Q.I]

x (years of experience)	y (salary in thousands)	$(x-\bar{x})$	$(y-\bar{y})$	$(x-\bar{x})(y-\bar{y})$	$(x-\bar{x})^2$	$(y-\bar{y})^2$
3	30	-6.1	-25.4	154.94	+37.21	+645.16
8	57	-1.1	1.6	-1.76	+1.21	2.56
9	64	-0.1	8.6	-0.86	+0.01	73.96
13	72	3.9	16.6	64.74	15.21	275.56
3	36	-6.1	-19.4	118.34	+37.21	+376.36
6	43	-3.1	-12.4	38.44	+9.61	+153.76
11	59	1.9	3.6	6.84	3.61	12.96
21	90	11.9	34.6	411.74	141.61	1197.16
1	20	-8.1	-35.4	286.74	+65.61	1253.16
16	83	6.9	27.6	190.44	47.61	761.76
91	554			1269.6	358.9	4752.4

$\therefore \bar{x} = \frac{91}{10} = \underline{9.1}$

$\bar{y} = \frac{554}{10} = \underline{55.4}$

$\therefore b_{yx} = \frac{\Sigma(x_i-\bar{x})(y_i-\bar{y})}{\Sigma(x_i-\bar{x})^2}$

$= \frac{1269.6}{358.9} = 3.537 \approx \underline{\underline{3.54}}$

$\therefore b_{xy} = \frac{\Sigma(x_i-\bar{x})(y_i-\bar{y})}{\Sigma(y_i-\bar{y})^2}$

$= \frac{1269.6}{4752.4} = 0.267 \approx \underline{\underline{0.27}}$

$\therefore$ Regression line of Y on X is given by

$$y_i - \bar{y} = byx(x_i - \bar{x})$$

$$\therefore y_i - 55.4 = 3.54(x_i - 9.1)$$

$$\therefore y_i = 3.54x_i - 3.54(9.1) + 55.4$$

$$\therefore \boxed{y_i = 3.54x_i + 23.186}$$

$\therefore$ For year of experience is 10 (x) :-

$$y = 3.54(10) + 23.186$$

$$= \underline{58.586} \approx \underline{59 \text{ thousands}}$$

FOR EDUCATIONAL USE

II) Given :- Support = 50% = 3/5
Confidence = 70%

Trans Id	Items
1	Laptop, Mouse, Headphones, Pendrive, Speakers
2	Laptop, Headphones
3	Laptop, Mouse, Pendrive
4	Mouse, Speakers
5	Laptop, Pendrive

Soln:

Itemsets	Support Count
Laptop	4/5
Mouse	3/5
Headphones	2/5 → Discard
Pendrive	3/5
Speakers	2/5 → Discard

Itemsets	Support Count
Laptop, Mouse	2/5 → Discard
Laptop, Pendrive	3/5
Mouse, Pendrive	2/5 → Discard

Itemsets	Support Count
Laptop, Pendrive	3/5

∴ Confidence % of {Laptop} → {Pendrive}

$$= \frac{\text{Support(Laptop, Pendrive)}}{\text{Support(Laptop)}} = \frac{3/5}{4/5} = 0.75 = 75\%$$

FOR EDUCATIONAL USE

∴ Confidence % of {Pendrive} → {Laptop}

$$= \frac{Supp(Pendrive, Laptop)}{Supp(Pendrive)} = \frac{3/5}{3/5} = 1 = 100\%$$

Q.III

Fruit	Yellow	Sweet	Long	Total
Mango	350	450	0	650
Banana	400	300	350	400
others	50	100	50	150
Total	800	850	400	1200

Predict which type of fruit it is if it has the foll features :- Fruit{Yellow, Sweet, Long}

To find :- $P(Mango / Yellow, Sweet, Long)$

$P(Banana / Yellow, Sweet, Long)$

$P(Others / Yellow, Sweet, Long)$

∴ Using Naive Baye's Algorithm

$$\therefore P(Mango/tuple) = \frac{P(tuple/Mango) \cdot P(Mango)}{P(tuple)}$$

$$\therefore P(tuple/Mango) = P(Yellow/Mango) \times P(Sweet/Mango) \times P(Long/Mango)$$

$$= \frac{350}{650} \times \frac{450}{650} \times \frac{0}{650}$$

$$= 0$$

$$\therefore P(Mango/tuple) = 0$$

Illy, $P(\text{Banana}/\text{tuple}) = \dfrac{P(\text{tuple}/\text{Banana}) \cdot P(\text{Banana})}{P(\text{tuple})}$

$\therefore P(\text{tuple}/\text{Banana}) = P(\text{Yellow}/\text{Banana}) \times P(\text{Sweet}/\text{Banana}) \times P(\text{Long}/\text{Banana})$

$= \dfrac{400}{400} \times \dfrac{300}{400} \times \dfrac{350}{400}$

$= \underline{\underline{0.656}}$

$\therefore P(\text{Banana}) = \dfrac{400}{1200} = \underline{0.33}$

$\therefore P(\text{tuple}) = P(\text{Mango}) \cdot P(\text{Yellow}/\text{Mango}) + P(\text{Banana}) \cdot P(\underset{\text{tuple}}{\overset{\text{Sweet, Long}}{}}/\text{Banana}) + P(\text{Others}) \cdot P(\text{tuple}/\text{Others})$

No need to calculate

$\therefore P(\text{Banana}/\text{tuple}) = \dfrac{0.656 \times 0.33}{P(\text{tuple})} = \dfrac{0.218}{P(\text{tuple})}$

Illy, $P(\text{Others}/\text{tuple}) = \dfrac{P(\text{tuple}/\text{Others}) \cdot P(\text{Others})}{P(\text{tuple})}$

$\therefore P(\text{tuple}/\text{Others}) = P(\text{Yellow}/\text{Others}) \times P(\text{Sweet}/\text{Others}) \times P(\text{Long}/\text{Others})$

$= \dfrac{50}{150} \times \dfrac{100}{150} \times \dfrac{50}{150}$

$= \underline{0.0740}$

$\therefore P(\text{Others}) = \dfrac{150}{1200} = \underline{0.125}$

Euclidean Distance $d(i,j) = \sqrt{|x_{i_1} - x_{j_1}|^2 + |x_{i_2} - x_{j_2}|^2 \ldots}$

$\therefore P(\text{Others}/\text{tuple}) = \dfrac{0.0740 \times 0.125}{P(\text{tuple})}$

$= \dfrac{0.00925}{P(\text{tuple})}$

Therefore, since the $P(\text{Banana}/\text{tuple})$ is maximum i.e. $\dfrac{0.218}{P(\text{tuple})}$, the Fruit is classified as Banana.

Q.IV]

Object	x	y	
A	1	1	→ Centroid 1
B	2	1	
C	4	3	→ Centroid 2
D	5	4	

Generate clusters using K-Means Clustering ($K = 2$).

$C_1(1,1)$ $C_2(4,3)$

Object	x	y	d_1	d_2	Cluster
A	1	1	0	3.61	1
B	2	1	1	2.83	1
C	4	3	3.61	0	2
D	5	4	5	1.41	2

New Centroids $C_1 = (1.5, 1)$; $C_2 = (4.5, 3.5)$

Object	x	y	d_1	d_2	Cluster
A	1	1	0.5	4.3	1
B	2	1	0.5	3.54	1
C	4	3	3.20	0.70	2
D	5	4	4.61	0.70	2