Machine Learning

Assignment 5.3

Submitted By: Ranji Raj

November 23, 2020

Formulas used

$$Entropy(S) = \sum_{i=1}^{c} -p_i \log_2 p_i$$

$$Gain(S, A) = Entropy(S) - Entropy(A)$$

$$SplitInformation(S, A) = -\sum_{i=1}^{c} \frac{|S_i|}{|S|} \log_2 \frac{|S_i|}{|S|}$$

$$GainRatio(S, A) = \frac{Gain(S, A)}{SplitInformation(S, A)}$$

a) perfect split, hence, the split points 2.5, 3.5, 5.5, 9.5

Instance	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
x	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
t	0	0	0	1	0	0	1	1	1	1	0	0	0	0	0

Table 1: Specifying split points 2.5, 3.5, 5.5, 9.5

Calculating system entropy,

$$Entropy(S) = -p_{t=1} \log_2 p_{t=1} - p_{t=0} \log_2 p_{t=0}$$

$$Entropy(S) = -\frac{5}{15} \log_2 \frac{5}{15} - \frac{10}{15} \log_2 \frac{10}{15} = \mathbf{0.9182}$$

Calculating entropy of attributes,

$$Entropy(x < 2.5) = -\frac{0}{3} \log_2 \frac{0}{3} - \frac{3}{3} \log_2 \frac{3}{3} = 0.0$$

$$Entropy(x \geq 2.5, x < 3.5) = -\frac{1}{1}\log_2\frac{1}{1} - \frac{0}{1}\log_2\frac{0}{1} = 0.0$$

$$Entropy(x \geq 3.5, x < 5.5) = -\frac{0}{2}\log_2\frac{0}{2} - \frac{2}{2}\log_2\frac{2}{2} = 0.0$$

$$Entropy(x \geq 5.5, x < 9.5) = -\frac{4}{4}\log_2\frac{4}{4} - \frac{0}{4}\log_2\frac{0}{4} = 0.0$$

$$Entropy(x \geq 9.5) = -\frac{0}{5}\log_2\frac{0}{5} - \frac{5}{5}\log_2\frac{5}{5} = 0.0$$

So,

$$Gain(S, a) = Entropy(S) - \sum_{i=1}^{5} Entropy(a) = 0.9182 - 0 = \mathbf{0.9182}$$

Calculating Split Information,

$$SplitInformation(S, a) = -\frac{3}{15}\log_2\frac{3}{15} - \frac{1}{15}\log_2\frac{1}{15} - \frac{2}{15}\log_2\frac{2}{15} - \frac{4}{15}\log_2\frac{4}{15} - \frac{5}{15}\log_2\frac{5}{15} = \mathbf{2.14}$$

Calculating Gain Ratio,

$$GainRatio(S, a) = \frac{Gain(S, a)}{SplitInformation(S, a)}$$

$$\therefore \quad GainRatio(S, a) = \frac{0.9182}{2.14} = \mathbf{0.4290}$$

b) the split points 5.5, 9.5

Instance	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
x	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
t	0	0	0	1	0	0	1	1	1	1	0	0	0	0	0

Table 2: Specifying split points 5.5, 9.5

Calculating entropy of attributes,

$$Entropy(x < 5.5) = -\frac{1}{6}\log_2\frac{1}{6} - \frac{5}{6}\log_2\frac{5}{6} = 0.65$$

$$Entropy(x \geq 5.5, x < 9.5) = -\frac{4}{4}\log_2\frac{4}{4} - \frac{0}{4}\log_2\frac{0}{4} = 0.0$$

$$Entropy(x \geq 9.5) = -\frac{0}{5}\log_2\frac{0}{5} - \frac{5}{5}\log_2\frac{5}{5} = 0.0$$

So,

$$Gain(S, b) = Entropy(S) - \sum_{i=1}^{3} Entropy(b) = 0.9182 - (0.65 * \frac{6}{15}) = \mathbf{0.65}$$

Calculating Split Information,

$$SplitInformation(S,b) = -\frac{6}{15}\log_2\frac{6}{15} - \frac{4}{15}\log_2\frac{4}{15} - \frac{5}{15}\log_2\frac{5}{15} = \mathbf{1.56}$$

Calculating Gain Ratio,

$$GainRatio(S,b) = \frac{Gain(S,b)}{SplitInformation(S,b)}$$

$$\therefore \quad GainRatio(S,b) = \frac{0.65}{1.56} = \mathbf{0.4167}$$

c) the split point 9.5

Instance	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
x	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
t	0	0	0	1	0	0	1	1	1	1	0	0	0	0	0

Table 3: Specifying split point 9.5

Calculating entropy of attributes,

$$Entropy(x < 9.5) = -\frac{5}{10}\log_2\frac{5}{10} - \frac{5}{10}\log_2\frac{5}{10} = 1.0$$

$$Entropy(x \geq 9.5) = -\frac{0}{5}\log_2\frac{0}{5} - \frac{5}{5}\log_2\frac{5}{5} = 0.0$$

So,

$$Gain(S,c) = Entropy(S) - \sum_{i=1}^{2} Entropy(c) = 0.9182 - (1.0 * \frac{10}{15}) = \mathbf{0.251}$$

Calculating Split Information,

$$SplitInformation(S,c) = -\frac{10}{15}\log_2\frac{10}{15} - \frac{5}{15}\log_2\frac{5}{15} = 0.9182$$

Calculating Gain Ratio,

$$GainRatio(S,c) = \frac{Gain(S,c)}{SplitInformation(S,c)}$$

$$\therefore \quad GainRatio(S,c) = \frac{0.251}{0.9182} = \mathbf{0.273}$$