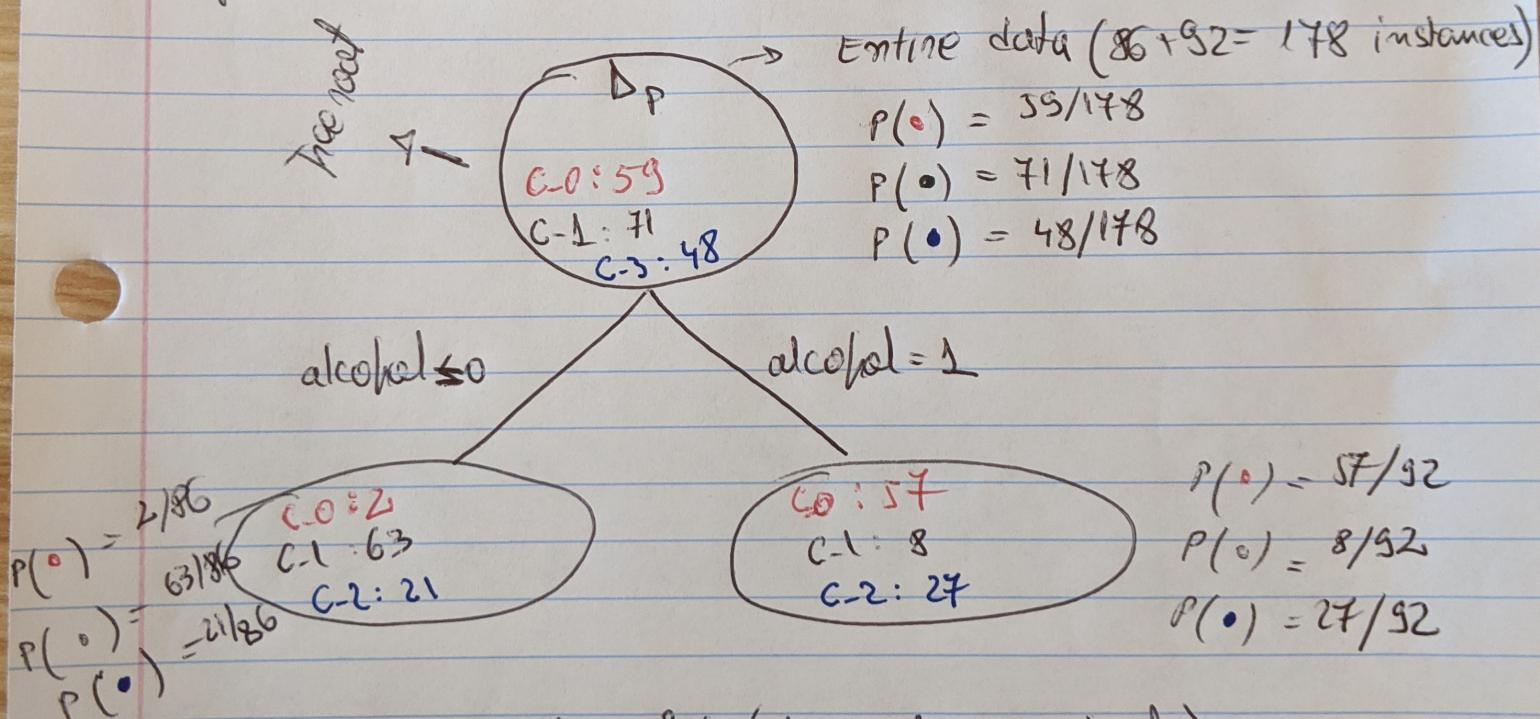


Step 2: Define contingency table

	cultivar-0	cultivar-1	cultivar-2	Total
feature=alcohol	2	63	21	86
	57	8	27	92

Step 2: Draw decision tree:



Step 3: Compute $iG(\Delta_P, f=\text{alcohol})$

By def: $iG(\Delta_P, f=\text{alcohol}) =$

$$= J(\Delta_P) - \left[\frac{N_{\text{alcohol}=0}}{N_P} \cdot J(\Delta_{\text{alcohol}=0}) + \right.$$

$$\left. \frac{N_{\text{alcohol}=1}}{N_P} \cdot J(\Delta_{\text{alcohol}=1}) \right]$$

Note: $\Delta_{\text{alcohol}=0}$ { child nodes
 $\Delta_{\text{alcohol}=1}$ }

Step 4: Compute $J(\Delta_p)$

$$\begin{aligned} \text{(eq 1)} \quad J(\Delta_p) &= - \left[P(\text{class} = c_0 | \Delta_p) \cdot \log_2 P(\text{class} = c_0 | \Delta_p) + \right. \\ &\quad P(\text{class} = c_1 | \Delta_p) \cdot \log_2 P(\text{class} = c_1 | \Delta_p) + \\ &\quad \left. P(\text{class} = c_2 | \Delta_p) \cdot \log_2 P(\text{class} = c_2 | \Delta_p) \right] \\ &= - \left[\frac{59}{178} \cdot \log_2 \left(\frac{59}{178} \right) + \frac{71}{178} \cdot \log_2 \left(\frac{71}{178} \right) + \frac{48}{178} \cdot \log_2 \left(\frac{48}{178} \right) \right] \end{aligned}$$

Step 5: Compute $J(\Delta_{\text{alcohol}} = 0)$ and $J(\Delta_{\text{alcohol}} = 1)$:

$$\begin{aligned} \text{(eq 2)} \quad J(\Delta_{\text{alcohol}} = 0) &= - \left[\frac{2}{86} \cdot \log_2 \left(\frac{2}{86} \right) + \frac{63}{86} \cdot \log_2 \left(\frac{63}{86} \right) + \right. \\ &\quad \left. + \frac{21}{86} \cdot \log_2 \left(\frac{21}{86} \right) \right] \end{aligned}$$

$$\text{(eq 3)} \quad J(\Delta_{\text{alcohol}} = 1) = - \left[\frac{57}{92} \cdot \log_2 \left(\frac{57}{92} \right) + \frac{35}{92} \cdot \log_2 \left(\frac{35}{92} \right) + \frac{27}{92} \cdot \log_2 \left(\frac{27}{92} \right) \right]$$

Step 6: Compute info gain (iG):

$$iG(\Delta_p, f = \text{alcohol}) = (\text{eq 1}) - \left[\frac{86}{178} \cdot (\text{eq 2}) + \frac{92}{178} \cdot (\text{eq 3}) \right]$$

Q₁: What if we've more features (e.g. ash, malic-acid...)

A₁: Repeat steps 1-6 for each feature

Q₂: How do we know which feature is the most informative?

A₂: The one with the highest IG

Q₃: What do we do after the first split?

A₃: It's an iterative process. The child node becomes Δp ; repeat steps 1-6 for all remaining features to find the second best feature; ... and so on