PCA (Principal Component Analysis)

Q- Given the following data, use PCA to reduce the dimension from 2 to 1.

x	y
4	11
8	4
13	5
7	14

Step 1 :- Dataset :

No. of features, $n = 2$

No. of samples, $N = 4$

Step 2 :- Computation of mean of variables :

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

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

Step 3 :- Computation of covariance matrix :

ordered pairs are : (x,x) (x,y) (y,x) (y,y)

Covariance of all ordered pairs

$$\mathrm{cov}(x,x) = \frac{1}{N-1}\sum_{k=1}^{N}(x_{ik}-\bar{x}_i)(x_{jk}-\bar{x}_j)$$

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

$$= \frac{1}{3}\left[16+0+25+1\right]$$

$$= 42/3 = 14$$

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

$$= -11$$

$$\mathrm{cov}(y,x) = -1$$

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

$= 23$

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

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

Step 4:- Eigen value and Eigen vector
Normalized eigen vector.

i) Eigen value $|S - \lambda I| = 0$

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

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

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

$\Rightarrow \lambda = 30.3849, \; 6.6151$

$\Rightarrow \lambda_1 = 30.3849$ and $\lambda_2 = 6.6151$

$\Downarrow$

First principal component.

ii) Eigen vector of λ_1

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

$\begin{bmatrix} 14-\lambda_1 & -11 \\ -11 & 23-\lambda \end{bmatrix} \begin{bmatrix} U_1 \\ U_2 \end{bmatrix} = 0.$

$\begin{bmatrix} (14-\lambda_1)(u_1) - 11u_2 \\ -11u_1 + (23-\lambda)u_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

$$(14-\lambda_1)\,u_1 - 11u_2 = 0$$

$$-11u_1 + (23-\lambda)\,u_2 = 0.$$

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

when $t = 1$, $u_1 = 11$ & $u_2 = 14 - \lambda_1$

Eigen vector U_1 of $\lambda_1 = \begin{bmatrix} 11 \\ 14-\lambda_1 \end{bmatrix} = \begin{bmatrix} 11 \\ 14-30.3841 \end{bmatrix}$

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

iii) Normalize the eigen vector U_1

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

$$= \begin{bmatrix} 0.5574 \\ -0.8313 \end{bmatrix}$$

For λ_2, $e_2 = \begin{bmatrix} 0.8303 \\ 0.5574 \end{bmatrix}$

Step 5:- Derive new dataset

First Principal PC1 component
P_{11}
P_{12}
P_{13}
P_{14}

$$\Rightarrow P_{11} = e_1^T \begin{bmatrix} 4-8 \\ 11-8.5 \end{bmatrix}$$

$$= \begin{bmatrix} 0.5574 & -0.8313 \end{bmatrix} \begin{bmatrix} -4 \\ 2.5 \end{bmatrix}$$

$$= -4.3052$$

$$P_{12} = \begin{bmatrix} 0.5544 & -0.8313 \end{bmatrix} \begin{bmatrix} 8-8 \\ 4-8.5 \end{bmatrix}$$

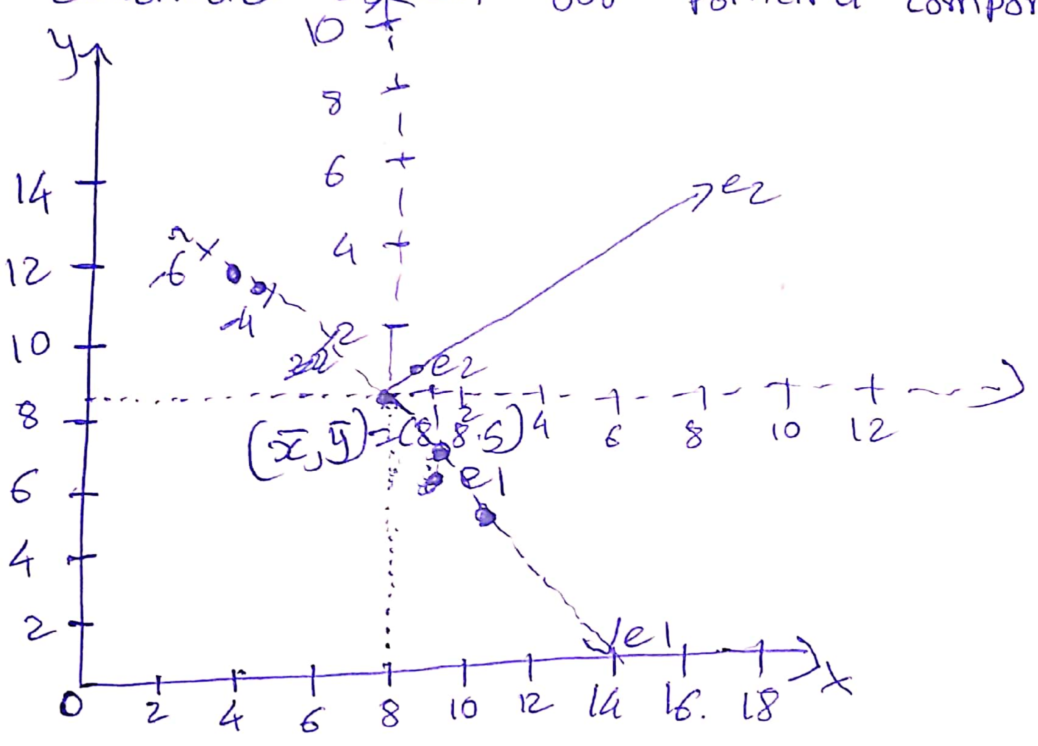
$$= 3.7361$$

Similarly $P_{13} = 5.6928$

$P_{14} = -5.1238.$

PC1
-4.3052
3.7361
5.6928
-5.1238

Coordinate system for principal components



Plotted point of PC1 on e1