* Problem 1:-

① Depth 1 decision tree:-

$$Gain(S, F) = Entropy(S) - \sum_{v \in Values(F)} \frac{|Sv|}{|S|} Entropy(Sv)$$

$$Entropy(decision) = P_+ \log_2 P_+ + P_- \log_2 P_-$$

1- Entropy for the dataset
   (+8, -6)

$$E(s) = -\left(\frac{8}{14}\right) \log_2 \frac{8}{14} - \left(\frac{6}{14}\right) \log_2\left(\frac{6}{14}\right) = 0.985$$

2- Attribute :→ Early registration (early, late)

$$- S early [+4, -2] → E(s)_{early} = -\frac{4}{6} \log_2\left(\frac{4}{6}\right) - \frac{2}{6} \log_2\left(\frac{2}{6}\right) = 0.918$$

$$- S late [+4, -4] → E(s)_{late} = -\frac{4}{8} \log_2\left(\frac{4}{8}\right) - \frac{4}{8} \log_2\left(\frac{4}{8}\right) = 1.0$$

$$G Gain(S, Early registration) = Entropy(s) - \sum_{v \in (Early, late)} \frac{|Sv|}{S} Entropy(Sv)$$

$$G Gain(S, Early registration) = 0.985 - \left(\frac{6}{14} * 0.918 + \frac{8}{14} * 1\right)$$

$$= 0.02 → ①$$

3. Attribute :→ Finished homework (Complete, unComplete)

$$- S Complete [+5, -2] → E(s)_{Complete} = -\frac{5}{7} \log_2\left(\frac{5}{7}\right) - \frac{2}{7} \log_2\left(\frac{2}{7}\right) = 0.863$$

$$- S unComplete [+3, -4] → E(s)_{unComplete} = -\frac{3}{7} \log_2\left(\frac{3}{7}\right) - \frac{4}{7} \log_2\left(\frac{4}{7}\right) = 0.985$$

$$G Gain(s, Finished homework) = 0.985 - \left(0.863 * \frac{7}{14} + 0.985 * \frac{7}{14}\right) = 0.061$$

4- Attribute $\rightarrow$ Senior (Senior, nonsenior)

$\rightarrow S_{Senior}$ $(+5,-3) \rightarrow E(S)_{Senior} = -\frac{5}{8}\log_2\left(\frac{5}{8}\right) - \frac{3}{8}\log_2\left(\frac{3}{8}\right) = 0.954$

$\rightarrow S_{nonsenior}$ $(+3,-3) \rightarrow E(S)_{nonsenior} = -\frac{3}{8}\log_2\left(\frac{3}{8}\right) - \frac{3}{8}\log_2\left(\frac{3}{8}\right) = 1$

$\rightarrow$ Gain$(S, Senior) = 0.985 - \left(1 * \frac{6}{14} + 0.954 * \frac{8}{14}\right) = 0.011$

5- Attribute $\rightarrow$ likes Coffee (yes, No)

$\rightarrow S_{yes Coffe}$ $(+3,-1) \rightarrow E(S)_{yes} = -\frac{3}{4}\log_2\left(\frac{3}{4}\right) - \frac{1}{4}\log_2\left(\frac{1}{4}\right) = 0.811$

$\rightarrow S_{No Coffee}$ $(+5,-5) \rightarrow E(S)_{No} = -\frac{5}{10}\log_2\left(\frac{5}{10}\right) - \frac{5}{10}\log_2\left(\frac{5}{10}\right) = 1$

$\rightarrow$ Gain$(S, likes Coffee) = 0.985 - \left(1 * \frac{10}{14} + \frac{4}{14} * 0.811\right) = 0.039$

6- Attribute $\rightarrow$ liked the last homework (liked, unliked)

$\rightarrow S_{liked}$ $(+5, -4) \rightarrow E(S)_{liked} = -\frac{5}{9}\log_2\left(\frac{5}{9}\right) - \frac{4}{9}\log_2\left(\frac{4}{9}\right) = 0.99$

$\rightarrow S_{unliked}$ $(+3, -2) \rightarrow E(S)_{unliked} = -\frac{3}{5}\log_2\left(\frac{3}{5}\right) - \frac{2}{5}\log_2\left(\frac{2}{5}\right) = 0.97$

$\rightarrow$ Gain$(S, liked the last homework) = 0.985 - \left(0.99 * \frac{9}{14} + \frac{5}{14} * 0.97\right)$
$= 0.002$

$\therefore$ The root has the maximum gain

$\because$ Gain$(S, Early registration) = 0.02$
Gain$(S, Finished homework) = 0.061$
Gain$(S, Senior) = 0.011$
Gain$(S, likes Coffee) = 0.039$
Gain$(S, liked the last homework) = 0.002$

∴ The root is the finished homework

↳ finished homework

Yes → [+5, -2]

No → [+3, -4]

finished homework

finished ————— unfinished

[+5, -2]   [?]   [?]   [+3, -4]
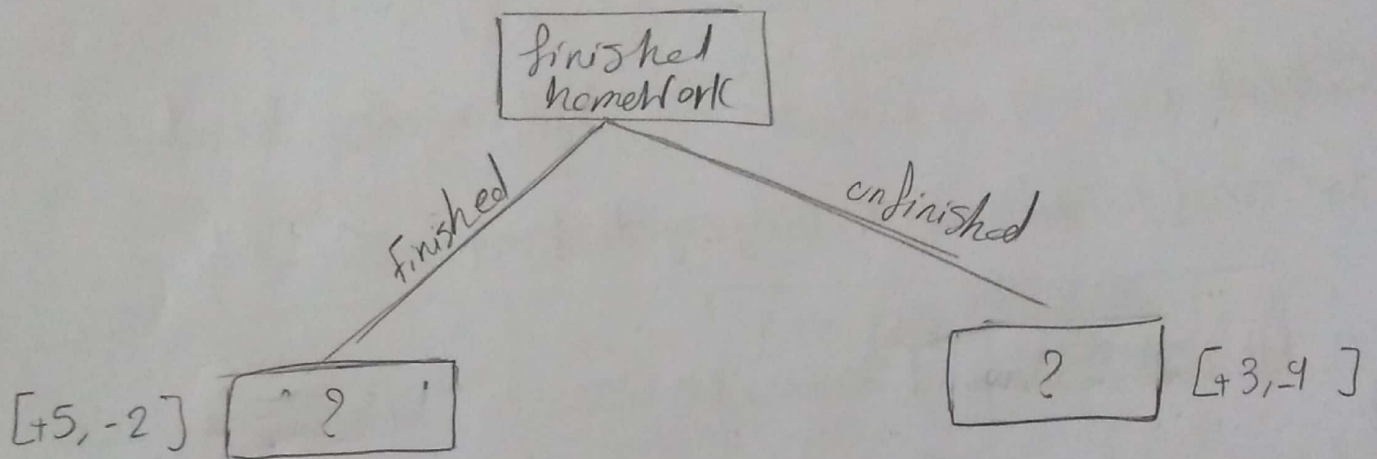
finished homework	Early registration	Senior	Likes Coffee	liked the last homework	A
1	1	0	0	1	1
1	1	1	0	1	1
1	0	1	0	1	0
1	0	1	0	0	1
1	0	0	1	1	1
1	1	1	0	0	1
1	0	1	1	1	0

① Entropy of finished homework (+5, -2)

$$E(S)_{finished homework} = -\frac{5}{7} log_2\left(\frac{5}{7}\right) - \frac{2}{7} log_2\left(\frac{2}{7}\right)$$

$$= 0.86$$

② Attribute: Early registration (early, late)

↳ S_{Early} (+3, -0) → $E(S)_{early} = -\frac{3}{3} log_2\left(\frac{3}{3}\right) = 0$

↳ S_{late} (+2, -2) → $E(S)_{late} = -\frac{2}{4} log_2\left(\frac{2}{4}\right) - \frac{2}{4} log_2\left(\frac{2}{4}\right) = 1$

↳ Gain (finished homework, early registration) $= 0.86 - 1 * \frac{4}{7}$

$$= 0.288$$

③ Attribute: Senior (Senior, Non senior)

↳ S_{senior} (+3, -2) ⟶ $E(S)_{senior} = -\frac{3}{5} log_2 \frac{3}{5} - \frac{2}{5} log_2\left(\frac{2}{5}\right) = 0.97$

↳ $S_{Nonsenior}$ (+2, -0) ⟶ $E(S)_{Nonsenior} = 0$

↳ Gain (finished homework, Senior) $= 0.86 - 0.97 * \frac{5}{7} = 0.167$

④ Attribute: likes coffee (Yes, No)

↳ $S_{yes coffee}$ (+1, -1) ⟹ $E(S)_{yes coffee} = -\frac{1}{2} log_2 \frac{1}{2} - \frac{1}{2} log_2 \frac{1}{2} = 1$

↳ $S_{No coffee}$ (+4, -1) ⟹ $E(S)_{No coffee} = -\frac{4}{5} log_2 \frac{4}{5} - \frac{1}{5} log_2 \frac{1}{5} = 0.72$

↳ Gain (finished homework, coffee) $= 0.86 - \left(1 * \frac{2}{7} + 0.72 * \frac{5}{7}\right)$

$$= 0.06$$

⑤ Attribute: Liked the homework (Liked, Unliked)

↳ S liked (+3, -2) → $E(s)_{liked} = -\frac{3}{5} \log_2 \frac{3}{5} - \frac{2}{5} \log_2 (\frac{2}{5}) = 0.97$

↳ S unliked (+2, 0) → $E(s)_{unliked} = 0$

↳ Gain (finished homework, liked the last homework)

$$= 0.86 - (0.97 * \frac{5}{7}) = 0.167$$

∴ The Node from the finished homework is Early registration

unfinished homework	Early registration	Senior	Likes coffee	Liked the last homework	A
0	0	1	0	0	0
0	0	1	1	1	1
0	1	0	0	1	0
0	0	1	0	1	1
0	1	0	0	0	0
0	0	0	0	1	0
0	1	0	1	0	1

① Entropy of unfinished homework (+3, -4)

$$E(s)_{unfinished} = -\frac{3}{7} \log_2 \frac{3}{7} - \frac{4}{7} \log_2 \frac{4}{7} = 0.985$$

② Attribute : Early registration (early, late)

$S_{early}(+1,-2) \rightarrow E(S)_{early} = -\frac{1}{3}\log_2\frac{1}{3} - \frac{2}{3}\log_2\frac{2}{3} = 0.918$

$S_{late}(+2,-2) \rightarrow E(S)_{late} = 1$

G Gain (unfinished homework, early registration)

$= 0.985 - (1 * \frac{4}{7} + \frac{3}{7} * 0.918) = 0.02$

③ Attribute: Senior (Senior, Non senior)

$S_{Senior}(+2, -1) \rightarrow E(S)_{Senior} = -\frac{2}{3}\log_2\frac{2}{3} - \frac{1}{3}\log\frac{1}{3} = 0.918$

$S_{Nonsenior}(+1, -3) \rightarrow E(S)_{Nonsenior} = -\frac{1}{4}\log\frac{1}{4} - \frac{3}{4}\log\frac{3}{4} = 0.811$

G Gain (unfinished homework, Senior) $= 0.985 - (0.811 * \frac{4}{7} + \frac{3}{7} * 0.918)$

$= 0.128$

④ Attribute : Likes Coffee (yes, No)

$S_{yes}(+2/-0) \rightarrow E(S)_{yes} = -\frac{2}{2}\log\frac{2}{2} = 0$

$S_{No}(+1,-4) \rightarrow E(S)_{No} = -\frac{1}{5}\log_2\frac{1}{5} - \frac{4}{5}\log_2\frac{4}{5} = 0.72$

G Gain (unfinished homework, likes Coffee)

$= 0.985 - (0.72 * \frac{5}{7}) = 0.47$

⑤ Attribute: liked the last homework (liked, unliked)

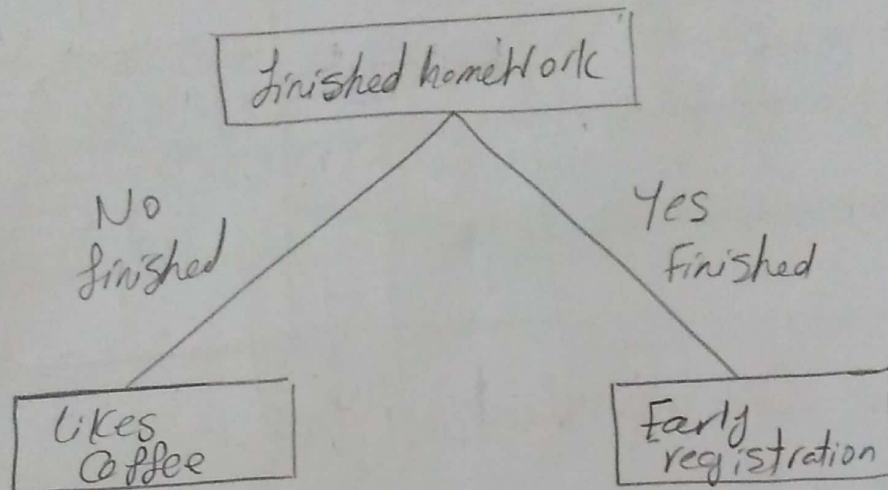$S_{liked}(+2,-2) \rightarrow E_{liked} \rightarrow = 1$

$S_{unliked}(+1,-2) \rightarrow E_{unliked} \rightarrow -\frac{1}{3}\log_2\frac{1}{3} - \frac{2}{3}\log_2\frac{2}{3}$

$= 0.918$

G Gain (unfinished homework, liked the last) $= 0.985 - (0.918 * \frac{3}{7} + \frac{4}{7} * 1) = 0.02$

∴ The node of the unfinished homework is likes coffee

∴ The first depth decision tree is :-

Finished homework
— No finished → Likes Coffee
— Yes Finished → Early registration

2 The depth 2 :-

↳ Early registration (yes)

Early registration	finished homework	Senior	Likes Coffee	Lost homework	A
1	1	0	0	1	1
1	1	1	0	1	1
1	1	1	0	0	1

$$E(early) = -\frac{3}{3} \log_2 \frac{3}{3} = 0$$

∴ from the entropy and the table → early registra
and finished homework gives class A = 1

Early registration	finished homeWork	Senior	lost homeWork	A
O	1	1	1	O
O	1	1	O	1
O	1	O	1	1
O	1	1	1	O

① Entropy of late registration (+2, -2)

$\therefore E(s) = 1$

② Attribute: Senior (Senior, Non senior)

$\delta Senior (+1, -2) = -\frac{1}{3} \log_2 \frac{1}{3} - \frac{2}{3} \log_2 \frac{2}{3} = 0.918$

$\delta Non senior (+1, -0) \Longrightarrow E_{Non senior} = 6$

↳ Gain (finished homeWork + early registration, Senior)

$= 1 - (0.918 * \frac{3}{4}) = 0.311$

③ Attribute: last homeWork (liked, unliked)

$\delta liked (+1, -2) \rightarrow E(s) = -\frac{1}{3} \log_2 \frac{1}{3} - \frac{2}{3} \log_2 \frac{2}{3} = 0.918$

$\delta unliked (+1, -0) \longrightarrow E(s) = 0$

↳ Gain (finished + late, last homeWork) $= 1 - (0.918 * \frac{3}{4}$

$1 = (0.918 * \frac{3}{4}) = 0.311$

∵ Information gain for both (Senior and lasthomeWork) is the same → ∴ We Can Choose any of them to be the child node → Implementation Code choose the first feature Comes → ∴ let the Senior be the Child node →

if yes $E_{lasthome} = 0.918 - (\frac{2}{3} \times 0 + \frac{1}{3} \times 0) = 0.918$

if No $E_{unliked} = 0.918 - 0$ = 0.918

$E_{unlikedfast} = 0 \rightarrow class\ A = 1$

homeWork (1,0)

* Likes Coffee Child nodes :-

① Yes

unfinished homeWork	likes Coffee	last homeWork	A
0	1	↑	↓
0	1	0	1

Entropy Likescoffee = 0

∴ from entropy and table, the Node of likes Coffee

is class A

② No

unfinished homeWork	unlike Coffee	last homeWork	A
0	0	0	0
0	0	1	0
0	0	0	0
0	0	1	0

⟹ The Child node is class A = 0

* The Decision tree

Finished homework
$E(S) = 0.86$

No (unfinished)

Yes (finished)

Likes Coffee

Early registration

No

Yes

No

Yes

Not A
$(A=0)$

$A=1$

Senior

$A=1$

No

Yes

$A=1$