



United International University (UIU)

Dept. of Computer Science & Engineering (CSE)

Class Test II: Trimester: Summer - 2019

Course Code: CSI 415, Course Title: Pattern Recognition, Sec: B

Total Marks: 20

Duration: 30 Minutes

Answer all questions. Figures in the right-hand margin indicates full marks.

Question 1:

20

Find the Attribute-Value and Class-Value (AVC) Table for the attribute, A_i (where A_i is Income if the last digit of your Student ID is odd number else it will be Age). Then please find the Gain Ratio of the attribute, A_i . Please write the necessary equations for finding Gain Ratio.

Training Data Table:

No.	Income	Age	Education	Marital Status	Usage
1	Low	Old	University	Married	Low
2	Medium	Young	College	Single	Medium
3	Low	Old	University	Married	Low
4	High	Young	University	Single	High
5	Low	Old	University	Married	Low
6	High	Young	College	Single	Medium
7	Medium	Young	College	Married	Medium
8	Medium	Old	High School	Single	Low
9	High	Old	University	Single	High
10	Low	Old	High School	Married	Low
11	Medium	Young	College	Married	Medium
12	Medium	Old	High School	Single	Low
13	High	Old	University	Single	High
14	Low	Old	High School	Married	Low
15	Medium	Young	College	Married	Medium

Pattern CT-02 Solve 1 Section-B

Page → 1

Ans To The Q.No → 1

Working With Age Column (Making Avc-Table)

Attribute "Age" Values	Class Values		
	Low	Medium	High
	old	0	2
Young	0	5	1
	7	5	3

The Formulas

- $Info(D) = - \sum_{i=1}^n P_i \log_2 P_i$
- $Info_A(D) = \sum_{j=1}^n \frac{|D_j|}{|D|} \times Info(D)$
- $Gain(A) = Info(D) - Info_A(D)$
- $Split Info_A(D) = - \sum_{j=1}^n \frac{|D_j|}{|D|} \times \log_2 \left(\frac{|D_j|}{|D|} \right)$
- $Gain Ratio(A) = \frac{Gain(A)}{Split Info_A(D)}$

The Calculation

$$\begin{aligned}
 Info(D) &= - \frac{7}{15} \log_2 \left(\frac{7}{15} \right) - \frac{5}{15} \log_2 \left(\frac{5}{15} \right) - \frac{3}{15} \log_2 \left(\frac{3}{15} \right) \\
 &= -0.46 \log_2 (0.46) - 0.33 \log_2 (0.33) - 0.2 \log_2 (0.2) \\
 &= -0.46 \times (-1.12) - 0.33 \times (-1.59) - 0.2 \times (-2.32) \\
 &= 1.5039
 \end{aligned}$$

$$\begin{aligned}
 \text{Info}_{\text{Age}}^{(D)} &= \frac{\frac{9}{15} \times \left(-\frac{7}{9} \log_2\left(\frac{7}{9}\right) - \frac{0}{9} \log_2\left(\frac{0}{9}\right) - \frac{2}{9} \log_2\left(\frac{2}{9}\right) \right)}{\text{old}} \\
 &+ \frac{\frac{6}{15} \times \left(-\frac{0}{6} \log_2\left(\frac{0}{6}\right) - \frac{5}{6} \log_2\left(\frac{5}{6}\right) - \frac{1}{6} \log_2\left(\frac{1}{6}\right) \right)}{\text{young}} \\
 &= 0.6 \left(-0.77 \log_2(0.77) - 0 \log_2(0) - 0.22 \log_2(0.22) \right) + \\
 &0.4 \left(-0 \log_2(0) - 0.83 \log_2(0.83) - 0.16 \log_2(0.16) \right) \\
 &= 0.6 (0.29 + 0 + 0.48) + 0.4 (0 + 0.22 + 0.42) \\
 &= 0.6 \times 0.77 + 0.4 \times 0.64 \\
 &= 0.718
 \end{aligned}$$

$$\begin{aligned}
 \text{Gain}(\text{Age}) &= \text{Info}(D) - \text{Info}_{\text{Age}}^{(D)} \\
 &= 1.5039 - 0.718 \\
 &= 0.7859
 \end{aligned}$$

$$\begin{aligned}
 \text{Split Info}_{\text{Age}}^{(D)} &= \frac{-\frac{9}{15} \log_2\left(\frac{9}{15}\right)}{\text{old}} - \frac{\frac{6}{15} \log_2\left(\frac{6}{15}\right)}{\text{young}} \\
 &= -0.6 \log_2(0.6) - 0.4 \log_2(0.4) \\
 &= -0.6 \times (-0.73) - 0.4 \times (-1.32) \\
 &= 0.438 + 0.528 \\
 &= 0.966
 \end{aligned}$$

$$\begin{aligned}
 \text{Gain Ratio (A)} &= \frac{\text{Gain}(\text{Age})}{\text{Split Info}_{\text{Age}}^{(D)}} = \frac{0.7859}{0.966} \\
 &= 0.81
 \end{aligned}$$

Ans. -

Looking With Income Column (Making AIC-Table)

Attribute Values	Class Values			
	Low	Medium	High	
Low	5	0	0	5
Medium	2	4	0	6
High	0	1	3	4
	7	5	3	

The Formulas

$$a. \text{Info}(D) = - \sum_{i=1}^n p_i \log_2 p_i$$

$$b. \text{Info}_A(D) = \sum_{j=1}^n \frac{|D_j|}{|D|} \times \text{Info}(D)$$

$$c. \text{Gain}(A) = \text{Info}(D) - \text{Info}_A(D)$$

$$d. \text{Split Info}_A(D) = - \sum_{j=1}^n \frac{|D_j|}{|D|} \times \log_2 \left(\frac{|D_j|}{|D|} \right)$$

$$e. \text{Gain Ratio}(A) = \frac{\text{Gain}(A)}{\text{Split Info}_A(D)}$$

The Calculation

$$\text{Info}(D) = - \frac{7}{15} \log_2 \left(\frac{7}{15} \right) - \frac{5}{15} \log_2 \left(\frac{5}{15} \right) - \frac{3}{15} \log_2 \left(\frac{3}{15} \right)$$

$$= - 0.46 \log_2(0.46) - 0.33 \log_2(0.33) - 0.2 \log_2(0.2)$$

$$= - 0.46 \times (-1.12) - 0.33 \times (-1.59) - 0.2 \times (-2.32)$$

$$= 1.5039$$

$$\begin{aligned} \text{Info ID)} \\ \text{Income} &= \frac{5}{15} \times \left(-\frac{5}{5} \log_2 (5/5) - \frac{0}{5} \log_2 (0/5) - \frac{0}{5} \log_2 (0/5) \right) + \\ &\quad \frac{6}{15} \times \left(-\frac{2}{6} \log_2 (2/6) - \frac{4}{6} \log_2 (4/6) - \frac{0}{6} \log_2 (0/6) \right) + \\ &\quad \frac{4}{15} \times \left(-\frac{0}{4} \log_2 (0/4) - \frac{1}{4} \log_2 (1/4) - \frac{3}{4} \log_2 (3/4) \right) \end{aligned}$$

Low Medium High

$$= 0.33 (-1 \times 0 + 0 + 0) + 0.4 (0.52 + 0.39 + 0) + 0.26 (0 + 0.5 + 0.31)$$

$$= 0 + 0.364 + 0.2106$$

$$= 0.5746$$

$$\begin{aligned} \text{Gain (Income)} &= \text{Info ID)} - \text{Info Income ID)} \\ &= 1.5039 - 0.5746 \\ &= 0.9293 \end{aligned}$$

$$\begin{aligned} \text{Split Info ID)} \\ \text{Income} &= \frac{-\frac{5}{15} \log_2 (5/15)}{\text{Low}} - \frac{\frac{6}{15} \log_2 (6/15)}{\text{Medium}} \\ &\quad - \frac{\frac{4}{15} \log_2 (4/15)}{\text{High}} \\ &= -0.33 \times (-1.59) - 0.4 \times (-1.32) - 0.26 \times (-1.94) \\ &= 0.52 + 0.528 + 0.5044 \\ &= 1.5524 \end{aligned}$$

$$\begin{aligned} \text{Gain Ratio (A)} &= \frac{\text{Gain (Income)}}{\text{Split Info Income ID)} \\ &= \frac{0.9293}{1.5524} \\ &= 0.59 \end{aligned}$$

Ans ✓