BRAC UNIVERSITY

Inspiring Excellence

Assignment 02

CSE422: Artificial Intelligence

Reshad Ul Karim

Student ID: 22201594

Section: 16

Submitted to

Dr. Swakkhar Shatabda

Professor, Department of Computer Science and Engineering

BRAC University

Answer to the question. 1

(a)

(I)

Age (A)	Gender	Smartphone (S)	Count	P(A, G, S)
Young	Male	Nokia	14	0.14
Young	Male	Samsung	8	0.08
Young	Male	Apple	14	0.14
Young	Female	Nokia	8	0.08
Young	Female	Samsung	12	0.12
Young	Female	Apple	9	0.09
Old	Male	Nokia	7	0.07
Old	Male	Samsung	5	0.05
Old	Male	Apple	11	0.11
Old	Female	Nokia	1	0.01
Old	Female	Samsung	3	0.03
Old	Female	Apple	8	0.08
			100	1

(II) $P(\text{not Apple} \mid \text{Old Age}) = \dfrac{P(\text{not Apple} \cap \text{Old Age})}{P(\text{Old Age})}$

$$= \frac{0.07+0.05+0.01+0.03}{0.07+0.05+0.11+0.01+0.03+0.08}$$

$\therefore P(\text{not Apple} \mid \text{Old Age}) = \dfrac{16}{35} = 0.457$ [Ans]

(iii) $S \perp A \mid G = \text{female}$

$$P(S \mid A, G = \text{female}) = P(S \mid G = \text{female}) \cancel{\cdot} P(A \mid G = \text{female})$$

$$P(S, A, \underset{=\text{female}}{G}) = P(S \mid A, G = \text{female}) \cdot P(A \mid G = \text{female}) \cdot P(G = \text{female})$$

$$\frac{P(S, A, G = \text{female})}{P(G = \text{female})} = P(S \mid G = \text{female}) \cdot P(A \mid G = \text{female})$$

$$\Rightarrow P(S\ A \mid G = \text{female}) = P(S \mid G = \text{female}) \cdot P(A \mid G = \text{female})$$

$$\Rightarrow P(S = \text{Nokia}, A = \text{young} \mid G = \text{female}) = \frac{0.08}{0.12 + 0.08 + 0.09 + 0.01 + 0.03 + 0.08} = \frac{8}{41}$$

$$\therefore P(S = \text{Nokia} \mid G = \text{female}) = \frac{0.01 + 0.08}{0.41} = \frac{9}{41}$$

$$\therefore P(A = \text{young} \mid G = \text{female}) = \frac{0.08 + 0.12 + 0.09}{0.41} = \frac{29}{41}$$

$S = \text{Nokia}$, $A = \text{young}$

$$\therefore P(S \cdot A \mid G = \text{female}) = P(S \mid G = \text{female}) \cdot P(A \mid G = \text{female})$$

$$\frac{8}{41} = \frac{9}{41} \times \frac{29}{41}$$

but $\frac{8}{41} \neq \frac{261}{1681}$

$P(S = \text{Samsung}, A = \text{young} \mid G = \text{female}) = \frac{12}{41}$ ③

$P(S = \text{Samsung} \mid G = \text{female}) = \frac{15}{41}$

$P(A = \text{young} \mid G = \text{female}) = \frac{29}{41}$

but, $\frac{12}{41} \neq \frac{15}{41} \times \frac{29}{41}$

$\Rightarrow \frac{12}{41} \neq \frac{435}{1681}$

Again,

$P(S = \text{Apple}, A = \text{young} \mid G = \text{female}) = \frac{9}{41}$

$P(S = \text{Apple} \mid G = \text{female}) = \frac{17}{41}$

$P(A = \text{young} \mid G = \text{female}) = \frac{29}{41}$

but, $\frac{9}{41} \neq \frac{17}{41} \times \frac{29}{41}$

$\therefore \frac{9}{41} \neq \frac{493}{1681}$

$P(S = \text{Nokia}, A = \text{old} \mid G = \text{female}) = \frac{1}{41}$

$P(S = \text{Nokia} \mid G = \text{female}) = \frac{9}{41}$

$P(A = \text{Old} \mid G = \text{female}) = \frac{12}{41}$

but, $\frac{1}{41} \neq \frac{9}{41} \times \frac{12}{41}$

$\therefore \frac{1}{41} \neq \frac{108}{1681}$

$$P(S = \text{Samsung}, A = \text{old} \mid G = \text{female}) = \frac{3}{41}$$

but, $\frac{3}{41} \neq \frac{15}{41} \times \frac{12}{41}$

$$\Rightarrow \frac{3}{41} \neq \frac{180}{1681}$$

$$P(S = \text{Apple}, A = \text{old} \mid G = \text{female}) = \frac{8}{41}$$

but $\frac{8}{41} \neq \frac{17}{41} \times \frac{12}{41}$

$$\therefore \frac{8}{41} \neq \frac{204}{1681}$$

So, we can see for each condition of age and smartphone

$$P(S \mid A, G = \text{female}) \neq P(S \mid G = \text{female}) \cdot P(A \mid G = \text{female})$$

So smartphone is not conditionally independent to age given gender is female.

(iv) $P(S = \text{samsung} \mid A = \text{old}) = \frac{8}{35}$

so expected number $= \frac{8}{35} \times 30$

$= 6.857$

≈ 6 users.

(b)

$$P(A|B,C) = P(B|A,C)$$

$$P(A|C) = P(B|C)$$

$$P(A|B,C) = \frac{P(A,B,C)/P(C)}{P(B,C)/P(C)} = \frac{P(A,B|C)}{P(B|C)}$$

$$P(B|A,C) = \frac{P(A,B,C)/P(C)}{P(A,C)/P(C)} = \frac{P(A,B|C)}{P(A|C)}$$

We know,

$$P(A|C) = P(B|C)$$

So, $$\frac{P(A,B|C)}{P(B|C)} = \frac{P(A,B|C)}{P(A|C)}$$

$$\Rightarrow \frac{P(A,B|C)}{P(A|C)} = \frac{P(A,B|C)}{P(A|C)}$$

$\therefore$ L.H.S. = R.H.S. [Proved]

Answer to the question. 2

H = HMPV positive

$\overline{H}$ = HMPV negative

T = Test Positive

$\overline{T}$ = Tested Negative

$P(\overline{T}|H) = 0.07$

$P(\overline{T}|\overline{H}) = 0.88$

$P(\overline{H}) = 0.86,$

$P(H) = 0.14$

$P(T) = 0.07 \cancel{\times} \times 0.14 + 0.88 \times 0.86 = 0.766$

$P(T|H) = 0.93)$

$P(T|\overline{H} = 0.12)$

(a) $P(H|\overline{T}) = \dfrac{P(\overline{T}|H)\cdot P(H)}{P(\overline{T})} = \dfrac{0.07 \times 0.14}{0.7666}$

$= 0.0128$ [Ans] 2

(b) $P(H|\overline{T}\cdot T\cdot T\cdot T_3) = \dfrac{P(\overline{T}\cdot T\cdot T\cdot T|H)\cdot P(H)}{P(\overline{T}\cdot T\cdot T\cdot T)}$

$\Rightarrow P(\overline{T}\cdot T\cdot T\cdot T|H) = 0.07 \times [0.93]^3 = 0.0563$

$\therefore P(\overline{T}\cdot T\cdot T\cdot T|\overline{H}) = 0.88 \times [0.12]^3 = 0.00152$

$\therefore P(\overline{T}\cdot T\cdot T\cdot T) = 0.0563 \times 0.14 + 0.00152 \times 0.86$

$= 0.00919$

$$So,\ P(H \mid \bar{T}\cdot T\cdot T\cdot T) = \frac{0.0563 \times 0.14}{0.00919} = 0.8578$$

$$So,\ P(H \mid \bar{T}\cdot T\cdot T\cdot T) = 0.8578$$

$$P(\bar{H} \mid \bar{T}\cdot T\cdot T\cdot T) = 0.1422$$

∴ So it is more likely he is HMPV positive.

Answer to the question. 3

Exam Score Range	Pass	Fail	Total
Above 80%	8	2	10
Between 50% and 80%	5	5	10
Below 50%	1	9	10
	14	16	30

(a) Entropy = 0.918

Entropy for above 80%.

$$E_1 = -\frac{8}{10}\log_2\frac{8}{10} - \frac{2}{10}\log_2\frac{2}{10} = 0.722$$

Entropy for between 50% and 80%.

$$E_2 = -\frac{5}{10}\log_2\frac{5}{10} - \frac{5}{10}\log_2\frac{5}{10} = 1.$$

Entropy for below 50%.

$$E_3 = -\frac{1}{10}\log_2\frac{1}{10} - \frac{9}{10}\log_2\frac{9}{10} = 0.469$$

Weighted Entropy

$E_{after} = \frac{10}{30} \times 0.722 + \frac{10}{30} \times 1 + \frac{10}{30} \times 0.469$

$= 0.730$

So, information gain $= 0.918 - 0.730 = 0.188$

(b)

Attendance	Pass	Fail	Total
High Attendance	1	2	3
Moderate	0	7	7
Total	1	9	10

For High Attendance,

$E_1 = -\frac{1}{3}\log_2\frac{1}{3} - \frac{0}{7}\log_2\frac{0}{7} = 0.9183$

For Moderate Attendance

$E_2 = \frac{0}{7}\log_2\frac{0}{7} - \frac{7}{7}\log_2\frac{7}{7} = 0$

So, weighted entropy

$E_{after} = \frac{3}{10} \times 0.9183 + 0 = 0.2755$

(c)

Entropy for below 50% $= 0.469$

Information gain $= 0.469 - 0.2755$

$= 0.1935$

(d) (i) False. It can be used for regression tasks as well.

(ii) True. This ensures we get the most information from the split.

(iii) False. Entropy measures the impurity and disorder of the dataset. The higher the impurity, the more uncertain the split is.

Answer to the question. 4

(a) Weight is a continuous variable, therefore it is a regression problem.

(b) Learning rate helps us control the size of the step taken by the algorithm. A learning rate that decreases over time lets a model take larger steps toward the optimum when training begins — helping it quickly escape flat regions or shallow traps — and then gradually switches to smaller, more cautious updates as it approaches a minimum. Early on, the higher rate encourages fast progress across the loss surface, while later the reduced rate prevents overshooting and dampens oscillation around optimum, allowing finer adjustments. This "big then small approach not only speeds up convergence in the initial phases but also improves stability and precision at end, making it more likely that model will settle into a good solution rather than bouncing indefinitely or stagnating too soon.

(c) Gradient descents works by adding the negative of the gradient to the parameter.

$$\theta_{new} = \theta_{old} - \alpha \nabla J(\theta)$$

As we can see at point 1, the gradient of the error function $J(\theta)$ is positive. But according to the equation of gradient descent we have to add the negative of the gradient. So, we subtract α.gradient from the old parameter which takes to point 2, as it is closer to maxima. The negative in the formula ensure we move in the correct direction for example, towards the minima.

(d) $\alpha = 0.1$

$m_1 = 0.4$, $m_2 = 0.5$, $c = 0.6$

$\hat{y} = m_1x_1 + m_2x_2 + c \quad \therefore \hat{y} = 0.4x_1 + 0.5x_2 + 0.6$

Now, for $(2, 3, 4)$

$\hat{y} = 0.4 \times 2 + 0.5 \times 3 + 0.6 = 2.9$

$\therefore$ Error $= y - \hat{y} = 4 - 2.9 = 1.1$

for $(3, 4, 5)$

$\hat{y} = 0.4 \times 3 + 0.5 \times 4 + 0.6 = 3.8$

$\therefore$ Error $= y - \hat{y} = 5 - 3.8 = 1.2$

for $(4, 5, 6)$

$\hat{y} = 0.4 \times 4 + 0.5 \times 5 + 0.6 = 4.7$

$\therefore$ Error $= 6 - 4.7 = 1.3$

Loss function: $\frac{1}{m}\sum_{i=1}^{m}(y-\hat{y})^2$

$$\frac{\partial J}{\partial w_0} = -\frac{2}{m}\sum_{i=1}^{m}(y-\hat{y})$$

$$\frac{\partial J}{\partial w_1} = -\frac{2}{m}\sum_{i=1}^{m}(y-\hat{y})x_1$$

$$\frac{\partial J}{\partial W} = -\frac{2}{m}\sum_{i=1}^{m}(y-\hat{y})x_2$$

x_1	x_2	error	error*x_1	error*x_2
2	3	1.1	2.2	3.3
3	4	1.2	3.6	4.8
4	5	1.3	5.2	6.5
		3.6	11.0	14.6

$$\nabla J(c) = -\frac{2}{3}\times 3.6 = -2.4$$

$$\nabla J(m_1) = -\frac{2}{3}\times 11 = -\frac{22}{3}$$

$$\nabla J(m_2) = -\frac{2}{3}\times 14.6 = -\frac{146}{15}$$

$$m_{1\,new} = 0.4 - 0.1\times\left(-\frac{22}{3}\right) = 1.133$$

$$m_{2\,new} = 0.5 - 0.1\times\left(-\frac{146}{15}\right) = 1.4733$$

$$c_{new} = 0.6 - 0.1\times(-2.4) = 0.84$$

$\therefore$ $c = 0.84$,

$m_1 = 1.133$

$m_2 = 1.473$ [Ans]