066116 0.7933884
```

```
[ ] install.packages("ROCR")
    require(ROCR)
```

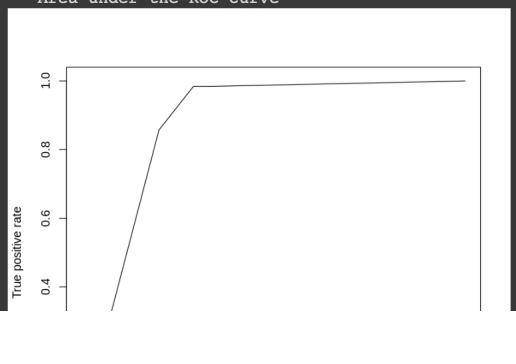
```
Installing package into ‘/usr/local/lib/R/site-library’
(as ‘lib’ is unspecified)

also installing the dependencies ‘bitops’, ‘gtools’, ‘caTools’, ‘gplots’


Loading required package: ROCR
```

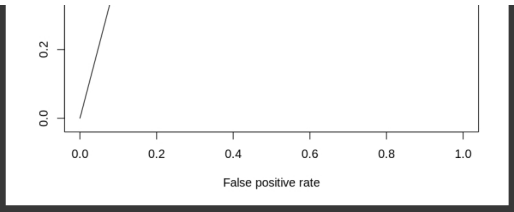
```
preds <- tree2[,2]
plot(performance(prediction(preds,hmeqtest$BAD),'tpr','fpr'))
```




```
[ ] pred <- prediction(preds,hmeqtest$BAD)
    perf <- performance(pred, "tpr","fpr")
    plot(perf)
    auc <- performance(pred, "auc")
    auc
```

```
A performance instance
  'Area under the ROC curve'
```






```
[ ] install.packages("ipred")
    library(ipred)
    library(rpart)
```

```
Installing package into ‘/usr/local/lib/R/site-library’
(as ‘lib’ is unspecified)

also installing the dependencies ‘listenv’, ‘parallelly’, ‘future’, ‘globals’, ‘future.apply’, ‘progressr’, ‘numDeriv’, ‘SQUAREM’, ‘lava’,
```

```
[ ] bagg.hmeq <- bagging(BAD~.-ID, data=hmeqtrain,
                         nbag=1000,
                         control=rpart.control(minsplit=10),
                         coob=T)
    bagg.hmeq
```

```
Bagging classification trees with 1000 bootstrap replications

Call: bagging.data.frame(formula = BAD ~ . - ID, data = hmeqtrain,
    nbag = 1000, control = rpart.control(minsplit = 10), coob = T)

Out-of-bag estimate of misclassification error:  0.158
```

```
[ ] install.packages("caret")
    library(caret)
```

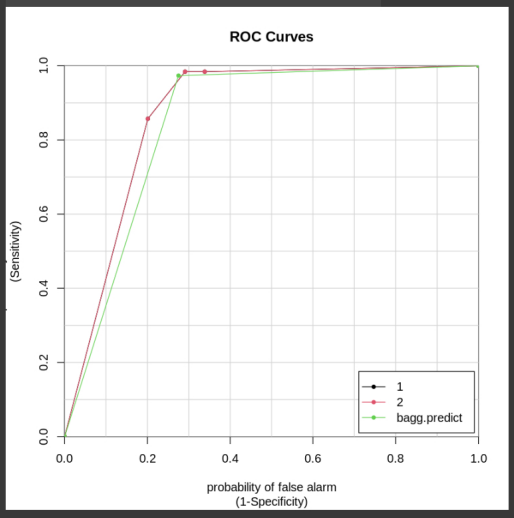
```
Installing package into ‘/usr/local/lib/R/site-library’
(as ‘lib’ is unspecified)

also installing the dependencies ‘proxy’, ‘iterators’, ‘gower’, ‘hardhat’, ‘timeDate’, ‘e1071’, ‘foreach’, ‘ModelMetrics’, ‘pROC’, ‘recipes

Loading required package: ggplot2

Loading required package: lattice

Warning message in system("timedatectl", intern = TRUE):
“running command 'timedatectl' had status 1”
```

```
[ ] bagg.predict <- predict(bagg.hmeq, hmeqtest, type="class")
```

```
[ ] library("caTools")
    colAUC(cbind(tree2, bagg.predict),hmeqtest$BAD, plotROC=T)
```

A matrix: 1 × 3 of type dbl

	1	2	bagg.predict
1 vs. 2	0.8719661	0.8719661	0.8492063




```
[ ] install.packages("ipred")
    library(ipred)
    library(rpart)
```

```
Installing package into '/usr/local/lib/R/site-library'
(as 'lib' is unspecified)
```

```
[ ] bagg.hmeq <- bagging(BAD~.-ID, data=hmeqtrain,
                         nbag=1000,
                         control=rpart.control(minsplit = 10),
                         coob=T)
    bagg.hmeq
```

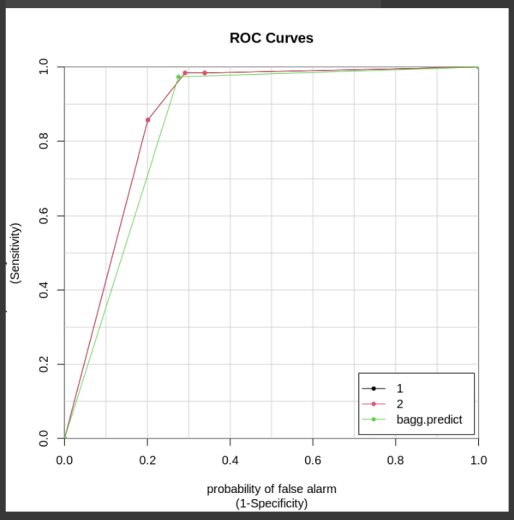
```
Bagging classification trees with 1000 bootstrap replications

Call: bagging.data.frame(formula = BAD ~ . - ID, data = hmeqtrain,
    nbag = 1000, control = rpart.control(minsplit = 10), coob = T)

Out-of-bag estimate of misclassification error:  0.157
```

```
[ ] bagg.predict <- predict(bagg.hmeq,hmeqtest,type="class")
```

```
colAUC(cbind(tree2,bagg.predict),hmeqtest$BAD,plotROC=T)
```

A matrix: 1 × 3 of type dbl

	1	2	bagg.predict
1 vs. 2	0.8719661	0.8719661	0.8492063



✓ 0초 오전 12:44에 완료됨