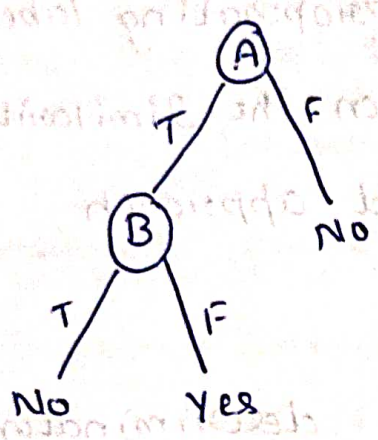
Fundamentals of Machine Learning

10/09/2021

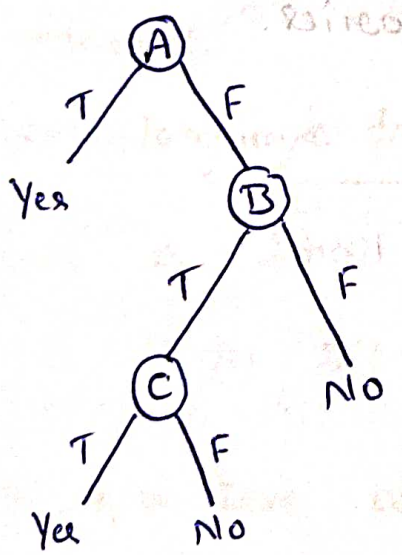
Assignment -2

1. Develop the decision tree for the following

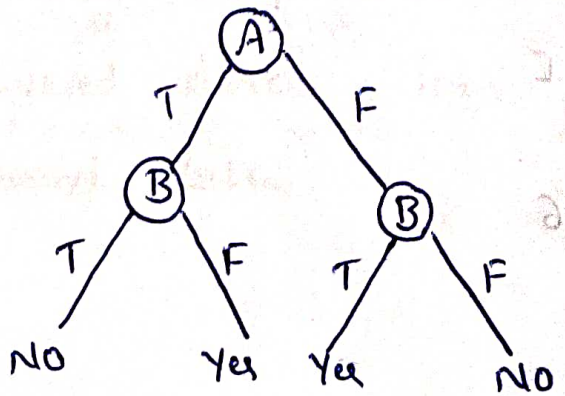
a) $A \wedge \neg B$



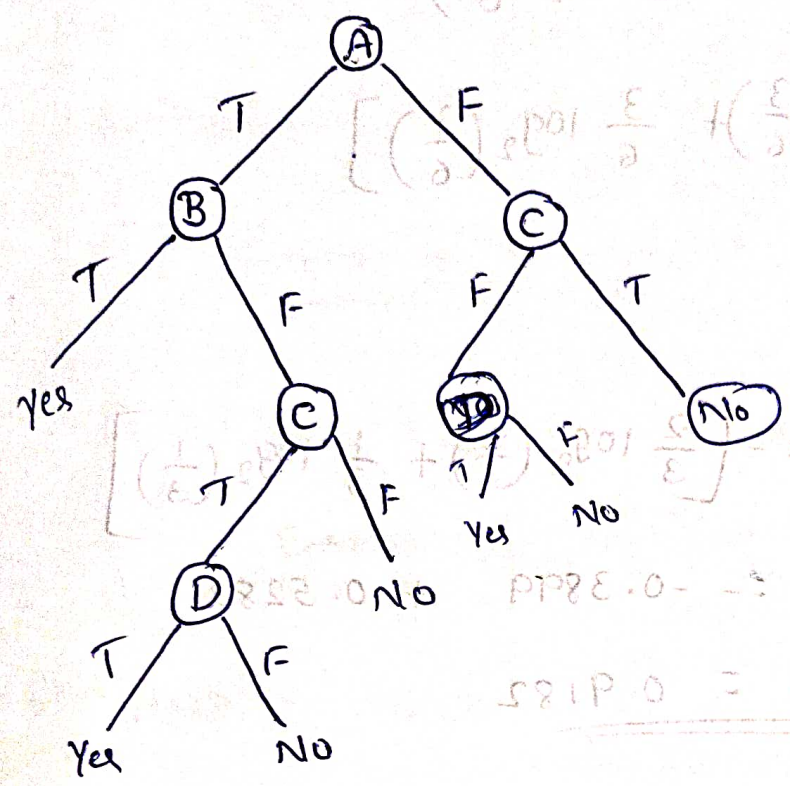
b) $A \vee [B \wedge C]$



c) A XOR B $(A \oplus B)$



A) $[A \wedge B] \vee [C \wedge D]$



2. Develop the decision tree using ID3 algoritm and Gini Index methods for the following data set. Also, construct the decision tree using gain ratio approach.

Instance	Attribute_1	Attribute_2	class
1	T	T	+
2	T	T	+
3	T	F	−
4	F	F	+
5	F	T	−
6	F	T	−

Entropy $(S) = -\left(P_{\oplus} \log_2 P_{\oplus} + P_{\ominus} \log_2 P_{\ominus}\right)$

$S : [3+, 3-]$ (+ve, -ve) $= -\left[\frac{3}{6}\log_2\left(\frac{3}{6}\right) + \frac{3}{6}\log_2\left(\frac{3}{6}\right)\right]$

Entropy $(S) = 1.0$

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

$= -(-0.3899) + (-0.5283)$

Entropy $(S_T) = 0.9182$

$S_F = [1+, 2-]$

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

Entropy for $S_F = 0.9182$

$$\text{Gain}(S, a_1) = \text{Entropy}(S) - \sum_{\forall \text{ value}} \frac{|S_v|}{|S|} \text{Entropy}(S_v)$$

$$= 1 - \left[\frac{3}{6} \times 0.9182 + \frac{3}{6} \times 0.9182\right]$$

$\text{Gain}(S, a_1) = 0.0818$

Attribute 2 :-

a_2 = values (attribute_2) = T, F

S_T = 4 True = [2+, 2-] = $-\left[\frac{2}{4}\log_2\left(\frac{2}{4}\right) + \frac{2}{4}\log_2\left(\frac{2}{4}\right)\right]$

Entropy for S_T = 1.0

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

Entropy (S_F) = 1.0

$$Gain(S, a_2) = E(S) - \sum_{v \in \text{value}} \frac{|S_v|}{|S|} \text{Entropy}(S_v)$$

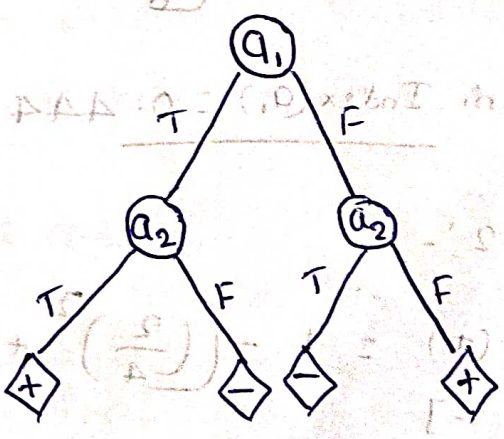
$$= 1 - \left[\frac{4}{6} \log \times 1 + \frac{2}{6} \times 1\right]$$

$Gain(S, a_2) = 0$

Decision Tree - ID_3 algorithm

Gain (S, a_1) = 0.0818 ✓

Gain (S, a_2) = 0



Gini Index Approach

Attribute-1 :-

$$\text{Gini Index} = \sum_{v \in \text{value}} \frac{|S_v|}{|S|} \text{Gini}(S_v)$$

$$\text{Gini}(T) = 1 - \sum_{i=1}^{c} (P_i)^2$$

[2+, 1-]

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

$$\text{Gini}(T) = 0.444$$

$$\text{Gini}(F) = 1 - \left[\left(\frac{1}{3}\right)^2 + \left(\frac{2}{3}\right)^2\right]$$

[1+, 2-]

$$\text{Gini}(F) = 0.444$$

$$\text{Gini Index}(a_1) = \left[\frac{3}{6} \times 0.444 + \frac{3}{6} \times 0.444\right]$$

$$\text{Gini Index}(a_1) = 0.444$$

Attribute-2 :-

$$\text{Gini}(T) = 1 - \left[\left(\frac{2}{4}\right)^2 + \left(\frac{2}{4}\right)^2\right]$$

[2+, 2-]

$$\text{Gini}(T) = 0.5$$

$$\text{Gini}(F) = 1 - \left[\left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^2\right]$$

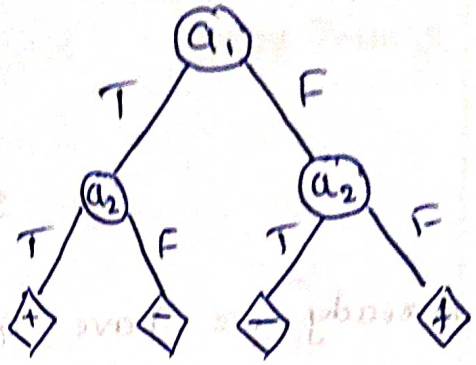
[1+, 1-]

$$\text{Gini}(F) = 0.5$$

Gini Index $= \left[\frac{4}{6}(0.5) + \frac{2}{6}(0.5)\right]$

Gini Index $(a_2) = 0.5$

Decision tree - Gini Index Approach



Gini Index $(a_1) = 0.444$ ✓

Gini Index $(a_2) = 0.5$

Constructing decision tree using Gain ratio approach

Attribute - 1:-

$$\text{Split information}(S, a_1) = -\sum_{i=1}^{c} \frac{|S_i|}{|S|} \log_2 \frac{|S_i|}{|S|}$$

$$S.I\ (S, a_1) = -\left[\frac{3}{6}\log_2\left(\frac{3}{6}\right) + \frac{3}{6}\log_2\left(\frac{3}{6}\right)\right]$$

$$= -[(-1) + (-1)]$$

$$S.I\ (S, a_1) = 1$$

Already we have Gain $(S, a_1) = 0.0818$

$$\text{Gain ratio}(S, a_1) = \frac{\text{Gain}(S, a_1)}{\text{Split Information}(S, a_1)}$$

$$= \frac{0.0818}{1}$$

Gain ratio $(S, a_1) = 0.0818$

Attribute_2:-

$$\text{Split Information}(S, a_2) = -\left[\frac{4}{6}\log_2\left(\frac{4}{6}\right) + \frac{2}{6}\log_2\left(\frac{2}{6}\right)\right]$$

$$= 0.9184$$

$$= -(-0.3899) + (-0.5283)$$

$$S.I\ (S, a_2) = 0.918$$

Already we have $\text{Gain}(S, a_2) = 0$

$$\text{Gain ratio}(S, a_2) = \frac{\text{Gain}(S, a_2)}{\text{Split information}(S, a_2)}$$

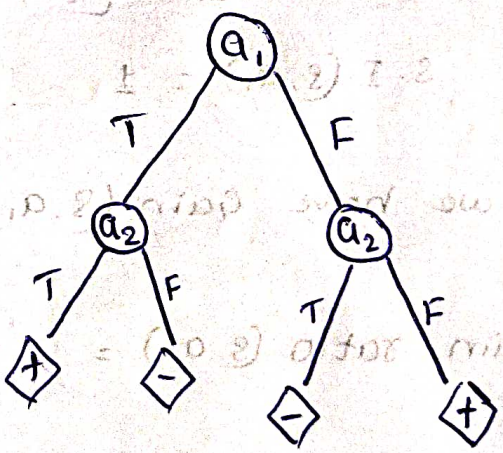
$$= \frac{0}{0.918}$$

$$\text{Gain ratio}(S, a_2) = 0$$

Decision tree for Gain ratio approach

$\text{Gain ratio}(S, a_1) = 0.0818$ ✓

$\text{Gain ratio}(S, a_2) = 0$



FML

Submitted by :- Karthik. C

Reg. No :- 240158016

Assignment_2 :- Decision tree