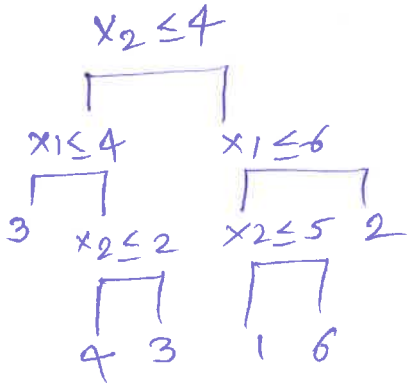
Anusha muppalla
16286311

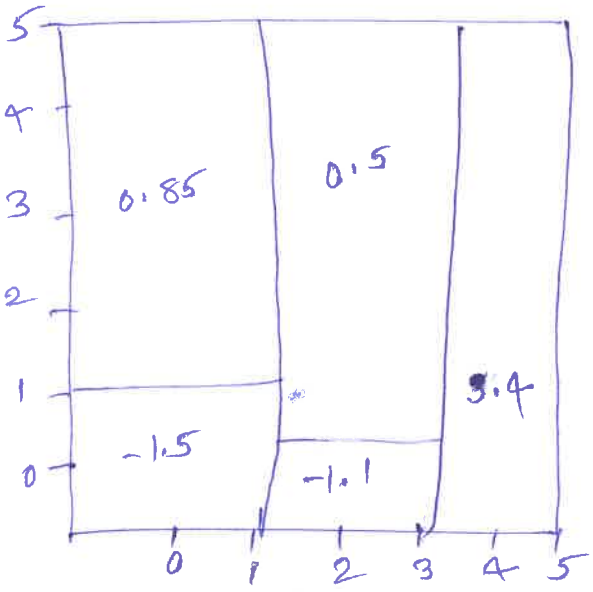
HW-7

1)

a) from given,



b) from given,



3) a
By observing above, majority vote approach × (RED) as it is the most commonly occuring class among 10 predictions, since there are more predictions that are ~~>~~ 0.5 than there are estimates that are ~~>~~ > 0.5 (0.5, 0.8, 0.85, 0.9) i.e 6 for Red vs 4 for Green

the avg probability approach takes the average of the 10 estimates. the average from this set of estimates is 0.55, resulting in a final classification of Green since the average estimate is > 0.5

4)

step1 :- Age :- age of each participant

step2 :- F=0 :- mean of all ages

$$F_0 = 35.36$$

step3 :- pseudo Residual 0 = Age - F0

ID	Age	F0	Pseudo Residual 0
1	12	35.36	-23.36
2	13	35.36	-22.36
3	14	35.36	-21.36
4	16	35.36	-19.36
5	20	35.36	-15.36
6	24	35.36	-11.36
7	32	35.36	-3.36
8	45	35.36	9.64
9	65	35.36	29.64
10	73	35.36	37.64
11	75	35.36	39.64

step4:- h_0 = mean of all ages in each leaf.

Root

{-23.36, -22.36, -21.36, -19.36, -15.36, -11.36, -3.36, 9.64, 29.64, 37.64, 39.64}

product A == N (F)

{-23.36, -22.36, -21.36, -15.36, -3.36}

product A == Y (T)

{-19.36, -11.36, 9.64, 29.64, 37.64, 39.64}

↓

mean value for each leaf ⇒ -17.16

↓

mean value for each leaf ⇒ 14.30

ID	AGE	F_0	Pseudo Residual 0	f_{10}
1	12	35.36	-23.36	-17.16
2	13	35.36	-22.36	-17.16
3	14	35.36	-21.36	-17.16
4	16	35.36	-19.36	14.30
5	20	35.36	-15.36	-17.16
6	24	35.36	-11.36	14.30
7	32	35.36	-3.36	-17.16
8	45	35.36	9.64	14.30
9	65	35.36	29.64	14.30
10	73	35.36	37.64	14.30
11	75		39.64	14.30

steps:- gamma 0 = 1 (by default)

step6:- F1 = F0 + h0

ID	AGE	F0	pseudo Residual 0	h0	gamma0	F1
1	12	35.36	-23.36	-17.16	1	18.2
2	13	35.36	-22.36	-17.16	1	18.2
3	14	35.36	-21.36	-17.16	1	18.2
4	16	35.36	-19.36	14.30	1	21.06
5	20	35.36	-15.36	-17.16	1	18.2
6	24	35.36	-11.36	14.30	1	21.06
7	32	35.36	-3.36	-17.16	1	18.2
8	45	35.36	9.64	14.30	1	21.06
9	65	35.36	29.64	14.30	1	21.06
10	73	35.36	37.64	14.30	1	21.06
11	75	35.36	39.64			

step7:- pseudo Residual 1 = Pesudo Residual 0 - h0

pseudo Residual1
-6.2
-5.2
-4.2
-33.66
1.8
-25.66
13.8
-4.66
15.34
23.34
25.34

step8:- h_1 = mean of all ages in each leaf

{ -6.2, -5.2, -4.2, -33.66, 1.8, -25.66, 13.8, -4.66, 15.34, 23.34, 25.34 }

product B == N(F)

{ -33.66, -25.66, 13.8, -4.66, 23.34, 25.34 }

mean = -0.25

product B == Y(T)

{ -6.2, -5.2, -4.2, 1.8, 15.34 }

mean = 0.308

step9:- gamma 1 = 1

∴

ID	F1	Pseudo Residual 1	h1	gamma 1
1	18.2	-6.2	0.308	1
2	18.2	-5.2	0.308	1
3	18.2	-4.2	0.308	1
4	18.2	-33.66	-0.25	1
5	18.2	1.8	0.308	
6	18.2	-25.66	-0.25	1
7	18.2	13.8	-0.25	1
8	18.2	-4.66	-0.25	1
9	18.2	15.34	0.38	1
10	18.2	23.34	-0.25	1
11	18.2	25.34	-0.25	1

step 10:- F2 = F1 + h1

ID	AGE	F0	PR0	h0	gamma0	F1	PR1	h1	gamma1	F2
1	12	35.36	-23.36	-17.16	1	15.2	-6.2	0.308	1	18.508
2	13	35.36	-22.36	-17.16	1	18.2	-5.2	0.308	1	18.508
3	14	35.36	-21.36	-17.16	1	18.2	-4.2	0.308	1	18.508
4	16	35.36	-19.36	14.30	1	21.06	-33.66	+0.25	1	20.81
5	20	35.36	-15.36	-17.16	1	18.2	1.8	0.38	1	18.508
6	24	35.36	-11.36	14.30	1	21.06	-25.66	-0.25	1	20.81
7	32	35.36	-3.36	-17.16	1	18.2	13.8	-0.25	1	20.81
8	45	35.36	9.64	14.30	1	21.06	-4.66	-0.25	1	20.81
9	65	35.36	29.64	14.30	1	21.06	15.34	0.38	1	18.508
10	73	35.36	37.64	14.30	1	21.06	23.34	-0.25	1	20.81
11	75	35.36	39.64	14.30	1	21.06	25.34	-0.25	1	20.81

(2)

Consider the Gini index, classification error, and entropy in a simple classification setting with two classes. Create a single plot that displays each of these quantities as a function of ˆpm1. The x-axis should display ˆpm1, ranging from 0 to 1, and the y-axis should display the value of the Gini index, classification error, and entropy. (Note: you may use R to make this plot

```
pm <- seq(0, 1, 0.001)

# Classification error
class_error <- 1 - pmax(pm, 1-pm)

# Gini index
gini_index <- 2*pm*(1-pm)

# Cross entropy
cross_entropy <- -((pm*log(pm))+((1-pm)*log(1-pm)))

matplot(pm, cbind(class_error, gini_index, cross_entropy), col=c("black", "green", "orange
legend("topright",pch=19,col=c("black", "green", "orange"), legend=c("Classification_Error
```

Console Terminal Jobs

```
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

  Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[Workspace loaded from ~/.RData]

> pm <- seq(0, 1, 0.001)
> # Classification error
> class_error <- 1 - pmax(pm, 1-pm)
> # Gini index
> gini_index <- 2*pm*(1-pm)
> # Cross entropy
> cross_entropy <- -((pm*log(pm))+((1-pm)*log(1-pm)))
> matplot(pm, cbind(class_error, gini_index, cross_entropy), col=c("black", "green", "orange"),
 ylab = "Split Criterion")
> legend("topright",pch=19,col=c("black", "green", "orange"), legend=c("Classification_Error_Rat
e","Gini_Index","Cross-entropy"))
>
```