Entropy & Information Gain (Decision Tree)

Instance	f_1	f_2	f_3	Target Class
1	T	T	1.0	Yes
2	T	T	6.0	Yes
3	T	F	5.0	No
4	F	F	4.0	Yes
5	F	T	7.0	No
6	F	T	3.0	No
7	F	F	8.0	No
8	T	F	7.0	Yes
9	F	T	5.0	No

$P(yes) = 4/9$ $\quad P(no) = 5/9$

$$E(S) = -\sum_{i=1}^{n} P_i \log_2 P_i$$

formula

$$\log_2 a = \frac{\log_{10} a}{\log_{10} 2}$$

$\therefore$

$$E(S) = -P_{yes}\log_2(P_{yes}) - P_{no}\log_2(P_{no})$$

$$= 0.52 + 0.4711$$

$$= 0.9911$$

$\therefore E(S) = 0.9911$

$$\text{Info Gain}(f_1) = E(S) - \sum_{u \in (T,F)} \frac{|S_u|}{|S|} \cdot E(S_u)$$

weighted avg of f_1

- calculating weighted Avgs.

for (f_1):

f_1	Yes	No
T	3	1
F	1	4

$$E(S_T) = -\left(P_{T_{yes}}\log_2(P_{T_{yes}})\right) - \left(P_{T_{no}}\log_2(P_{T_{no}})\right)$$

$$E(S_F) = -\left(P_{F_{yes}}\log_2(P_{F_{yes}})\right) - \left(P_{F_{no}}\log_2(P_{F_{no}})\right)$$

$\frac{|S_T|}{|S|} = \frac{4}{9}$

$\frac{|S_F|}{|S|} = \frac{5}{9}$

Weighted Avg $f_1 = \frac{4}{9} \cdot E(S_T) + \frac{5}{9} E(S_F)$

$\Rightarrow \frac{4}{9}\left[-\left(\frac{3}{4}\right)\left(\log\frac{3}{4}\right) - \left(\frac{1}{4}\right)\log\left(\frac{1}{4}\right)\right]$

$+$

$\frac{5}{9}\left[-\left(\frac{1}{5}\right)\left(\log\frac{1}{5}\right) - \left(\frac{4}{5}\right)\log\left(\frac{4}{5}\right)\right]$

$\Rightarrow 0.444(-0.811) + 0.555(-0.722)$

$\Rightarrow -0.76094$

$\therefore$ Weighted avg of $f_1 = -0.76094$

$\therefore$ Info gain of $f_1 = E(S) - \text{wavg}(f_1)$

$= 0.9911 - 0.76094$

$= 0.23017$

$\therefore$ Info gain $(f_1) = 0.2301$

0.2295 original