

21-51-01

Q (a) (i) tree

(i)

(iii)

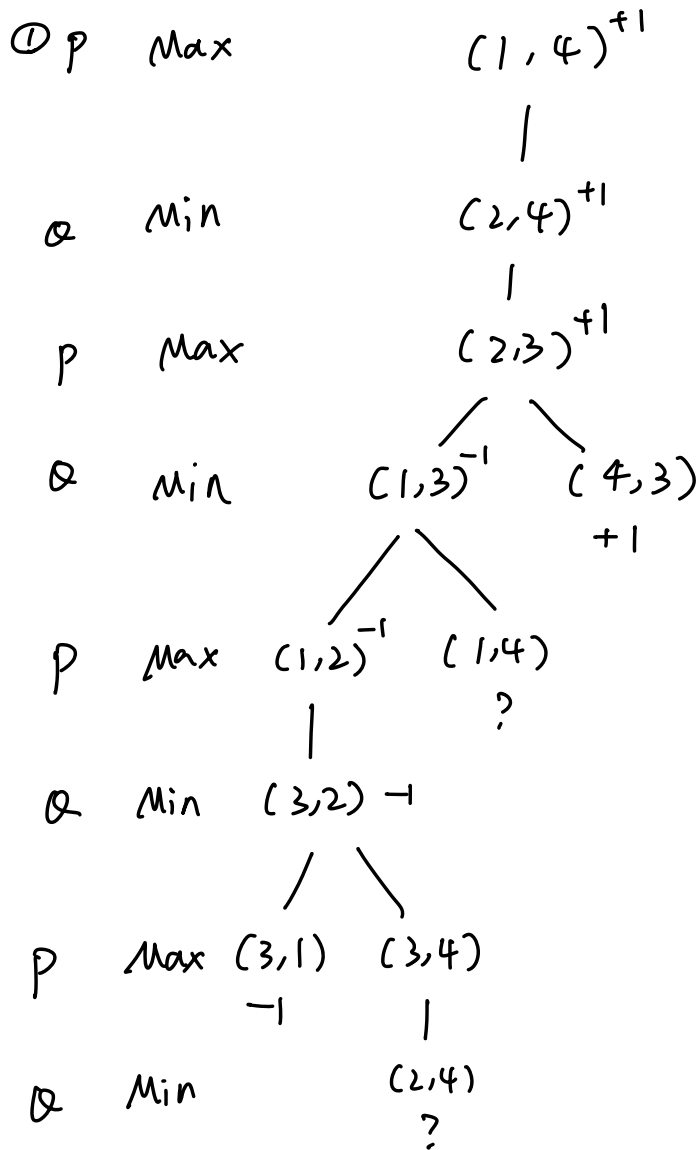
(b) (i)

(ii)

(iii)

(iv)

Solution (a) (i) State space search tree



(ii) Annotate ?

(iii) ① we do not expand "?" which prevents an infinite tree.

② ignore the "?" branch

③ If all children of node are "?" we treated it as value 0 and

propagates upward.

$$\begin{aligned} \text{cb) ci) } I_{\text{info}}(D) &= -\sum_{i=0}^m p_i \log_2 p_i \\ &= -\frac{4}{9} \log_2 \frac{4}{9} - \frac{5}{9} \log_2 \frac{5}{9} \\ &= 0.9911 \end{aligned}$$

$$\begin{aligned} \text{cii) } I_{\text{info}_A}(D) &= \sum_{j=1}^v \frac{|D_j|}{|D|} \times I_{\text{info}}(D_j) \\ &= \frac{4}{9} I_{\text{info}}(D_{\text{yes}}) + \frac{5}{9} I_{\text{info}}(D_{\text{no}}) \\ &= \frac{4}{9} \times \left[ -\frac{3}{4} \log_2 \frac{3}{4} - \frac{1}{4} \log_2 \frac{1}{4} \right] + \frac{5}{9} \times \left[ -\frac{1}{5} \log_2 \frac{1}{5} - \frac{4}{5} \log_2 \frac{4}{5} \right] \\ &= 0.7616 \end{aligned}$$

$$G_{\text{ain}}(A) = I_{\text{info}}(D) - I_{\text{info}_A}(D) = 0.2295$$

$$\begin{aligned} I_{\text{info}_B}(D) &= \frac{5}{9} I_{\text{info}}(D_{\text{yes}}) + \frac{4}{9} I_{\text{info}}(D_{\text{no}}) \\ &= \frac{5}{9} \left[ -\frac{2}{5} \log_2 \frac{2}{5} - \frac{3}{5} \log_2 \frac{3}{5} \right] + \frac{4}{9} \times \left[ -\frac{2}{4} \log_2 \frac{2}{4} - \frac{2}{4} \log_2 \frac{2}{4} \right] \\ &= \frac{5}{9} \times 0.9710 + \frac{4}{9} \\ &= 0.9839 \end{aligned}$$

$$G_{\text{ain}}(B) = 0.9911 - 0.9839 = 0.0072$$

$$\begin{aligned}
 \text{ciii) } Gini(D) &= 1 - \sum_{i=1}^2 p_i^2 \\
 &= 1 - \left(\frac{4}{9}\right)^2 - \left(\frac{5}{9}\right)^2 \\
 &= 0.4938
 \end{aligned}$$

$$\begin{aligned}
 \textcircled{2} \text{ } Gini_A(D) &= \frac{|D_1|}{|D|} Gini(D_1) + \frac{|D_2|}{|D|} Gini(D_2) \\
 &= \frac{4}{9} Gini(D_{yes}) + \frac{5}{9} Gini(D_{no}) \\
 &= \frac{4}{9} \left[ 1 - \left(\frac{3}{4}\right)^2 - \left(\frac{1}{4}\right)^2 \right] + \frac{5}{9} \times \left[ 1 - \left(\frac{1}{5}\right)^2 - \left(\frac{4}{5}\right)^2 \right] \\
 &= \frac{4}{9} \times \frac{3}{8} + \frac{5}{9} \times \frac{8}{25} \\
 &= \frac{1}{6} + \frac{8}{45} \\
 &= 0.3444
 \end{aligned}$$

$$\begin{aligned}
 \textcircled{3} \Delta Gini(A) &= Gini(D) - Gini_A(D) \\
 &= 0.1494
 \end{aligned}$$

$$\begin{aligned}
 \textcircled{4} \text{ } Gini_B(D) &= \frac{|D_1|}{|D|} Gini(D_1) + \frac{|D_2|}{|D|} Gini(D_2) \\
 &= \frac{5}{9} Gini(D_{yes}) + \frac{4}{9} Gini(D_{no}) \\
 &= \frac{5}{9} \times \left[ 1 - \left(\frac{2}{5}\right)^2 - \left(\frac{3}{5}\right)^2 \right] + \frac{4}{9} \times \left[ 1 - \left(\frac{1}{2}\right)^2 - \left(\frac{1}{2}\right)^2 \right] \\
 &= \frac{5}{9} \times 0.48 + \frac{4}{9} \times 0.5 \\
 &= 0.4889
 \end{aligned}$$

$$\begin{aligned}
 \textcircled{5} \Delta Gini(B) &= Gini(D) - Gini_B(D) \\
 &= 0.0049
 \end{aligned}$$

(iii) ID3 selects the attribute with the highest information gain.

$$\text{Gain}(A) > \text{Gain}(B)$$

attribute A is chosen for the root node.