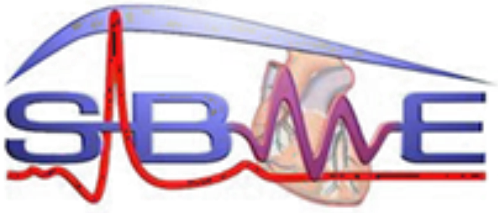
Systems & Biomedical Engineering Department Faculty of Engineering

Cairo University



Machine Learning Assignment #3

Gehad Mohamed Ahmed Ali Mohamed |Sec: 1|B.N: 28

Problem1: 1]

NO.: ROX Date :

Entropy = $-[p_+ \log p_+ + p_- \log p_-]$

Information gain = $E(S) - \sum \frac{S_v}{S} E[S_v]$

$[8_{+ve}, 6_{-ve}]$ ← entropy

$E(S) = -\left[\frac{8}{14}\log_2\frac{8}{14} + \frac{6}{14}\log_2\frac{6}{14}\right] = 0.985$

Early registeration: $1\left[{}_{+ve}\frac{4}{6}, {}_{-ve}\frac{2}{6}\right], 0\left[\frac{4}{+ve\,8}, \frac{4}{8_{ve}}\right]$

$E(1) = -\left[\frac{4}{6} * \log\frac{4}{6} + \frac{2}{6}\log\frac{2}{6}\right] = 0.9182$

$E(0) = -\left[\frac{4}{8} * \log\frac{4}{8} + \frac{4}{8}\log\frac{4}{8}\right] = 1$

IG =

$= 0.985 - \left[\frac{6}{14} * 0.9182 + \frac{8}{14} * 1\right]$

$= 0.02057$

Finished homework II: $1\left[\frac{5}{+ve}, \frac{2}{7_{-ve}}\right], 0\left[\frac{3}{+ve}, \frac{4}{7_{-ve}}\right]$

$E(1) = -\left[\frac{5}{7}\log_2\frac{5}{7} + \frac{2}{7}\log_2\frac{2}{7}\right]_7 = 0.86312$

$E(0) = -\left[\frac{3}{7}\log_2\frac{3}{7} + \frac{4}{7}\log_2\frac{4}{7}\right] = 0.985$

IG =

$0.985 - \left[\frac{7}{14} * 0.86312 + \frac{7}{14} * 0.985\right]$

$= 0.0609$

Date: ROXI التاريخ:

Senior: 1 $\left[\underset{+ve}{\frac{5}{8}}, \underset{-ve}{\frac{3}{8}}\right]$, 0 $\left[\underset{+ve}{\frac{3}{6}}, \underset{-ve}{\frac{3}{6}}\right]$

$E(1) = -\left[\frac{5}{8}\log_2\frac{5}{8} + \frac{3}{8}\log_2\frac{3}{8}\right] = 0.9544$

$E(0) = -\left[\frac{3}{6}\log_2\frac{3}{6} + \frac{3}{6}\log_2\frac{3}{6}\right] = 1$

$IG = 0.985 - \left[\left(\frac{8}{14} * 0.9544\right) + \left(\frac{6}{14} * 1\right)\right]$

$= 0.011057$

Likes Coffee: 1 $\left[\underset{+ve}{\frac{3}{4}}, \underset{-ve}{\frac{1}{4}}\right]$, 0 $\left[\underset{+ve}{\frac{5}{10}}, \underset{-ve}{\frac{5}{10}}\right]$

$E(1) = -\left[\frac{3}{4} * \log_2\frac{3}{4} + \frac{1}{4} * \log_2\frac{1}{4}\right] = 0.811$

$E(0) = -\left[\frac{5}{10} * \log_2\frac{5}{10} + \frac{5}{10} * \log_2\frac{5}{10}\right] = 1$

$IG = 0.985 - \left[\frac{4}{14} * 0.811 + \frac{10}{14} * 1\right]$

$= 0.039$

Date :

liked the last homework: (1) $\left[\underset{+ve\ g}{5}, \underset{-ve\ g}{4}\right]$

(0) $\left[\underset{+ve\ 5}{3}, \underset{-ve\ 5}{2}\right]$

$E(1) = -\left[\frac{5}{9}\log_2\frac{5}{9} + \frac{4}{9}\log_2\frac{4}{9}\right] = 0.991$

$E(0) = -\left[\frac{3}{5}\log\frac{3}{5} + \frac{2}{5}\log\frac{2}{5}\right] = 0.97$

$IG = 0.985 - \left[\frac{9}{14} * 0.991 + \frac{5}{14} * 0.97\right]$

$= 0.0015$

∴ finished homework II ## .

هذا حالة من الحالتين وهنرسها

When Finished homework II = (1)

ER	Finished HII	S	LC	I+LH	A
1	1	0	0	1	1
1	1	1	0	1	1
0	1	1	0	1	0
0	1	1	0	0	1
0	1	0	1	1	1
1	1	1	0	0	1
0	1	1	1	1	0

$$E(S) = -\left[\frac{5}{7} * \log_2 \frac{5}{7} + \frac{2}{7} \log \frac{2}{7}\right]$$

$$= \boxed{0.863}$$

ER: 1 [+ve 3, -ve 0], 0 [+ve 2/4, -ve 2/4]

$$E(1) = -\left[1 \log_2 1 + \text{zero} \log_2 0\right] = \text{zero}$$

$$E(0) = -\left[\frac{1}{2} \log_2 \frac{1}{2} + \frac{1}{2} \log_2 \frac{1}{2}\right] = 1$$

$$IG = 0.863 - \left[\text{zero} * \frac{3}{7} + 1 * \frac{4}{7}\right]$$

$$= \boxed{0.2915}$$

NO.: ROX

Date :

S: 1 $\left[\frac{3}{+ve\,5}, \frac{2}{-ve\,5}\right]$, 0 $\left[\frac{2}{+ve\,2}, 0\right]_{-ve}$

$E(1) = -\left[\frac{3}{5}\log_2\frac{3}{5} + \frac{2}{5}\log_2\frac{2}{5}\right] = 0.97$

$E(0) = -[1\log_2 1 + zero\log_2 zero] = 0$

$IG = 0.863 - \left[\frac{5}{7} * 0.97 + zero\right]$

$= \boxed{0.17014}$

LC: 1 $\left[\frac{1}{+ve\,2}, \frac{1}{-ve\,2}\right]$, 0 $\left[\frac{4}{+ve\,5}, \frac{1}{-ve\,5}\right]$

$E(1) = -\left[\left[\frac{1}{2}\log_2\frac{1}{2}\right] + \frac{1}{2}\log_2\frac{1}{2}\right] = \boxed{1}$

$E(0) = -\left[\frac{4}{5}\log_2\frac{4}{5}\right] + \left[\frac{1}{5} * \log_2\frac{1}{5}\right]$

$= \boxed{0.72}$

IG =

$0.863 - \left[\frac{2}{7} * 1 + \left(\frac{5}{7} * 0.72\right)\right]$

$= \boxed{0.063}$

Date:

ROXI

التاريخ:

Late, Late, Late → 1 [3/5 (+ve), 2/5 (-ve)], 0 [2/2 (+ve), 0 (-ve)]

$E(1) = -\left[\frac{3}{5}\log_2\frac{3}{5} + \frac{2}{5}\log_2\frac{2}{5}\right] = 0.97$

$E(0) = -\left[1\log_2 1 + \text{zero}\log_2 0\right] = \text{zero}$

$IG = 0.863 - \left[\frac{5}{7} * 0.97 + \text{zero} * \frac{2}{7}\right]$

$= 0.17$

∴ Early Registeration when Firth → (1)

when (Finished homework II = zero)

ER	FH	S	LC	L+LH	A
0	0	1	0	0	0
0	0	1	1	1	1
0	0	1	1	1	0
1	0	0	0	1	1
0	0	1	0	0	0
1	0	0	0	1	0
0	0	0	0	0	1
1	0	0	1		

NO.: ROXI رقم:
Date : التاريخ:

$$E(S) = -\left[\frac{3}{7}\log\frac{3}{7} + \frac{4}{7}\log_2\frac{4}{7}\right]$$

$$= \boxed{0.98522}$$

ER: $1\left[\frac{1}{+ve\,3}, \frac{2}{-ve\,3}\right], 0\left[\frac{2}{+ve\,4}, \frac{2}{-ve\,4}\right]$

$$E(1) = -\left[\frac{1}{3}\log_2\frac{1}{3} + \frac{2}{3}\log_2\frac{2}{3}\right] = 0.918$$

$$E(0) = -\left[\frac{1}{2}\log_2\frac{1}{2} + \frac{1}{2}\log_2\frac{1}{2}\right] = 1$$

$$IG = 0.985 - \left[\frac{3}{7} * 0.918 + \frac{4}{7} * 1\right]$$

$$\boxed{0.0201}$$

S: $1\left[\frac{2}{+ve\,3}, \frac{1}{-ve\,3}\right], 0\left[\frac{1}{+ve\,4}, \frac{3}{-ve\,4}\right]$

$$E(1) = -\left[\frac{2}{3}\log_2\frac{2}{3} + \frac{1}{3}\log_2\frac{1}{3}\right]$$

$$= 0.918$$

$$E(0) = -\left[\frac{1}{4}\log_2\frac{1}{4} + \frac{3}{4}\log_2\frac{3}{4}\right] =$$

$$0.811$$

$$IG = 0.985 - \left[0.918 * \frac{3}{7} + 0.811 * \frac{4}{7}\right]$$

$$= \boxed{0.128}$$

LC: 1 [2 (+ve), 0 (-ve)], 0 [1 (+ve 5), 4 (-ve 5)]

$E(1) = -[1 \log_2 1 + \text{zero} \log_2 \text{zero}] = \text{zero}$

$E(0) = -\left[\frac{1}{5}\log_2\frac{1}{5} + \frac{4}{5}\log_{2.5}\frac{4}{5}\right] = 0.72$

$IG = 0.985 - \left[\text{zero} + \frac{5}{7} * 0.72\right]$

$= 0.47$

LtH: 1 [2 (+ve), 2 (-ve)], 0 [1 (+ve 3), 2 (-ve 3)]

$E(1) = -\left[\frac{1}{2}\log_2\frac{1}{2} + \frac{1}{2}\log_2\frac{1}{2}\right] = 1$

$E(0) = -\left[\frac{1}{3}\log_2\frac{1}{3} + \frac{2}{3}\log_2\frac{2}{3}\right] = 0.918$

$IG = 0.985 - \left[1 * \frac{4}{7} + 0.918 * \frac{3}{7}\right]$

$= \boxed{0.0201}$

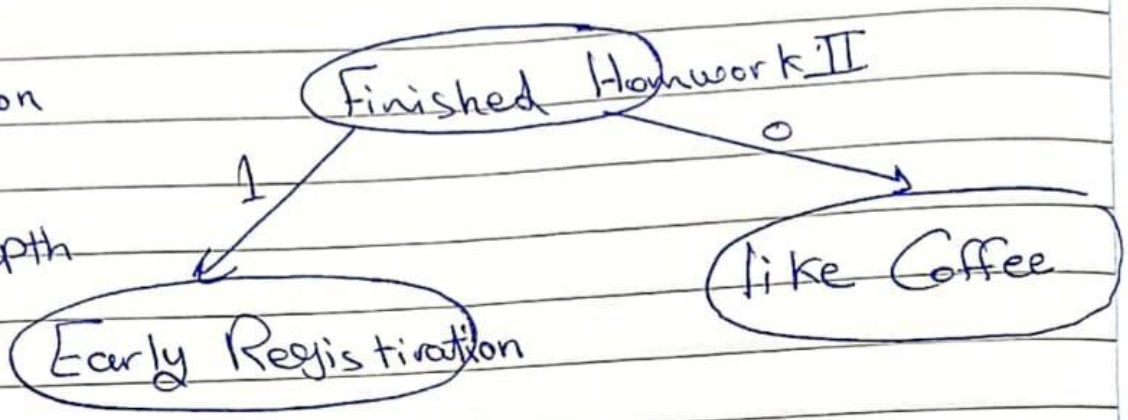
72

∴ LC is the ~~head node~~ Child node for

NO.:
Date:
ROXI
التاريخ:

∴ [Like Coffee] is the child for finished Homework [II] when it's equal to [Zero]

1st Decision tree for 1st depth



when finished Ho → 0 → like Coffee → 0

ER	S	LC	LtLH	A
0	1	0	0	0
1	0	0	1	0
0	1	0	1	1
1	0	0	0	0
0	0	0	1	0

$$E(S) = -\left[\frac{1}{5}\log_2\frac{1}{5} + \frac{4}{5}\log_2\frac{4}{5}\right] = 0.72$$

ER: 1[+ve 0, -ve 2], 0[+ve 1/3, -ve 2/3]

$$E(1) = -[0\log\text{zero} + 1\log_2 1] = \text{Zero}$$

$$E(0) = -\left[\frac{1}{3}\log_2\frac{1}{3} + \frac{2}{3}\log_2\frac{2}{3}\right] = 0.918$$

$$IG = 0.72 - [\text{zero} * \frac{2}{5} + 0.918 * \frac{3}{5}) = 0.1692$$

Date: ROXI التاريخ:

S: 1 $\underset{+ve}{\left[\frac{1}{2}\right.}, \underset{-ve}{\left.\frac{1}{2}\right]}$, 0 $\underset{+ve}{\left[0\right.}, \underset{-ve}{\left.\frac{3}{3}\right]}$

$E(1) = -\left[\frac{1}{2}\log_2\frac{1}{2} + \frac{1}{2}\log_2\frac{1}{2}\right] = 1$

$E(0) = -[\text{zero} * \log\text{zero} + \log 1] = \text{zero}$

$IG = 0.72 - [1 * \frac{2}{5} + \text{zero}] = 0.32$

HLH: 1 $\underset{+ve}{\left[\frac{1}{3}\right.}, \underset{-ve}{\left.\frac{2}{3}\right]}$, 0 $\underset{+ve}{\left[0\right.}, \underset{-ve}{\left.\frac{2}{2}\right]}$

$E(1) = -\left[\frac{1}{3}\log_2\frac{1}{3} + \frac{2}{3}\log_2\frac{2}{3}\right] = 0.918$

$E(0) = -[0\log_2\text{zero} + 1\log_2 1] = \text{zero}$

$IG = 0.72 - [\frac{3}{5} * 0.918 + \text{zero}]$

$= 0.1692$

Senior is the child for like Coffee when it's equal to zero

when like Coffee goes to 1

NO.: ROXI رقم:
Date : التاريخ:

ER	S	LC	LtLH	A
0	1	1	1	1
1	0	1	0	1

$E(S) = -\left[\frac{2}{2}\log 1 + \text{Zero}\,(\log_2 \text{zero})\right] = \text{Zero}$

S: 1[1 (+ve), 0 (-ve)], 0[1 (+ve), 0 (-ve)]

$E(1) = -[1 \log 1 + \text{zer}\ ] = 0$

$E(0) = \text{Zero}$

$IG = \text{zero}$

ER:

$E(1) = 0$, $E(0) = 0$, $IG = \text{zero}$

ITH: $IG = \text{zero}$

∴ Leaf node with value = 1 (Class (value))

when like Coffee goes to 1

Date: التاريخ:

when finished homework II goes through 1 then Early registeration
↳ when it goes through [1] zero. — (Firstly)
↳ when it goes through [0]. — secondly

Firstly ER	S	LC	LtLH	A
1	0	0	1	1
1	1	0	1	1
1	1	0	0	1

E(ER=1) = $-[\frac{1}{3}\log_2 1 + \text{zero}\log\text{zero}]$ = zero

S: 1[+ve 1, -ve 0], 0[+ve 1, -ve 0]

E(1) = zero IG =
E(0) = zero

LC: 1[0, 0], 0[1, 0]
E(1) = zero E(0) = 0 IG = 0

LtLH: 1[1, 0], 0[1, 0]
IG: zero

∴ Leaf Node = [1] ##

Secondly ER → zero

ER	FHII	S	LC	LtLH	A
0		1	0	0	0
0		1	1	1	1
0		1	0	1	1
0		0	0	1	0

$$E(ER=0) = -\left[\frac{1}{2}\log_2\frac{1}{2} + \frac{1}{2}\log_2\frac{1}{2}\right] = 1$$

$$S: 1\underset{+ve}{\left[\frac{2}{3}\right.}, \underset{-ve}{\left.\frac{1}{3}\right]}, 0\underset{+ve}{\left[0\right.}, \underset{-ve}{\left.1\right]}$$

$$E(1) = -\left[\frac{2}{3}\log_2\frac{2}{3} + \frac{1}{3}\log_2\frac{1}{3}\right] = 0.918$$

$$E(0) = -\left[\text{zero}\log_2\text{zero} + 1\log_2 1\right] = \text{zero}$$

$$IG = 1 - \left[-0.918 * \frac{3}{4} + \text{zero}\right] = \boxed{0.315}$$

$$LC: 1\underset{+ve}{\left[1\right.}, \underset{-ve}{\left.0\right]}, 0\left[\frac{1}{3}, \frac{2}{3}\right]$$

$$E(1) = -\left[1\log 1 + \text{zero}\log_2\text{zero}\right] = \text{zero}$$

$$E(0) = -\left[\frac{1}{3}\log_2\frac{1}{3} + \frac{2}{3}\log_2\frac{2}{3}\right] = 0.918$$

$$IG = 1 - \left[\text{zero} * \frac{1}{4} + \frac{3}{4} * 0.918\right]$$

$$= 0.315$$

$$LtLH: 1\left[\frac{2}{3}, \frac{1}{3}\right], 0\left[0, 1\right]$$

$$E(1) = 0.918, \quad E(0) = \text{zero}$$

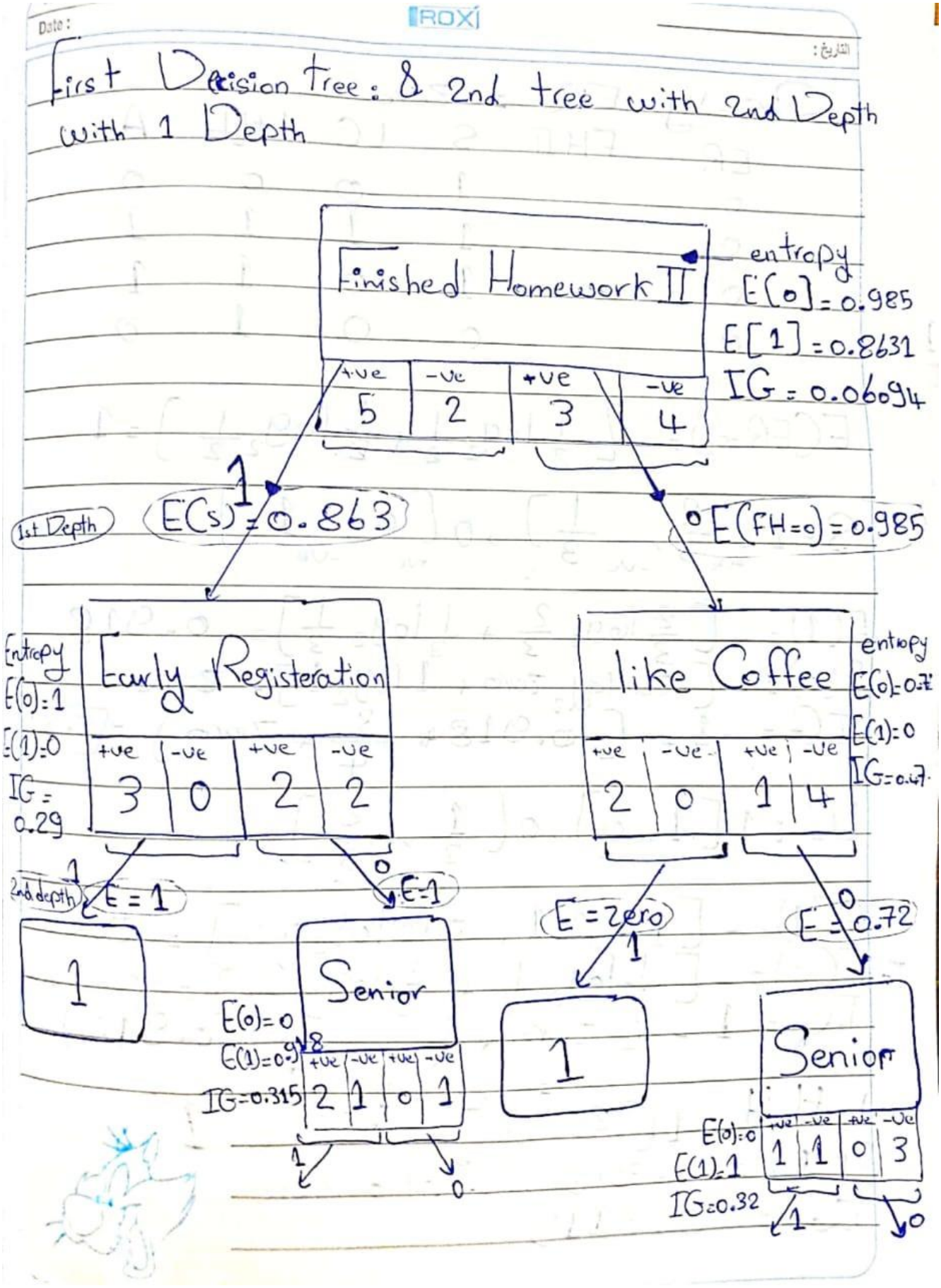
$$IG = 0.3115$$

Date :

ROXi

التاريخ :

First Decision Tree : & 2nd tree with 2nd Depth with 1 Depth



Problem1: 2]

CART would build a less deep tree than ID3, CART looks for splits that minimize the prediction squared error (the least–squared deviation). The prediction in each leaf is based on the weighted mean for node. It has the following advantages, CART can easily handle both numerical and categorical variables. CART algorithm will itself identify the most significant variables and eliminate non-significant ones. CART can easily handle outliers. ID3, as an "Iterative Dichotomiser," is for binary classification only CART, or "Classification And Regression Trees," is a family of algorithms (including, but not limited to, binary classification tree learning). With rpart(), you can specify method='class' or method='anova', but rpart can infer this from the type of dependent variable (i.e., factor or numeric).
Loss functions used for split selection. ID3, as other comments have mentioned, selects its splits based on Information Gain, which is the reduction in entropy between the parent node and (weighted sum of) children nodes. CART, when used for classification, selects its splits to achieve the subsets that minimize Gini impurity.