

# Classification - Decision Tree Solved Problems

Ques: Information Gain:-

$$I(P, n) = -\frac{P}{S} \log_2 \frac{P}{S} - \frac{n}{S} \log_2 \frac{n}{S} \quad \text{--- (1)}$$

$$S = P + n$$

$$E(A) = \sum_{i=1}^n \frac{P_i + n_i}{P + n} I(P_i, n_i) \quad \text{--- (2)}$$

$$\text{Gain}(A) = I(P, n) - E(A) \quad \text{--- (3)}$$

$$\log_2 x = \frac{\log_{10} x}{\log_{10} 2}$$

Ques: for the following medical data create a decision tree.

Sore Throat	Fever	Swollen Glands	Congestion	Head ace	Diagnosis
Y	Y	Y	Y	Y	High
N	N	N	Y	Y	Mild
Y	Y	N	Y	N	Negative
Y	N	Y	N	N	H
N	Y	N	Y	N	N
N	N	N	Y	N	M
N	N	Y	N	N	H
Y	N	N	Y	Y	M
N	Y	N	Y	Y	N
Y	Y	N	Y	Y	N

$$\text{Sample space} = H + M + N = 3 + 3 + 4 = 10$$

Info gain

$$I(P, n) = - \left[ \frac{3}{10} \log_2 \left( \frac{3}{10} \right) + \frac{3}{10} \log_2 \left( \frac{3}{10} \right) + \frac{4}{10} \log_2 \left( \frac{4}{10} \right) \right]$$

$$= - [0.6 \log_2 (0.3) + 0.4 \log_2 (0.4)]$$

$$= - \left[ 0.6 \frac{\log_{10}(0.3)}{\log_{10} 2} + 0.4 \frac{\log_{10}(0.4)}{\log_{10} 2} \right]$$

$$= 1.038 + 0.5272 = 1.562$$

Finding splitting attributes (Select attribute with highest gain)

(i) ~~Some~~ High

(i) 'Some' thread

	H	M	N
Y	2	1	2
N	1	2	2

E (Some thread)

$$I(\text{Yes}) = - \left[ \frac{2}{5} \log_2 \left( \frac{2}{5} \right) + \frac{1}{5} \log_2 \left( \frac{1}{5} \right) + \frac{2}{5} \log_2 \left( \frac{2}{5} \right) \right]$$

$$= 1.52$$

$$I(\text{No}) = 1.52$$

$$\text{Entropy (Sore Throat)} = 0.5 \times 1.52 + 0.5 \times 1.52 \\ = 1.52$$

$$E(S.T) = \text{Info Gain} + P \left[ \frac{-\gamma}{-N} \right] +$$

$$\text{Gain (A)} = 1.562 - 1.52 = 0.05$$

Gain of S.T is 0.05.

(i) Fever : 0.72 (2)

(ii) Swollen : 0.88 (1)

(iv) Congestion : 0.45 (3)

(v) Headache : 0.05

