Date: Page:

Mohamed Saleh

20150134

Assignment (2)

(a) $Entropy(Y) = -\frac{P}{N+P}\log\frac{P}{N+P} - \frac{N}{P+N}\log\left(\frac{N}{P+N}\right) = -\frac{5}{8}\log\frac{5}{8} - \frac{3}{8}\log\frac{3}{8}$

$= 0.954$

For A_1

	P_i	N_i	$I(P_i,N_i)$
1	3	0	0
0	2	3	0.97

$I(P_i,N_i) = -\frac{2}{5}\log\frac{2}{5} - \frac{3}{5}\log\frac{3}{5} = 0.97$

$\therefore entropy(A_1) = \sum \frac{P_i+N_i}{P+N} I(P_i,N_i)$

$= \frac{3}{8} \times 0 + \frac{5}{8} \times 0.97 = 0.61$

$\therefore IG(A_1) = 0.954 - 0.61 = 0.344$

For A_2

	P_i	N_i	$I(P_i,N_i)$
1	2	2	1
0	3	1	0.81

$I(P_i,N_i) = -\frac{3}{4}\log\frac{3}{4} - \frac{1}{4}\log\frac{1}{4} = 0.81$

$entropy = \frac{2+2}{5+3} \times 1 + \frac{3+1}{5+3} \times 0.81 = 0.905$

$\therefore IG(A_2) = 0.954 - 0.905 = 0.049$

For A_3:

	P_i	N_i	$I(P_i,N_i)$
1	3	1	0.81
0	2	2	1

Same as A_2 $\therefore IG = 0.049$

For A_4:

	P_i	N_i	$I(P_i,N_i)$
1	3	1	0.81
0	2	2	1

Same as in A_2, A_3 $\therefore IG = 0.049$

So we choose A_1 to be the root Node with the ma

entropy = 0.344

So we will remove from the whole table the rows that have $A_1 = 1$

we need to find a new Node

H&A

Date: Page:

Example	A_2	A_3	A_4	y
X_3	1	0	0	1
X_4	1	1	0	0
X_6	1	0	1	0
X_7	0	1	1	1
X_8	0	1	0	0

Entropy (y) = 0.97

for A_2

	P_i	N_i	$I(P_i, N_i)$
0	1	1	0
1	3	2	0.918

$\therefore$ Entropy(A_2) $= \frac{2+1}{5} * 0.918 = 0.55 \Rightarrow IG = 0.97 - 0.55 = 0.42$

for A_3

	P_i	N_i	$I(P_i, N_i)$
0	2	1	0.9
1	3	1	0.918

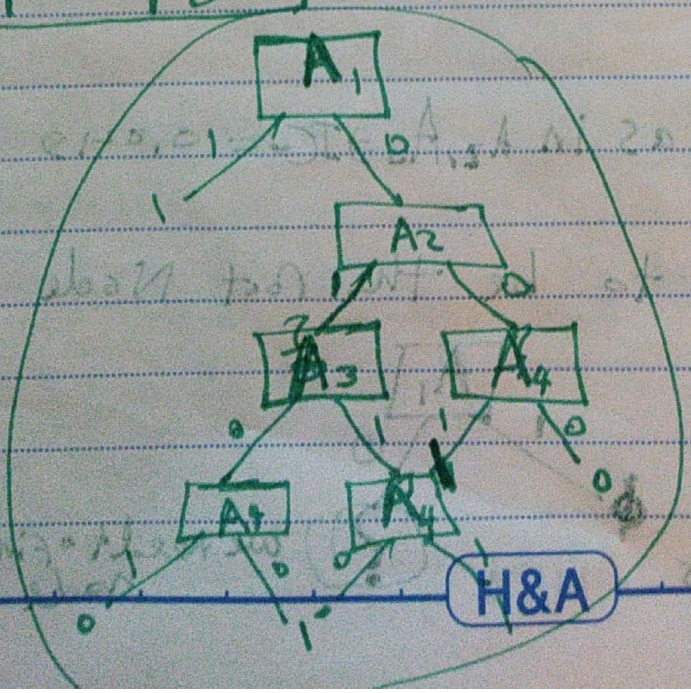
Same, IG = 0.42

for A_4:

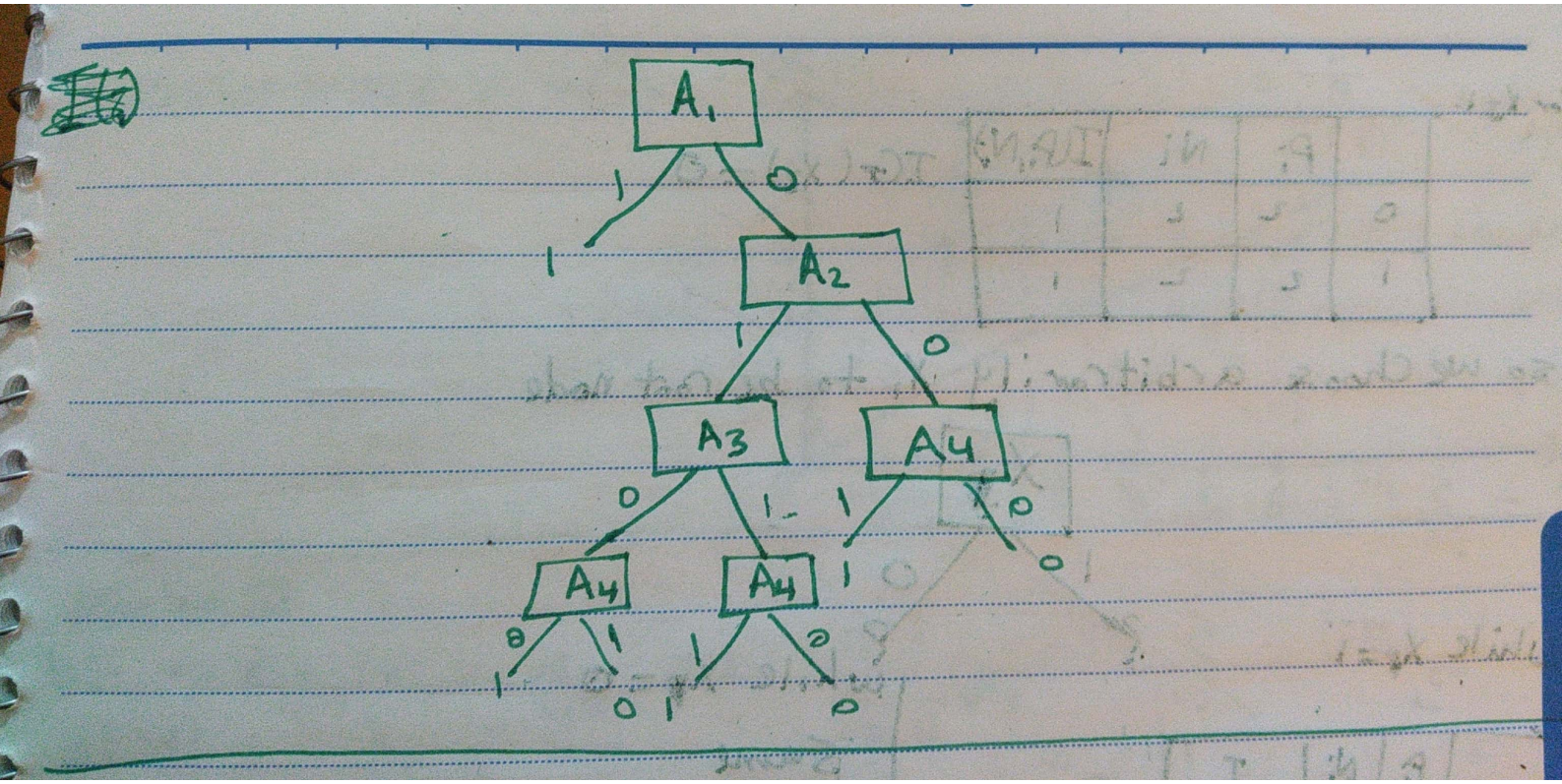
	P_i	N_i	$I(P_i, N_i)$
0	1	2	0.9
1	4	1	0

IG = 0.42

So we choose A_2



H&A



The decision list :-

if $A_1 == 1$
then output = 1
else if $A_1 = 0$
if A_2 if $A_2 = 0$
if $A_4 = 0$
then output = 0
else if $A_4 = 1$
then output = 1
else if $A_2 = 1$
if if $A_3 = 0$
if $A_4 = 0$
then output = 1
else if $A_4 = 1$
then out = 0
else if $A_3 = 1$
if $A_4 = 1$
then out = 1
else if $A_4 = 0$
then out = 0