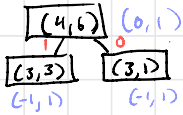
4

① $IG(D_P, f) = I(D_P) - \sum_{j=1}^{m} \frac{N_j}{N_P} \cdot I(D_j)$

$I_H(t) = -\sum_{i=1}^{c} P(i|t) \cdot \log_2(P(i|t))$

x_1



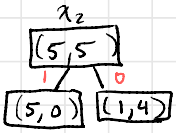
x_1: $I_H(D_P) = -\left(\frac{4}{10} \cdot \log_2\left(\frac{4}{10}\right) + \frac{6}{10}\log_2\left(\frac{6}{10}\right)\right) = 0.971$

$I_H(L) = -\left(\frac{3}{6} \cdot \log_2\left(\frac{3}{6}\right) + \frac{3}{6} \cdot \log_2\left(\frac{3}{6}\right)\right) = 1$

$I_H(R) = -\left(\frac{3}{4} \cdot \log_2\left(\frac{3}{4}\right) + \frac{1}{4} \cdot \log_2\left(\frac{1}{4}\right)\right) = 0.811$

$IG_H(tree) = 0.971 - \frac{6}{10}(1) - \frac{4}{10}(0.811) = 0.0466$

x_2



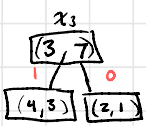
x_2: $I_H(D_P) = -\left(\frac{5}{10} \cdot \log_2\left(\frac{5}{10}\right) + \frac{5}{10} \cdot \log_2\left(\frac{5}{10}\right)\right) = 1$

$I_H(L) = -\left(\frac{5}{5}\log_2\left(\frac{5}{5}\right) + \frac{0}{5} \cdot \log_2\left(\frac{0}{5}\right)\right) = 0$

$I_H(R) = -\left(\frac{1}{5}\log_2\left(\frac{1}{5}\right) + \frac{4}{5}\log_2\left(\frac{4}{5}\right)\right) = 0.722$

$IG_H(tree) = 1 - \frac{5}{10}(0) - \frac{5}{10}(0.722) = 0.639$

x_3



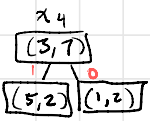
x_3: $I_H(D_P) = -\left(\frac{3}{10} \cdot \log_2\left(\frac{3}{10}\right) + \frac{7}{10}\log_2\left(\frac{7}{10}\right)\right) = 0.881$

$I_H(L) = -\left(\frac{4}{7}\log\left(\frac{4}{7}\right) + \frac{3}{7}\log_2\left(\frac{3}{7}\right)\right) = 0.985$

$I_H(R) = -\left(\frac{2}{3} \cdot \log_2\left(\frac{2}{3}\right) + \frac{1}{3}\log_2\left(\frac{1}{3}\right)\right) = 0.918$

$IG_H(Tree) = 0.881 - \frac{7}{10}(0.985) - \frac{3}{10}(0.918) = -0.0839$

x_4



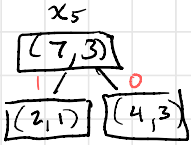
x_4: $I_H(D_P) = -\left(\frac{3}{10} \cdot \log_2\left(\frac{3}{10}\right) + \frac{7}{10}\log_2\left(\frac{7}{10}\right)\right) = 0.881$

$I_H(L) = -\left(\frac{5}{7}\log_2\left(\frac{5}{7}\right) + \frac{2}{7} \cdot \log_2\left(\frac{2}{7}\right)\right) = 0.863$

$I_H(R) = -\left(\frac{1}{3}\log_2\left(\frac{1}{3}\right) + \frac{2}{3} \cdot \log_2\left(\frac{2}{3}\right)\right) = 0.918$

$IG_H(tree) = 0.881 - \frac{7}{10}(0.863) - \frac{3}{10}(0.918) = 0.0015$

x_5



x_5: $I_H(D_P) = -\left(\frac{7}{10}\log_2\left(\frac{7}{10}\right) + \frac{3}{10}\log_2\left(\frac{3}{10}\right)\right) = 0.881$

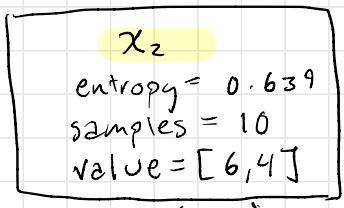
$I_H(L) = -\left(\frac{2}{3} \cdot \log_2\left(\frac{2}{3}\right) + \frac{1}{3}\log_2\left(\frac{1}{3}\right)\right) = 0.918$

$I_H(R) = -\left(\frac{4}{7}\log_2\left(\frac{4}{7}\right) + \frac{3}{7}\log_2\left(\frac{3}{7}\right)\right) = 0.985$

$IG(tree) = 0.881 - \frac{3}{10}(0.918) - \frac{7}{10}(0.985) = -0.0839$

- Since the information gain is maximized for feature x_2 ($IG = 0.639$), the feature "x_2: is long?" should be used as the root node.

(2) DECISION TREE



- value = [class 0, class 1]
- when $x_2 = 1$, $y = -1$ so its a pure leaf meaning we continue spliting when $x_2 = 0$

1 → samples = 5, value = [5,0]

0 → x_1, entropy = 0.0466, samples = 5, value = [1,4]

- next we will split based on feature x_1 since its the second largest maximum IG

1 → samples = 3, value = [0,3]

when $x_1 = 1$, $y = 1$ for all so no further splitting is needed.

0 → x_4, entropy = 0.0015, samples = 2, value = [1,1]

Here we split the data based on x_4 since it is the third largest IG

1 → samples = 1, value = [1,0]

0 → samples = 1, value = [0,1]

final leaf nodes that can no longer be split.

samples when $x_2 = 0$

x_1: know author?	x_2: is long?	x_3: has research?	x_4: has grade?	x_5: has lottery?	t: read?
0	0	1	1	0	-1
1	1	0	1	0	-1
0	1	1	1	1	-1
1	1	1	1	0	-1
0	1	0	0	0	-1
1	0	1	1	1	1
0	0	1	0	0	1
1	0	0	0	0	1
1	0	1	1	0	1
1	1	1	1	1	-1

x_1: know author?	x_2: is long?	x_3: has research?	x_4: has grade?	x_5: has lottery?	t: read?
0	0	1	1	0	-1
1	1	0	1	0	-1
0	1	1	1	1	-1
1	1	1	1	0	-1
0	1	0	0	0	-1
1	0	1	1	1	1
0	0	1	0	0	1
1	0	0	0	0	1
1	0	1	1	0	1
1	1	1	1	1	-1