—— ** problem

PCA → principle Component Analysis

used to
→* Reducing the Dimentionality / feature extension also

* Important point

* example : 100 variables are present
(suppose) * it reduces the 100 variable into 50 variable

---

step 1 : Data set

step 2 : Computation of Mean of variable

step 3 : Computation of covariance matrix.

step 4 : Eigen value, Eigen vector of Normalised Eigen vector

step 5 : New Dataset.

---

problem 1 :

features	Eg 1	Eg 2	Eg 3	Eg 4
x	4	8	13	7
y	11	4	5	14

Solution:

step 1 : Data set

No. of samples (N) = 4

No. of features (n) = 2

Step 2 : Computation of mean of variable

$$\bar{x} = \frac{4+8+13+7}{4}$$

$$\boxed{\bar{x} = 8}$$

$$\bar{y} = \frac{11+4+5+14}{4}$$

$$\boxed{\bar{y} = 8.5}$$

Step 3 : Covariance matrix (S)

Ordered pair : (x,x) (x,y) (y,y) (y,x)

$$\text{cov}(x,x) = \frac{1}{N-1} \Sigma ((x_i - \bar{x})^2$$

$$= \frac{1}{3}\left((4-8)^2 + (8-8)^2 + (13-8)^2 + (7-8)^2 \right)$$

$$= \frac{1}{3}(42)$$

$$\boxed{\text{cov}(x,x) = 14}$$

$$\text{cov}(y,y) = \frac{1}{N-1} \Sigma (y_i - \bar{y})^2$$

$$= \frac{1}{3}\left[(11-8.5)^2 + (4-8.5)^2 + (5-8.5)^2 + (14-8.5)^2 \right]$$

$$= \frac{1}{3}(69)$$

$$\boxed{\text{cov}(y,y) = 23}$$

$$\text{cov}(x,y) = \frac{1}{N-1}\left(\Sigma\left((x_i - \bar{x}) * (y_i - \bar{y}\right)\right)$$

$$= \frac{1}{3}\left((4-8)(11-8.5) + (8-8)(4-8.5) + (13-8)(5-8.5) + (7-8)(14-8.5) \right)$$

$$= \frac{-33}{3}$$

$$= -11$$

$$\boxed{\text{cov}(x,y) = \text{cov}(y,x) = -11}$$

Covariance matrix

$$S = \begin{bmatrix} cov(x,x) & cov(x,y) \\ cov(y,x) & cov(y,y) \end{bmatrix}$$

$$= \begin{bmatrix} 14 & -11 \\ -11 & 23 \end{bmatrix}$$

step 4: Eigen value, Eigen vector, Normalized Eign vector

i. Eigen value :

$$\det(S - \lambda I) = 0$$

$n = 2 \quad n \times n =$

$$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\det\left(\begin{bmatrix} 14 & -11 \\ -11 & 23 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}\right) = 0$$

$$\det\begin{bmatrix} 14-\lambda & -11 \\ -11 & 23-\lambda \end{bmatrix} = 0$$

$= 14(23) - 14\lambda - 23\lambda + \lambda^2 - 121$

$= \lambda^2 - 37\lambda + 322 - 121$

$= \lambda^2 - 37\lambda + 201$

$$(14-\lambda)(23-\lambda) - (11 \times 11) = 0$$

$$\lambda^2 - 37\lambda + 201 = 0$$

$$= \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

$$= \frac{37 \pm \sqrt{37^2 - 4(1)(201)}}{2}$$

Scanned with CamScanner

$$= \frac{37 \pm \cancel{234}\sqrt{1369 - 804}}{2}$$

$1/2$

$\sqrt{\ }$

$$= \frac{37 \pm \sqrt{565}}{2}$$

$$= \frac{37 \pm 23.76}{2}$$

$$= \frac{37 + 23.76}{2} \quad (\text{or}) \quad \frac{37 - 23.76}{2}$$

$$= \frac{60.76}{2} \quad (\text{or}) \quad \frac{13.24}{2}$$

$$\lambda_1 = 30.3849 \qquad \lambda_2 = 6.6151$$

take always largest value $\rightarrow \lambda_j$

ii) Eigen vector

$$(S - \lambda I) U = 0 \qquad U = \begin{bmatrix} U_1 \\ U_2 \end{bmatrix}$$

$$\begin{bmatrix} 14-\lambda & -11 \\ -11 & 23-\lambda \end{bmatrix} \begin{bmatrix} U_1 \\ U_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$(14-\lambda) U_1 - 11 U_2 = 0$ —— ①

$-11 U_1 + (23-\lambda) U_2 = 0$ —— ②

Eq^n ① $\quad 14 U_1 - \lambda U_1 - 11 U_2 = 0$

Eq^n ② $\quad -11 U_1 + 23 U_2 - \lambda U_2 = 0$

$$11 U_2 = (14-\lambda) U_1$$

$$\frac{U_1}{11} = \frac{U_2}{14-\lambda} \quad \text{—— ①}$$

$$11 U_1 = (23-\lambda) U_2$$

$$\frac{U_1}{23-\lambda} = \frac{U_2}{11} \quad \text{—— ②}$$

$$\frac{U_1}{11} = \frac{U_2}{14-\lambda} = t$$

we know that $\boxed{t=1}$ → Assume

$$\frac{U_1}{11} = 1$$

$$U_1 = 11$$

$$\frac{U_2}{14-\lambda} = \cancel{\frac{11}{11}}$$

$$\boxed{U_2 = 14-\lambda}$$

$$U = \begin{bmatrix} U_1 \\ U_2 \end{bmatrix} = \begin{bmatrix} 11 \\ 14-\lambda \end{bmatrix} = \begin{bmatrix} 11 \\ 14 - 30.3849 \end{bmatrix} = \begin{bmatrix} 11 \\ -16.3849 \end{bmatrix}$$

3. Normalized Eigenvector

$$\begin{bmatrix} \frac{U_1}{\sqrt{(U_1^2)^2 + U_2^2}} \\ \frac{U_2}{\sqrt{U_1^2 + U_2^2}} \end{bmatrix} = \begin{bmatrix} \frac{11}{\sqrt{11^2 + (-16.384)^2}} \\ \frac{-16.384}{\sqrt{11^2 + (-16.384)^2}} \end{bmatrix} \cdot \begin{bmatrix} 0.5574 \\ -0.8303 \end{bmatrix} = e$$

step 5 : New dataset :

	P_{11}	P_{12}	P_{13}	P_{14}
x	4	8	13	7
y	11	4	5	14

$$P_{11} = e_1^T \begin{bmatrix} x_1 - \bar{x} \\ y_1 - \bar{y} \end{bmatrix}$$

$$P_{11} = \begin{bmatrix} 0.5574 & -0.8303 \end{bmatrix} \begin{bmatrix} 4-8 \\ 11-8.5 \end{bmatrix}$$

$$= \begin{bmatrix} 0.5574 & -0.8303 \end{bmatrix} \begin{bmatrix} -4 \\ 2.5 \end{bmatrix}$$

$$= -2.2296 - 2.07575$$

$$= -4.30535$$

$$P_{12} = \begin{bmatrix} 0.5574 & -0.8303 \end{bmatrix} \begin{bmatrix} 0 \\ -4.5 \end{bmatrix}$$

$$= 3.73635$$

$$P_{13} = \begin{bmatrix} 0.5574 & -0.8303 \end{bmatrix} \begin{bmatrix} 5 \\ -3.5 \end{bmatrix}$$

$$= 2.787 + 2.90605$$

$$= 5.69305$$

$$P_{14} = \begin{bmatrix} 0.5574 & \cancel{-0} -0.8303 \end{bmatrix} \begin{bmatrix} -1 \\ 5.5 \end{bmatrix}$$

$$= -0.5574 - 4.56665$$

$$= -5.12405$$

Scanned with CamScanner