How to Compute Principal Components

Q: Given the following data, use PCA to reduce the dimension from 2 to 1.

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

Solution: Dimension of this dataset is 2. we need to reduce it from 2 to 1.

Step 1: Dataset

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

No of features, $n = 2$

No of Samples, $N = 4$.

Step 2: Find the mean of variables

$$\bar{x} = \frac{\sum x_i}{N} = \frac{4+8+13+7}{4} = 8$$

$$\bar{y} = \frac{\sum y_i}{N} = \frac{11+4+5+14}{4} = 8.5$$

Step 3: Computation of Covariance Matrix.

Write the ordered Pairs.

Ordered Pairs are:

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

x, y

(i) Covariance of all ordered pairs

$$Cov(x, x) = \frac{1}{N-1} \sum_{K=1}^{N} (x_{ik} - \bar{x}_i)(x_{jk} - \bar{x}_j)$$

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

$$= 14$$

If two variable are same then,

$$Cov(x, x) = \frac{1}{(N-1)} \sum_{K=1}^{N} (x_1 - \bar{x})^2$$

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

$$= -11$$

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

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

$$= 23$$

Using all these covariance values construct the covariance matrix.

(ii) Covariance Matrix

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

Step 4: Construct Eigen value, Eigen vector & Normalized Eigen vector

(i) Eigen value

Solve the equation

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

Identity matrix.

Covariance matrix

Eigen value

Here we know the S is 2x2 covariance matrix, so we have to consider 2x2 identity matrix.

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

$$\therefore \lambda I = \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}.$$

$$\therefore \det(S - \lambda I) = \det\left(\begin{bmatrix} 14 & -11 \\ -11 & 23 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}\right)$$

$$= \det\left(\begin{bmatrix} 14-\lambda & -11 \\ -11 & 23-\lambda \end{bmatrix}\right)$$

Now we have to find the determinant

$$\det(S - \lambda I) = (14-\lambda)(23-\lambda) - (-11)(-11) = 0$$

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

Roots of the equation

$$\frac{1}{2a}\sqrt{b^2 - 4ac}$$

$$\Rightarrow \lambda = 30.3849,\ 6.6151$$

Arrange the λ in decending order such that $\lambda_1 > \lambda_2$

$\therefore$ we get $\lambda_1 = 30.3849$ $\longrightarrow$ Largest, Gives the First Principal Component

$\lambda_2 = 6.6151$

(ii) Now we are going to find the Eigen vector of λ_1

Eigen vector of λ_1

we have equation

$$(\underbrace{S}_{\text{Cov matrix}} - \underbrace{\lambda_1}_{\text{largest Eigen value}} I) \underbrace{U_1}_{\text{Eigen vector of } \lambda_1} = 0$$

we need to find the U_1 from the above equation.

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

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

This gives

$$\Rightarrow \left. \begin{aligned} (14-\lambda_1)u_1 - 11u_2 &= 0 \\ -11u_1 + (23-\lambda_1)u_2 &= 0 \end{aligned} \right\}$$

Find value of u_1 & u_2 from these linear equation.

Consider the First linear equation. So we have,

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

when $t=1$, we get

$$u_1 = 11$$
$$u_2 = 14-\lambda_1$$

So our Eigenvector U_1 of λ_1 = $\begin{bmatrix} 11 \\ 14-\lambda_1 \end{bmatrix}$ $[\because \lambda_1 = 30.3849]$

$$= \begin{bmatrix} 11 \\ 14-30.3849 \end{bmatrix} = \begin{bmatrix} 11 \\ -16.3849 \end{bmatrix}$$

(iii) Now we have to normalize the eigen vector U_1

$$e_1 = \begin{bmatrix} \dfrac{11}{\sqrt{11^2+(-16.3849)^2}} \\ \dfrac{-16.3849}{\sqrt{11^2+(-16.3849)^2}} \end{bmatrix} = \begin{bmatrix} 0.5574 \\ -0.8303 \end{bmatrix}$$

Similarly unit eigen vector U_2 for λ_2

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

Step 5: Derive the New Dataset

Consider now only one dimension, ie 1st principal component.

	Sample 1	Sample 2	Sample 3	Sample 4
First Principal Component (PC_1)	P_{11}	P_{12}	P_{13}	P_{14}

$$P_{11} = e_1^T \begin{bmatrix} x_1 - \bar{x} \\ y_1 - \bar{y} \end{bmatrix} = \begin{bmatrix} 0.5574 & -0.8303 \end{bmatrix} \begin{bmatrix} 4-8 \\ 11-8.5 \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} 0.5574 & -0.8303 \end{bmatrix} \begin{bmatrix} -4 \\ 2.5 \end{bmatrix}$$

$$\boxed{P_{11} = -4.3052}$$

In the same way Find P_{12}

$$P_{12} = \begin{bmatrix} 0.5574 & -0.8303 \end{bmatrix} \begin{bmatrix} x_2 - \bar{x} \\ y_2 - \bar{y} \end{bmatrix} = \begin{bmatrix} 0.5574 & -0.8303 \end{bmatrix} \begin{bmatrix} 8-8 \\ 4-8.5 \end{bmatrix}$$

$$\boxed{P_{12} = 3.7361}$$

Similar find P_{13} & P_{14}

$$\therefore \boxed{P_{13} = 5.6928}$$
$$\boxed{P_{14} = -5.1238}$$

So Finally our dataset will be like this

	Sample 1	Sample 2	Sample 3	Sample 4
PC1	-4.3052	3.7361	5.6928	-5.1238

Now we are going to draw the coordinate system for Principal Components.