PCA

① Given the data table, reduce the dimension from 2 to 1 using principal component analysis. (PCA)

Features	Samples N ; N=1,2,3...			
	N_1	N_2	N_3	N_4
x_1	4	8	13	7
x_2	11	4	5	14

So, Performing PCA

Step 1. calculating mean ~~vector~~ from x_1 and x_2 features and form vector.

Step 2. calculate the co-variance matrix '$\underline{\underline{S}}$'

Step 3. Calculate the Eigen values of the covariance matrix.

$$S = \begin{bmatrix} cov(x_i, x_p) \end{bmatrix}$$

⇒ where, p is the no. of features;

$p = 2$; x_1 & x_2

$$S = \begin{bmatrix} cov(x_1, x_1), & cov(x_1, x_2) \\ cov(x_2, x_1), & cov(x_2, x_2) \end{bmatrix}$$

$$\cancel{cov(x_1, x_2) = \frac{1}{N-1} \sum_{K=1}^{N} (x_{1K} - \bar{x}_1)(x_{2K} - \bar{x}_2)}$$

$$cov(x_1, x_2) = \frac{1}{N-1} \sum_{K=1}^{N} (x_{1K} - \bar{x}_1)(x_{2K} - \bar{x}_2)$$

Step 4: Eigen Values of the covariance matrix

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

Where I is Identity Matrix

Step 5. Computation of eigenvectors,

$$u = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

Step 6. Check the large eigen value ; between λ_1 & λ_2

Step 7. Find unit eigen vector of large eigen value.

Step 8. Compute final PCA from eigen vectors.

Step 9. Perform transformation

Eigen values :- characteristic roots ; Non-zero character Directions

Eigen vectors :- The direction representation; Removes Noise.

Rule :: A nonzero vector is an eigen vector if there is a number λ such That

$$\boxed{Ax = \lambda x}$$

λ = Eigen value
a scalar value means it has only one component

Performing
Step 1. calculating the mean μ_1 & μ_2

N = 4

$$\mu_1 = \bar{X}_1 ; \quad \bar{X}_1 = \frac{1}{4}(4+8+13+7) = 8$$

$$\mu_2 = \bar{X}_2 ; \quad \bar{X}_2 = \frac{1}{4}(11+4+5+14) = 8.5$$

So, Mean vector $= \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} 8 \\ 8.5 \end{bmatrix}$

Step 2. Calculating covariance Matrix 'S'.

$$S = \begin{bmatrix} cov(x_1, x_1) , & cov(x_1, x_2) \\ cov(x_2 x_1) & cov(x_2 x_2) \end{bmatrix}$$

Performing $(x_1, x_1) = \frac{1}{N-1}\sum_{K=1}^{N}(x_{1K} - \bar{x}_1)(x_{1K} - \bar{x}_1)$

$$= \frac{1}{4-1}\sum_{K=1}^{4}(x_{1K} - \bar{x}_1)^2$$

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

$$\text{conv}(x_1, x_2) = \frac{1}{N-K} \sum_{K=1}^{N} (x_{1K} - \bar{x}_1)(x_{2K} - \bar{x}_2)$$

$$= \frac{1}{3} \sum_{K=1}^{4} (x_{1K} - \bar{x}_1)(x_{2K} - \bar{x}_2)$$

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

$$= -11$$

$$\text{conv.}(x_2\, x_1) = \text{conv.}(x_1, x_2) = -11$$

$$\text{conv.}(x_2, x_2) = \frac{1}{N-1} \sum_{K=1}^{N} (x_{2K} - \bar{x}_2)(x_{2K} - \bar{x}_2)$$

$$= \frac{1}{3} \sum_{K=1}^{4} (x_{2K} - \bar{x}_2)^2$$

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

$$= 23$$

so, final covariance Matrix for S is

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

Step 4 Finding eigen values of co-variance Matrix; where determinant $(S - \lambda I) = 0$

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

$$\Rightarrow (14-\lambda)(23-\lambda) - (-11)^2 = 0$$

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

So, $\lambda = \frac{1}{2}(37 \pm \sqrt{565})$

$$\lambda = \frac{1}{2}\left(37 \pm \sqrt{565}\right)$$

$$= 30.3849;\ 6.6151$$

$$\lambda_1 = 30.3849$$

$$\lambda_2 = 6.6151$$

Step 5. computing the eigenvectors

$$u = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

$$\Rightarrow (S - \lambda I)u = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} 14-\lambda & -11 \\ -11 & 23-\lambda \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = 0$$

$$\Rightarrow \begin{bmatrix} (14-\lambda)u_1 & -11u_2 \\ -11u_1 & (23-\lambda)u_2 \end{bmatrix} = 0$$

$$\Rightarrow (14-\lambda)u_1 - 11u_2 = 0 \quad \text{— (eq}^n\text{ 1)}$$

$$\Rightarrow -11u_1 + (23-\lambda)u_2 = 0 \quad \text{— (eq}^n\text{ 2)}$$

Now, from eq^n; we can get,

$$(14-\lambda)u_1 - 11u_2 = 0$$

$$\Rightarrow (14-\lambda)u_1 = 11u_2$$

$$\Rightarrow \frac{u_1}{11} = \frac{u_2}{14-\lambda} = t$$

So, $u_1 = 11t$; $u_2 = (14-\lambda)t$

If we assume $t=1$; then $u_1 = 11$;
$u_2 = 14-\lambda$

So, $u = \begin{bmatrix} 11 \\ 14-\lambda \end{bmatrix}$

Now checking larger Eigen value;

$\lambda_1 > \lambda_2$
30.3849 6.6151

So, we will consider $\lambda_1 = 30.3849$

Now, finding unit of eigenvector (e_1) and compute length of u_1

$$\|u_1\| = \sqrt{11^2 + (14-\lambda)^2}$$
$$= \sqrt{(11)^2 + (14-30.3849)^2}$$
$$= 19.7348.$$

$$e_1 = \begin{bmatrix} 11/\|u_1\| \\ (14-\lambda_1)/\|u_1\| \end{bmatrix} = \begin{bmatrix} 11/19.73 \\ (14-30.38)/19.73 \end{bmatrix}$$

$$= \begin{array}{cc} \begin{bmatrix} 0.5574 \\ -0.8303 \end{bmatrix} & \begin{array}{c} u_1 \\ u_2 \end{array} \end{array}$$

Thus the ~~PCA~~ PCA will be $\begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$

$$\begin{matrix} u_1 \\ u_2 \end{matrix} = \begin{bmatrix} 0.5574 \\ 0.8303 \end{bmatrix}$$

The feature vector (4, 11) gets transformed;

to = Transpose of Eigen Vector x [feature vector - Mean vector]

$$= \begin{bmatrix} 0.5574 \\ 0.8303 \end{bmatrix}^T \times \left(\begin{bmatrix} 4 \\ 11 \end{bmatrix} - \begin{bmatrix} 8 \\ 8.5 \end{bmatrix} \right)$$

$$= (0.5574 - 0.8303) \times \begin{bmatrix} -4 \\ 2.5 \end{bmatrix}$$

$$= \cancel{4.03} \quad 4.30535$$