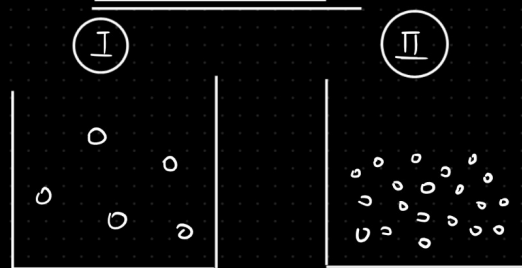


f_1 f_2 target

Class	Gender	Stay in hostel
9	M	Yes
10	F	No
8	F	Yes
8	F	No
9	M	Yes
10	M	No
11	F	Yes
11	M	Yes
8	F	Yes
9	M	No
11	M	No
11	M	Yes
10	F	No
10	M	Yes

Entropy

Randomness



High Randomness

Entropy $\Rightarrow - \sum_{i=1}^c p_i \log_2 p_i$

Probability

Computational expensive

Documentation

\rightarrow default = gini

\downarrow Entropy \rightarrow measure of impurity

Information Gain $\rightarrow E(\text{target}) - \text{Weighted Average}$

\downarrow high

f_1

$IG(\text{class}) \Rightarrow 0.98522 - 0.8574 = 0.1278$

f_2

$IG(\text{Gender}) \Rightarrow 0.98522 - 0.914 = 0.0712$

class \rightarrow Root Node

Gini Index
IG

for all the
features \rightarrow
in a given
dataset

$$\begin{aligned} E(\text{target}) &\Rightarrow -p_y \log_2 p_y - p_N \log_2 p_N \\ &\Rightarrow -\frac{8}{14} \log_2 \frac{8}{14} - \frac{6}{14} \log_2 \frac{6}{14} \\ &\Rightarrow \underline{\underline{0.98522}} \end{aligned}$$

$$\begin{aligned} E(\text{class} = 8) &\Rightarrow -\frac{2}{3} \log_2 \frac{2}{3} - \frac{1}{3} \log_2 \frac{1}{3} \\ &\Rightarrow \underline{\underline{0.918}} \end{aligned}$$

$$\begin{aligned} E(\text{class} = 9) &\Rightarrow -\frac{2}{3} \log_2 \frac{2}{3} - \frac{1}{3} \log_2 \frac{1}{3} \\ &\Rightarrow \underline{\underline{0.918}} \end{aligned}$$

$$\begin{aligned} E(\text{class} = 10) &\Rightarrow -\frac{1}{4} \log_2 \frac{1}{4} - \frac{3}{4} \log_2 \frac{3}{4} \\ &\Rightarrow \underline{\underline{0.811}} \end{aligned}$$

$$\begin{aligned} E(\text{class} = 11) &\Rightarrow -\frac{3}{4} \log_2 \frac{3}{4} - \frac{1}{4} \log_2 \frac{1}{4} \\ &\Rightarrow \underline{\underline{0.811}} \end{aligned}$$

weighted

$$\underline{\text{average (class)}} \Rightarrow \frac{3}{14} * 0.918 +$$

$$\frac{3}{14} * 0.918 + \frac{4}{14} * 0.811 + \frac{4}{14} * 0.811$$

$$\Rightarrow \underline{\underline{0.8574}}$$

$$\begin{aligned} & - P_Y \log_2 P_Y - P_N \log_2 P_N \\ E(\text{Gender} = \text{male}) & \Rightarrow -\frac{5}{8} \log_2 \frac{5}{8} - \frac{3}{8} \log_2 \frac{3}{8} \\ & \Rightarrow \underline{\underline{0.954}} \end{aligned}$$

$$\begin{aligned} E(\text{Gender} = \text{female}) & \Rightarrow -\frac{3}{6} \log_2 \frac{3}{6} - \frac{3}{6} \log_2 \frac{3}{6} \\ & \Rightarrow \underline{\underline{1}} \end{aligned}$$

$$\begin{aligned} \underline{\text{weighted average}} & \Rightarrow \frac{8}{14} * 0.954 + \frac{6}{14} * 1 \\ & \Rightarrow \underline{\underline{0.974}} \end{aligned}$$