

$$(Q1) \quad E(S) = \sum_{i=1}^C -p_i \log_2 p_i$$

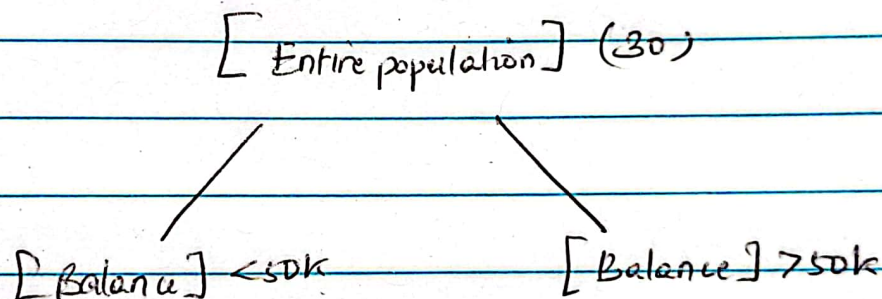
In this example:

$$-3/10 \log_2 (3/10) - 7/10 \times \log_2 (7/10) \approx 0.88$$

Information gain:

$$IG(Y, X) = E(Y) - E(Y|X)$$

Decision Tree:



$$E(P) = 16/30 \log_2 (16/30) - 14/30 \log_2 (14/30) \approx 0.99$$

$$E(B < 50k) \approx 0.39$$

$$E(B > 50k) \approx 0.79$$

$$\begin{aligned} E(\text{Balance}) &= 13/30 \times 0.39 + 17/30 \times 0.79 \\ &= 0.62 \end{aligned}$$

Info Gain :

$$I_{\text{B}}(P, B) \Rightarrow 0.99 - 0.62$$

$$\Rightarrow 0.37$$