Computation of PCA →

① Standardization of data :

x_1	x_2	$(x_1 - \bar{x}_1)$	$(x_2 - \bar{x}_2)$
2.5	2.4		
0.5	0.7		
2.2	2.9		
1.9	2.2		
3.1	3.0		
2.3	2.7		
2.0	1.6		
1.0	1.1		
1.5	1.6		
1.1	0.9		

$\bar{x}_1 = 1.81$

$\bar{x}_2 = -1.91$

→ Subtract the ~~data~~ mean from the corresponding data component to recenter the dataset.

The dataset after recentering : (zero Mean Data)

x_1	x_2
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

← Adjusted dataset will have mean zero.

(2) compute the covariance Matrix :

$$C = \begin{bmatrix} cov(x_1,x_1) & cov(x_1,x_2) \\ cov(x_2,x_1) & cov(x_2,x_2) \end{bmatrix}$$

$$cov(x,y) = \frac{\sum(x-\bar{x})(y-\bar{y})}{n-1}$$

x	y	A $(x-\bar{x})$	B $(y-\bar{y})$	A·B	A^2	B^2
2.5	2.4	0.69	0.49	.3381	.4761	.2401
0.5	0.7	-1.31	-1.21	1.5851	1.7161	1.4641
2.2	2.9	0.39	0.99	.3861	.1521	.9801
1.9	2.2	0.09	0.29	0.0261	.0081	.0841
3.1	3.0	1.29	1.09	1.4061	1.6641	1.1881
2.3	2.7	0.49	0.79	.3871	.2401	0.6241
2.0	1.6	0.19	-0.31	-.0589	0.0361	0.0961
1.0	1.1	-0.81	-0.81	.6561	0.6561	.6561
1.5	1.6	-0.31	-0.31	.0961	0.0961	.0961
1.1	0.9	-0.71	-1.01	.7171	0.5041	1.0201
$\bar{x}=1.81$	$\bar{y}=1.91$			5.539	5.549	6.449

$cov(x,y) = \frac{5.539}{9} = .615444$

$cov(x,x) = \frac{5.549}{9} = 0.616555$

$cov(y,y) = \frac{6.449}{9} = 0.716555$

$$C = \begin{bmatrix} 0.616556 & .615444 \\ 0.615444 & .716556 \end{bmatrix}$$

(3) compute Eigen values λ and eigenvectors x of covariance matrix.

$|A - \lambda I| = 0$

$= |C - \lambda I| = 0$

Eigen values :- $\lambda_1 = 1.28403$, $\lambda_2 = 0.0490834$

Eigen vectors →

for $\lambda_1 = 1.28403$ $\begin{bmatrix} .677873 \\ 0.735179 \end{bmatrix}$

for $\lambda_2 = 0.0490834$ $\begin{bmatrix} 0.735179 \\ -0.677873 \end{bmatrix}$

Total sample variance = (sum of eigen values)

$= 1.2840 + 0.0490 = 1.333$

Var.	Eigenvector 1	Eigenvector 2
x_1	.678	.735
x_2	.735	-.678
Eigenvalue	1.2840	0.0490
% of total Variance	1.2840/1.333 = 96.3%	0.490/1.333 = 3.7%

It can be seen that approximately 96% of the total variance is concentrated in eigenvector 1 and 4% in eigenvector 2.

4) Sort Eigenvalues in descending order :-

Arrange Eigenvectors by Eigenvalues - From highest to lowest. this gives the component in order of significance.

The eigenvector with the highest eigenvalue is the principal component of the dataset.
components of lesser significance can be ignored

- Some information is lost, but if eigenvalues are small, then not much is lost. If some components are left out, then the final data-set will have less dimensions than the original.

(5) Form a feature vector and construct the projection matrix W. →

Projection matrix W is calculated by taking the eigenvectors that are choosen and forming a matrix with them in columns.

Here, there are two eigenvectors, so there are 2 choices of matrices.

→ A feature vector can be formed with both the ~~eigenvalues~~ eigenvectors :

$$W_1 = \begin{bmatrix} .677873 & .735179 \\ .735179 & -.677873 \end{bmatrix}$$

→ If the smaller, less significant component is left out, then there will be a single column :

$$W_2 = \begin{bmatrix} .677873 \\ .735179 \end{bmatrix}$$

Imp.
**

(6) Transform the original dataset :

Dataset after subtracting the mean

The new dataset is derived by taking $Z = XW$.

Case 1 : $W_1 = \begin{bmatrix} .677873 & .735179 \\ .735179 & -.677873 \end{bmatrix}$

Final Data (Z_1) = XW_1

dataset after subtracting the mean.