Example modified and adapted from an example from geeks for geeks example.

Mathematical Approach to PCA

The main guiding principle for **Principal Component Analysis** is FEATURE EXTRACTION i.e. "Features of a data set should be less as well as the similarity between each other is very less." In PCA, a new set of features are extracted from the original features which are quite dissimilar in nature. Therefore, an **n-dimensional feature space** gets transformed into an **m-dimensional feature space**, where the dimensions (new features) are orthogonal (uncorrelated) to each other.

Concept of Orthogonality: This concept is based on linear algebra. **Vector Space** is a set of vectors. They can be represented as a linear combination of the smaller set of vectors called **BASIS VECTORS.** Therefore, any vector 'v' in a vector space can be represented as:

$$v = \sum_{i=1}^{n} a_i u_i$$

where ***a*** represent 'n' scalars and u represents the basis vectors. Basis vectors are orthogonal to each other. The orthogonality of the vectors can be thought of as an extension of the vectors being perpendicular in a 2-D vector space**.** Therefore, our feature vector (data-set) can be transformed into a set of principal components (just like the basis vectors).

Objectives of PCA:

1. The new features are distinct i.e. the covariance between the new features (in case of PCA, they are the principal components) is **0**.

2. The principal components are generated in order of the variability in the data that it captures. Hence, the first principal component should capture the maximum variability, the second one should capture the next highest variability etc.

3. The sum of the variance of the new features / the principal components should be equal to the sum of the variance of the original features.

Working of PCA:

PCA works on a process called **Eigenvalue Decomposition** of a covariance matrix of a data set. The steps are as follows:

- First, calculate the covariance matrix of a data set.

- Then, calculate the eigenvectors of the covariance matrix.

- The eigenvector having the highest eigenvalue represents the direction in which there is the highest variance. So, this will help in identifying the first principal component.
- The eigenvector having the next highest eigenvalue represents the direction in which data has the highest remaining variance, which is also orthogonal to the first direction. Therefore, this helps in identifying the second principal component.
- Like this, identify the top 'k' eigenvectors having top 'k' eigenvalues to get the 'k' principal components.

Numerical for PCA:

Consider the following dataset

x1	**2.5**	**0.5**	**2.2**	**1.9**	**3.1**	**2.3**	**2.0**	**1.0**	**1.5**	**1.1**
x2	**2.4**	**0.7**	**2.9**	**2.2**	**3.0**	**2.7**	**1.6**	**1.1**	**1.6**	**0.9**

Step 1: Standardize the Dataset

Mean for x_1 = 1.81 = x_{1mean}

Mean for x_2 = 1.91 = x_{2mean}

We will change the dataset using the standard scaler.

$x_{1new} = x_1 - x_{1mean}$	0.69	-1.31	0.39	0.09	1.29	0.49	0.19	-0.81	-0.31	-0.71
$x_{2new} = x_2 - x_{2mean}$	0.49	-1.21	0.99	0.29	1.09	0.79	-0.31	-0.81	-0.31	-1.01

Step 2: Find the Eigenvalues and Eigenvectors.

$$\text{Correlation Matrix c} = C = \left(\frac{X \cdot X^T}{N-1}\right)$$

where, X is the Dataset Matrix (a 10 by 2 matrix in this example)

X^T is the transpose of the X (a 2 by 10 matrix in this example) and N is the number of elements = 10

Therefore,

$$C = \left(\frac{X \cdot X^T}{10-1}\right) = \left(\frac{X \cdot X^T}{9}\right)$$

{Therefore, to calculate the Correlation Matrix, we have to do the multiplication of the Dataset Matrix with its transpose}

$$C = \begin{bmatrix} 0.616556 & 0.615444 \\ 0.615444 & 0.716556 \end{bmatrix}$$

Using the equation, **| C – λI | = 0– equation (i)** where { \lambda is the eigenvalue and I is the Identity Matrix }

Therefore, solving equation (i)

$$\begin{bmatrix} 0.616556 & 0.615444 \\ 0.615444 & 0.716556 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\begin{vmatrix} 0.616556 - \lambda & 0.615444 \\ 0.615444 & 0.716556 - \lambda \end{vmatrix} = 0$$

Taking the determinant of the left side, we get

$$0.44180 – 0.616556\lambda – 0.716556\lambda + \lambda^2 – 0.37877 = 0$$

$$\lambda^2 – 1.33311\lambda + 0.06303 = 0$$

Refresher:

The determinant of a 2×2 matrix is

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc.$$

Factoring that equation, we get two values for λ, that are **(λ_1) = 1.28403** and **(λ_2) = 0.0490834**. Now, we must find the eigenvectors for the eigenvalues λ_1 **and** λ_2.

To find the eigenvectors from the eigenvalues, we will use the following approach:

First, we will find the eigenvectors for the eigenvalue 1.28403 by using the equation C·X=λ·X

$$\begin{bmatrix} 0.616556 & 0.615444 \\ 0.615444 & 0.716556 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} = 1.28403 \cdot \begin{bmatrix} x \\ y \end{bmatrix}$$
$$\begin{bmatrix} 0.616556x + 0.615444y \\ 0.615444x + 0.716556y \end{bmatrix} = \begin{bmatrix} 1.28403x \\ 1.28403y \end{bmatrix}$$

Solving the matrices, we get

0.616556x + 0.615444y = 1.28403x

x = 0.922049 * y

NOTE: x and y belong to the matrix X, so if we put y = 1, x comes out to be 0.922049. Therefore, the updated X matrix is as follows:

$$X = \begin{bmatrix} 0.922049 \\ 1 \end{bmatrix}$$

Up until this point, we haven't reached the eigenvectors. To get there, however, we must do a bit of modifications to the X matrix. They are as follows:

A. Find the square root of the sum of the squares of the element in X matrix i.e.

$$\sqrt{0.922049^2 + 1^2} = \sqrt{0.850174 + 1} = \sqrt{1.850174} = 1.3602$$

B. Now divide the elements of the X matrix by the number 1.3602 (just found that value in step A)

$$\begin{bmatrix} \frac{0.922049}{1.3602} \\ \\ \frac{1}{1.3602} \end{bmatrix} = \begin{bmatrix} 0.67787 \\ 0.73518 \end{bmatrix}$$

At this point, we have the eigenvectors for the eigenvalue λ_1, they are 0.67787 and 0.73518

Secondly, we will find the eigenvectors for the eigenvalue 0.0490834 by using the equation {Same approach as of previous step)

$$\begin{bmatrix} 0.616556 & 0.615444 \\ 0.615444 & 0.716556 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} = 0.0490834 \cdot \begin{bmatrix} x \\ y \end{bmatrix}$$

$$\begin{bmatrix} 0.616556x + 0.615444y \\ 0.615444x + 0.716556y \end{bmatrix} = \begin{bmatrix} 0.0490834x \\ 0.0490834y \end{bmatrix}$$

Solving the matrices, we get

0.616556x + 0.615444y = 0.0490834x

y = -0.922053

NOTE: x and y belong to the matrix X, so if we put x = 1, y comes out to be -0.922053. Therefore, the updated X matrix is as follows:

$$X = \begin{bmatrix} 1 \\ -0.922053 \end{bmatrix}$$

Up until this point, we haven't reached the eigenvectors. To get there, however, we must do a bit of modifications to the X matrix. They are as follows:

A. Find the square root of the sum of the squares of the elements in X matrix i.e.

$$\sqrt{1^2 + (-0.922053)^2} = \sqrt{1 + 0.85018} = \sqrt{1.85018} = 1.3602$$

B. Now divide the elements of the X matrix by the number 1.3602 (just found that value in Step A)

$$\begin{bmatrix} \frac{1}{1.3602} \\ \\ \frac{-0.922053}{1.3602} \end{bmatrix} = \begin{bmatrix} 0.735179 \\ -0.677873 \end{bmatrix}$$

So now we found the eigenvectors for the eigenvector λ_2, they are 0.735176 and -0.677873

Sum of eigenvalues ($\lambda 1$) and ($\lambda 2$) = 1.28403 + 0.0490834 = 1.33 = Total Variance {Majority of variance comes from $\lambda 1$ }

Step 3: Arrange Eigenvalues

The eigenvector with the highest eigenvalue is the Principal Component of the dataset. Therefore, in this case, eigenvectors of λ_1 are the principal components.

{Basically in order to complete the numerical we have to only solve till this step, but if we have to prove why we have chosen that particular eigenvector we have to follow the steps from 4 to 6}

Step 4: Form Feature Vector

$$\begin{bmatrix} 0.677873 & 0.735179 \\ 0.735179 & -0.677879 \end{bmatrix}$$ This is the FEATURE VECTOR for Numerical

Where first column are the eigenvectors of λ1 & second column are the eigenvectors of λ2

Step 5: Transform Original Dataset

Use the equation Z = X V

$$\begin{bmatrix} 0.69 & 0.49 \\ -1.31 & -1.21 \\ 0.39 & 0.99 \\ 0.09 & 0.29 \\ 1.29 & 1.09 \\ 0.49 & 0.79 \\ 0.19 & -0.31 \\ -0.81 & -0.81 \\ -0.31 & -0.31 \\ -0.71 & -1.01 \end{bmatrix} \cdot \begin{bmatrix} 0.677873 & 0.735179 \\ 0.735179 & -0.677879 \end{bmatrix} = \begin{bmatrix} 0.8297008 & 0.17511574 \\ -1.77758022 & -0.14285816 \\ 0.99219768 & -0.38437446 \\ 0.27421048 & -0.13041706 \\ 1.67580128 & 0.20949934 \\ 0.91294918 & -0.17528196 \\ -1.14457212 & -0.04641786 \\ -0.43804612 & -0.01776486 \\ -1.22382.62 & 0.16267464 \end{bmatrix} = Z$$

Step 6: Reconstructing Data

Use the equation X = Z∗V^T where (V^T is Transpose of V), X = Row Zero Mean Data

$$\begin{bmatrix} 0.8297008 & 0.17511574 \\ -1.77758022 & -0.14285816 \\ 0.99219768 & -0.38437446 \\ 0.27421048 & -0.13041706 \\ 1.67580128 & 0.20949934 \\ 0.91294918 & -0.17528196 \\ -1.14457212 & -0.04641786 \\ -0.43804612 & -0.01776486 \\ -1.22382.62 & 0.16267464 \end{bmatrix} \cdot \begin{bmatrix} 0.677873 & 0.735179 \\ 0.735176 & -0.677879 \end{bmatrix} = \begin{bmatrix} 0.6899999766573 & 0.4899999834233 \\ -1.3099999556827 & -1.2099999590657 \\ 0.389999968063 & 0.9899999665083 \\ 0.0899999969553 & 0.2899999901893 \\ 0.61212695653593 & 0.35482096313253 \\ 0.4899999834233 & 0.7899999732743 \\ 0.189999935723 & -0.309999995127 \\ -0.8099999725977 & -0.8099999725977 \\ -0.3099999895127 & -0.3099999895127 \\ -0.7099999759807 & -1.0099999658317 \end{bmatrix}$$

So in order to reconstruct the original data, we follow:

Row Original DataSet = Row Zero Mean Data + Original Mean

$$\begin{bmatrix} 0.6899999766573 & 0.4899999834233 \\ -1.3099999556827 & -1.2099999590657 \\ 0.389999968063 & 0.9899999665083 \\ 0.0899999969553 & 0.2899999901893 \\ 0.61212695653593 & 0.35482096313253 \\ 0.4899999834233 & 0.7899999732743 \\ 0.189999935723 & -0.309999995127 \\ -0.8099999725977 & -0.8099999725977 \\ -0.3099999895127 & -0.3099999895127 \\ -0.7099999759807 & -1.0099999658317 \end{bmatrix} + \begin{bmatrix} 1.81 & 1.91 \end{bmatrix} = \begin{bmatrix} 2.49 & 2.39 \\ 0.5 & 0.7 \\ 2.19 & 2.89 \\ 1.89 & 2.19 \\ 3.08 & 2.99 \\ 2.30 & 2.7 \\ 2.01 & 1.59 \\ 1.01 & 1.11 \\ 1.5 & 1.6 \\ 1.1 & 0.9 \end{bmatrix}$$

So for the eigenvectors of first eigenvalue, data can be reconstructed similar to the original dataset. Thus we can say that the Principal Component of the dataset is $\lambda 1$ is 1.28403 followed by $\lambda 2$ **that is 0.0490834**