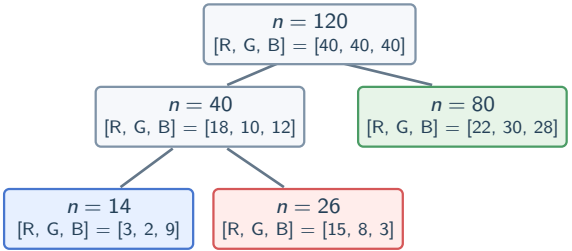
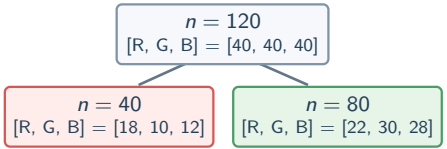
Cost-complexity pruning: which tree wins?

Tree A · 3 leaves



Tree B · 2 leaves



For each value of α, which tree minimizes this expression?

$$T_{\text{best}} = \arg\min_{T \subseteq T_0} \left(\left[\sum_{m=1}^{|T|} \frac{N_m}{N} \sum_{k=1}^{K} \hat{p}_{mk}(1 - \hat{p}_{mk}) \right] + \alpha |T| \right)$$

$|T| = 3$ **(A)**, 2 **(B)** $\quad K = 3 \quad N = 120 \quad \alpha \in \{0.01,\ 0.05\}$

Leaf colors indicate the predicted class; counts are shown in the order [red, green, blue].