Parshvanath Charitable Trust's
A. P. SHAH INSTITUTE OF TECHNOLOGY
(Approved by AICTE New Delhi & Govt. of Maharashtra, Affiliated to University of Mumbai)
(Religious Jain Minority)

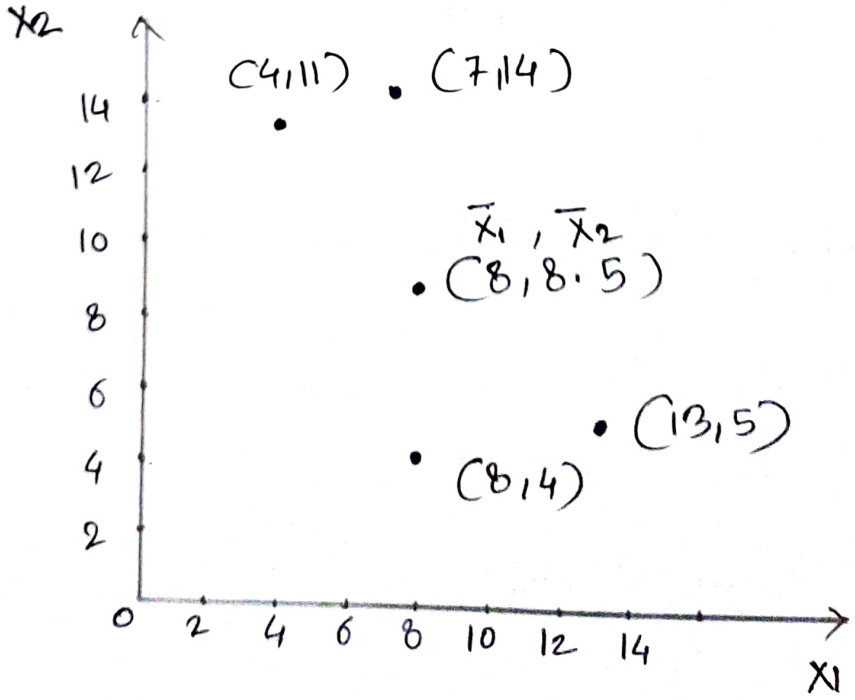
DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
(ARTIFICIAL INTELLIGENCE & MACHINE LEARNING)

Poincipal Component Analysis

① Given the data in a table, reduce the dimensions from 2 to 1 using PCA.

Feature	1	2	3	4
X_1	4	8	13	7
X_2	11	4	5	14

Ans: The following graph shows the scatter plot of the given data points.



step 1: Calculate mean

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

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

step 2: Calculation of the covariance matrix.

$$Cov(X_1, X_1) = \frac{1}{n-1}(4-8)^2 + (8-8)^2 + (13-8)^2 + (7-8)^2$$

$$= 14$$

$$Cov(X_2, X_2) = \frac{1}{n-1}(11-8.5)^2 + (4-8.5)^2 + (5-8.5)^2 + (14-8.5)^2$$

$$= 23$$

$$Cov(X_1, X_2) = \frac{1}{3}((4-8)(11-8.5) + (8-8)(4-8.5) + (13-8)(5-8.5) + (7-8)(14-8.5))$$

$$= -11$$

$$Cov(X_2, X_1) = -11$$

∴ The covariance matrix is,

$$S = \begin{bmatrix} cov(X_1 X_1) & cov(X_1 X_2) \\ cov(X_2 X_1) & cov(X_2 X_2) \end{bmatrix}$$

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

step 3: Eigen values of the covariance matrix.

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

$$= (14-\lambda)(23-\lambda) - (-11) * (-11)$$

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

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

$$= 30.38, 6.6151$$

$$\therefore \lambda_1 = 30.38$$

$$\lambda_2 = 6.6151$$

step 4 : Compute eigenvector

Compute eigenvector corresponding to largest eigenvalue.

Consider eigenvector $= \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$

$$\therefore S - \lambda I X = \begin{bmatrix} 14-\lambda_1 & -11 \\ -11 & 23-\lambda_1 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

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

By above matrix, we will get 2 equations,

$$(14-\lambda_1)u_1 - 11u_2 = 0$$
$$-11u_1 - (23-\lambda_1)u_2 = 0$$

Choose $\lambda_1 = 30.38$

$$\therefore -16.38u_1 - 11u_2 = 0$$
$$-11u_1 + (-7.38)u_2 = 0$$

above equations are not independant.

$$\therefore \quad \frac{u_1}{11} = \frac{u_2}{-16.38} = t$$

$$\therefore \quad u_1 = 11t$$

$$u_2 = -16.38t$$

Take $t = 1$

∴ eigenvector corresponding to eigenvalue λ is,

$$U = \begin{bmatrix} 11 \\ -16.38 \end{bmatrix}$$

- length of $U = \|U\| = \sqrt{11^2 + (-16.38)^2}$

$$= 19.73$$

∴ Unit eigenvector $e_1 = \begin{bmatrix} 11/\|U\| \\ -16.38/\|U\| \end{bmatrix}$

$$= \begin{bmatrix} 11/19.73 \\ -16.38/19.73 \end{bmatrix}$$

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

step 5 : Compute principal components

$$e^T \begin{bmatrix} x_{1k} - \bar{x}_1 \\ x_{2k} - \bar{x}_2 \end{bmatrix} = \begin{bmatrix} 0.5574 & -0.8303 \end{bmatrix} \begin{bmatrix} 4-8 \\ 11-8.5 \end{bmatrix}$$

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

$$= (0.5574 * -4) - 0.8303(2.5)$$

$$= -4.30535$$

$$e^T \begin{bmatrix} x_{1k} - \bar{x}_1 \\ x_{2k} - \bar{x}_2 \end{bmatrix} = \begin{bmatrix} 0.5574 & -0.8303 \end{bmatrix} \begin{bmatrix} 8-8 \\ 4-8.5 \end{bmatrix}$$

$$= 0.5574(0) - 0.8303(-4.5)$$

$$= 3.7363$$

$$e^T \begin{bmatrix} x_{1k} - \bar{x}_1 \\ x_{2k} - \bar{x}_2 \end{bmatrix} = \begin{bmatrix} 0.5574 & -0.8303 \end{bmatrix} \begin{bmatrix} 13-8 \\ 5-8.5 \end{bmatrix}$$

$$= 0.5574(5) - 0.8303(-3.5)$$

$$= 2.787 + 2.91$$

$$= 5.693$$

Parshvanath Charitable Trust's
A. P. SHAH INSTITUTE OF TECHNOLOGY
(Approved by AICTE New Delhi & Govt. of Maharashtra, Affiliated to University of Mumbai)
(Religious Jain Minority)

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
(ARTIFICIAL INTELLIGENCE & MACHINE LEARNING)

$$e^T \begin{bmatrix} x_{1k} - \bar{x}_1 \\ x_{2k} - \bar{x}_2 \end{bmatrix} = \begin{bmatrix} 0.5574 & -0.8303 \end{bmatrix} \begin{bmatrix} 7-8 \\ 14-8.5 \end{bmatrix}$$

$$= 0.5574(-1) - 0.8303(5.5)$$

$$= -0.5574 - 4.567$$

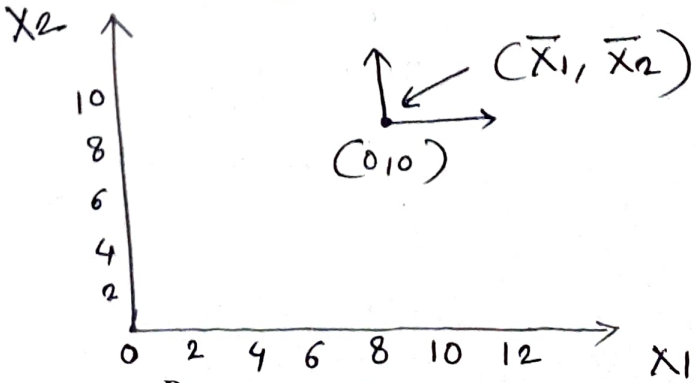
$$= -5.124$$

The results of above calculation are listed in below table.

X_1	4	8	13	7
X_2	11	4	5	14
First Principal Components	-4.3053	3.7363	5.693	-5.124

step 6: Shift (project) data points on PC (first)

First, shift the origin (0,0) to the mean.



Now draw the eigenvector from this origin

$$e = \begin{bmatrix} 0.55 \\ -0.83 \end{bmatrix}$$