Principle Component Analysis - Dimensionality Reduction in ML

- Given the data in Table, reduce the dimension from 2 to 1 using the Principal Component Analysis (PCA) Algorithm.

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

Sol :-

Step 1: Calculate Mean

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

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

Step 2 :- Calculation of the covariance matrix.

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

$$Cov(X_1, X_1) = \frac{1}{N-1} \sum_{K=1}^{N} (X_{1K} - \bar{X}_1)(X_{1K} - \bar{X}_1)$$

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

$$Cov(X_1, X_1) = 14$$

$$Cov(X_1, X_2) = \frac{1}{N-1} \sum_{K=1}^{N} (X_{1K} - \overline{X}_1)(X_{2K} - \overline{X}_2)$$

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

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

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

$$Cov(X_2, X_2) = \frac{1}{N-1} \sum_{K=1}^{N} (X_{2K} - \overline{X}_2)(X_{2K} - \overline{X}_2)$$

$$= \frac{1}{3}\Big((11-8.5)^2 + (4-8.5)^2 + (5-8.5)^2 + (14-8.5)^2 \Big)$$

$$= 23$$

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

Step 3:- Eigenvalues of the covariance matrix.

The characteristic equation of the covariance matrix is,

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

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

$$0 = (14-\lambda)(23-\lambda) - (-11)(-11)$$
$$= 14(23-\lambda) - \lambda(23-\lambda) - 121$$
$$= 322 - 14\lambda - 23\lambda + \lambda^2 - 121$$
$$= \lambda^2 - 37\lambda + 201$$

$\lambda_1 = 30.3849 \quad \lambda_2 = 6.6151$

Step 4 :- Computation of the eigenvectors

$$U = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \qquad \begin{bmatrix} 0 \\ 0 \end{bmatrix} = (S - \lambda I)U$$

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

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

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

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

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

$$U_1 = \begin{bmatrix} 11 \\ 14-\lambda \end{bmatrix}$$

- To find a unit eigenvector, we compute the length of U_1 which is given by, ④

$$\|U_1\| = \sqrt{11^2 + (14-\lambda)^2}$$

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

$$e_1 = \begin{bmatrix} 11/\|U_1\| \\ (14-\lambda_1)/\|U_1\| \end{bmatrix}$$

$$= \begin{bmatrix} 11/19.7348 \\ (14-30.3849)/19.7348 \end{bmatrix}$$

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

$$\|U_2\| = \sqrt{11^2 + (14-6.6151)^2}$$

$$e_2 = \begin{bmatrix} 11/\|U_2\| \\ (14-\lambda_2)/\|U_2\| \end{bmatrix}$$

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

Step 5 : Computation of first principal component

$$e_1^T \begin{bmatrix} X_{1k} - \bar{X}_1 \\ X_{2k} - \bar{X}_2 \end{bmatrix}$$

$$e_1^T \begin{bmatrix} X_{1k} - \bar{X}_1 \\ X_{2k} - \bar{X}_2 \end{bmatrix} = \begin{bmatrix} 0.5574 & -0.8303 \end{bmatrix} \begin{bmatrix} X_{11} - \bar{X}_1 \\ X_{21} - \bar{X}_2 \end{bmatrix}$$

$$= (0.5574)(X_{11} - \bar{X}_1) - 0.8303(X_{21} - \bar{X}_2)$$
$$= (0.5574)(4-8) - 0.8303(11-8.5)$$
$$= -4.30535$$

$$\cancel{e_1^T}$$

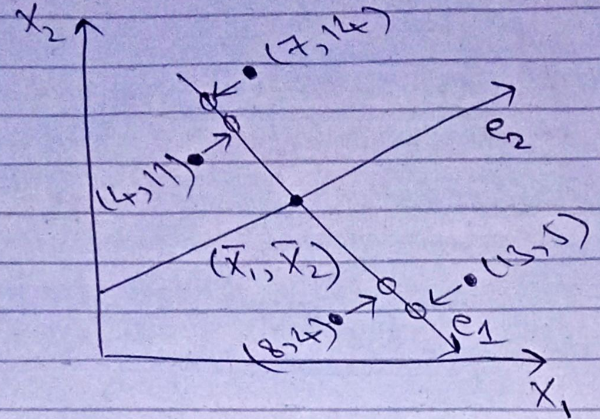
$$= (0.5574)(X_{12} - \bar{X}_1) - 0.8303(X_{22} - \bar{X}_2)$$
$$= (0.5574)(8-8) - 0.8303(4-8.5)$$
$$= 3.7361$$

$$= (0.5574)(13-8) - (0.8303)(5-8.5)$$
$$= 5.6928$$

$$= (0.5574)(7-8) - (0.8303)(14-8.5)$$
$$= -5.1238$$

Feature	Ex 1	Ex 2	Ex 3	Ex 4
X_1	4	8	13	7
X_2	11	4	5	14
First Principle components	-4.3052	3.7361	5.6928	-5.1238

Step 6 :- Geometrical meaning of first principal components.



—— X —— X ——