ECE 5464 SP24 - Prof. Jones - HW2

Part I

First, calculating the entropy of target feature in given data, i.e. "Edible".

Entropy is given by,

$$H(t,D) = -\sum_{l \in levels(t)} \left[P(t=l) \log_2 P(t=l) \right] \quad \text{--- } 1$$

$$\Rightarrow -\left[P(t=0) \log_2 P(t=0)\right] - \left[P(t=1) \log_2 (t=1)\right]$$

$$= -\left[\frac{7}{24} \log_2 \frac{7}{24} + \frac{17}{24} \log_2 \frac{17}{24}\right]$$

$$\therefore H(t,D) = 0.87086 \text{ bits.}$$

Now we need to calculate the remainder for each feature,

remainder is given by,

$$rem(d,D) = \sum_{l \in levels(d)} \frac{|D_{d=l}|}{|D|} \times H(t, D_{d=l}) \quad \text{--- } 2$$

, where, $H(t, D_{d=l})$ is the entropy of target feature given the condition of each feature.

$$\Rightarrow rem(White, D) = \frac{|D_{White=0}|}{|D|} \times H(t, D_{White=0}) + \frac{|D_{White=1}|}{|D|} \times H(t, D_{White=1})$$

$$= -\frac{14}{24}\left[\frac{4}{14} \times \log_2 \frac{4}{14} + \frac{10}{14}\log_2\frac{10}{14}\right] - \frac{10}{24}\left[\frac{3}{10}\log_2\frac{3}{10} + \frac{7}{10}\log_2\frac{7}{10}\right]$$

$$= -\frac{7}{12}\left[-0.863\right] - \frac{5}{12}\left[-0.881\right]$$

$\therefore$ rem(White, D) = 0.8705 bits

Now for the 'Tall' attribute,

From eq^n (2).

$$\text{rem}(\text{Tall}, D) = \frac{|D_{Tall=0}|}{|D|} \times H(t, D_{Tall=0}) + \frac{|D_{Tall=1}|}{|D|} \times H(t, D_{Tall=1})$$

$$= -\frac{10}{24}\left[\frac{4}{10}\log_2\frac{4}{10} + \frac{6}{10}\log_2\frac{6}{10}\right] - \frac{14}{24}\left[\frac{3}{14}\log_2\frac{3}{14} + \frac{11}{14}\log_2\frac{11}{14}\right]$$

$\Rightarrow$ rem(Tall, D) = 0.84137 bits

For "frilly" attribute, from eq^n (2).

$$\text{rem}(\text{Frilly}, D) = \frac{|D_{Frilly=0}|}{|D|} \times H(t, D_{Frilly=0}) + \frac{|D_{Frilly=1}|}{|D|} \times H(t, D_{Frilly=1})$$

$$= -\frac{16}{24}\left[\frac{3}{16}\log_2\frac{3}{16} + \frac{13}{16}\log_2\frac{13}{16}\right] - \frac{8}{24}\left[\frac{4}{8}\log_2\frac{4}{8} + \frac{4}{8}\log_2\frac{4}{8}\right]$$

$\therefore$ rem(frilly, D) = 0.79740 bits

Now, to calculate the Information gain, we need to pick the maximum amongst information gain from each attribute.

Information Gain is given by,

$$IG(d,D) = H(t,D) - rem(d,D) \quad \text{(3)}$$

$$\Rightarrow IG(White, D) = H(Edible, D) - rem(White, D).$$

$$\Rightarrow 0.87086 - 0.8705 = 0.0003 \text{ bits.}$$

$$IG(Tall, D) = H(Edible, D) - rem(Tall, D)$$

$$\Rightarrow 0.87086 - 0.84137 = 0.02949 \text{ bits}$$

$$IG(frilly, D) = H(Edible, D) - rem(frilly, D).$$

$$\Rightarrow 0.87086 - 0.79740 = 0.0734 \text{ bits}$$

$\therefore$ $IG(Frilly, D)$ is the highest, thus we select "frilly" as the 1^{st} decision.

Now, for the second decision, we ~~make~~ calculate Entropy of target feature while "frilly = 0" and then while "frilly = 1"

$$H(Edible, D_{frilly=0}) = -\sum_{levels \in (0,1)} P(t=l) \log_2 P(t=l)$$

$$= -\left[\frac{3}{16}\log_2\frac{3}{16} + \frac{13}{16}\log_2\frac{13}{16}\right] = 0.6962 \text{ bits}$$

Now, we need to calculate remainders of remaining 2 features,

$$\Rightarrow rem(White, D_{Frilly=0}) = \frac{|D_{F=0,White=0}|}{|D|} \times H(Edible, D_{F=0,W}) + \frac{|D_{F=0,White=1}|}{|D|} \times H(Edible, D_{F=0,W})$$

$$= -\frac{9}{16}\left[\frac{3}{9}\log_2\frac{3}{9} + \frac{6}{9}\log_2\frac{6}{9}\right] - \frac{7}{16}\left[\frac{0}{7}\log_2\frac{0}{7} + \frac{7}{7}\log_2\frac{7}{7}\right]$$

~~= 0.6462 bits.~~

= 0.5165 bits

Now for "Tall" attribute,

$$\Rightarrow rem(Tall, D_{Frilly=0}) = -\frac{7}{16}\left[\frac{3}{7}\log_2\frac{3}{7} + \frac{4}{7}\log_2\frac{4}{7}\right] - \frac{9}{16}\underbrace{\left[\frac{0}{9}\log_2\frac{0}{9} + \frac{9}{9}\log_2\frac{9}{9}\right]}_{0}$$

$= 0.4310$ bits

Now to calculate information gain for 2nd Decision,

using eqn 3,

$$IG(White, D_{Frilly=0}) = H(Edible, D_{Frilly=0}) - rem(White, D_{Frilly=0})$$

$= 0.1797$ bits

$$IG(Tall, D_{Frilly=0}) = H(Edible, D_{Frilly=0}) - rem(Tall, D_{Frill=0})$$

$= 0.2625$ bits

$\therefore$ $IG(Tall, D_{Frilly=0})$ gives highest info gain, so we select attribute "Tall" for second decision.

Now for last decision under Frilly = 0, we do same calculation,
while, Frilly = 0 and Tall = 0.

$$\Rightarrow H(Edible, D_{F=0,T=0}) = -\left[\frac{3}{7}\log_2\frac{3}{7} + \frac{4}{7}\log_2\frac{4}{7}\right]$$

$$= 0.98522 \text{ bits.}$$

$$\Rightarrow rem(white, D_{F=0,T=0}) = -\frac{4}{7}\underbrace{\left[\frac{3}{4}\log_2\frac{3}{4} + \frac{1}{4}\log_2\frac{1}{4}\right]}_{(\text{for } W=0)} - \frac{3}{7}\underbrace{\left[\frac{0}{3}\log_2\frac{0}{3} + \frac{3}{3}\log_2\frac{3}{3}\right]}_{(\text{for } W=1)}$$

$$= 0.46358 + \underline{\underline{0}} \text{ bits}$$

∴ We can conclude these on frilly = 0 path,

if T=0, W=1 ⇒ Edible = 1

T=0, W=0 ⇒ Edible = mostly 0 (low entropy)

T=1, W = does not matter ⇒ Edible = 1

Now doing the same calculation again for Frilly = 1 path,

$$\Rightarrow H(Edible, D_{F=1}) = -\left[\frac{4}{8}\log_2\frac{4}{8} + \frac{4}{8}\log_2\frac{4}{8}\right]$$

$$= 1 \text{ bits.}$$

$$\Rightarrow rem(White, D_{F=1}) = -\frac{5}{8}\underbrace{\left[\frac{1}{5}\log_2\frac{1}{5} + \frac{4}{5}\log_2\frac{4}{5}\right]}_{(\text{for } W=0)} - \frac{3}{8}\underbrace{\overbrace{\left[\frac{0}{3}\log_2\frac{0}{3} + \frac{3}{3}\log_2\frac{3}{3}\right]}^{0}}_{(\text{for } W=1)}$$

$$= 0.4512 \text{ bits}$$

$\Rightarrow$ rem(Tall, $D_{F=1}$) $= -\frac{3}{8}\left[\frac{1}{3}\log_2\frac{1}{3} + \frac{2}{3}\log_2\frac{2}{3}\right] - \frac{5}{8}\left[\frac{3}{5}\log_2\frac{3}{5} + \frac{2}{5}\log_2\frac{2}{5}\right]$

$= 0.9508$ bits

Now, Information gains :-

$IG(\text{White}, D_{F=1}) = 1 - 0.4512 = 0.5488$ bits

$IG(\text{Tall}, D_{F=1}) = 1 - 0.9508 = 0.0492$ bits.

$\therefore$ $IG(\text{White}, D_{F=1})$ is greater and we select that for our 2nd decisia in "frilly = 1" path.

Now, calculating same things for frilly = 1, White = 0.

$\Rightarrow H(\text{Edible}, D_{F=1,W=0}) = -\left[\frac{1}{5}\log_2\frac{1}{5} + \frac{4}{5}\log_2\frac{4}{5}\right] = 0.7219$ bits

$\Rightarrow$ rem(Tall, $D_{F=1,W=0}$) $= -\frac{2}{5}\overbrace{\left[\frac{0}{2}\log_2\frac{0}{2} + \frac{2}{2}\log_2\frac{2}{2}\right]}^{0} - \frac{3}{5}\left[\frac{1}{3}\log_2\frac{1}{3} + \frac{2}{3}\log_2\frac{2}{3}\right]$

(for T = 0) (for T = 1)

$= 0.55097$ bits.

$\therefore$ We can conclude the following on "frilly = 1" path.

if W = 0, T = 0 $\Rightarrow$ Edible = 1

W = 0, T = 1 $\Rightarrow$ Edible = mostly 1

W = 1, T = doesn't matter $\Rightarrow$ Edible = 0

Now constructing the tree with the calculated information:-