PRINCIPAL COMPONENT ANALYSIS :- (PCA)

1. Every data set, to be used for ML model, have multiple attributes/dimensions – many of which might have similarity with each other.

2. In general, any ML algo. performs better as the no. of related attributes/features reduced.

3. i.e. a key to the success of ML lies in the fact that features are less in no. as well as in similarity b/w each other is very less.

4. This is the main guiding philosophy of PCA technique of feature extraction.

5. In PCA, a new set of features are extracted from the original features which are quite dissimilar in nature.

6. So, an n-dimensional feature space gets transformed to an m-dimensional feature space, where the dimensions are orthogonal to each other i.e. completely independent of each other.

7. The feature vector can be transformed to a vector space of the basis vectors which are termed as Principal Components.

8. These principal comp. just like basis vectors are orthogonal to each other.

9. So a feature ~~of~~ vector set (which may have similarity with each other) is transformed to a set of principal components (which are completely unrelated.)

10. However, the principal comp. capture the variability of the original feature space.

The objective of PCA is to make the transformation in such a way that

① The new features are distinct i.e. the covariance b/w the new features i.e. principal components is 0.

② The principal comp. are generated in order of their variability in data that it captures. Hence, the first principal component should capture the max. variability, the second Princ. comp. should Capture the next highest variability etc.

③ The sum of variance of the new features or the princ. comp. = sum of variance of original features.

PCA works based on a process c/d eigenvalue decomp. of a covariance matrix of a data set. Below are steps to be followed.

① First, calculate the covariance matrix of data set.

② then, calculate the eigen values of the cov. matrix.

③ The eigenvector having highest eigenvalue represents the direction in which there is the highest variance. ∴ So, this will help in identifying Ist Principal Component.

④ The eigenvector having next highest eigenvalue represents the direction in which data has the highest remaining variance & also orthogonal to the first direction. So, this helps in identifying 2nd Principal Comp.

⑤ Like this, identify the top 'K' eigenvectors having top 'K' eigenvalues so as to get the 'K' principal comp.

PCA – Algorithm Steps:-

Step 1: Read/Scan Dataset Features' vectors as

Features	Eg.1	Eg.2	----	Eg.N
X_1	X_{11}	X_{12}	--	X_{1N}
X_2	X_{12}	X_{22}	---	X_{2N}
⋮	⋮	⋮		⋮
X_n	X_{n1}	X_{n2}	---	X_{nN}

Step 2:- Compute the means of the variables

Mean of X_i

$$\overline{X_i} = \frac{1}{N}(X_{i1} + X_{i2} - - - X_{iN})$$

Step 3:- Calculate the covariance matrix

→ Covariance of all the ordered pairs (X_i, X_j)

→ $Co(X_i, X_j) = \frac{1}{N-1} \sum_{k=1}^{N} (X_{ik} - \overline{X_i})(X_{jk} - \overline{X_j})$

→ Construct $n \times n$ matrix S called co-variance matrix

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

Step 4: Calculate the eigenvalues and normalised eigen vectors of the covariance matrix

→ To find eigen values, solve the equations

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

→ We get n roots $\lambda_1, \lambda_2 --- \lambda_n$ (eigen values)

→ Now arrange :- $\lambda_1 > \lambda_2 > ---- \lambda_n$

→ For each eigen value, the corresponding eigen vector is a vector

$$U = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} \quad (n \times 1)\text{ matrix}$$

such that $(S - \lambda' I)U = 0$ [$\lambda' \to$ eigen value]

→ Normalise the eigen vector

→ Divide the vector, U by its length

i.e. Normalised eigen vector will be

$$e_i = \frac{U_i}{\|U_i\|}$$

where $\|U\| = \sqrt{u_1^2 + u_2^2 + -- u_n^2}$

→* The unit eigen vector corresponding to the largest eigen value is the first principal component.

Step 5 :- Derive new dataset

→ New dataset with reduced dimension is

Features	Example-1	Example-2	----	Example-N
PC_1	P_{11}	P_{12}	----	P_{1N}
PC_2	P_{21}	P_{22}	----	P_{2N}
$\vdots$	$\vdots$	$\vdots$		$\vdots$
PC_n	P_{n1}	P_{2n}	----	P_{nN}

such that

$$P_{ij} = e_j^T \begin{bmatrix} x_{1j} - \overline{x}_1 \\ x_{2j} - \overline{x}_2 \\ \vdots \\ x_{nj} - \overline{x}_n \end{bmatrix}$$

Problem :1 Given the following data, use PCA to reduce the dimensions from 2 to 1.

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

Sol.^{n} :- Step1:- Read/Scan Dataset :-

No. of features, $n=2$

No. of samples, $N=4$

Step2: Computation of mean of variables

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

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

Step3:- Computation of co-variance matrix

Ordered pairs are as (x, y)

(x,x), (x,y), (y,x), (y,y)

i) $Cov(x,x) = \frac{1}{N-1} \sum_{K=1}^{N} (x_{iK} - \overline{x}_i)(x_{jK} - \overline{x}_j)$

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

$$= 14$$

i.e. if 2 variables are same, then $Cov(x,x) = Variance(x)$

ii) $Cov(x,y) = \frac{1}{4-1}\left[(4-8)(11-8.5) + (8-8)(4-8.5) + (13-8)(5-8.5) + (7-8)(14-8.5)\right]$

$= -11$

iii) $Cov(y,x) = Cov(x,y) = -11$

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

$= 23$

④ <u>Co-variance Matrix</u> :- $n \times n$

$$S = \begin{bmatrix} Cov(x,x) & Cov(x,y) \\ Cov(y,x) & Cov(y,y) \end{bmatrix} = \begin{bmatrix} 14 & -11 \\ -11 & 23 \end{bmatrix}$$

<u>Step 4</u> :- Compute Eigen value, vector & Normalise eigen vector

i) Eigen value :-

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

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

$$\det. \begin{bmatrix} 14-\lambda & -11 \\ -11 & 23-\lambda \end{bmatrix} = 0$$

$$\{14-\lambda\}\{23-\lambda\} - \{-11\}\{-11\} = 0$$

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

Solving it, we get $\lambda = 30.3849,\ 6.6151$

∴ Arrange eigen values $\lambda_1 > \lambda_2$

$\lambda_1 = 30.3849 \longrightarrow$ First Principal Component

$\lambda_2 = 6.6151$

ii) Eigen vector of λ_1

$$(S - \lambda I) U_1 = 0$$

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

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

$$\rightarrow (14-\lambda_1)u_1 - 11u_2 = 0 \quad -① $$

$$\& \quad -11u_1 + (23-\lambda_1)u_2 = 0 \quad -②$$

Using Cramer's Rule

$$\frac{u_1}{11} = \frac{u_2}{14-\lambda_1} = t$$

when, $t=1$ $\quad u_1 = 11 \quad \& \; u_2 = 14-\lambda_1$

$\rightarrow$ Eigen vector U_1 of $\lambda_1 = \begin{bmatrix} 11 \\ 14-\lambda_1 \end{bmatrix} = \begin{bmatrix} 11 \\ 14-30.3841 \end{bmatrix} = \begin{bmatrix} 11 \\ -16.3849 \end{bmatrix}$

iii) Normalise Eigen Vector U_1

$$e_1 = \begin{bmatrix} 11/\sqrt{11^2+(-16.38)^2} \\ -16.38/\sqrt{11^2+(-16.38)^2} \end{bmatrix} = \begin{bmatrix} 0.5574 \\ -0.8303 \end{bmatrix}$$

Similarly for λ_2 $\quad e_2 = \begin{bmatrix} 0.8303 \\ 0.5574 \end{bmatrix}$

Step 5 :- Derive New Dataset

	from eg1	eg2	eg3	eg4
First Principal Component PC_1	P_{11}	P_{12}	P_{13}	P_{14}

$$P_{11} = e_1^T \begin{bmatrix} 4-8 \\ 11-8.5 \end{bmatrix} = \begin{bmatrix} 0.5574 & -.8303 \end{bmatrix} \begin{bmatrix} -4 \\ -2.5 \end{bmatrix}$$

$$P_{11} = -4.3052$$

Similarly $P_{12} = \begin{bmatrix} 0.5574 & -0.8303 \end{bmatrix} \begin{bmatrix} 8-8 \\ 4-8.5 \end{bmatrix}$

$$P_{12} = 3.7361$$

→ $P_{13} = 5.6928$

→ $P_{14} = -5.1238$

	Eg1	Eg2	Eg3	Eg4
PC_1	-4.3052	3.7361	5.6928	-5.1238

* Coordinate System for Principal Components