1) Principal Component Analysis (PCA) is Feature Extraction ie.,

"Features of a data set should be less as well as the Similarity between each other is very less.

Working of PCA:

PCA works on the process called Eigenvalue Decomposition of a Covariance matrix of a data set.

→ standardize the dataset.

→ First, Calculate the Covariance matrix of a data set

→ Calculate the Eigen Vectors of the Covariance matrix

→ the EigenVector having the highest eigen Value represents the direction in which there is the highest Variance. So this will help in identifying the first principal Component

→ The eigen Vector having the highest EigenValue represents the direction in which data has the highest remaining variance and also orthogonal to the first direction. So, this helps in identifying the second principal Component

→ Like this, identify the top 'k' eigenVectors having top 'k' eigen Values to get the 'k' principal Components

Numerical:

Consider the following data set

$$X_1 : \{2.5,\ 0.5,\ 2.2,\ 1.9\}$$

$$X_2 : \{2.4,\ 0.7,\ 2.9,\ 2.2\}$$

$$X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$$

Step 1: standardize data set

Mean of X_1 : 1.775

Mean of X_2 : 2.05

Change the data set:

$$X_1 : X_1 - \text{mean} = \{0.725,\ -1.275,\ 0.425,\ 0.125\}$$

$$X_2 : X_2 - \text{mean} = \{0.35,\ -1.35,\ 0.85,\ 0.15\}$$

$$X : \begin{bmatrix} 0.725 & -1.275 & 0.425 & 0.125 \\ 0.35 & -1.35 & 0.85 & 0.15 \end{bmatrix}$$

Step 2: Find Eigen Value and Eigen Vector.

Correlation matrix : $C = \left(\dfrac{X \cdot X^T}{n-1} \right)$

here $n = 4$

$$C = \frac{1}{4-1} \begin{bmatrix} 0.725 & -1.275 & 0.425 & 0.125 \\ 0.35 & -1.35 & 0.85 & 0.15 \end{bmatrix} \begin{bmatrix} 0.725 & 0.35 \\ -1.275 & -1.35 \\ 0.425 & 0.85 \\ 0.125 & 0.15 \end{bmatrix}$$

$$= \frac{1}{3} \begin{bmatrix} 2.3475 & 2.355 \\ 2.355 & 2.69 \end{bmatrix}$$

$$C = \begin{bmatrix} 0.749167 & 0.785 \\ 0.785 & 0.8967 \end{bmatrix}$$

Using the Equation, $|C - \lambda I| = 0$

where, $\lambda =$ Eigen Value

$I =$ Identity matrix

$$\left| \begin{bmatrix} 0.749167 & 0.785 \\ 0.785 & 0.8967 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right| = 0$$

$$\begin{vmatrix} 0.749167 - \lambda & 0.785 \\ 0.785 & 0.8967 - \lambda \end{vmatrix} = 0$$

$$(0.749167 - \lambda)(0.8967 - \lambda) - (0.785)(0.785) = 0$$

$$\lambda^2 - 1.64597\lambda + 0.67178 - 0.616225 = 0$$

$$\lambda^2 - 1.64597\lambda + 0.5556 = 0$$

$$\lambda_1 = 1.17185$$

$$\lambda_2 = 0.474124$$

where λ_1 & λ_2 are two Eigen Values.

To find Eigen Vector from Eigen Values,

we had a formulae

$$C \cdot M = \lambda \cdot M$$

where: $M = \begin{bmatrix} x \\ y \end{bmatrix}$

where C is Correlation matrix.

First we find the Eigen vector for the Eigen value 1.17185

$$\begin{bmatrix} 0.749167 & 0.785 \\ 0.785 & 0.8967 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1.17185\, x \\ 1.17185\, y \end{bmatrix}$$

$$\Rightarrow 0.785x + 0.8967y = 1.17185y$$

$$0.785x = 0.27515y$$

$$\boxed{x = 0.35051y}$$

$$M = \begin{bmatrix} 0.35051 \\ 1 \end{bmatrix}$$

find the square root of the sum of the squares of the element in ~~Y~~ M

$$t = \sqrt{(0.35051)^2 + (1)^2} = 1.05965$$

B found eigenvector divide the Elements of the M ~~x~~ matrix by the number 1.3602

Eigenvector, (0.67787 and 0.73518)

$$\frac{1}{t} M = \begin{bmatrix} 0.67787 \\ 0.73518 \end{bmatrix}$$

find the Eigen vector for the Eigen Value 0.474124

$$\begin{bmatrix} 0.7491672 & 0.785 \\ 0.785 & 0.8967 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 0.474124 \begin{bmatrix} x \\ y \end{bmatrix}$$

$$0.785x + 0.8967y = 0.474124y$$

$$0.785x = -0.42258y$$

$$x = \frac{-0.42258}{0.785} y$$

$$\boxed{x = -0.53832y}$$

$$M = \begin{bmatrix} 1 \\ -1.8576 \end{bmatrix}$$

find the square root of the sum of the squares of the ~~number~~ Element in M

$$\sqrt{(1)^2 + (-1.8576)^2} = 2.1097$$

~~Now~~ divide the Elements of the M matrix by the number 2.1097

$$\text{Eigen vector} = \begin{bmatrix} 0.474 \\ 0.881 \end{bmatrix}$$

Sum of Eigen Values (λ_1) & (λ_2) = $1.17185 + 0.474124$

$= 1.645974$

= Total variance

(Majority of variance comes from λ_1)

Z