

2.4.1.6 Gini Index

$$Gini(D) = 1 - \sum_{i=1}^2 p_i^2 \quad \text{Binary split only}$$

$$Gini_A(D) = \frac{|D_1|}{|D|} Gini(D_1) + \frac{|D_2|}{|D|} Gini(D_2)$$

$$\Delta Gini(D) = Gini(D) - Gini_A(D)$$

Q: $Gini(D) = ?$ $Gini_A(D) = ?$

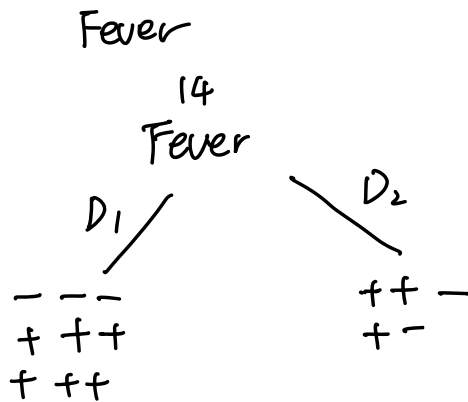
PID	Fever	Cough	Sore Throat	Tiredness	Flu
1	no	yes	no	yes	-
2	no	yes	no	no	-
3	mild	yes	no	yes	+
4	yes	mild	no	yes	+
5	yes	no	yes	yes	+
6	yes	no	yes	no	-
7	mild	no	yes	no	+
8	no	mild	no	yes	-
9	no	no	yes	yes	+
10	yes	mild	yes	yes	+
11	no	mild	yes	no	+
12	mild	mild	no	no	+
13	mild	yes	yes	yes	+
14	yes	mild	no	no	-

Solution ① $Gini(D)$

$$Gini(D) = 1 - \left(\frac{9}{14}\right)^2 - \left(\frac{5}{14}\right)^2 = 0.459$$

② Fever

A binary split of D into D_1 "fever $\in \{no, mild\}$ "
 D_2 : "fever $\in \{yes\}$ "



$$\begin{aligned}
 Gini_{Fever}(D) &= \frac{9}{14} Gini(D_1) + \frac{4}{14} Gini(D_2) \\
 &= \frac{9}{14} \left[1 - \left(\frac{6}{9}\right)^2 - \left(\frac{3}{9}\right)^2 \right] + \frac{4}{14} \left[1 - \left(\frac{3}{5}\right)^2 - \left(\frac{2}{5}\right)^2 \right] \\
 &= 0.4229 \\
 &= 0.423
 \end{aligned}$$

$$\begin{aligned}
 \Delta Gini(Fever) &= 0.459 - 0.423 \\
 &= 0.036
 \end{aligned}$$

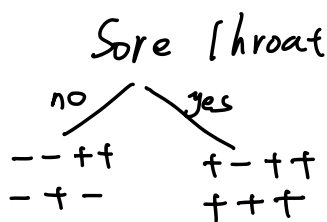
③ Cough

A binary split of D into D_1 "cough $\in \{no, mild\}$ "
 D_2 : "cough $\in \{yes\}$ "

$$\begin{aligned} Gini_{cough}(D) &= \frac{10}{14} Gini(D_1) + \frac{4}{14} Gini(D_2) \\ &= \frac{10}{14} \left[1 - \left(\frac{7}{10}\right)^2 - \left(\frac{3}{10}\right)^2 \right] + \frac{4}{14} \left[1 - \left(\frac{2}{4}\right)^2 - \left(\frac{2}{4}\right)^2 \right] \\ &= 0.443 \end{aligned}$$

$$\begin{aligned} \Delta Gini(Cough) &= 0.459 - 0.443 \\ &= 0.016 \end{aligned}$$

④ Sore Throat



$$\begin{aligned} Gini_{S.T.}(D) &= \frac{7}{14} Gini(D_{no}) + \frac{7}{14} Gini(D_{yes}) \\ &= \frac{1}{2} \left[1 - \left(\frac{4}{7}\right)^2 - \left(\frac{3}{7}\right)^2 \right] + \frac{1}{2} \left[1 - \left(\frac{6}{7}\right)^2 - \left(\frac{1}{7}\right)^2 \right] \\ &= 0.3673 \end{aligned}$$

$$\Delta Gini(Sore Throat) = 0.459 - 0.367 = 0.092$$

⑤ Tiredness

Tiredness	
no	yes
- - + +	- + + +
+ -	- + + +

$$\begin{aligned}
 Gini_{T.}(D) &= \frac{6}{14} Gini(D_{no}) + \frac{8}{14} Gini(D_{yes}) \\
 &= \frac{3}{7} \left[1 - \left(\frac{3}{6}\right)^2 - \left(\frac{3}{6}\right)^2 \right] + \frac{4}{7} \left[1 - \left(\frac{6}{8}\right)^2 - \left(\frac{2}{8}\right)^2 \right] \\
 &= 0.4286
 \end{aligned}$$

$$\Delta Gini(\text{Sore Throat}) = 0.459 - 0.429 = 0.03$$

⑥

F.	C.	S.T.	T.
0.036	0.016	0.092	0.03

Δ
cough is the smallest value.

Split by cough result in max
reduction in Gini index