"Assignment #3"

"Decision Trees"

problem # 1

1. Decide the root :-

[1] Early Restination

$S = [+8, -6]$

$Entropy(S) = \frac{-8}{14}\log_2\frac{8}{14} - \frac{6}{14}\log_2\frac{6}{14} = 0.985$

$S_{(0)} = [-4, +4]$ → equal

$= Entropy[S_0] = \frac{-4}{8}\log_2\frac{4}{8} - \frac{4}{8}\log_2\frac{4}{8} = 1$

$S_{(1)} = [+4, -2]$

$Entropy[S_1] = \frac{-4}{6}\log_2\frac{4}{6} - \frac{2}{6}\log_2\frac{2}{6} = 0.918$

$gain(S, Early\ registination)$
$= Entropy(S) - \sum \frac{|S_v|}{|S|} Entropy(S_v)$

$gain = 0.985 - \frac{8}{14}*1 - \frac{6}{14}*0.918 = 0.02$

[2] Finished home work

$S = [+8, -6]$

$Entropy(S) = \frac{-8}{14}\log\frac{8}{14} - \frac{6}{14}\log\frac{6}{14} = 0.985$

$S_{(0)} = [-4, +3]$

$Entropy(S) = \frac{-4}{7}\log\frac{4}{7} - \frac{3}{7}\log\frac{3}{7} = 0.985$

$S_{(1)} = [+5, -2]$

$Entropy(S_1) = \frac{-5}{7}\log_2\frac{5}{7} - \frac{2}{7}\log_2\frac{2}{7} = 0.863$

$gain = 0.985 - \frac{7}{14}*0.985 - \frac{7}{14}*0.863 = 0.061$

[3] Senior

$S = [+8, -6]$

$Entropy(S) = \frac{-8}{14}\log\frac{8}{14} - \frac{6}{14}\log\frac{6}{14} = 0.985$

$S_{(0)} = [-3, +3]$ → equal

$Entropy(S_0) = 1$

$S_{(1)} = [+5, -3]$

$Entropy(S_1) = \frac{-5}{8}\log\frac{5}{8} - \frac{3}{8}\log\frac{3}{8} = 0.954$

$gain(S, Senior) = 0.985 - \frac{6}{14}*1 - \frac{8}{14}*0.954$
$= 0.01128$

[4] Likes Cafee

$S = [+8, -6]$

$Entropy(S) = \frac{-8}{14}\log\frac{8}{14} - \frac{6}{14}\log\frac{6}{14} = 0.985$

$S_{(0)} = [+5, -5]$ → equal

$Entropy(S) = 1$

$S_{(1)} = [+3, -1]$

$Entropy(S_1) = \frac{-3}{4}\log\frac{3}{4} - \frac{1}{4}\log\frac{1}{4} = 0.811$

$gain(S, likes\ cafe) = 0.985 - \frac{10}{14}*1 - \frac{4}{14}*0.811$
$= 0.039$

5. Liked the last home work :-

$S_0 = [+3, -2]$

Entropy $(S_0) = -\frac{3}{5}\log_2\frac{3}{5} - \frac{2}{5}\log_2\frac{2}{5}$
$= 0.97$

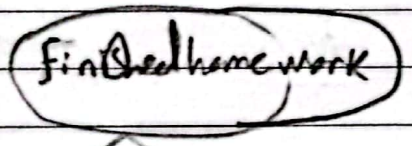
$S_1 = [+5, -4]$

Entropy $(S_1) = -\frac{5}{9}\log_2\frac{5}{9} - \frac{4}{9}\log_2\frac{4}{9} = 0.99$

* gain (S, liked last home work)

$= 0.985 - \frac{5}{14} * 0.97 - \frac{9}{14} * 0.99 =$
0.00214

Finished home work ~~Early registeration~~ has the most informative gain ∴ root.



1, 2, 4, 5, 8, 11, 12 — 3, 6, 7, 9, 10, 13, 14

at ① Finished home work

$S = [5, -2]$

Entropy $(S) = -\frac{5}{7}\log_2\frac{5}{7} - \frac{2}{7}\log_2\frac{2}{7} = 0.863$

Early registeration

$S = [+3, 0]$

S_0 (1)

Entropy $(S_1) = -\frac{3}{3}\log_2\frac{3}{3} - 0 = 0$

R

$S_{(0)} = [2, -2]$

Entropy $(S_0) = -\frac{2}{4}\log_2\frac{2}{4} - \frac{2}{4}\log_2\frac{2}{4} = 1$

gain (Finished, Early reg.)

$= 0.863 - \frac{3}{7} * 0 - \frac{4}{7} * 1 = 0.292$

* Seniors $[S_0, S_1]$

$S_0 [+2, 0]$

Entropy $= 0$

$S_1 [+3, -2]$

Entropy $= -\frac{3}{5}\log_2\frac{3}{5} - \frac{2}{5}\log_2\frac{2}{5} = 0.971$

gain (1, Senior) $= 0.863 - \frac{5}{7} * 0.971$
$- 0 = 0.169$

* likes Coffe

$S_0 = [+4, -1]$

Entropy $= -\frac{4}{5}\log_2\frac{4}{5} - \frac{1}{5}\log_2\frac{1}{5} = 0.722$

$S_1 [1, -1]$ → equal

Entropy $(S) = 1$

gain (1, likes coffe)

$= 0.863 - \frac{5}{7} * 0.722 - \frac{2}{7} * 1 = 0.062$

* liked last home work

$S_0 [0, +2]$

Entropy $(S_1 = 0$

$S_1 [3, -2]$

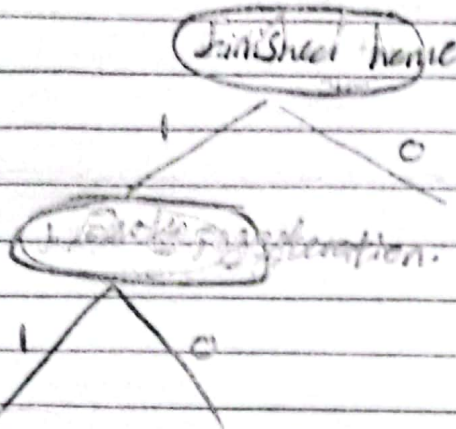
Entropy $(S_1) = 0.971$

gain (1, liked) $= 0.863 - \frac{5}{7} * 0.971 - 0$

$= 0.169.$

Early registeration.

→ ~~Finished last homework~~ has higher information gain.



all gives A → 1

$S = [+3, 0^{-}]$

Entropy(S) = 0

(a) [0] branch.

$S[Early, 0] = [+2, -2]$ equal

Entropy = 1

Senior

S_0 [1, 0]

∴ Entropy = 0

S_1 [+1, −2]

$Entropy = -\frac{1}{3}\log_2\frac{1}{3} - \frac{2}{3}\log_2\frac{2}{3} = 0.918$

Entropy (Early(0), Senior) =

$1 - \frac{3}{4} \times 0.918 - 0 = 0.312$

Liked last homework

$S_0 = [1, 0]$ ∴ Entropy = 0

$S_1 = [1, -2]$ ∴ entropy = 0.918

$gain = 1 - \frac{3}{4} \times 0.918 - \frac{1}{4} \times 0 = 0.312$

Dikes offc

$S_0 = [1, 0]$ equal

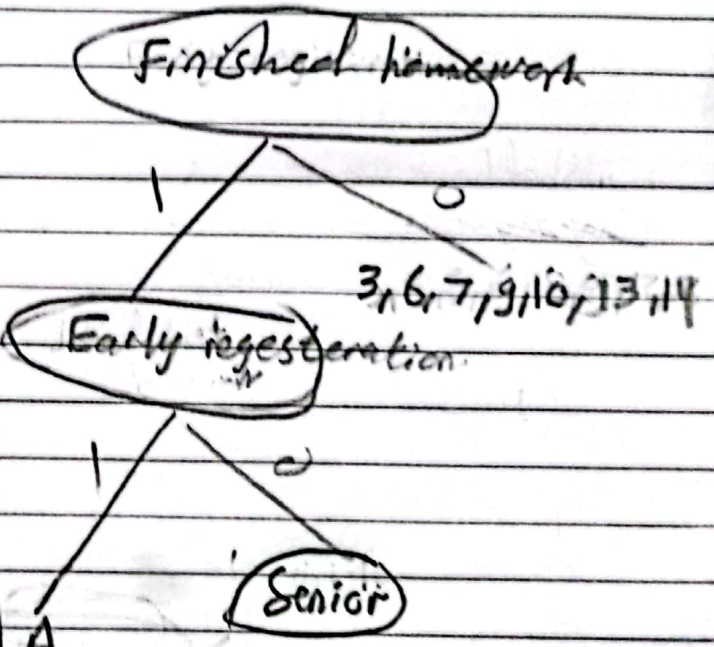
∴ Entropy = 1

$S_1 = [+1, -1]$ equal

entropy = 1

$gain = 1 - \frac{2}{4} \times 1 - \frac{2}{4} \times 1 = 0$

∴ ~~Finished homework~~ Senior has higher gain.



* Finished homework (a)

$S = [3, -4]$

Entropy $= \frac{-3}{7}\log_2\frac{3}{7} - \frac{4}{7}\log_2\frac{4}{7} = 0.985$

* Early registration

$S_{(a)} = [+1, -2]$

entropy $= \frac{-1}{3}\log\frac{1}{3} - \frac{2}{3}\log\frac{2}{3} = 0.918$

$S_{(a)} = [+2, -2]$ equal

$\therefore$ Entropy $= 1$

gain $= 0.985 - \frac{3}{7} * 0.918 - \frac{4}{7} * 1 = 0.02$

* Senior

$S_{(1)} = [+2, -1]$

Entropy $= \frac{-2}{3}\log_2\frac{2}{3} - \frac{1}{3}\log_2\frac{1}{3} = 0.918$

$S_{(a)} = [+1, -3]$

Entropy $= \frac{-1}{4}\log_2\frac{1}{4} - \frac{3}{4}\log_2\frac{3}{4} = 0.811$

gain $= 0.985 - \frac{3}{7} * 0.918 - \frac{4}{7} * 0.811$

$= 0.128$

* Likes Coffee

$S_{(1)} = [+2, 0]$

entropy $= 0$

$S_{(a)} = [+1, -4]$

entropy $= \frac{-1}{5}\log_2\frac{1}{5} - \frac{4}{5}\log_2\frac{4}{5} = 0.722$

gain $= 0.985 - 0 - \frac{5}{7} * 0.722 = 0.469$

[1] Liked last homework

$S_{(1)} = [2, -2]$

entropy $= 1$

$S_{(a)} = [+1, -2]$

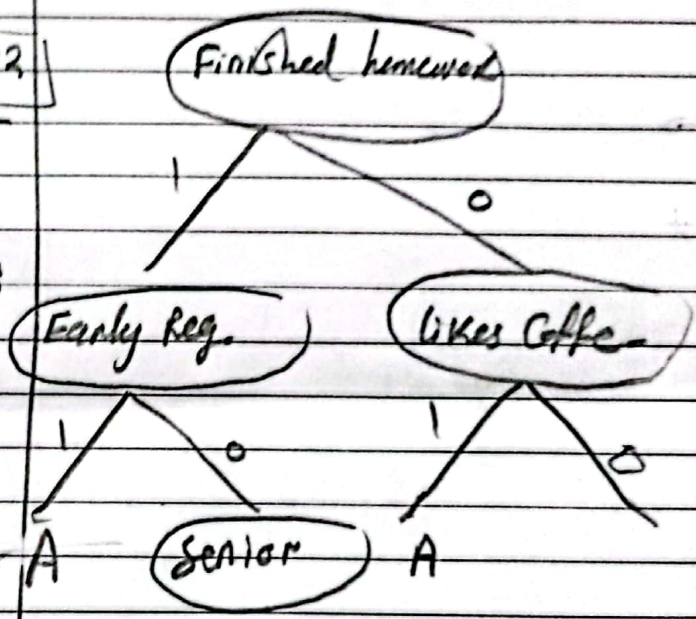
entropy $= \frac{-1}{3}\log_2\frac{1}{3} - \frac{2}{3}\log_2\frac{2}{3}$

$= 0.918$

gain $= 0.985 - \frac{4}{7} - \frac{3}{7} * 0.918$

$= 0.02$

$\therefore$ Likes Coffe



→ Likes Coffe at ① → A

→ Likes Coffee (a)

$S = [+1, -4]$

entropy $= \frac{-1}{5}\log_2\frac{1}{5} - \frac{4}{5}\log_2\frac{4}{5} = 0.72$

[4] Early Registeration

$S_{(1)} = [0, -2]$

entropy = 0

$S_{(0)} = [+1, -2]$

$E = 0.918$

$gain = 0.72 - 0 - \frac{3}{5} * 0.918 = \boxed{0.171}$

[5] Senior

$S_{(1)} = [+1, -1]$ equal

$E = 1$

$S_{(0)} = [0, -3]$

$E = 0$

$gain = 0.72 - \frac{2}{5} - 0 = \boxed{0.322}$

* Liked last homework.

$S_{(1)} = [+1, -2]$

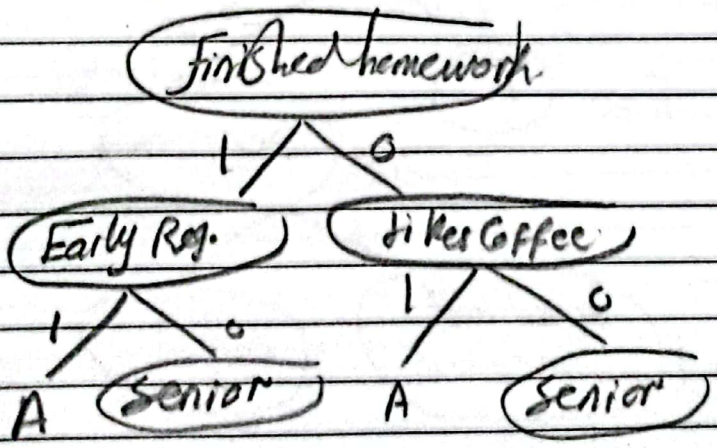
$E = 0.918$

$S_{(0)} = [0, -2]$

$E = 0$

$gain = 0.72 - \frac{3}{5} * 0.918 - 0 = \boxed{0.171}$

∴ Senior



[5] Finished —1→ Early —0→ Senior —1→

$S = [+1, -2]$

$Entropy = -\frac{1}{3}\log\frac{1}{3} - \frac{2}{3}\log\frac{2}{3} = 0.918$

* Likes Coffee

$S_{(1)} = [0, -1]$

$E = 0$

$S_{(0)} = [+1, -1]$

$E = 1$

$gain = 0.918 - 0 - \frac{2}{3} = \boxed{0.251}$

* Liked last homework

$S_{(1)} = [0, -2]$

$E = 0$

$S_{(0)} = [1, 0]$

$E = 0$

$\therefore gain = 0.918 - 0 - 0 = \boxed{0.918}$

∴ Liked last homework

Senior —0→

$S = [1, 0]$

$E = 0$ ∴ A

Finished —0→ Likes —0→ Senior —1→

Senior

$S_{(1)} = [1, -1]$

$E = 1$

* early

$S_{(1)} = [0, 0]$ $E = 0$

$S_{(0)} = [1, -1]$ $E = 1$

$gain = 1.0 - \frac{2}{2} * 1 = 0$

* Liked last homework

$S_{(1)} = [1, 0]$

$E = 0$

$S_{(0)} = [0, -1]$

$E = 0$

$\therefore$ gain = 1

2. Liked last homework

Senior → 0

$S\ [0, -3]$

$E = 0$

$\therefore$ not A

$\therefore$

Tree.