IS5152 Assignment 4 Wang Ruicong A0244297W

1. (a) GINI index $= 1 - (\frac{3}{5})^2 - (\frac{2}{5})^2 = \frac{12}{25} = 0.48$

(b) ① If we split by University Education {Yes} vs {No}

② If we split ~~us~~ by Income Level

(a) {High} vs {Medium, Low} (b) {Low} vs {High, Medium}

~~(b) {Medium} vs {High, Low}~~

③ If we split by Profession

(a) {Self-employed} vs {White-collar, Blue-collar}

(b) {White-collar} vs {Self-employed, Blue-collar}

(c) {Blue-collar} vs {Self-employed, White-collar}

There are total 6 possible splits

(c) For University Education

$$I_G(q_1, q_2) = \frac{3}{5} \times (1 - (\frac{2}{3})^2 - (\frac{1}{3})^2) + \frac{2}{5} \times 0 = \frac{4}{15}$$

For Income level (a)

$$I_G(q_1, q_2)_H = \frac{2}{5} \times (1 - (\frac{1}{2})^2 - (\frac{1}{2})^2) + \frac{3}{5} \times (1 - (\frac{2}{3})^2 - (\frac{1}{3})^2) = \frac{7}{15}$$

~~For Income level (b)~~

$$\cancel{I_G(q_1, q_2)_M = \frac{2}{5} \times (1 - (\frac{1}{2})^2 - (\frac{1}{2})^2) + \frac{3}{5} \times (1 - (\frac{2}{3})^2 - (\frac{1}{3})^2) = \frac{7}{15}}$$

For Income level (b)

$$I_G(q_1, q_2)_2 = \frac{1}{5} \times 0 + \frac{4}{5} \times (1 - (\frac{1}{2})^2 - (\frac{1}{2})^2) = \frac{2}{5}$$

We find that $\frac{4}{15} < \frac{7}{15}$ and $\frac{4}{15} < \frac{2}{5}$

Spliting Universiy Education will result in lower impurity

Therefore, Income level will not be selected for splitting the root node.

2. (a) There are 4 tasks not success and 6 tasks success

$$\text{Entropy} = -\frac{4}{10}\log_2\left(\frac{4}{10}\right) - \frac{6}{10}\log_2\left(\frac{6}{10}\right) = 0.971$$

(b) We sort the data based on experience

Experience	6	8	10	12	14	15	18	25	29	30
Task Success	0	0	1	1	0	0	1	1	1	1

There are three possible thresholds to be choosen to split : 9, 13, 16.5

For 9, Information gain $= 0.971 - \frac{2}{10}\times 0 - \frac{8}{10}\times\left(-\frac{2}{8}\log_2\frac{2}{8} - \frac{6}{8}\log_2\frac{6}{8}\right) \approx 0.322$

For 13, Information gain $= 0.971 - \frac{4}{10}\times\left(-\frac{2}{4}\log_2\frac{2}{4} - \frac{2}{4}\log_2\frac{2}{4}\right) - \frac{6}{10}\times\left(-\frac{2}{6}\log_2\frac{2}{6} - \frac{4}{6}\log_2\frac{4}{6}\right) \approx 0.02$

For 16.5 Information gain $= 0.971 - \frac{6}{10}\times\left(-\frac{2}{6}\log_2\frac{2}{6} - \frac{4}{6}\log_2\frac{4}{6}\right) - \frac{4}{10}\times 0 \approx 0.42$

0.42 is the maximum value, thus we use 16.5 to split

For $Exp \le 16.5$, Entropy $= -\frac{2}{6}\log_2\frac{2}{6} - \frac{4}{6}\log_2\frac{4}{6} = 0.9183$

There are two possible thresholds : 9 and 13

For 9 Information gain $= 0.9183 - \frac{2}{6}\times 0 - \frac{4}{6}\times\left(-\frac{2}{4}\log_2\frac{2}{4} - \frac{2}{4}\log_2\frac{2}{4}\right) \approx 0.2516$

For 13 Information gain $= 0.9183 - \frac{2}{6}\times 0 - \frac{4}{6}\times\left(-\frac{2}{4}\log_2\frac{2}{4} - \frac{2}{4}\log_2\frac{2}{4}\right) \approx 0.2516$

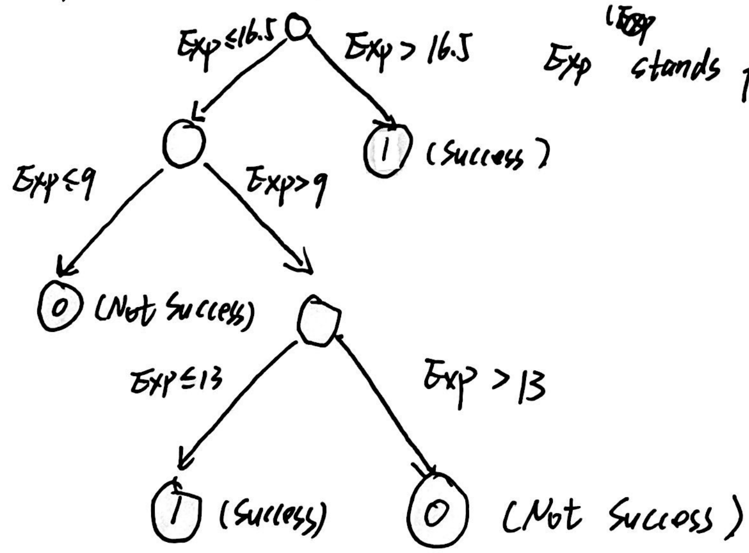
The information gain is same for 9 and 13. we select 9 randomly

For $Exp \le 16.5$ and $Exp > 9$, Entropy $= -\frac{2}{4}\log_2\frac{2}{4} - \frac{2}{4}\log_2\frac{2}{4} = 1$

We choose 13 ~~to~~ as the threshold.

Therefore, the complete decition tree is shown in the following

Exp stands for Experience.



(c) There are $2^{4-1} - 1 = 7$ possible splits. The possible splits are :

① {A} vs {B, C, D} ② {B} vs {A, C, D} ③ {C} vs {A, B, D} ④ {D} vs {A, B, C}
⑤ {A, B} vs {C, D} ⑥ {A, C} vs {B, D} ⑦ {A, D} vs {B, C}

3、(a) $w_0^T x_1 = -\frac{3}{2} < 0$ predict class 0 correct

$w_0^T x_2 = -\frac{11}{4} < 0$ predict class 0 correct

$w_0^T x_3 = -\frac{5}{2} < 0$ predict class 0 correct

$w_0^T x_4 = \frac{1}{2} > 0$ predict class 1 correct

$w_0^T x_5 = -\frac{1}{2} < 0$ predict class 0 incorrect

$w_0^T x_6 = 0$ predict class 0 incorrect

The accuracy is $\frac{4}{6} \approx 66.67\%$

(b) $\Delta w = \eta \cdot \left((t_1 - o_1) x_1 + (t_2 - o_2) x_2 + (t_3 - o_3) x_3 + (t_4 - o_4) x_4 + (t_5 - o_5) x_5 + (t_6 - o_6) x_6 \right)$

$$= 0.01 \times \left[\frac{3}{2}\begin{pmatrix} 3 \\ 1 \\ 1 \end{pmatrix} + \frac{11}{4}\begin{pmatrix} 4 \\ -0.5 \\ 1 \end{pmatrix} + \frac{5}{2}\begin{pmatrix} 3 \\ -1 \\ 1 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} 1 \\ 3 \\ 1 \end{pmatrix} + \frac{3}{2}\begin{pmatrix} 3 \\ 3 \\ 1 \end{pmatrix} + 1 \times \begin{pmatrix} 2 \\ 3 \\ 1 \end{pmatrix} \right]$$

$$= \begin{bmatrix} 0.3 \\ 0.06625 \\ 0.0975 \end{bmatrix}$$

$$w_1 = w_0 + \Delta w = \begin{bmatrix} -\frac{1}{2} \\ \frac{1}{2} \\ -\frac{1}{2} \end{bmatrix} + \begin{bmatrix} 0.3 \\ 0.06625 \\ 0.0975 \end{bmatrix} = \begin{bmatrix} -0.2 \\ 0.56625 \\ -0.4025 \end{bmatrix}$$

(c) before update: Sum of squared error

$$= \frac{1}{2}\left[(0-(-\tfrac{3}{2}))^2 + (0-(-\tfrac{11}{4}))^2 + (0-(-\tfrac{5}{2}))^2 + (1-\tfrac{1}{2})^2 + (1-(-\tfrac{1}{2}))^2 + (1-0)^2 \right]$$

$= 9.78125$

after update.

$o_1 = w_1^T x_1 = -0.43625$ $o_2 = w_1^T x_2 = -1.485625$ $o_3 = w_1^T x_3 = -1.56875$

$o_4 = w_1^T x_4 = 1.09625$ $o_5 = w_1^T x_5 = 0.69625$ $o_6 = w_1^T x_6 = 0.89625$

Therefore after update sum of squared error is $\frac{1}{2}\sum_{i=1}^{6}(t_i - o_i)^2 = 2.48533$

(d) $o_1 < 0$ predict class 0 correct

$o_2 < 0$ predict class 0 correct

$o_3 < 0$ predict class 0 correct

$o_4 > 0$ predict class 1 correct

$o_5 > 0$ predict class 1 correct

$o_6 > 0$ predict class 1 correct

The accuracy after weight update is $\frac{6}{6} = 100\%$

4. $x_1 = \begin{pmatrix} 0.4 \\ 0.43 \end{pmatrix}$ $\quad w_0^T x_1 = [0.5 \ \ -0.5] \begin{bmatrix} 0.4 \\ 0.43 \end{bmatrix} = -0.015$

$$o = \sigma(w_0^T x_1) = \frac{1}{1+e^{-w_0^T x_1}} = \frac{1}{1+e^{0.015}} = 0.49625 \ , \quad t = 0$$

$$\cancel{\delta = (t-o)(1-o)o}$$

$$\delta = (t-o)(1-o)\,o = -o^2 \cdot (1-o) = -0.124$$

$$w_1 = w_0 + \eta \delta x_1 = \begin{bmatrix} 0.5 \\ -0.5 \end{bmatrix} + 1 \times (-0.124) \times \begin{bmatrix} 0.4 \\ 0.43 \end{bmatrix}$$

$$= \begin{bmatrix} 0.45038 \\ -0.55334 \end{bmatrix}$$