MUAAZ SHOAIB
FA20-BCS-074
ML Assignment 01
02-Mar-2023
Decision Tree Classification Task

Q:- Entropy for each feature

Class level entropy:-

Total = 6, positive = 3, negative = 3

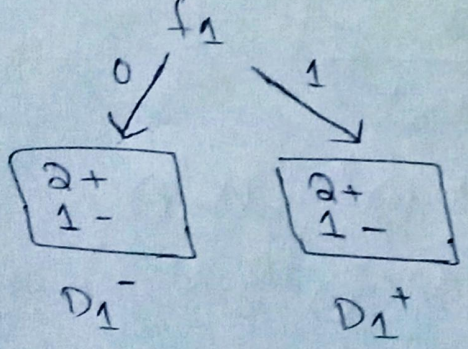
$H = -p \log_2 p - (1-p) \log_2 (1-p)$

P = proportion of positive examples in data set

$H = -\frac{3}{6} \log_2 \frac{3}{6} - (1-\frac{3}{6}) \log_2 (1-\frac{3}{6})$

$H = 1$

Entropy of f1



H for D_1^- $= -p \log_2 p - (1-p) \log_2 (1-p)$

$= -\frac{2}{3} \log_2 (\frac{2}{3}) - (1-\frac{2}{3}) \log_2 (1-\frac{2}{3})$

$= 0.92$

H for $D_1^+ = -p\log_2 p - (1-p)\log_2(1-p)$

$= -\frac{2}{3}\log_2\left(\frac{2}{3}\right) - \left(1-\frac{2}{3}\right)\log_2\left(1-\frac{2}{3}\right)$

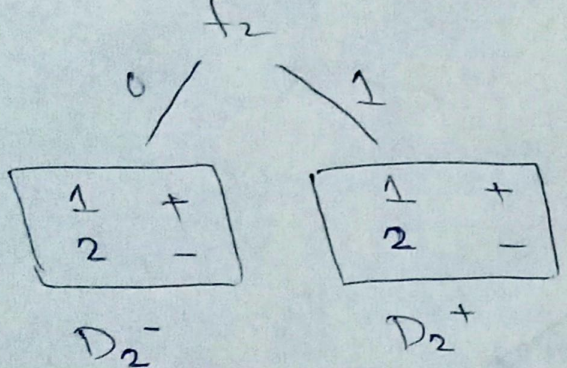
$= 0.92$

$AE_1 = ?$

$AE1 = \left(\frac{3}{6} \times 0.92\right) + \left(\frac{3}{6} \times 0.92\right)$

$AE1 = 0.92$

Entropy of f_2



H for $D_2^- = -p\log_2 p - (1-p)\log_2(1-p)$

$= -\frac{1}{3}\log_2\left(\frac{1}{3}\right) - \left(1-\frac{1}{3}\right)\log_2\left(1-\frac{1}{3}\right)$

$= 0.92$

H for $D_2^+ = -p\log_2 p - (1-p)\log_2(1-p)$

$= -\frac{1}{3}\log_2\left(\frac{1}{3}\right) - \left(1-\frac{1}{3}\right)\log_2\left(1-\frac{1}{3}\right)$

$= 0.92$

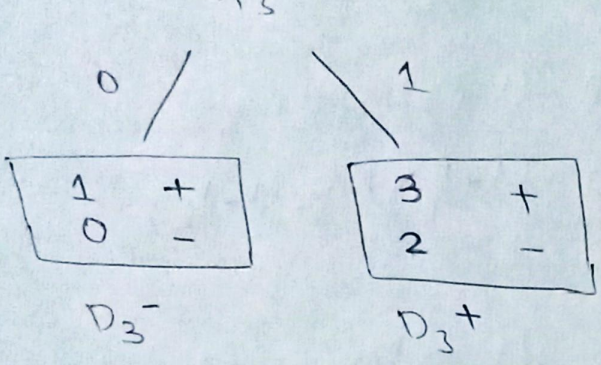
FA20-BCS-074

$AE_2 = ?$

$AE_2 = \left(\frac{3}{6} \times 0.92\right) + \left(\frac{3}{6} \times 0.92\right)$

$\boxed{AE_2 = 0.92}$

Entropy of f_3



H for $D_3^- = -p\log_2 p - (1-p)\log_2(1-p)$

$= -1\log_2(1) - (1-1)\log_2(1-1)$

$= 0 - 0\log_2(0)$

$= 0 - 0 = 0$

H for $D_3^+ = -p\log_2 p - (1-p)\log_2(1-p)$

$= -\frac{3}{5}\log_2\frac{3}{5} - \left(1-\frac{3}{5}\right)\log_2\left(1-\frac{3}{5}\right)$

$= 0.97$

$AE_3 = ?$

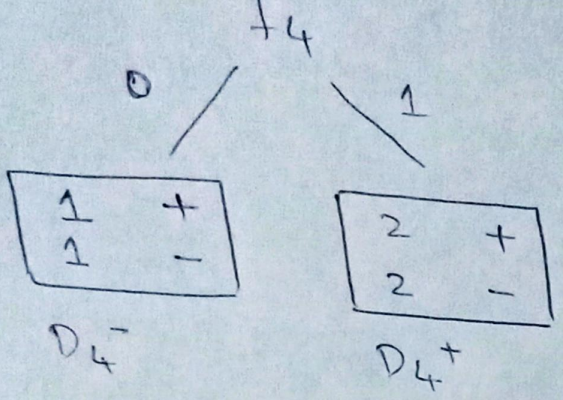
$AE_3 = \left(\frac{1}{6} \times 0\right) + \left(\frac{5}{6} \times 0.97\right)$

$\boxed{AE_3 = 0.81}$

FA20-BCS-074

(4)

Entropy of f_4



H for D_4^- $= -p\log_2 p - (1-p)\log_2(1-p)$

$= -\frac{1}{2}\log_2\left(\frac{1}{2}\right) - \left(1-\frac{1}{2}\right)\log_2\left(1-\frac{1}{2}\right)$

$= 1$

H for D_4^+ $= -p\log_2 p - (1-p)\log_2(1-p)$

$= -\frac{2}{4}\log_2\frac{2}{4} - \left(1-\frac{2}{4}\right)\log_2\left(1-\frac{2}{4}\right)$

$= 1$

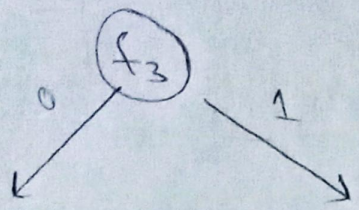
$AE_4 = ?$

$AE_4 = \left(\frac{2}{6} \times 1\right) + \left(\frac{4}{6} \times 1\right)$

$\boxed{AE_4 = 1}$

$AE_1 = 0.92$
$AE_2 = 0.92$
$AE_3 = 0.81$
$AE_4 = 1$

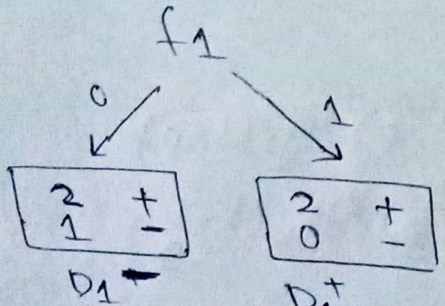
The best feature for split is f_3



0

f_1	f_2	f_3	f_4	y
0	1	1	0	0
1	0	1	1	1
1	1	1	0	1
0	0	1	1	1
0	1	1	1	0

Entropy for f_1



H for D_1^- $= -p\log_2 p - (1-p)\log_2(1-p)$

$= -\frac{2}{3}\log_2\frac{2}{3} - (1-\frac{2}{3})\log_2(1-\frac{2}{3})$

$= 0.91$

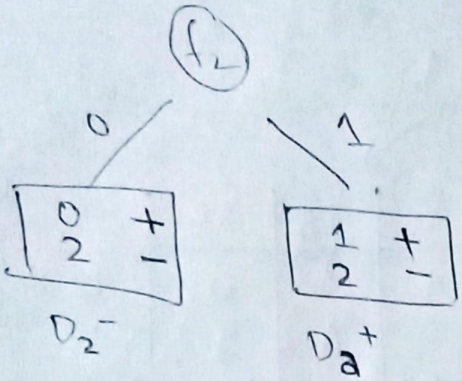
H for $D_1^+ = -p\log_2 p - (1-p)\log_2(1-p)$

$= 0$

$$AE_1 = \left(\frac{3}{5} \times 0.91\right) + \left(\frac{2}{5} \times 0\right)$$

$$\boxed{AE_1 = 0.55}$$

Entropy for f_2



H for $D_2^- = -p \log_2 p - (1-p) \log_2 (1-p)$

$$= -\frac{0}{2} \log_2 \frac{0}{2} - \left(1 - \frac{0}{2}\right) \log_2 \left(1 - \frac{0}{2}\right)$$

$$= (0)(\log_2(0/2)) - (0)(\log_2(1))$$

$$= 0 - 0$$

$$= 0$$

H for $D_2^+ = -p \log_2 p - (1-p) \log_2 (1-p)$

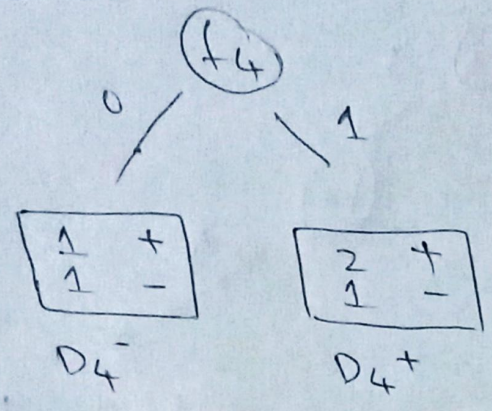
$$= -\frac{1}{3} \log_2 \left(\frac{1}{3}\right) - \left(1 - \frac{1}{3}\right) \log_2 \left(1 - \frac{1}{3}\right)$$

$$= 0.92$$

$$AE_2 = \left(\frac{2}{5} \times 0\right) + \left(\frac{3}{5} \times 0.92\right)$$

$$\boxed{AE_2 = 0.55}$$

Entropy for f_4



H for $D_4^- = -p\log_2 p - (1-p)\log_2(1-p)$

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

$= 1$

H for $D_4^+ = -p\log_2 p - (1-p)\log_2(1-p)$

$= -\frac{2}{3}\log_2(\frac{2}{3}) - (1-\frac{2}{3})\log_2(1-\frac{2}{3})$

$= 0.92$

$AE_4 = \frac{(1 \times 2}{5}) + (0.92 \times$

$= (\frac{2}{5} \times 1) + (\frac{3}{5} \times 0.92)$

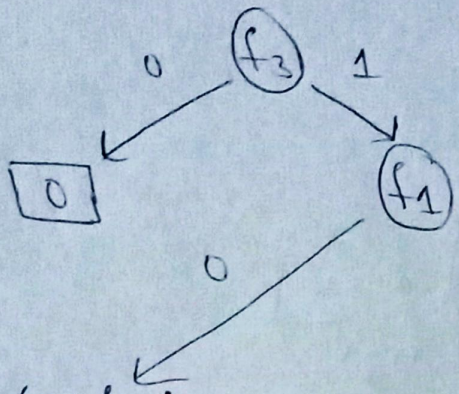
$\boxed{AE_4 = 0.95}$

$AE_1 = 0.55$

$AE_2 = 0.55$

$AE_4 = 0.95$

Arbitrarily decide to split on f_1



f_1	f_2	f_3	f_4	y
0	1	1	0	0
0	0	1	1	1
0	1	1	1	1

f_1	f_2	f_3	f_4	y
1	0	1	1	1
1	1	1	0	1