

① a) 4 attributes

b) Numerical attributes

c) $x_1 = \{10, 60, 10, 90\}$

$x_3 = \{30, 60, 30, 40\}$

$$\text{distance} = \sqrt{(10-30)^2 + (60-50)^2 + (10-30)^2 + (90-40)^2}$$

$$= \sqrt{400 + 100 + 400 + 2500}$$

$$= \sqrt{3400}$$

$$= 10\sqrt{34} \approx 58.3095$$

d) $x_2 = \{20, 50, 40, 70\}$

$$\text{length} = \sqrt{9400} = 10\sqrt{94} \approx 95.954$$

e) $x_2 = \{20, 50, 40, 70\}$

$x_4 = \{20, 50, 20, 60\}$

$$\cos(\theta) = \frac{\overline{x}_2 \cdot \overline{x}_4}{\|x_2\| \|x_4\|} = \frac{\overline{x}_2 \cdot \overline{x}_4}{10\sqrt{94} \times 10\sqrt{69}} = \frac{7900}{8053.571}$$

$$= 0.98f$$

f) Yes, so all features would make the distance calculations more accurate

g) using Min-Max Normalization , value = $\frac{x - \min(x)}{\max(x) - \min(x)}$

x_1	0	1	0	1
Data matrix = Scaled	0.5	0	1	0.75
x_2	1	0	$\frac{2}{3}$	0.375
x_3	0.5	0	$\frac{1}{3}$	0.625
x_4	0	1	$\frac{3}{3}$	0

h) ~~c~~ Distance = $\sqrt{1+1+\frac{4}{9}+\frac{25}{64}} = 1.684$

d) Length = 1.346

e) $\cos(\theta) = 0.9013$

③ a)

ID	a1	a2	a4
1	10	60	90
2	20	50	70
3	30	50	40
4	20	50	60
5	10	60	20

c) Mean a1 = $\frac{90}{5} = 18$

mean a2 = 54

mean a4 = 54

Mean Vector = (18 54 54)

Q) $Z = \begin{bmatrix} -8 & 6 & 36 \\ 2 & -4 & 16 \\ 12 & -4 & -14 \\ 2 & -4 & 6 \\ -8 & 6 & -44 \end{bmatrix}$

NOTES

e) $\sigma_x^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$

$$\sigma_{(x,y)} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

Cov Matrix $\begin{bmatrix} 56 & -32 & -12 \\ -32 & 24 & 16 \\ -12 & 16 & 744 \end{bmatrix}$

~~f)~~ eigenvector = $\begin{pmatrix} 75.7729965 \\ 3.687 \\ 744.941 \end{pmatrix}$

eigen vectors $\begin{bmatrix} 0.8503 & -0.526 & -0.0164 \\ -0.5264 & -0.85 & -0.02146 \\ 0.0026 & -0.02689 & 0.9996 \end{bmatrix}$

$$\text{ii) Explained Variance} = \frac{\sum_{i=1}^r \lambda_i}{\sum_{i=1}^n \lambda_i} = 0.9$$

NOTES

Now we need at least 2

$$\text{i) Projection Matrix} = U_r^T$$

$$\text{ii) Project the instance} = U_r^T \cdot Z$$

④	X_1	X_2
	6	-4
	-3	5
	-2	6
	7	-3

~~⑤~~

$$\text{data} = \begin{bmatrix} 0.5 & 4.5 & 2.5 \\ 2.2 & 1.5 & 0.1 \\ 3.9 & 3.5 & 1.1 \\ 0.5 & 1.9 & 4.9 \\ 0.5 & 3.2 & 1.2 \end{bmatrix}$$

iii) Explained Variance
 $= 0.8344$

~~⑥~~

$$\text{Centered data} = \begin{bmatrix} -1.71 & 1.58 & 0.54 \\ 0.36 & -1.42 & -1.86 \\ 2.06 & 0.88 & -0.86 \\ 0.26 & -1.02 & 2.94 \\ -1.34 & 0.28 & -0.76 \end{bmatrix}$$

~~⑦~~

$$\text{d) COV} = \begin{bmatrix} 1.6064 & -0.4148 & -0.2764 \\ -0.4148 & 1.1936 & -0.0432 \\ -0.2764 & -0.0432 & 2.7424 \end{bmatrix}$$

~~⑧~~

$$\text{b) i) Trace} = 1.6064 + 1.1936 + 2.7424 = 5.5424$$

Nile
STATIONERIES

$$a+b+c = 0.5424, \quad a = 0.9175$$

$$a = 0.505 \times b$$

$$b = 0.648 \times c$$

$$b = 1.8168$$

$$c = 2.808$$