Ans 3)

We know that $H(q) = -\sum_{i=1}^{K} p_{qK} \log_2 p_{qK}$

For the parent node 'q' ⇒ Entrophy

For a binary feature (Yes/No) $(H(q))$ $= -\sum_{i=1}^{2} p_{qK} \log p_{qK}$

$= -p(Yes/q) \cdot \log_2 (p(Yes/q))$

$\quad - p(No/q) \cdot \log_2 (p(No/q))$

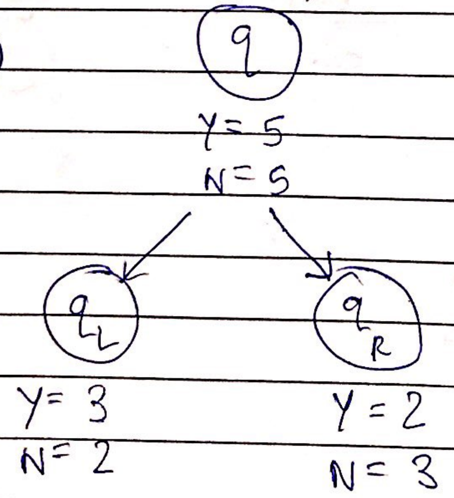
For child, let $q_L \rightarrow$ left child and $q_r \rightarrow$ right child

∴ Decrease in Entrophy between parent & child nodes $= H(q) - \sum_{i=1}^{K} p_{qk} \log_2 p_{qK}$

$$= H(q) - \Big[p_L(Yes) \cdot \log_2 (p_L(Yes)) + p_L(No) \cdot \log_2 (p_L(No))$$

$$+ p_R(Yes) \cdot \log_2 (p_R(Yes)) + p_R(No) \cdot \log_2 (p_R(No)) \Big]$$

eg.)



$H(q) = -\left[\frac{5}{10} \log_2 \frac{5}{10} + \frac{5}{10} \log_2 \frac{5}{10} \right]$

$= 1$

$H(q_L) = -\left[\frac{3}{5} \log_2 \frac{3}{5} + \frac{2}{5} \log_2 \frac{2}{5} \right]$

$= \cancel{-[0.78 + 1.32]}$

$= -[0.44 + 0.53]$

$$= 0.97$$
$$= H(q_R)$$

∴ IG i.e drop in error $= H(q) - \frac{N_L}{N_q} \cdot H(q_L) - \frac{N_R}{N_q} \cdot H(q_R)$

$$= 1 - \frac{5}{10}(0.97) - \frac{5}{10}(0.97)$$

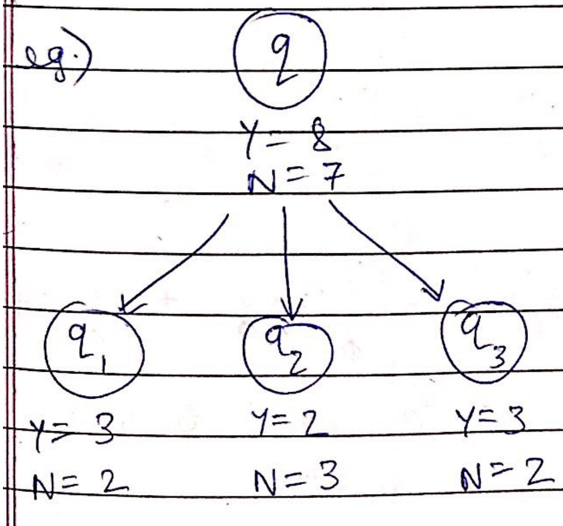
$$= 0.03$$

∴ As can be seen from the above example, for a binary feature $H(q) \in [0, 1]$
And also entrophy of child $\leq$ entrophy of parent as the size of the dataset splits.

∴ $\boxed{\text{Drop in error for Binary feature} \leq 1}$

Same applies to multiway branching as well.

eg.)



$$H(q) = -\left[\frac{8}{15}\log_2\frac{8}{15} + \frac{7}{15}\log_2\frac{7}{15}\right]$$

$$\approx 1$$

$$H(q_1) = 0.97 = H(q_3) = H(q_2)$$

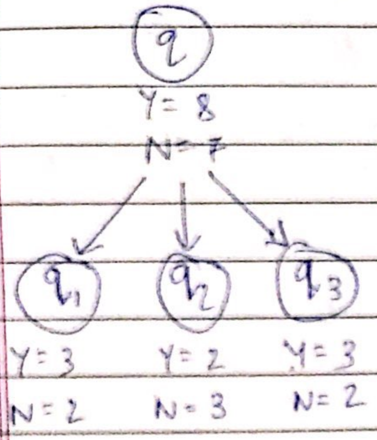
∴ Drop in error i.e IG $= H(q) - \frac{5}{15}(H(q_1)) - \frac{5}{15}(H(q_2)) - \frac{5}{15}(H(q_3))$

$$= 0.03$$

Scanned by CamScanner

Ans 4) Consider the following example :

Example ①



$$H(q) = -\left[\frac{8}{15}\log_2\left(\frac{8}{15}\right) + \frac{7}{15}\log_2\left(\frac{7}{15}\right)\right]$$

$$\approx 1$$

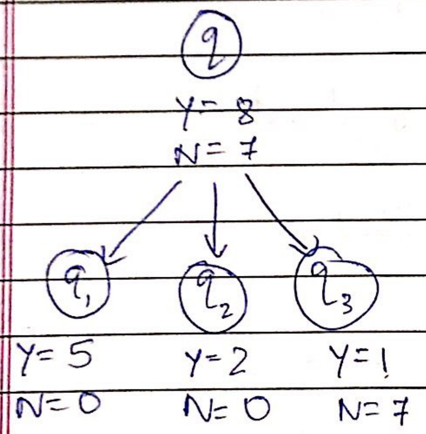
$$H(q_1) = H(q_2) = H(q_3) = 0.97$$

$$\therefore \text{Gain} = H(q) - \left[\frac{5}{15}H(q_1) + \frac{5}{15}H(q_2) + \frac{5}{15}H(q_3)\right]$$

$$= 1 - \frac{1}{3}(3 \times 0.97)$$

$$= 0.03$$

Example ②



$$H(q) \approx 1$$

$$H(q_1) = -\left[\frac{5}{5}\log_2\frac{5}{5}\right] - \left[\frac{0}{5}\log_2\frac{0}{5}\right]$$

$$= 0$$

$$H(q_2) = -\left[\frac{2}{2}\log\frac{2}{2} + \frac{0}{2}\log\frac{0}{2}\right]$$

$$= 0$$

$$H(q_3) = -\left[\frac{1}{8}\log\frac{1}{8} + \frac{7}{8}\log\frac{7}{8}\right]$$

$$= -[0.375 + 0.166]$$

$$= 0.54$$

$$\therefore \text{Gain} = 1 - \left(\frac{5}{15}(0) + \frac{2}{15}(0) + \frac{8}{15}(0.54)\right)$$

$$= 0.71$$

Date : / /

Entrophy is the measure of impurity of a child node.

As can be seen from example 1 & 2 :

Entrophy $\propto \dfrac{1}{\text{purity of } \cancel{\text{child}} \text{ nodes}}$

Since $Gain(E_q, V) = I(q) - \sum_{i=1}^{|V|} \frac{N_i}{N_q} I(i)$

$I(q)$ → Entrophy of parent

$\sum_{i=1}^{|V|} \frac{N_i}{N_q} I(i)$ → sum of proportion of entrophy of each child node

∴ To maximize Gain ⇒ child nodes should be as pure as possible ⇒ Entrophy of child nodes should be less

∴ Gain $\propto \dfrac{1}{\text{Sum of Entrophies of child nodes}}$.

Hence, maximizing gain is equivalent to minimizing impurity of children

Ans 5) Let P_{qK} → probability of getting class 'K' at node 'q'

$P_{qK'}$ → probability of getting all classes <u>except</u> 'K' at node 'q'.

Example

At node 'q' : There are 5 ~~classes~~ instances of 'a', 5 of 'b' and 5 of 'c', where a, b and c are output classes.

$$\therefore P_{q(a)} = \frac{5}{15} = P_{q(b)} = P_{q(c)} = \frac{1}{3}$$

$$\therefore Gini(q) = \sum_{i=1}^{3} P_{qK}(1 - P_{qK})$$

$$= \frac{1}{3}\left(1-\frac{1}{3}\right) + \frac{1}{3}\left(1-\frac{1}{3}\right) + \frac{1}{3}\left(1-\frac{1}{3}\right)$$

$$= \frac{1}{3}\times\frac{2}{3} + \frac{1}{3}\times\frac{2}{3} + \frac{1}{3}\times\frac{2}{3}$$

$$= P_a \times (P_b + P_c) + P_b(P_a + P_c) + P_c(P_a + P_b)$$

$$= \sum_{K \neq K'} P_{qK} \cdot P_{qK'}$$

∴ For a classification problem with $M > 2$, we can equivalently represent:

$$Gini(q) = \sum_{i=1}^{M} P_{qK}(1 - P_{qK}) = \sum_{K \neq K'} P_{qK} \cdot P_{qK'}$$