PCA - used for dimensionality reduction in ML
↳ high dim data to low dim.

Q Given the data in the table, reduce the dimension from 2 to 1 using PCA algorithm

Feature	Eg1	Eg2	Eg3	Eg4
X1	4	8	13	7
X2	11	4	5	14

Feature 1 → x_1
Feature 2 → x_2

step 1:

Calculate Mean

$$\overline{X_1} = \frac{4+8+13+7}{4} = 8$$

$$\overline{X_2} = \frac{11+4+5+14}{4} = 8.5$$

step 2:

Calculate covariance matrix

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

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

N → no. of egs. = 4

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

$$= 14$$

$$Cov(X1, X2) = \frac{1}{N-1} \sum_{k=1}^{N} (X_{1k} - \bar{X_1})(X_{2k} - \bar{X_2})$$

$$Cov(X1, X_2) = \frac{1}{3}((4-8) \cdot (11-8.5) + (8-8)$$

$$(4-8.5) + (13-8) \cdot (5-8.5) + (7-8) \cdot (14-$$

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

$$Cov(X_2, X1) = Cov(X1, X_2) = \underline{\underline{-11}}$$

$$Cov(X_2, X_2) = \frac{1}{3}((11-8.5)^2 + (4-8.5)^2 + (5-8.5)^2 + (14-8.5)^2)$$

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

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

<u>step3</u>: <u>Eigenvalues of the covariance matrix</u>.

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

S → covariance matrix I → Identity matrix

$$0 = det(S - \lambda I) \qquad \lambda I = \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}$$

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

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

deteemment

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

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

Calulating roots

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

step 4. Eigen vector calulation

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

u - eigen vector

S → covariance matrix   I - Identity matrix

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

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

· Use any one equation to get value

Consider 1st eqn.

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

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

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

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

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

Assume $t = 1$

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

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

(λ_1, λ_2)

Consider largest eigen value to calculate PCA.

here we want 1 principal component, we need one value

here we use $\lambda_1 = 30.3849$

$$u_1 = \begin{bmatrix} 11 \\ 14-\lambda_1 \end{bmatrix}$$

To find unit eigen vector, we compute length of u, which is ~~used~~ given by

$$\|u_1\| = \sqrt{11^2 + (14-\lambda_1)^2}$$

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

$$= \underline{\underline{19.7348}}$$

$$e_1 = \begin{bmatrix} 11/\|u_1\| \\ (14-\lambda)/\|u_1\| \end{bmatrix}$$

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

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

Similarly by taking λ_2 value as 6.6151

we will get $e_2 = \begin{bmatrix} 0.8303 \\ 0.5574 \end{bmatrix}$

Step 5: Computation of 1st principal compon...

for example 1

$$e_1^T \times \begin{bmatrix} X_{1k} - \overline{X_1} \\ X_{2k} - \overline{X_2} \end{bmatrix}$$

$$e_1^T \begin{bmatrix} X_{1k} - \overline{X_1} \\ X_{2k} - \overline{X_2} \end{bmatrix} = \begin{bmatrix} 0.5574 & -0.8303 \end{bmatrix} \times \begin{bmatrix} X_{11} - \overline{X_1} \\ X_{21} - \overline{X_2} \end{bmatrix}$$

$$= 0.5574 (X_{11} - \overline{X_1}) - 0.8303 (X_{21} - \overline{X_2})$$

$$= 0.5574 (4-8) - 0.8303 (11-8.5)$$

$$= -4.30535$$

for example 2

$$e_1^T \begin{bmatrix} X_{12} - \overline{X_1} \\ X_{22} - \overline{X_2} \end{bmatrix}$$

$$= 3.7361$$

Calculate same for example 3 & 4

Feature	Eg1	Eg2	Eg3	Eg4
X1	4	8	13	7
X2	11	4	5	14
First principal components	-4.3052	3.7361	5.6928	-5.1238

we have 2 feature X1, X2
we reduced that to the first principal component.