

## Decision Tree & Random Forests

Entropy of dataset :-

$$\begin{aligned} H(S) &= -p(\text{positive}) \cdot \log_2(p(\text{positive})) - p(\text{negative}) \cdot \log_2(p(\text{negative})) \\ &= -\frac{5}{10} \cdot \log_2\left(\frac{5}{10}\right) - \frac{5}{10} \log_2\left(\frac{5}{10}\right) \\ &= \log_2\left(\frac{10}{5}\right) = 1. \end{aligned}$$

$$H(S) = 1$$

# 1<sup>st</sup> iteration :

For each attribute, calculating information gain :-

(i) Smell

$$\begin{aligned} H(\text{smell} = \text{woody}) &= -\frac{3}{5} \log_2\left(\frac{3}{5}\right) - \frac{2}{5} \log_2\left(\frac{2}{5}\right) \\ &= 0.971 \end{aligned}$$

$$H(\text{smell} = \text{fruity}) = 0.971$$

Avg. entropy information for smell :-

$$\begin{aligned} I(\text{smell}) &= p(\text{woody}) \cdot H(\text{smell} = \text{woody}) + p(\text{fruity}) \cdot H(\text{smell} = \text{fruity}) \\ &= \left(\frac{1}{2} \times 0.971\right) \times 2 \\ &= 0.971. \end{aligned}$$

$$\begin{aligned}\text{Info gain (smell)} &= H(S) - I(\text{smell}) \\ &= 1 - 0.971\end{aligned}$$

(ii) Taste

$$H(\text{taste} = \text{sweet}) = -1 \log_2(1) = 0$$

$$H(\text{taste} = \text{salty}) = \left[ -\frac{1}{2} \log_2\left(\frac{1}{2}\right) \right] \times 2 = 1$$

$$H(\text{taste} = \text{sour}) = -1 \log_2(1) = 0$$

$$\begin{aligned}I(\text{taste}) &= \left(\frac{3}{10} \times 0\right) + \left(\frac{4}{10} \times 1\right) + \left(\frac{3}{10} \times 0\right) \\ &= 0.4\end{aligned}$$

$$\text{Info gain (taste)} = 1 - 0.4 = \boxed{0.6}$$

(iii) Portion

$$H(\text{portion} = \text{small}) = -\frac{1}{5} \log\left(\frac{1}{5}\right) - \frac{4}{5} \log\left(\frac{4}{5}\right)$$

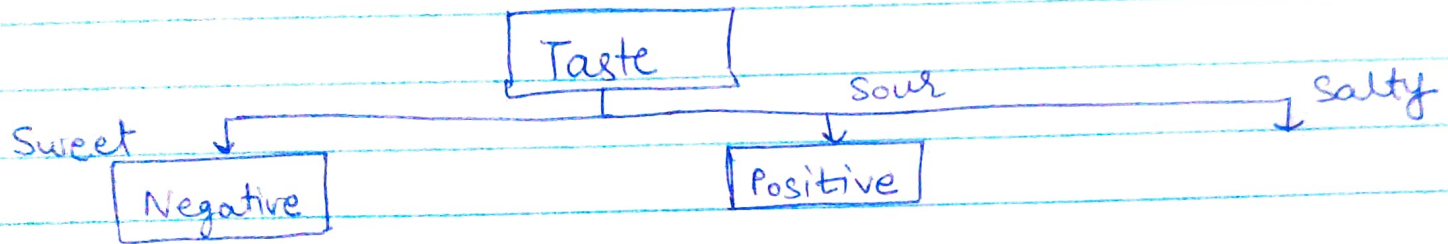
$$H(\text{---} = \text{large}) = -\frac{4}{5} \log\left(\frac{4}{5}\right) - \frac{1}{5} \log\left(\frac{1}{5}\right)$$

$$I(\text{portion}) = \left[ \frac{1}{2} \times \frac{\log 5 + 4 \log(5/4)}{5} \right] \times 2$$

$$\text{Info gain (portion)} = 1 - I(\text{portion}) = \boxed{0.278}$$

Info gain in taste is maximum so, we will split it into taste.

# 2<sup>nd</sup> iteration



$$\bullet H(\text{taste} = \text{sour}) = -1 \log(1) = 0$$

∴ When taste = sour, then review is positive (terminate).

$$\bullet H(\text{taste} = \text{sweet}) = -\frac{3}{3} \log\left(\frac{3}{3}\right) = 0$$

∴ Review is always negative, entropy = 0 (terminate).

$$\bullet H(\text{taste} = \text{salty}) = -\left[\frac{1}{2} \log\left(\frac{1}{2}\right) + \frac{1}{2} \log\left(\frac{1}{2}\right)\right] \times 2 = 1$$



(i) Smell

$$H(\text{salty, smell} = \text{woody}) = 0$$

$$H(\text{--- " --- fruity}) = -\frac{1}{2} \log\left(\frac{1}{2}\right) = 1$$

$$I(\text{salty, smell}) = \cancel{1 - 1} = 1$$

$$\text{Info gain} = 1 - 1 = \boxed{0}$$

(ii) Portion

$$H(\text{salty, portion} = \text{small}) = \left[ \frac{1}{2} \log\left(\frac{1}{2}\right) \right] \times 2 = 1$$

$$H(\text{--- " --- large}) = -1 \log(1) = 0$$

$$I(\text{salty, portion}) = \cancel{1} = 1$$

$$\text{Info gain}(\text{salty, portion}) = \boxed{1}$$

∴ We split through portion (max info gain).

