CS 5783 : Assignment 5

- Timmaraan Nyamagoudar

Q1.

Entropy is given by the formulae :-

$$E(s) = \sum_{i=1}^{c} -P_i \log_2(P_i)$$

Information gain is given by the formulae :-

$$IG(Y,X) = E(Y) - E(Y|X)$$

To get the best tree, we calculate the IG for entire decision tree. This gives us the fastest branch to desired output.

From given,

$P(Meh) = 5/10 = 0.5$

$P(Yummy) = 5/10 = 0.5$

Substituting, we get entropy

$$E(S) = \{-P_{meh} \cdot \log_2(P_{meh})\} - \{P_{yummy} \cdot \log_2(P_{yummy})\}$$

$$= \{-0.5 \log_2(0.5)\} - \{0.5 \log_2(0.5)\} = 0.5 + 0.5$$

$$= 1$$

This is an impure feature, as entropy is 1. It is thus difficult to make output prediction.

Q2.

Visual defects is chosen as root of decision tree.

To calculate information gain (IG) of decision tree.

We know, $P(\text{meh}) = P(\text{yummy}) = 0.5$

$E(\text{taste}) = 1$

For visual defects feature, the probabilities of observation are :-

		Meh	Yummy
$P(\text{some})$	$= 3/10$	3	0
$P(\text{none})$	$= 4/10$	2	2
$P(\text{many})$	$= 3/10$	0	3

Now we calculate $E(\text{taste} \mid \text{visual defects})$

$= P(\text{some}).\,E(3,0) + P(\text{none}).\,E(2,2) + P(\text{many}).\,E(0,3)$

$= \frac{3}{10}.\,E(3,0) + \frac{4}{10}.\,E(2,2) + \frac{3}{10}.\,E(0,3)$

$= 0.3.\left\{-\frac{3}{3}\log_2\left(\frac{3}{3}\right) - \frac{0}{3}\log_2\left(\frac{0}{3}\right)\right\} + 0.4.\left\{-\frac{2}{4}\log_2\left(\frac{2}{4}\right) - \frac{2}{4}\log_2\left(\frac{2}{4}\right)\right\}$

$+ \frac{3}{10}.\left\{-\frac{0}{3}\log_2\left(\frac{0}{3}\right) - \frac{3}{3}\log_2\left(\frac{3}{3}\right)\right\}$

$= 0.3(0) + 0.4(1) + 0.3(0)$

$E(\text{taste} \mid \text{visual defects}) = 0.4$

$IG = E(\text{taste}) - E(\text{taste} \mid \cancel{\text{given}}\ \text{visual defects})$

$= 1 - 0.4$

$= \underline{0.6}$

Q3. To calculate entropy of

H (Taste | Visual defect = some) and

H (Taste | Visual defect = none)

H (Taste | Visual defect = some)

$= P(\text{some}) \times E(3,0)$ (substituting values from previous question)

$= \frac{3}{10} \times \left\{ -\frac{3}{3} \log_2 \left(\frac{3}{3}\right) - \frac{0}{3} \log_2 \left(\frac{0}{3}\right) \right\}$

$= 0.3\,(0)$

$= 0$

H (Taste | Visual defect = none)

$= P(\text{none}) \times E(2,2)$

$= \frac{4}{10} \times \left\{ -\frac{2}{4} \log_2 \left(\frac{2}{4}\right) - \frac{2}{4} \log_2 \left(\frac{2}{4}\right) \right\}$

$= 0.4(1)$

$= 0.4$