

HW4

income	p_i	n_i	p_i, n_i
high	2	2	1
medium	4	1	0.7219
low	2	1	0.9182

$$I(2,2) = -\frac{2}{4} \log_2 \left(\frac{2}{4} \right) - \frac{2}{4} \log_2 \left(\frac{2}{4} \right) = 1$$

$$I(4,1) = -\frac{4}{5} \log_2 \left(\frac{4}{5} \right) - \frac{1}{5} \log_2 \left(\frac{1}{5} \right) = 0.7219$$

$$I(2,1) = -\frac{2}{3} \log_2 \left(\frac{2}{3} \right) - \frac{1}{3} \log_2 \left(\frac{1}{3} \right) = 0.9182$$

$$\text{Info}_{\text{income}}(D) = \frac{4}{12} I(2,2), \frac{5}{12} I(4,1), \frac{3}{12} I(2,1)$$

$$= \frac{4}{12} (1), \frac{5}{12} (0.7219), \frac{3}{12} (0.9182) = 0.8637$$

$$\text{Gain}(\text{income}) = \text{Info}(D) - \text{Info}_{\text{income}} = 0.9183 - 0.8637 = 0.0546$$

age	p_i	n_i	$I(p_i, n_i)$
< 20	2	2	1
21... 40	3	0	0
> 40	3	2	0.9182

$$I(2,2) = -\frac{2}{4} \log_2 \left(\frac{2}{4} \right) - \frac{2}{4} \log_2 \left(\frac{2}{4} \right) = 1$$

$$I(3,0) = -\frac{3}{3} \log_2 \left(\frac{3}{3} \right) - \frac{0}{3} \log_2 \left(\frac{0}{3} \right) = 0$$

$$I(3,2) = -\frac{3}{5} \log_2 \left(\frac{3}{5} \right) - \frac{2}{5} \log_2 \left(\frac{2}{5} \right) = 0.991$$

$$\text{Info}(D) = I(8,4) = -\frac{8}{12} \log_2 \left(\frac{8}{12} \right) - \frac{4}{12} \log_2 \left(\frac{4}{12} \right) = 0.9183$$

$$\text{Info}_{\text{age}}(D) = \frac{4}{12} I(2,2), \frac{3}{12} I(3,0), \frac{5}{12} I(3,2)$$

$$= \frac{4}{12} (1), \frac{3}{12} (0), \frac{5}{12} (0.991) = 0.7779$$

$$\begin{aligned} \text{Gain}(\text{age}) &= \text{Info}(D) - \text{Info}_{\text{age}}(D) \\ &= 0.9183 - 0.7779 \\ &= 0.1404 \end{aligned}$$