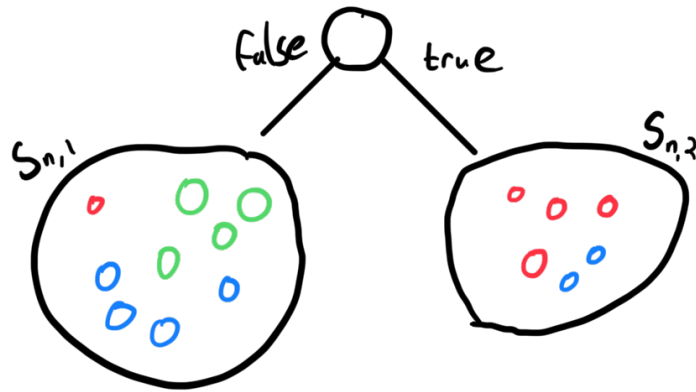


Decision trees



$$\text{total} = 4 + 6 = 10$$

$$S_{n,1} = 4$$

$$r = 1$$

$$g = 4$$

$$b = 4$$

$$S_{n,2} = 6$$

$$r = 4$$

$$g = 0$$

$$b = 2$$

$S_{n,1}$ Impurity Gini

$$= \left(\frac{1}{4}\right) \times \left(1 - \frac{1}{4}\right) + \frac{4}{4} \times \dots$$

Gini Impurity

for a node S

the Gini impurity H_G of S

is

$$\sum_i^n \frac{|S_i|}{|C_i|} \cdot \left(1 - \frac{|S_i|}{|C_i|}\right)$$

the entropy impurity H_E of S

is

$$-1 \cdot \sum \frac{|S_i|}{|C_i|} \cdot \log_2\left(\frac{|S_i|}{|C_i|}\right)$$

for a node S_i that splits into S_{i+1}, S_i

the Information Gain for S_i

$$= H(S_j) - \sum_{i=1}^2 \frac{|S_j^i|}{|S_j|} \cdot H(S_j^i)$$

