Decision Tree Classifier

Mathematics

Step 1: A) Gini Impurity

$$G = 1 - \sum_{i=1}^{n} P_i^2$$

P_i - propation of items in category i

n - total number of categories

B) Information Gain:

Information Gain = Impurity - Weigh

$$P_A = \frac{3}{5}, \quad P_B = \frac{2}{5}$$

Information Gain = Impurity − Weighted

$P_A = \frac{3}{5}$, $P_B = \frac{2}{5}$

$G = 1 - (P_A^2 + P_B^2) = 1 - \left(\left(\frac{3}{5}\right)^2 + \left(\frac{2}{5}\right)^2\right) = 1 - \left(\frac{9}{25} + \frac{4}{25}\right) = 1 - \frac{13}{25}$

Weighted Impurity = $\frac{\text{Size of Node}}{\text{Size of Parent}}$ × Imp

$H = -\sum_{i=1}^{n} P_i \log_2 P_i$

Information Gain = H_p − Weighted Average