Question 2 -

$$Info(D) = -\sum_{i=1}^{m} P_i \log_2(P_i)$$

$\therefore Info(D) = 0.9887$

$$Info_{Color}(D) = \left(\frac{3}{16}\right) I(1,2) + \left(\frac{13}{16}\right) I(8,5)$$

$= 0.9532$

$\therefore$ Gain (Color) $= 0.0355$ —— (1)

$$Info_{size}(D) = \left(\frac{8}{16}\right) I(6,2) + \frac{8}{16} I(3,5)$$

$= 0.8828$

$\therefore$ Gain (Size) $= 0.1059$ —— (2)

$$Info_{Shape}(D) = \left(\frac{4}{16}\right) I(3,1) + \left(\frac{12}{16}\right) I(6,6)$$

$= 0.9528$

$\therefore$ Gain (Shape) $= 0.0358$ —— (3)

From (1), (2) and (3) size has the highest gain.

$\therefore$ We split the data based on attribute size.

Scanned by CamScanner

A) Consider Size = Small

$\therefore$ $Info(D) = I(6,2)$
$= 0.8112$

$Info_{Color}(D) = \frac{6}{8} I(5,2) + \frac{2}{8} I(1,1)$
$= 0.7375$

$\therefore$ $Gain(Color) = 0.07368$ —— (4)

$Info_{Shape}(D) = \frac{6}{8} I(4,2) + \frac{2}{8} I(2,0)$
$= 0.68872$

$Gain(Shape) = 0.12247$ —— (5)

From (4) & (5), Shape has the highest gain.
$\therefore$ We split on shape.

i) Consider Shape = Regular.

$Info(D) = I(4,2)$
$= 0.91829$

$Info_{Color}(D) = \left(\frac{5}{6}\right) I(4,1) + \frac{1}{6} I(0,1)$
$= 0.6016$

Scanned by CamScanner

$\therefore$ Gain (Color) = 0.31668

$\therefore$ We split based on Color.

Thus, for Color = Green we can predict Edible? = '−' as it is the only value.

For Color = Yellow, as maximum occurance of Edible? is '+', we predict '+'.

ii) For Shape = Irregular, we predict that Edible? = '+'.

B) Consider Size = Large

$$Info(D) = I(3,5)$$
$$= 0.9544$$

$$Info_{Color}(D) = -\left(\frac{1}{8}\right) I(0,1) + \left(\frac{7}{8}\right) I(3,4)$$
$$= 0.8620$$

$\therefore$ Gain(Color) = 0.0924 —— (6)

$$Info_{Shape}(D) = \left(\frac{2}{8}\right) I(1,1) + \left(\frac{6}{8}\right) I(2,4)$$
$$= 0.9387$$

$\therefore$ Gain(Shape) = 0.0156 —— (7)

From (6) & (7), Color has highest gain.

$\therefore$ We split data based on Color.

Scanned by CamScanner

i) Consider Color = Yellow,

$$Info(D) = I(3,4)$$
$$= 0.9852$$

$$Info_{Shape}(D) = \left(\frac{1}{7}\right) I(1,0) + \left(\frac{6}{7}\right) I(2,4)$$
$$= 0.8571$$

$\therefore$ Gain (Shape) = 0.1280

$\therefore$ We split on Shape.

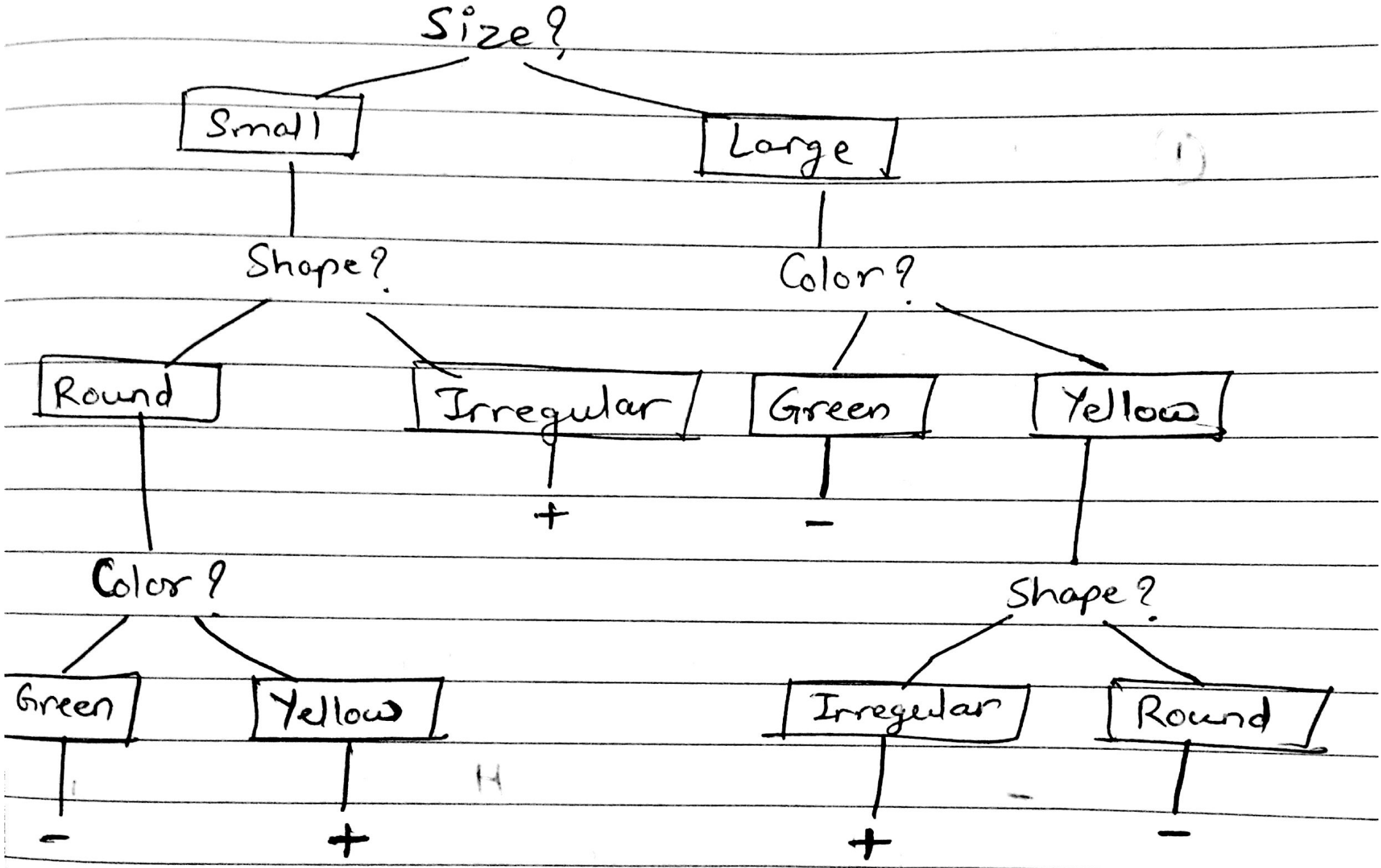
$\therefore$ We predict ~~that~~ for Shape = Irregular, Edible? = '+' as it is the only value.

And for shape = Regular, as maximum occurances are -ve, we predict Edible? = '-'.

ii) Consider, Color = Green,

As there is only one entry, we predict that Edible? = '-'.

Scanned by CamScanner



Scanned by CamScanner

Question 3 -

$\theta_{spam} = P(C_{spam})$

$$\boxed{\theta_{spam} = \frac{3}{7}}$$

$\theta_{secret|spam} = P(secret = 1 | C_{spam})$

$$= \frac{P(secret = 1, C_{spam})}{P(C_{spam})}$$

$$= \frac{2/7}{3/7}$$

$$\boxed{\theta_{secret|spam} = \frac{2}{3}}$$

Scanned by CamScanner

$$\theta_{secret|non\text{-}spam} = \frac{P(secret=1, non,spam)}{P(non\text{-}spam)}$$

$$= \frac{1/7}{4/7}$$

$$\boxed{\theta_{secret|non\text{-}spam} = \frac{1}{4}}$$

$$\theta_{sports|non\text{-}spam} = \frac{P(sports=1, C_{non\text{-}spam})}{P(C_{non\text{-}spam})}$$

$$= \frac{2/7}{4/7}$$

$$\boxed{\theta_{sports|non\text{-}spam} = \frac{1}{2}}$$

$$\theta_{dollar|spam} = \frac{P(dollar=1 \mid C_{spam})}{P(C_{spam})}$$

$$= \frac{1/7}{3/7}$$

$$\boxed{\theta_{dollar|spam} = \frac{1}{3}}$$

Scanned by CamScanner

Question 4 -

① Since α_7, α_8 & α_{20} are non-zero; x_7, x_8 & x_{20} are support vectors

Support vectors :- Proof

For a given sum, a hyper-plane is given as →

$$H = W \cdot x + b = \pm 1$$

If; $\vec{x}_k$ lies on H; then $\vec{x}_k$ is support vectors.

Substituting $x_{k=1} \rightarrow x_{k=20}$; we have x_7 in $\begin{pmatrix} -0.5639 \\ 0.8181 \end{pmatrix} \cdot x_7 + 0.6651 \approx 0.99$

$x_{18} \approx -0.998$

$x_{20} \approx -0.998$

∴ There are 3 support vectors x_7, x_{18} and x_{20}.

Scanned by CamScanner

(2) We know that,

$$\vec{w} = \sum_{i=1}^{N} \alpha_i y_i \vec{x}_i$$

Since $N = 20$ and just 3 data points have 'non-zero' alphas. We have to sum these 3 values.

$$\vec{W} = \alpha_7 y_7 \vec{x}_7 + \alpha_{18} y_{18} \vec{x}_{18} + \alpha_{20} y_{20} \vec{x}_{20}$$

$$\vec{W} = [(0.53, 0.77) \times 1 \times 0.4952]$$
$$+ [(2.05, -0.69) \times -1 \times 0.0459]$$
$$+ [(1.63, -0.91) \times -1 \times 0.4493]$$

$$\vec{W} = (0.2624, 0.3813) + (-0.094, 0.028)$$
$$+ (-0.7323, 0.4088)$$

$$\therefore \vec{W} = (-0.5639, 0.8181)$$

(3) $\cancel{b = \sum_{k:\alpha}}$

$$b = \sum_{k:\alpha_k \neq 0} (y_k - w' x_k) / N_k$$

Again; we just have 3 non-zero 'α' with data point. $N_k = 3$

$$3 \cdot b = y_7 - w' x_7 + y_{18} - w' x_{18}$$
$$+ y_{20} - w' x_{20}$$

Scanned by CamScanner

$$w = (-0.5639,\ 0.8181)$$

$$\therefore\ w' = \begin{pmatrix} -0.5639 \\ 0.8181 \end{pmatrix}$$

$$x_{k=7} = (x_{k_1},\ x_{k_2})' = (0.53,\ 0.77)'$$

$$= \begin{pmatrix} 0.53 \\ 0.77 \end{pmatrix}$$

$$x_{k=18} = (2.05,\ -0.62)'$$

$$= \begin{pmatrix} 2.05 \\ -0.62 \end{pmatrix}$$

$$x_{k=20} = (1.63,\ -0.91)'$$

$$= \begin{pmatrix} 1.63 \\ -0.91 \end{pmatrix}$$

$$3b = 1 - \begin{pmatrix} -0.5639 \\ 0.8181 \end{pmatrix} \cdot \begin{pmatrix} 0.53 \\ 0.77 \end{pmatrix}$$

$$+ (-1) - \begin{pmatrix} -0.5639 \\ 0.8181 \end{pmatrix} \cdot \begin{pmatrix} 2.05 \\ -0.62 \end{pmatrix}$$

$$+ (-1) - \begin{pmatrix} 0.5639 \\ 0.8181 \end{pmatrix} \cdot \begin{pmatrix} 1.63 \\ -0.91 \end{pmatrix}$$

$$\therefore\ 3b = 0.6689 + 0.6631 + 0.6635$$

$$\therefore\ \boxed{b = 0.6651}$$

Scanned by CamScanner

(4) $f(x) = w' x + b$

$w' = (-0.5639, 0.8181)' x + 0.6651$

$\therefore f(x) = \begin{pmatrix} -0.5639 \\ 0.8181 \end{pmatrix} \cdot x + 0.6651$

(5) $\vec{x}_k = (-1, 2)' = \begin{pmatrix} -1 \\ 2 \end{pmatrix}$

$f(-1,2) = \begin{pmatrix} -0.5639 \\ 0.8181 \end{pmatrix} \begin{pmatrix} -1 \\ 2 \end{pmatrix} + 0.6651$

$= 0.5639 + 1.6362 + 0.6651$

$\therefore f(-1,2) = 2.8652 > +1$

$\therefore x(-1,2)$ belongs to class label $+1$.

Scanned by CamScanner

Question 5 -

Given mutual information gain is,

$$I(x;y) = \sum_{x} \sum_{y} p(x,y) \log\left(\frac{p(x,y)}{p(x)\,p(y)}\right)$$

we know that, entropy is defined as,

$$H(p) = -\sum_{x \in X} p(x) \log_2(p(x))$$

Now, we can extend above definition of entropy for 'joint-probability' distribution

$$H(p) = -\sum_{x \in X} \sum_{y \in Y} p(x,y) \log(p(x,y)) \quad \text{—— (1)}$$

We can replace $H(p)$ by $H(x)$ assuming 'p' is ~~that~~ our underlying distribution. Also called conditional ~~probability~~ ~~entry~~ entropy; $H(Y|X)$ is amount of information needed to calculate y; if other party knows X.

$$\therefore H(Y|X) = \sum_{x \in X} p(x)\, H(Y|X=x)$$

$$= \sum_{x \in X} \sum_{y \in Y} p(x,y) \cdot \log(p(x,y)) \quad \text{—— (2)}$$

Scanned by CamScanner

From 'chain rule',

$$H(X,Y) = H(X) + H(Y|X)$$

$$= H(Y) + H(X|Y)$$

or,

$$H(X) - H(X|Y) = H(Y) - H(Y|X)$$

This difference is called "Information Gain".

$$H(X) - H(X|Y) = H(Y) - H(Y|X)$$

$$= \sum_{x,y} p(x,y) \log \left(\frac{p(x,y)}{p(x)\,p(y)} \right)$$

This is mutual information gain $I(X;Y)$.

$\therefore$ Information gain in decision tree is synonymous with mutual information gain between two random variables.

Scanned by CamScanner