Nicholas Garde
227006946
CS 633 ML
10/10/2024

1 Information Gain

(1)

$$InfoGain = H_{parent} - \frac{n_l H_l + n_r H_r}{n_l + n_r}$$

$$H_{root} = -(\frac{1}{3}log_3\frac{1}{3}) = 1$$

$$\frac{\partial f}{\partial y} = 3y^2 + \frac{1}{y} + x$$

$$H = -\sum^{labels} p(label)log_3 p(label)$$

(X1)
$(X_1 = 0)$

$$p_1 = 0, p_2 = 0.5, p_3 = 0.5$$

$$H_l = -2(0.5log_3 0.5) = log_3 2 \approx 0.63$$

$(X_1 = 1)$

$$p_1 = 0.5, p_2 = 0.25, p_3 = 0.25$$

$$H_r = -(0.5log_3 0.5 + 2 * 0.25log_3 0.25) = 1.5log_3 2 \approx 0.95$$

InfoGain(X1)

$$n_l = 2, n_r = 4$$

$$InfoGain = 1 - \frac{2log_3 2 + 4 * 1.5log_3 2}{6} = 1 - \frac{4log_3 2}{3} \approx 0.159$$

(X2)
$(X_2 = 0)$

$$p_1 = 0, p_2 = \frac{1}{3}, p_3 = \frac{2}{3}$$

$$H_l = -(\frac{1}{3}log_3\frac{1}{3} + \frac{2}{3}log_3\frac{2}{3}) = \frac{1}{3} - \frac{2}{3}log_3\frac{2}{3} \approx 0.579$$

$(X_2 = 1)$

$$p_1 = \frac{2}{3}, p_2 = \frac{1}{3}, p_3 = 0$$

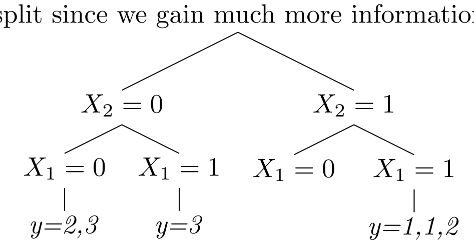
$$H_r = -(\frac{2}{3}log_3\frac{2}{3} + \frac{1}{3}log_3\frac{1}{3}) = \frac{1}{3} - \frac{2}{3}log_3\frac{2}{3} \approx 0.579$$

InfoGain(X2)

$$n_l = 3, n_r = 3$$

$$InfoGain = 1 - \frac{6(\frac{1}{3} - \frac{2}{3}log_3\frac{2}{3})}{6} = 1 - (\frac{1}{3} - \frac{2}{3}log_3\frac{2}{3}) \approx 0.913$$

(2)
We should use X_2 for the first split since we gain much more information from it than X_1



(3)
This tree would likely fail to classify $X_1 = 0, X_2 = 1$ because it has never seen data like that when it was built , the bin for it is empty. If I were to classify the datapoint based on this tree, I would label the bin 2. Our initial data shows that $X_1 = 0$ points are labelled 2 or 3, but no $X_2 = 1$ points are labelled 3 and only the one labelled 2, therefore I think it is reasonable to label this bin 2.

2 Conditional Entropy

(1)

$$H(X|Y) = -\sum_{m=1}^{M}\sum_{n=1}^{N} p_Y(y_m)p_{X|Y}(x_n|y_m)log_3(p_{X|Y}(x_n|y_m))$$

(X1)

$$p_1 = p_2 = p_3 = \frac{1}{3}$$

$$p_{1|X_1=0} = 0, p_{1|X_1=1} = \frac{1}{2}$$

$$p_{2|X_1=0} = \frac{1}{2}, p_{2|X_1=1} = \frac{1}{4}$$

$$p_{3|X_1=0} = \frac{1}{2}, p_{3|X_1=1} = \frac{1}{4}$$

H(X1)

$$-[p_1 p_{1|X_1=0} log_3 p_{1|X_1=0}+$$

$$p_1 p_{1|X_1=1} log_3 p_{1|X_1=1}+$$

$$p_2 p_{2|X_1=0} log_3 p_{2|X_1=0}+$$

$$p_2 p_{2|X_1=1} log_3 p_{2|X_1=1}+$$

$$p_3 p_{3|X_1=0} log_3 p_{3|X_1=0}+$$

$$p_3 p_{3|X_1=1} log_3 p_{3|X_1=1}]$$

$$0 + \frac{1}{3}\frac{1}{2}log_3\frac{1}{2} + \frac{1}{3}\frac{1}{2}log_3\frac{1}{2} + \frac{1}{3}\frac{1}{4}log_3\frac{1}{4} + \frac{1}{3}\frac{1}{2}log_3\frac{1}{2} + \frac{1}{3}\frac{1}{4}log_3\frac{1}{4} = 0.753$$

(X2)

$$p_1 = p_2 = p_3 = \frac{1}{3}$$

$$p_{1|X_1=0} = 0, p_{1|X_1=1} = \frac{2}{3}$$

$$p_{2|X_1=0} = \frac{1}{3}, p_{2|X_1=1} = \frac{1}{3}$$

$$p_{3|X_1=0} = \frac{2}{3}, p_{3|X_1=1} = 0$$

H(X2)

$$-[p_1 p_{1|X_2=0} log_3 p_{1|X_2=0}+$$

$$p_1 p_{1|X_2=1} log_3 p_{1|X_2=1}+$$

$$p_2 p_{2|X_2=0} log_3 p_{2|X_2=0}+$$

$$p_2 p_{2|X_2=1} log_3 p_{2|X_2=1}+$$

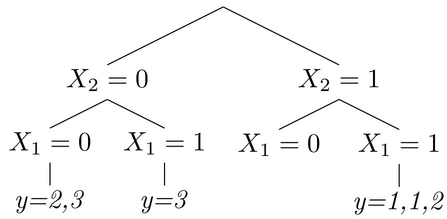
$$p_3 p_{3|X_2=0} log_3 p_{3|X_2=0}+$$

$$p_3 p_{3|X_2=1} log_3 p_{3|X_2=1}]$$

$$0 + \frac{1}{3}\frac{2}{3}log_3\frac{2}{3} + \frac{1}{3}\frac{1}{3}log_3\frac{1}{3} + \frac{1}{3}\frac{1}{3}log_3\frac{1}{3} + \frac{1}{3}\frac{2}{3}log_3\frac{2}{3} + 0 = 0.184$$

(2)

The entropy of X_1 is far greater than that of X_2. Since entropy is a measure of uncertainty, we will again use X_2 for the first split.



(3)

This tree would likely fail to classify $X_1 = 0, X_2 = 1$ because it has never seen data like that when it was built , the bin for it is empty. If I were to classify the datapoint based on this tree, I would label the bin 2. Our initial data shows that $X_1 = 0$ points are labelled 2 or 3, but no $X_2 = 1$ points are labelled 3 and only the one labelled 2, therefore I think it is reasonable to label this bin 2.