1)

Type

One | Two

Scale	Shade	Texture	Class
one	light	thin	A
Two	light	thin	A
Two	dark	thin	C
one	dark	thin	C

Scale	Shade	Text	Class
two	light	thin	B
two	dark	thin	B
one	light	thin	C

One:-

$Info(D) = I(2,2)$

$= 1$

Scale:-

$I(one) = A \rightarrow 1, \; C \rightarrow 1$

$I = 1$

$I(two) = A \rightarrow 1, \; C \rightarrow 1$

$I = 1$

$Info(scale) = \frac{2}{4} \times 1 + \frac{2}{4} \times 1 = 1$

$Gain(scale) = 1 - 1 = 0$

<u>shade:-</u>

$I(light) = A \rightarrow 2, \ C \rightarrow 0$

$I = 0$

$I(dark) = A \rightarrow 0, \ C \rightarrow 2$

$I = 0$

$Info(shade) = \frac{2}{4} \times 0 + \frac{2}{4} \times 0 = 0$

$Gain(shade) = 1 - 0 = 1$

<u>texture:-</u>

$I(thin) = A \rightarrow 2, \ C \rightarrow 2$

$I \Rightarrow -\frac{2}{4} \log\left(\frac{2}{4}\right) - \frac{2}{4} \log\left(\frac{2}{4}\right)$

$I = 1$

$Info(text) = \frac{4}{4} \times 1 = 1$

$Gain(text) = 1 - 1 = 0$

<u>two</u> :-

$$Info(D) = \frac{-2}{3}\log\left(\frac{2}{3}\right) - \frac{1}{3}\log\left(\frac{1}{3}\right)$$

$$= 0.389 + 0.528$$
$$= \underline{0.917}$$

<u>scale</u> :-

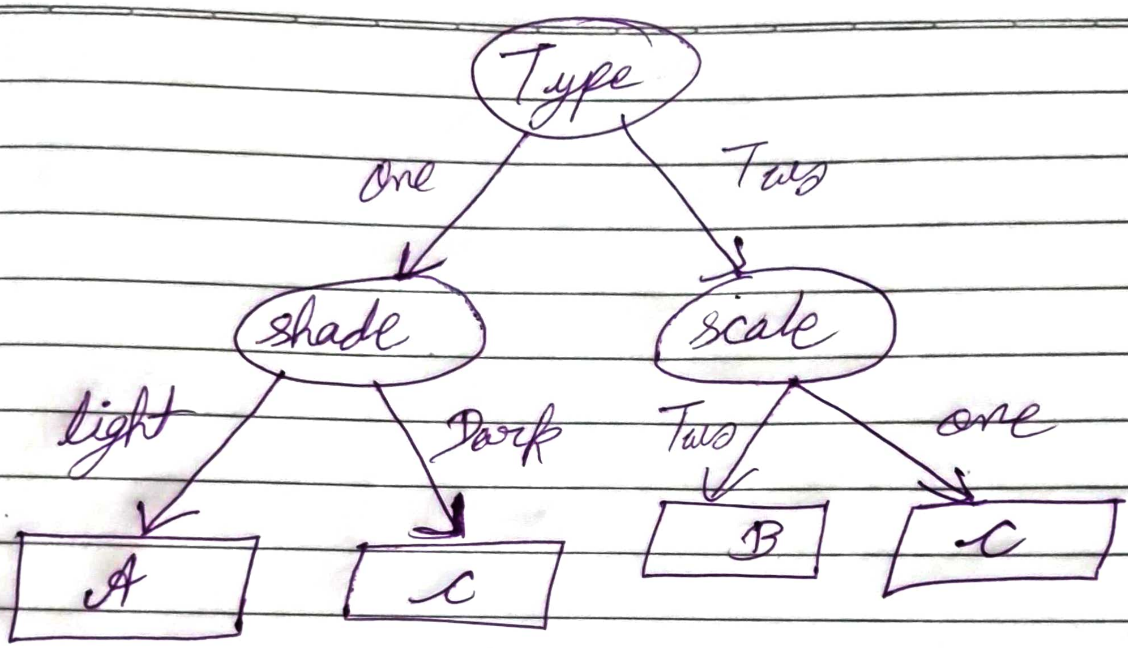
I(two) = $B \to 2$, $C \to 0$
$I = 0$

I(one) = $B \to 0$, $C \to 1$
$I = 0$

$$Info(scale) = \frac{2}{3} \times 0 + \frac{1}{3} \times 0 = 0$$

$$Gain(scale) = 0.917 - 0 = \underline{0.917}$$

There is no need to check others, since this is the maximum gain possible.


Type
One
Two
shade
scale
light
Dark
Two
one
A
C
B
C

Show that the entropy of a node never increases after splitting it into smaller successor nodes.

2

Answer:

Let $Y = \{y_1, y_2, \cdots, y_c\}$ denote the c classes and $X = \{x_1, x_2, \cdots, x_k\}$ denote the k attribute values of an attribute X. Before a node is split on X, the entropy is:

$$E(Y) = -\sum_{j=1}^{c} P(y_j) \log_2 P(y_j) = \sum_{j=1}^{c} \sum_{i=1}^{k} P(x_i, y_j) \log_2 P(y_j), \tag{4.1}$$

where we have used the fact that $P(y_j) = \sum_{i=1}^{k} P(x_i, y_j)$ from the law of total probability.

After splitting on X, the entropy for each child node $X = x_i$ is:

$$E(Y|x_i) = -\sum_{j=1}^{c} P(y_j|x_i) \log_2 P(y_j|x_i) \tag{4.2}$$

where $P(y_j|x_i)$ is the fraction of examples with $X = x_i$ that belong to class y_j. The entropy after splitting on X is given by the weighted entropy of the children nodes:

$$\begin{aligned}
E(Y|X) &= \sum_{i=1}^{k} P(x_i) E(Y|x_i) \\
&= -\sum_{i=1}^{k} \sum_{j=1}^{c} P(x_i) P(y_j|x_i) \log_2 P(y_j|x_i) \\
&= -\sum_{i=1}^{k} \sum_{j=1}^{c} P(x_i, y_j) \log_2 P(y_j|x_i), \qquad (4.3)
\end{aligned}$$

where we have used a known fact from probability theory that $P(x_i, y_j) = P(y_j|x_i) \times P(x_i)$. Note that $E(Y|X)$ is also known as the conditional entropy of Y given X.

To answer this question, we need to show that $E(Y|X) \leq E(Y)$. Let us compute the difference between the entropies after splitting and before splitting, i.e., $E(Y|X) - E(Y)$, using Equations 4.1 and 4.3:

$$\begin{aligned}
& E(Y|X) - E(Y) \\
&= -\sum_{i=1}^{k} \sum_{j=1}^{c} P(x_i, y_j) \log_2 P(y_j|x_i) + \sum_{i=1}^{k} \sum_{j=1}^{c} P(x_i, y_j) \log_2 P(y_j) \\
&= \sum_{i=1}^{k} \sum_{j=1}^{c} P(x_i, y_j) \log_2 \frac{P(y_j)}{P(y_j|x_i)} \\
&= \sum_{i=1}^{k} \sum_{j=1}^{c} P(x_i, y_j) \log_2 \frac{P(x_i) P(y_j)}{P(x_i, y_j)} \qquad (4.4)
\end{aligned}$$

To prove that Equation 4.4 is non-positive, we use the following property a logarithmic function:

$$\sum_{k=1}^{d} a_k \log(z_k) \leq \log\Big(\sum_{k=1}^{d} a_k z_k\Big), \qquad (4.5)$$

subject to the condition that $\sum_{k=1}^{d} a_k = 1$. This property is a special case of a more general theorem involving convex functions (which include the logarithmic function) known as Jensen's inequality.

By applying Jensen's inequality, Equation 4.4 can be bounded as follows:

$$\begin{aligned} E(Y|X) - E(Y) &\leq \log_2 \left[\sum_{i=1}^{k} \sum_{j=1}^{c} P(x_i, y_j) \frac{P(x_i)P(y_j)}{P(x_i, y_j)} \right] \\ &= \log_2 \left[\sum_{i=1}^{k} P(x_i) \sum_{j=1}^{c} P(y_j) \right] \\ &= \log_2(1) \\ &= 0 \end{aligned}$$

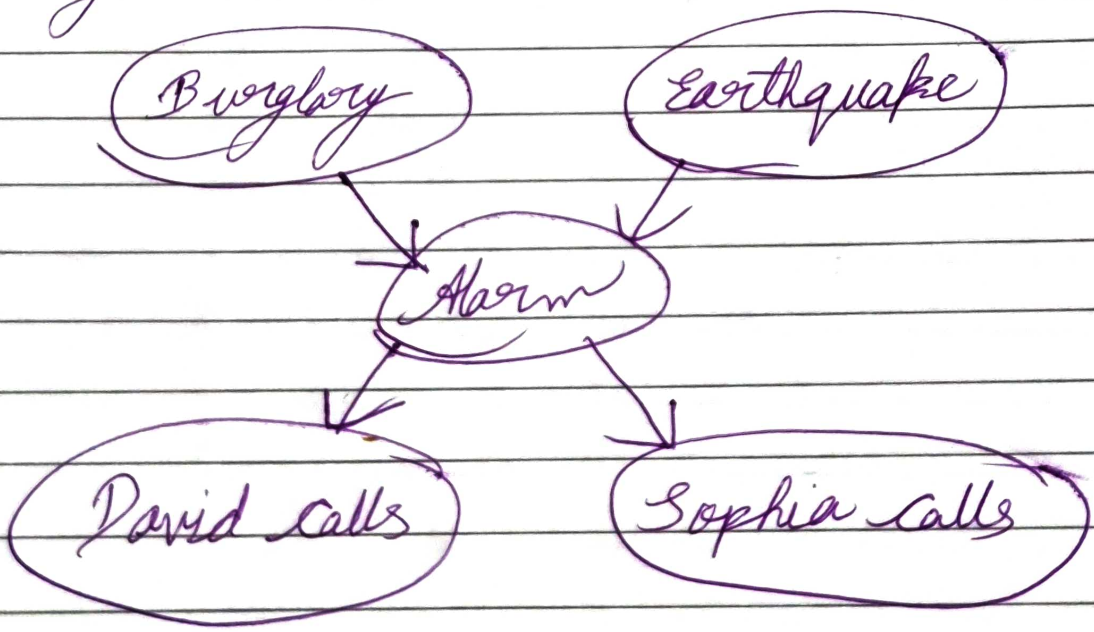
Because $E(Y|X) - E(Y) \leq 0$, it follows that entropy never increases after splitting on an attribute.

3) Naives Bayes theorem assumes conditional independence b/w different event whereas Bayes theorem does not.

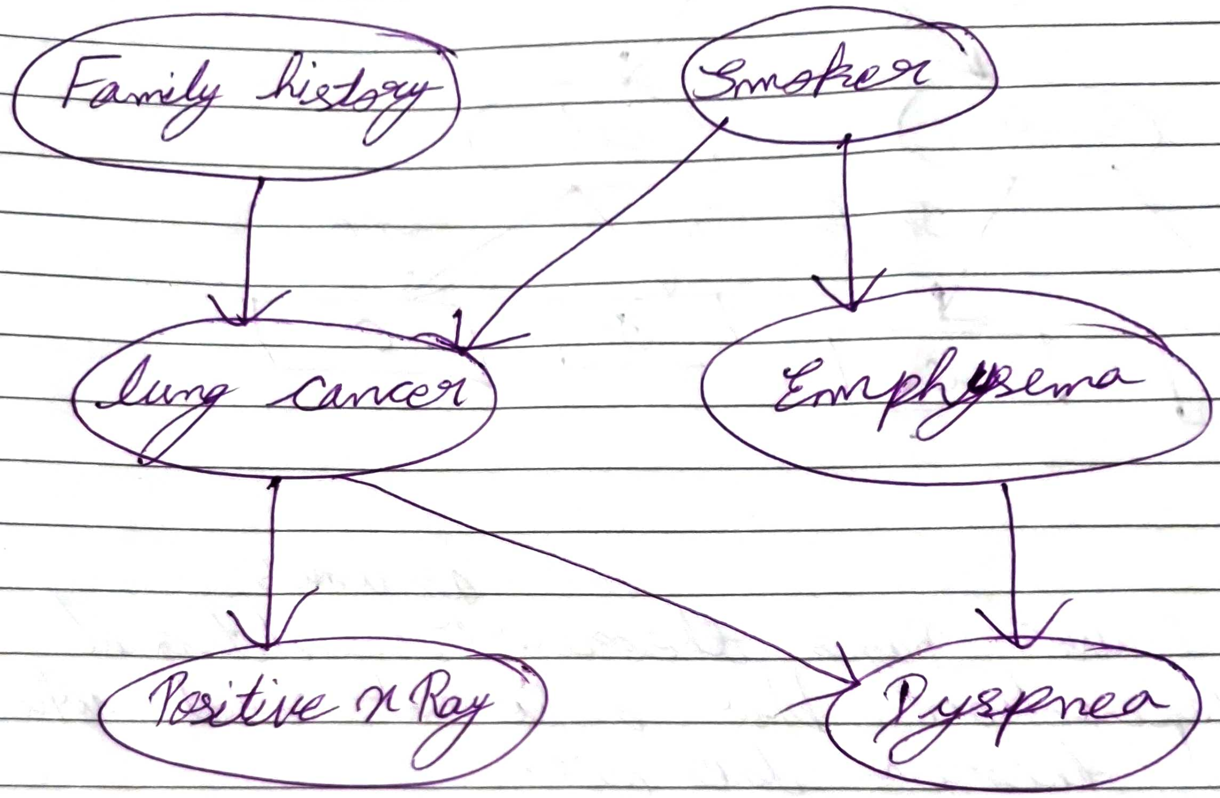
- This means the relationship b/w all i/p features are independent.

- Eg, where we'll use Bayesian belief network:-

i) Burglar Alarm:-



ii) Lung cancer :-



4) $E \Rightarrow$ investment = No, Travel = ~~No~~ Yes, Reading = Yes,
health = No

E_1 is investment = No
E_2 " Travel = Yes
E_3 " Reading = Yes
E_4 " Health = No

$$P(Male/E) = \frac{P(E_1/Male) \cdot P(E_2/Male)\, P(E_3/Male)\, P(E_4/Male) \cdot P(Male)}{P(E)}$$

$$= (1/6) \times (3/6) \times (1/6) \times (5/6) \times (6/10)$$

$$= 0.0069$$

$$P(Female/E) = \frac{P(E_1/Female)\, P(E_2/Female)\, P(E_3/Female)\, P(E_4/Female)\, P(Female)}{P(E)}$$

$$= (3/4) \times (2/4) \times (3/4) \times (1/4) \times (4/10)$$

$$= 0.02$$

Hence, Bayes Classification Algorithm classifies the sex of new tuple as Female. Ans