Homework 4

Charles Hwang

2/27/2020

Charles Hwang

Professor Matthews

STAT 351-001

16 March 2020

Problem 1

```
rm(list=ls())
mat <- matrix(c(4,9,3,0),ncol=2)
rownames(mat) <- c("Low","High")
colnames(mat) <- c("Nearby","Not Nearby")
mat
```

```
##      Nearby Not Nearby
## Low       4          3
## High      9          0
```

```
fisher.test(mat)
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  mat
## p-value = 0.0625
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##  0.000000 1.643973
## sample estimates:
## odds ratio
##          0
```

```
cat("We fail to reject H0 at a = .05. There is insufficient evidence (p = ",fisher.test(mat)$p.value,")
```

```
## We fail to reject H0 at a = .05. There is insufficient evidence (p = 0.0625) that there is a signifi
```

Problem 3

```
rm(list=ls()) # Chi-squared test warnings turned off to reduce clutter
library(gtools)
library(perm)
mat <- matrix(c(4,5,14,7),ncol=2)
rownames(mat) <- c("Made First","Missed First")
```

```
colnames(mat) <- c("Made Second","Missed Second")
mat
```

```
##              Made Second Missed Second
## Made First             4            14
## Missed First           5             7
```

```
mcnemar.test(mat)          # Problem 3a
```

```
##
##  McNemar's Chi-squared test with continuity correction
##
## data:  mat
## McNemar's chi-squared = 3.3684, df = 1, p-value = 0.06646
```

```
cat("We fail to reject H0 at a = .05. There is insufficient evidence (p = ",mcnemar.test(mat)$p.value,"
```

```
## We fail to reject H0 at a = .05. There is insufficient evidence (p = 0.06645742) that there the proba
```

```
chisq.test(mat)$statistic # Problem 3b
```

```
## X-squared
## 0.5357143
```

```
chisq.test(mat)$p.value
```

```
## [1] 0.4642143
```

```
as.data.frame(as.table(mat))
```

```
##           Var1          Var2 Freq
## 1   Made First   Made Second    4
## 2 Missed First   Made Second    5
## 3   Made First Missed Second   14
## 4 Missed First Missed Second    7
```

```
cat("This p-value (p = ",chisq.test(mat)$p.value,") is greater than the p-value for McNemar's test (p =
```

```
## This p-value (p = 0.4642143) is greater than the p-value for McNemar's test (p = 0.06645742). We fail
```

Problem 4

```
rm(list=ls())
library(boot)
library(tree)
summary(nuclear)
```

```
##       cost            date             t1              t2
##  Min.   :207.5   Min.   :67.17   Min.   : 7.00   Min.   :44.00
##  1st Qu.:310.3   1st Qu.:67.90   1st Qu.:11.75   1st Qu.:56.50
##  Median :448.1   Median :68.42   Median :13.00   Median :62.50
##  Mean   :461.6   Mean   :68.58   Mean   :13.75   Mean   :62.38
##  3rd Qu.:612.0   3rd Qu.:68.92   3rd Qu.:15.25   3rd Qu.:70.25
##  Max.   :881.2   Max.   :71.08   Max.   :22.00   Max.   :85.00
##       cap               pr               ne              ct
##  Min.   : 457.0   Min.   :0.0000   Min.   :0.00   Min.   :0.0000
##  1st Qu.: 745.0   1st Qu.:0.0000   1st Qu.:0.00   1st Qu.:0.0000
##  Median : 822.0   Median :0.0000   Median :0.00   Median :0.0000
```

```
##  Mean   : 825.4   Mean   :0.3125   Mean   :0.25   Mean   :0.4062
##  3rd Qu.: 947.2   3rd Qu.:1.0000   3rd Qu.:0.25   3rd Qu.:1.0000
##  Max.   :1130.0   Max.   :1.0000   Max.   :1.00   Max.   :1.0000
##        bw             cum.n             pt
##  Min.   :0.0000   Min.   : 1.000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.: 3.000   1st Qu.:0.0000
##  Median :0.0000   Median : 7.500   Median :0.0000
##  Mean   :0.1875   Mean   : 8.531   Mean   :0.1875
##  3rd Qu.:0.0000   3rd Qu.:12.500   3rd Qu.:0.0000
##  Max.   :1.0000   Max.   :21.000   Max.   :1.0000
```

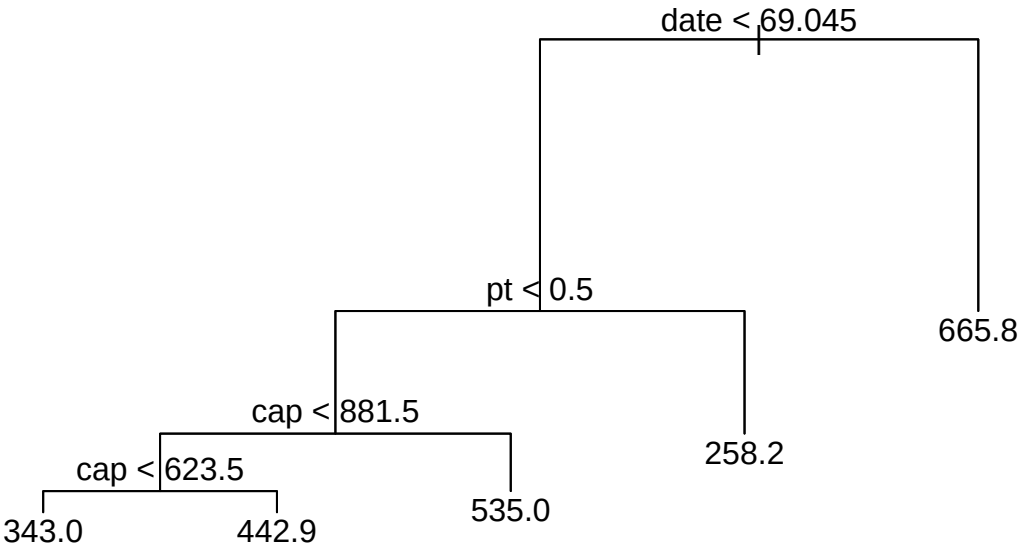
```
help(nuclear)
T <- tree(cost~.,data=nuclear)
T
```

```
## node), split, n, deviance, yval
##       * denotes terminal node
##
##  1) root 32 897200 461.6
##    2) date < 69.045 25 431800 404.4
##      4) pt < 0.5 19 256600 450.5
##        8) cap < 881.5 12 132500 401.3
##         16) cap < 623.5 5   8360 343.0 *
##         17) cap > 623.5 7  95090 442.9 *
##        9) cap > 881.5 7  44990 535.0 *
##      5) pt > 0.5 6   6512 258.2 *
##    3) date > 69.045 7  91700 665.8 *
```

```
predict(T)
```

```
##        1        2        3        4        5        6        7        8
## 442.8571 535.0014 535.0014 535.0014 535.0014 343.0140 442.8571 343.0140
##        9       10       11       12       13       14       15       16
## 442.8571 442.8571 343.0140 442.8571 343.0140 535.0014 442.8571 442.8571
##       17       18       19       20       21       22       23       24
## 665.7900 343.0140 665.7900 535.0014 535.0014 665.7900 665.7900 665.7900
##       25       26       27       28       29       30       31       32
## 665.7900 665.7900 258.2200 258.2200 258.2200 258.2200 258.2200 258.2200
```

```
plot(T)
text(T)
```

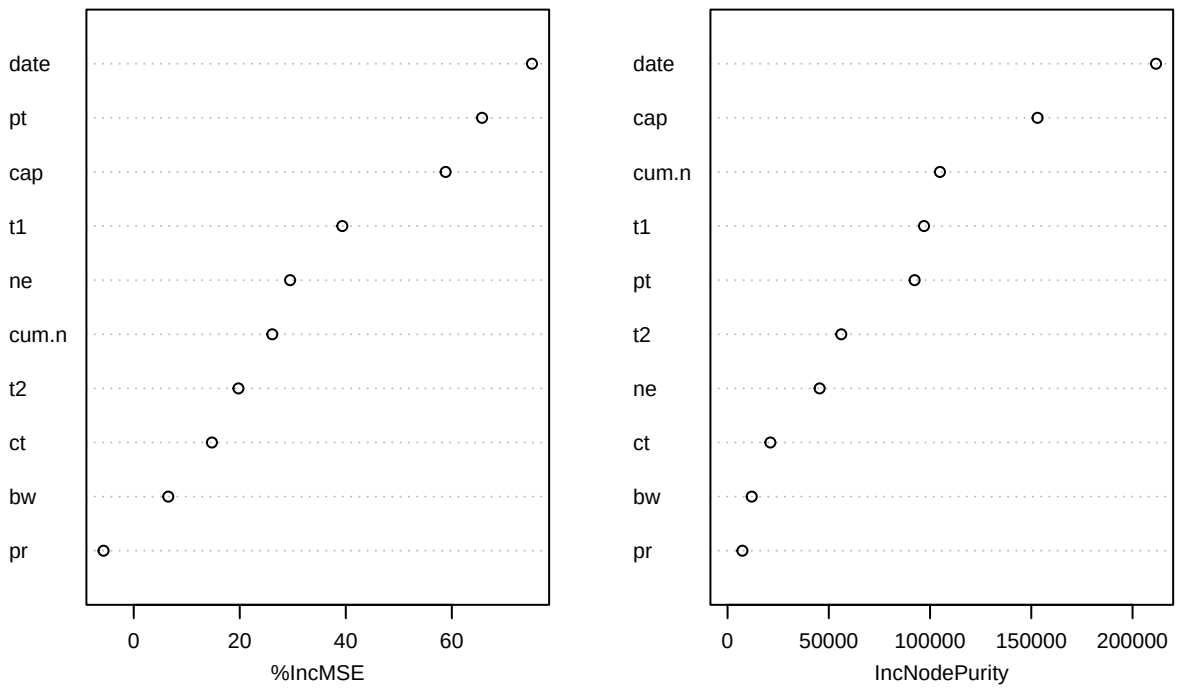


Problem 5

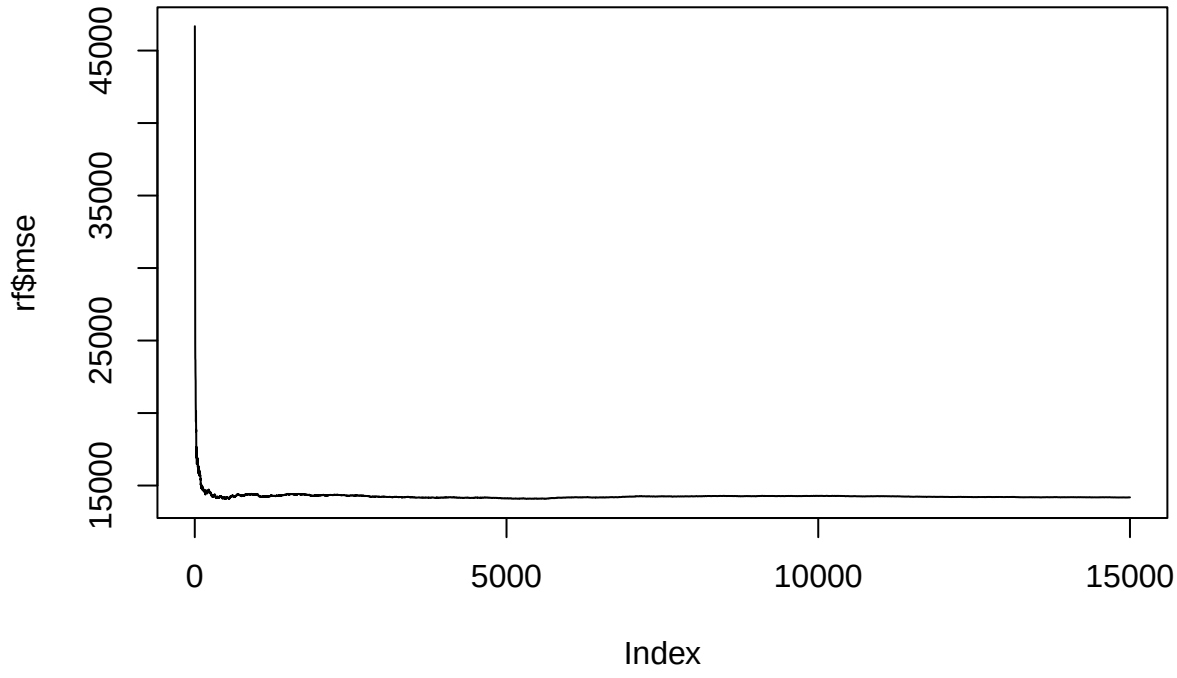
```
library(randomForest)
set.seed(1603)
rf <- randomForest(cost~.,data=nuclear,ntree=15000,importance=TRUE)
predict(rf,nuclear)
```

```
##        1        2        3        4        5        6        7        8
## 461.1838 464.8139 459.2161 595.5055 591.0104 392.0829 354.7571 371.8326
##        9       10       11       12       13       14       15       16
## 433.7896 552.9182 370.0421 433.6155 400.3616 495.1349 458.2751 414.8454
##       17       18       19       20       21       22       23       24
## 644.3733 386.7744 693.5187 500.9013 534.0825 646.5347 596.0820 592.4311
##       25       26       27       28       29       30       31       32
## 541.3617 676.9768 249.4214 297.1414 288.6339 291.2985 249.2143 293.5640
```

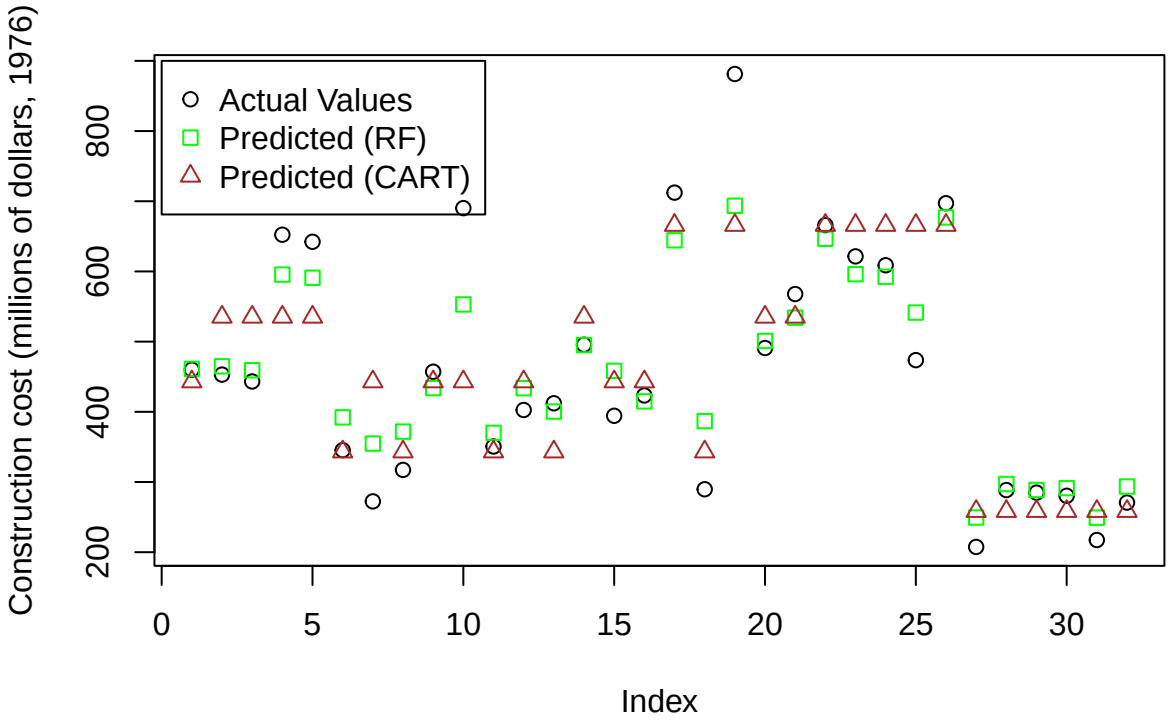
```
varImpPlot(rf,main="Variable Importance Plot",cex=.7) # Adding some graphs to visualize the random fore
```




```
plot(rf$mse,type="l")
```




```
plot(nuclear$cost,ylab="Construction cost (millions of dollars, 1976)")
points(predict(rf,nuclear),pch=0,col="green")
points(predict(T),pch=2,col="brown") # Comparing CART model with actual values and random forest
legend(0,900,c("Actual Values","Predicted (RF)","Predicted (CART)"),col=c("black","green","brown"),pch=
```


Actual Values
Predicted (RF)
Predicted (CART)
Construction cost (millions of dollars, 1976)
200
400
600
800
0
5
10
15
20
25
30
Index