PCA Problem - 1

I) Given the following data, use PCA to reduce the dimension from 2 to 1.

Feature	Ex1	Ex2	Ex 3	Ex 4
x	4	8	13	7
y	11	4	5	14

Ans:- Step1: Data set.

Feature	Ex1	Ex2	Ex 3	Ex4
x	4	8	13	7
y	11	4	5	14

No of features $n = 2$

No. of samples, $N = 4$

Step2:- Computation of mean of variables

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

$$\bar{y} = \frac{11+4+5+14}{4} = 8.5$$

Step3:- Computation of covariance Matrix.

Ordered pair = (x,x), (x,y) (y,x) (y,y)

(i) covariance of all ordered pairs.

$$\boxed{\begin{gathered} cov(x_i, x_j) = \frac{1}{N-1} \sum_{k=1}^{N} (x_{ik} - \bar{x}_i)(x_{jk} - \bar{x}_j) \\ \text{or} \\ cov(x,x) = \frac{1}{N-1} \sum_{k=1}^{N} (x_i - \bar{x})^2 \end{gathered}}$$

$$cov(x,x) = \frac{1}{4-1}\Big[(\overset{E_1}{4}-8)^2 + (\overset{Ex_2}{8}-8)^2 + (\overset{Ex_3}{13}-8)^2 + (\overset{Ex_4}{7}-14)^2 \Big] = 14$$

(x values only)

$$\operatorname{cov}(x,y) = \frac{1}{4-1}\Big[(\overset{x_1}{4}-\overset{\bar{x}}{8})(\overset{y_1}{11}-\overset{\bar{y}}{8.5}) + (\overset{x_2}{8}-8)(\overset{y_2}{4}-8.5) + (13-8)(5-8.5) + (7-8)(14-8.5)\Big]$$

$$= \underline{\underline{-11}}$$

$$\operatorname{cov}(y,x) = \underline{\underline{-11}}$$

$$\operatorname{cov}(y,y) = \frac{1}{4-1}\Big[(\overset{y_1}{11}-\overset{\bar{y}}{8.5})^2 + (\overset{y_2}{4}-\overset{\bar{y}}{8.5})^2 + (\overset{y_3}{5}-8.5)^2 + (\overset{y_4}{4}-8.5)^2\Big]$$

$$= \underline{\underline{23}}$$

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

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

Step 4:- Find Eigen Value, Eigenvector and Normalized Eigen vector

(i) Eigen Value

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

$$\lambda I = \underline{\underline{\begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}}}$$

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

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

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

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

$$\det(S-\lambda I) \Rightarrow (14-\lambda)(23-\lambda) - (-11 \times -11) = 0.$$

$\lambda^2 - 37\lambda + 201 = 0.$ — ①

Solving the Quadratic Eqn (I)

$\lambda_1 = \underline{\underline{30.3849}}$ $\lambda_2 = \underline{\underline{6.6151}}$

(Eigen Values)

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

where $a = 1$
$b = -37$
$c = 201$

Arrange Eigen Values in descending order

$\lambda_1 = \underline{\underline{30.3849}} \Rightarrow$ first principal component.

$\lambda_2 = \underline{\underline{6.6151}}$ (we proceed wt λ_1)

(ii) Eigen vector of λ_1 (find U_1)

$(S - \lambda_1 I) U_1 = 0$

$$\begin{pmatrix} 14 - \lambda_1 & -11 \\ -11 & 23 - \lambda_1 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = 0.$$

$(14 - \lambda_1) u_1 + -11 (u_2) = 0$ — ①

$-11 u_1 + (23 - \lambda_1) u_2 = 0.$ — ②

consider eqn ①

$(14 - \lambda_1) u_1 = 11 (u_2)$

$$\frac{u_1}{11} = \frac{u_2}{(14 - \lambda_1)} = t$$

when $t = 1$, $u_1 = \underline{\underline{11}}$ $u_2 = \underline{\underline{14 - \lambda_1}}$

Eigen Vector U_1 of $\lambda_1 = \begin{bmatrix} 11 \\ 14 - \lambda_1 \end{bmatrix}$ || $\lambda_1 = 30.3849$

$= \begin{bmatrix} 11 \\ 14 - 30.3849 \end{bmatrix} = \begin{bmatrix} 11 \\ -16.3849 \end{bmatrix}$

(iii) Normalize the Eigen Vector U_1

$$e_1 = \begin{bmatrix} \dfrac{11}{\sqrt{11^2 + (-16.3849)^2}} \\ \dfrac{-16.3849}{\sqrt{11^2 + (-16.3849)^2}} \end{bmatrix} = \begin{bmatrix} 0.5574 \\ -0.8303 \end{bmatrix}$$

length

λ_2 (Not need) $e_2 = \begin{bmatrix} 0.8303 \\ 0.5574 \end{bmatrix}$ (follow the same steps of above).

step5 :- Derive new data set.

	Ex1	Ex2	Ex.3	Ex4
First principal PC1 Component	P_{11}	P_{12}	P_{13}	P_{14}

$P_{11} = e_1^T \begin{bmatrix} 4 - 8 \\ 11 - 8.5 \end{bmatrix}$ → mean $\bar{x}$, → mean $\bar{y}$

first point (4, 11)

$= \begin{bmatrix} 0.5574 & -0.8308 \end{bmatrix} \begin{bmatrix} -4 \\ 2.5 \end{bmatrix} = \underline{-4.3052}$

$P_{12} = e_1^T \begin{bmatrix} 8 - 8 \\ 4 - 8.5 \end{bmatrix} = \begin{bmatrix} 0.5574 - 0.8308 \end{bmatrix} \begin{bmatrix} 0 \\ -3.5 \end{bmatrix} = 3.736$

Second Point (8, 4)

$P_{13} = 5.6928$

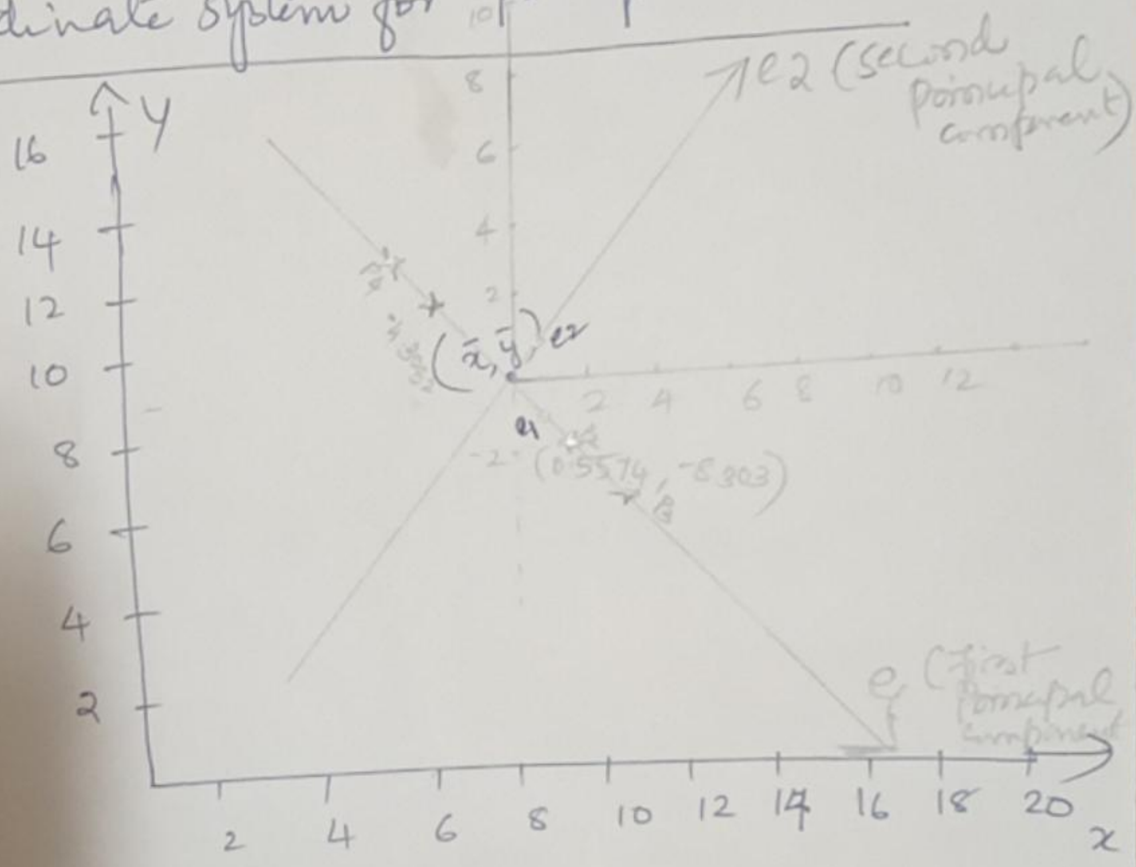
$P_{14} = -5.1238$

Final Dataset

	Ex1	Ex2	Ex3	Ex4
PC_1	-4.3052	3.7361	5.6928	-5.1238
	P_1	P_2	P_3	P_4

$\Rightarrow$ new dataset Dimension is 1

Draw coordinate system for principal component.



First mark; mean $(\bar{x}, \bar{y}) = (8, 8.5)$

Secondly mark e_1 and e_2

$$e_1 = \begin{bmatrix} 0.5574 \\ -0.8303 \end{bmatrix}$$

$$e_2 = \begin{bmatrix} 0.8303 \\ 0.5574 \end{bmatrix}$$