PRINCIPAL COMPONENT ANALYSIS

PROBLEM

Consider the following data, use PCA to reduce the dimension from 2 to 1.

Example	Feature 1 x	Feature 2 y
1	4	11
2	8	4
3	13	5
4	7	14

NO. OF FEATURES, $n = 2$

NO. OF SAMPLES, $N = 4$

STEP 1: 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 2: COMPUTATION OF COVARIANCE MATRIX

Ordered pairs are (x, x) (x, y) (y, x) (y, y)

$$\underbrace{\text{Cov}(x, x)}_{\text{Same variables.}} \xrightarrow{\text{formula}} = \frac{1}{N-1} \sum_{K=1}^{N} (x_i - \bar{x})^2$$

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

$$= \frac{1}{3} \left[(-4)^2 + (0)^2 + (5)^2 + (1)^2 \right]$$

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

$$= \frac{1}{3} \left[42 \right] \Rightarrow 14$$

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

$$\text{cov}(x, y) = \frac{1}{N-1} \sum_{K=1}^{N} (x_{iK} - \bar{x}_i)(x_{jK} - \bar{x}_j)$$

$x \rightarrow x_i$, $y \rightarrow x_j$

x - 1st value

y - 1st value

$$= \frac{1}{4-1} \Big[\big[(x_{i1} - \bar{x}_i)(x_{j1} - \bar{x}_j) \big]$$

x-2nd value

y-2nd value

$$+ \big[(x_{i2} - \bar{x}_i)(x_{j2} - \bar{x}_j) \big]$$

x-3rd value

y-3rd value

$$+ \big[(x_{i3} - \bar{x}_i)(x_{j3} - \bar{x}_j) \big]$$

x-4th value

y-4th value

$$+ \big[(x_{i4} - \bar{x}_i)(x_{j4} - \bar{x}_j) \big] \Big]$$

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

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

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

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

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

Covariance matrix $n \times n$ 2×2

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

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

STEP 3: EIGEN VALUE, EIGEN VECTOR, NORMALIZED EIGEN VALUE

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

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

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

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

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

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

$$\frac{1}{2a}\sqrt{b^2 - 4ac}$$

$a = 1 \quad b = 37$

$c = 201$

$$\lambda = 30.3849,\ 6.6151$$

$$\lambda 1 > \lambda_2$$

$\lambda 1 = 30.3849$, $\lambda 2 = 6.6151$

(ii) EIGEN VECTOR OF $\lambda 1$

$$(S - \lambda_1 I)\vec{U_1} \stackrel{?}{=} 0 \qquad U_1 = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

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

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

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

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

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

when $t = 1$

$u_1 = 11 \qquad u_2 = 14 - \lambda$

Eigen Vector U_1 of λ_1

$$= \begin{bmatrix} 11 \\ 14 - \lambda_1 \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} 11 \\ 14 - 30.3841 \end{bmatrix} = \begin{bmatrix} 11 \\ -16.3849 \end{bmatrix}$$

(iii) Normaliced the Eigen Vector v_1

$$e_1 = \begin{bmatrix} 11/\sqrt{11^2+(-16.36)^2} \\ \dfrac{-16.3849}{\sqrt{11^2+(-16.36)^2}} \end{bmatrix}$$

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

λ_2

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

STEP 4 : REDUCED DATASET

EXAMPLE	PRINCIPAL COMPONENT PC1
1	-4.3052
2	3.7361
3	5.6928
4	-5.1238

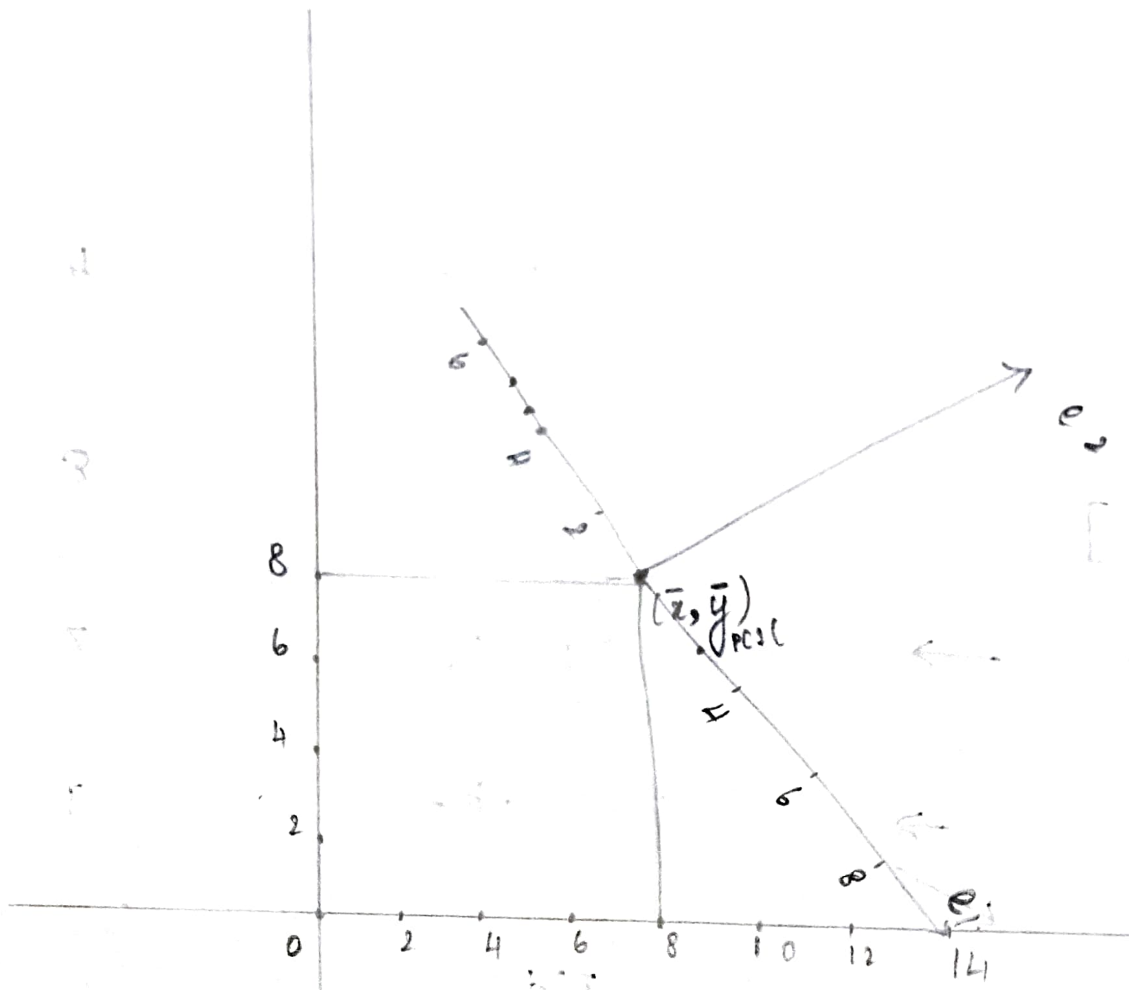
Reduced Dimension

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

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

$$e_1^T \begin{bmatrix} 8-8 \\ 4-8.5 \end{bmatrix}$$

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



$(\bar{x}, \bar{y})$
PC1
e_1
e_2