COEN 4890 Hw#1 | Zhou Shen

3. Decision Trees

1. $\langle X_1, X_2 \rangle \to Y$

$$H(Y) = -\sum_{i=1}^{k} P(Y=y_i)\log_2 P(Y=y_i) \quad \text{here, } k=2.$$

$$= -\sum_{i=1}^{2} P(Y=y_i)\log_2 P(Y=y_i) \qquad y_1: + \to \frac{12}{21} \qquad y_2: - \to \frac{9}{21}$$

$$= -\left(\frac{12}{21}\cdot\log_2\frac{12}{21} + \frac{9}{21}\cdot\log_2\frac{9}{21}\right) = -\frac{4}{7}\log_2\frac{4}{7} - \frac{3}{7}\log_2\frac{3}{7}$$

+12 +3

2. IG. $IG(X) = H(Y) - H(Y|X)$?

$$H(Y|X) = -\sum_{j=1}^{} P(X=x_j)\sum_{i=1}^{k} P(Y=y_i|X=x_j)\log_2 P(Y=y_i|X=x_j)$$

$$= -\frac{13}{21}\left(\frac{5}{13}\log_2\frac{5}{13} + \frac{8}{13}\log_2\frac{8}{13}\right) - \frac{8}{21}\left(\frac{7}{8}\log_2\frac{7}{8} + \frac{1}{8}\log_2\frac{1}{8}\right)$$

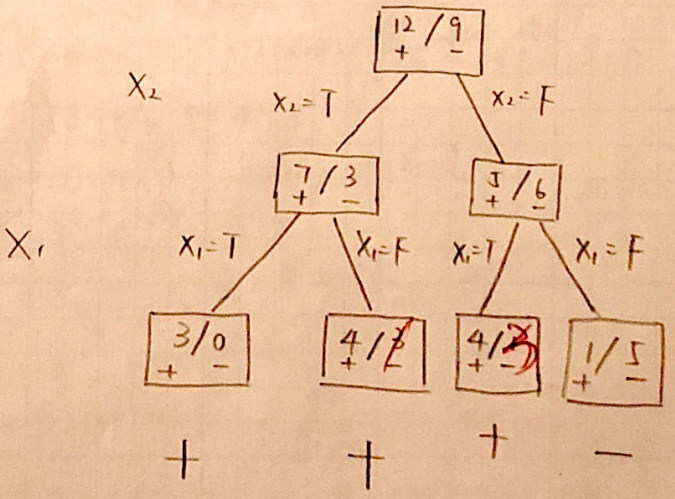
$\overline{X_1}$: F $\quad \overline{X_1}$ F&Y:+ $\quad \overline{X_1}$ F&Y:- $\quad X_1$: T $\quad X_1$ T&Y:+ $\quad X_1$ F&Y:-

Using Matlab to solve it, we get IG_1 is 0.7903,

IG_2 is 0.9265. IG_2 > IG_1

+2

3. From 2 we know that IG_2 is bigger, so we choose IG_2 firstly.



+7