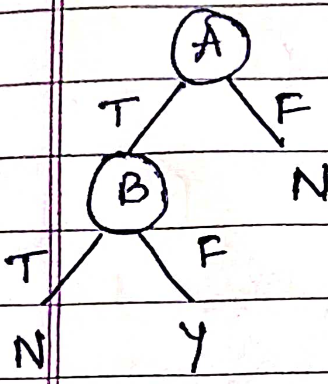
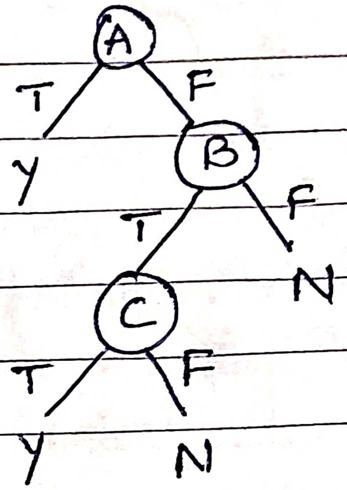
Construct decision tree for the following boolean expressions

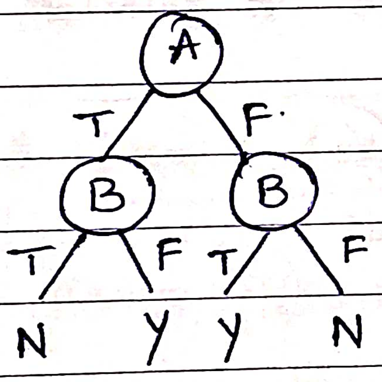
1. $A \wedge \neg B$



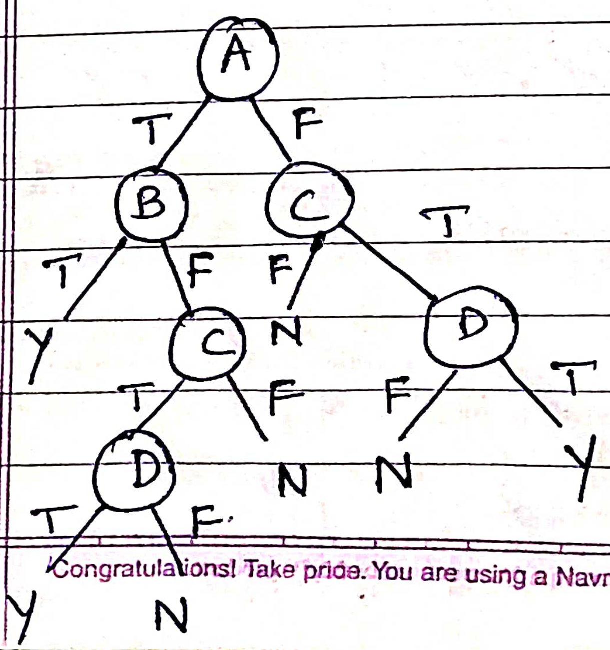

2. $A \vee (B \wedge C)$




3. $A \text{ XOR } B = (A \wedge \neg B) \vee (\neg A \wedge B)$




4. $(A \wedge B) \vee (C \wedge D)$



2.

| Instance | classification | a1 | a2 |
|---|---|---|---|
| 1 | + | T | T |
| 2 | + | T | T |
| 3 | − | T | F |
| 4 | + | F | F |
| 5 | − | F | T |
| 6 | − | F | T |

1. what is entropy of this collection of training examples with respect to the target function classification.

2. what is information gain of a1 & a2 relative to these training examples.

3. Draw decision tree for the given dataset

$$\boxed{\text{Entropy}(S) = \sum_{i=1}^{c} -P_i \log_2 P_i}$$

[c = no. of class labels
$P_i$ = Proportion of $i^{th}$ class]

for Target attribute classification

$$\text{Entropy}(3+, 3-) = -\frac{3}{6}\log_2 3/6 - \frac{3}{6}\log_2 3/6$$

$$= 1$$

$$\boxed{\text{Gain}(S, A) = \text{Entropy}(S) - \sum_{V \in \text{values}(A)} \frac{|S_V|}{|S|}\,\text{Entropy}(S_V)}$$

~~A = attribute~~, ~~Values(A) = Set of possible values of A~~

$S_V$ = Subset of S for which ~~attribute A has value V~~

Values $(a_1)$ = T, F

$S = [3+, 3-]$
$S_T = [2+, 1-]$
$S_F = [1+, 2-]$

$$\text{Gain}(S, a_1) = \text{Entropy}(S) - \sum_{V \in \{T,F\}} \frac{|S_V|}{|S|} \text{Entropy}(S_V)$$

$$= \text{Entropy}(S) - \frac{3}{6}\text{Entropy}(S_T) - \frac{3}{6}\text{Entropy}(S_F)$$

$$= 1 - \frac{3}{6}\left[-\frac{2}{3}\log_2 2/3 - \frac{1}{3}\log_2 1/3\right]$$
$$- \frac{3}{6}\left[-\frac{1}{3}\log_2 1/3 - \frac{2}{3}\log_2 2/3\right]$$

$$= 1 - \frac{3}{6}\left[-\frac{2}{3} * -0.585 - \frac{1}{3} * -1.585\right]$$
$$- \frac{3}{6}\left[-\frac{1}{3} * -1.585 - \frac{2}{3} * -0.585\right]$$

$$= 1 - \frac{3}{6} * 0.918 - \frac{3}{6} * 0.918$$

$$= 0.082$$

Congratulations! Take pride. You are ...

Values (a2) = T, F

$S = [3+, 3-]$

$S_T = [2+, 2-]$

$S_F = [1+, 1-]$

$$Gain(S, a2) = Entropy(S) - \sum_{v \in \{T, F\}} \frac{|S_v|}{|S|} Entropy(S_v)$$

$$= 1 - \frac{4}{6} * Entropy(S_T) - \frac{2}{6} * Entropy(S_F)$$

$$= 1 - \frac{4}{6} * 1 - \frac{2}{6} * 1$$

$$= 0$$

Maximum gain

$$\boxed{\begin{aligned} Gain(S, a1) &= 0.082 \\ Gain(S, a2) &= 0 \end{aligned}}$$

Decision tree