Machine Learning CS6375

Assignment-2 Decision Tree Induction.

— Bhanuteja Naga Venkata Sai
Thandava Koduri
bxk170230@utdallas.edu

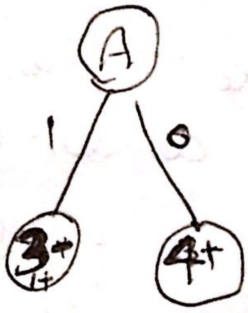
— Sashidhar Donthiri
sxd173730@utdallas.edu.

① (a) $Y = (\neg A \vee B) \wedge \neg(C \wedge A)$ | A, B, C are boolean

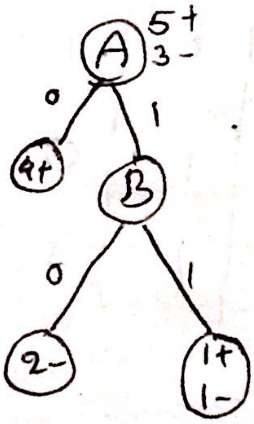
Truth table:

A	B	C	Y
1	1	1	0
1	1	0	1
1	0	1	0
0	1	1	1
1	0	0	0
0	0	1	1
0	1	0	1
0	0	0	1

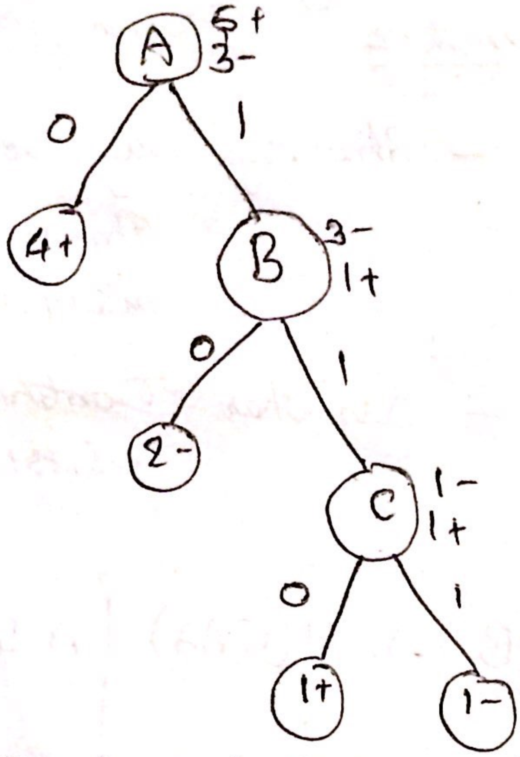
step ① since ∀ values of (0) A, there is value 1 in Y, consider A as root node,



Step ② split wrt B



step 2 split wrt. C :-



This is final Decision Tree.

(b) $(A \oplus B) \wedge C$

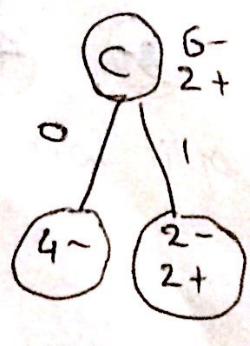
$\Downarrow$

$Y = (A'B \vee B'A) \wedge C$

step 1

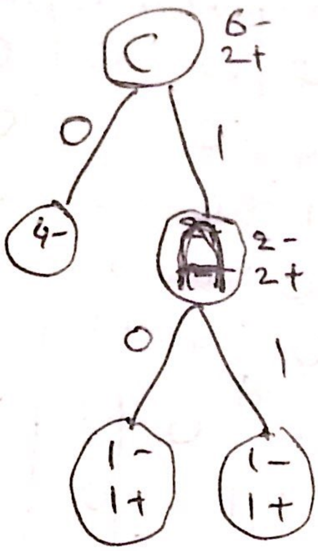
Since $\forall$ $C = 0 \Rightarrow Y = 0$,

so Consider C as root,

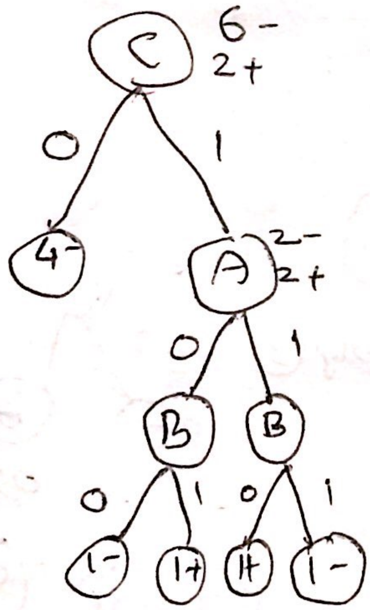


A	B	C	Y
0	0	0	0
0	0	1	0
0	1	0	0
1	0	0	0
0	1	1	1
1	1	0	0
1	0	1	1
1	1	1	0

step ② Consider A as next node, when C=1 because C=0 is pure.



step ③, since A=0 has equal nodevalues, Consider B for each of children nodes of A.



This is final decision tree

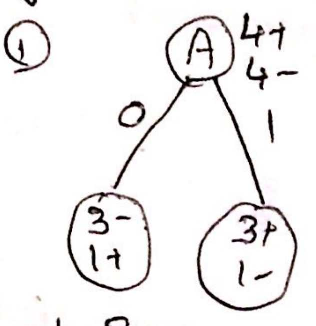
(c) $Y = (A \vee B) \wedge (B \vee C) \wedge (A \vee C)$

Truth Table:

Since here, the Y is symmetric, consider any feature as root node

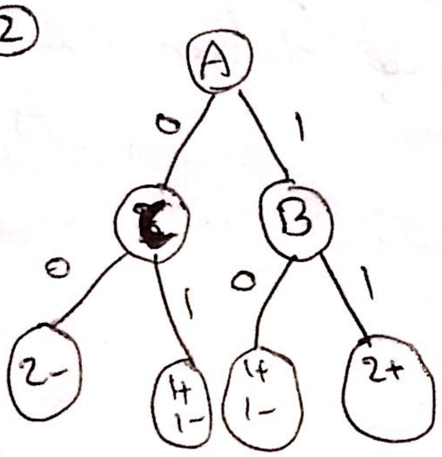
A	B	C	Y
0	0	0	0
0	0	1	0
1	0	0	0
0	1	0	0
0	1	1	1
1	0	1	1
1	1	0	1
1	1	1	1

①

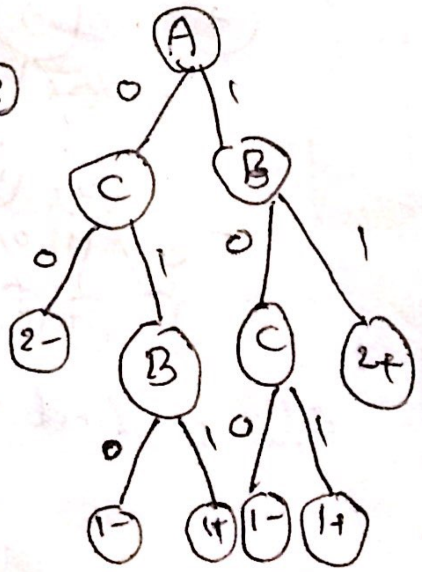


cost B

step ②



step ③



This is final decision Tree

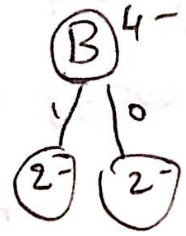
(d) $Y = (A \vee B) \wedge \neg A \wedge \neg B$

Truth table:

A	B	Y
0	0	0
~~0~~ 0	0	0
0	1	0
1	0	0
1	0	0
1	1	0
0	1	0
1	1	0

(d) $Y = A \vee B \wedge \neg A \neg B$

A	B	Y
0	0	0
0	1	0
1	0	0
1	1	0



Since, $Y = 0$ $\forall$ values of A & B

both A & B can be chosen as root node with no further classification needed.

A or B is final decision Tree.

(2)

Instance	X_1	X_2	X_3	Class
1	1	0	0	1
2	0	1	0	1
3	0	0	0	0
4	1	0	1	0
5	0	0	0	0
6	1	1	0	1
7	0	1	1	0
8	1	0	0	1
9	0	0	0	0
10	1	0	0	1

According to ID3, check Information gain for the attributes X_1, X_2, X_3.
Feature with Highest I.G. is root Node.

Information Gain, I.G = Entropy (parent) − avg. entropy (children)

$$\text{Entropy}(E) = \sum_{0}^{n} -P_i \log P_i$$

Number of +ve (1) in Class = 5

-ve (0) in class = 5

$$E_{parent} = -\left[\frac{5}{10}\log_2\frac{5}{10} + \frac{5}{10}\log\frac{5}{10}\right]$$

$$= +1$$

Entropy of X_1 =

$$H(Y/X_1) = -\sum P(x_1) \sum P(Y/x_1) \log P(Y/x_1)$$

$$= -\sum_{x_1} P(X_1) \sum_{y} H(Y/X = x_1)$$

$$H(Class/X_1=0) = -\left[\frac{4}{5}\log_2\frac{4}{5}+\frac{1}{5}\log_2\frac{1}{5}\right]$$

$$= 0.721928$$

$$H(Class/X_1=1) = -\left[\frac{4}{5}\log_2\frac{4}{5}+\frac{1}{5}\log_2\frac{1}{5}\right]$$

$$= 0.721928.$$

$$IG = 1 - H(Class/X_1)$$

$$= 1 - \left[\frac{5}{10}\times 0.7219 + \frac{5}{10}\times 0.7219\right]$$

$$= 1 - 0.7219 = 0.278$$

Entropy (X_2) =

$$H(Y/X_2) = -\sum_{X_2} P(X_2) \sum_{y} H(Y/X = X_2)$$

$$H(Class/X_2=0) = -\frac{4}{7}\log\frac{4}{7} - \frac{3}{7}\log\frac{3}{7}$$

$$= 0.9849$$

$$H(Class/X_2=1) = -\frac{2}{3}\log\frac{2}{3} - \frac{1}{3}\log 1/3$$

$$= 0.9182$$

$$H(Class/X_2) = \frac{7}{10}\times(.985) + \frac{3}{10}\times(.918)$$

$$= .965$$

$$IG = 1 - .965 = 0.035$$

Entropy of X_3

$$H(Y/X_3=0) = -\frac{3}{8}\log 3/8 - \frac{5}{8}\log 5/8$$

$$= 0.9543$$

$$H(Y/X_3=1) = -\left[\frac{0}{2}\log 0/2 + \frac{2}{2}\log 2/2\right] = 0$$

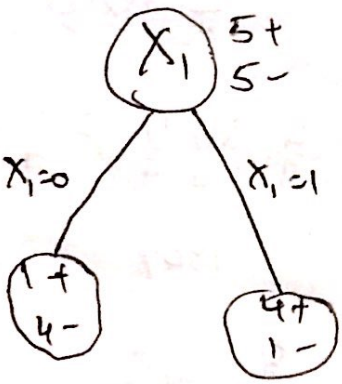
$$H(Class/X_3) = 0.9543 \times \frac{8}{10}$$

$$= 0.7635$$

$$IG = 1 - 0.7635$$

$$= 0.2365$$

$\Rightarrow$ $IG(X_1)$ is highest, so X_1 is root node.



Here, we have to decide X_2/X_3 as our left or right nodes based on the information gain ($X_1=0$ & $X_1=1$ do not give pure nodes)

For Left node:- Entropy when (X_1) is 0

$$H(X_1=0/X_2=0) = -\left[\frac{3}{3}\log 3/3\right] = 0$$

$$H(X_1=0/X_2=1) = -\left[\frac{1}{2}\log\frac{1}{2} + \frac{1}{2}\log\frac{1}{2}\right]$$

$$= 1$$

Entropy $H(X_1=0/X_2) = \frac{3}{5}\times 0 + \frac{2}{5}\times 1$

$= 0.4$

$\therefore$ IG $= 0.32198$
(X_2 wrt $X_1=0$)

(left side)

Entropy of X_3 when $X_1=0$

$\Rightarrow$ $H(X_1=0/X_3=0) = -\left[\frac{1}{4}\log\frac{1}{4} + \frac{3}{4}\log 3/4\right]$

$= 0.8112$

$H(X_1=0/X_3=1) = -[1/1 \log 1] = 0$

Entropy, $H(X_1=0/X_3) = \frac{4}{5}\times .8112 + 0$

$= .649.$

IG (X_3 wrt $X_1=0$) $= 0.7219 - 0.649$

$= 0.0729$

$\therefore$ Since IG of X_2 is more, when $X_1=0$, X_2 is on left side of tree

Now consider, $X_1=1$ (right node)

Entropy $(X_1=1) = 0.7219$

$H(X_1=1/X_2=0) = -\left[\frac{1}{4}\log\frac{1}{4} + \frac{3}{4}\log 3/4\right]$

$= 0.81125$

$H(X_1=1 \mid X_2=1) = -\frac{1}{1}\log 1/1 = 0$

$H(X_1=1 \mid X_2) = \frac{4}{5}(.81125) + \frac{1}{5}(0)$

$= 0.649$

IG on X_2 when $(X_1=1)$

$= 0.7219 - 0.649$

$= 0.0729$

Entropy of X_3 when $X_1 = 1$ (right side)

$H(X_1=1) = 0.7219$

$H(X_1=1 \mid X_3=0) = -\frac{4}{4}\log_2 4/4 = 0$

$H(X_1=1 \mid X_3=1) = -\left[\frac{1}{1}\log 1\right] = 0$

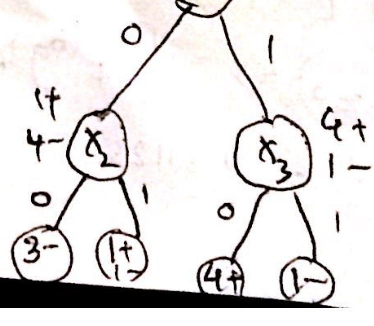
$H(X_1=0 \mid X_3) = \frac{4}{5} \times 0 + \frac{1}{5} \times 0$

$= 0$

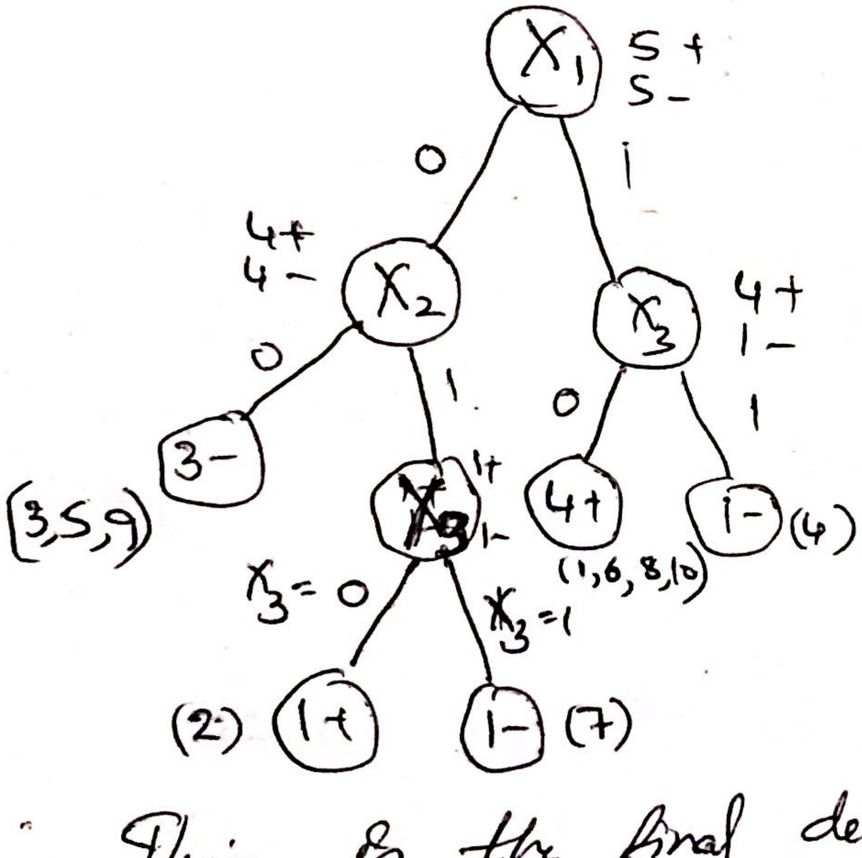
I.G on X_3 wrt $X_1=1 \Rightarrow 0.729 - 0$

$= 0.729$

Since IG on X_3 is more on right side $(X_1=1)$, split it on X_3,



From the obtained tree, we observe that the left leaf of X_1 is not pure when $X_2 = 1$, so we will have to split it again, and since the leaf node obtained by splitting X_1 on X_3 is pure, no further splitting is needed.



∴ This is the final decision Tree.