- PCA is a dimensionality reduction technique
- It finds new dimensions, called Principal Component, from the given feature set.

Algorithm:

1. Calculate & Remove mean
2. Calculate covariance matrix
3. Calculate Eigen values & vectors of the covariance matrix
4. Arrange Eigen vector in descending order of Eigen values.
5. Eigen vectors are the principal components.
6. Remove Eigen vectors with small Eigen values to reduce dimensions.
7. Represent data on principal component axis.

A	B	C	D	E	F	G
x_1	x_2	$x_1-\mu_{x_1}$	$x_2-\mu_{x_2}$	$(x_1-\mu_{x_1})^2$	$(x_2-\mu_{x_2})^2$	$(x_1-\mu_{x_1})\cdot(x_2-\mu_{x_2})$
1	1	-2.5	-2.375	6.25	5.64	5.937
1	2	-2.5	-1.375	6.25	1.89	3.437
3	2	-0.5	-1.375	0.25	1.89	0.687
3	4	-0.5	0.625	0.25	0.39	-0.3125
4	3	0.5	-0.375	0.25	0.14	-0.1875
5	4	1.5	0.625	2.25	0.39	0.9375
5	6	1.5	2.625	2.25	6.89	3.9375
6	5	2.5	1.625	2.25	2.64	4.0625
				3.428	2.839	2.642

① Calculate & Remove mean

$$\mu_{x_1} = \frac{\sum x_{1i}}{n} = 3.5 \quad \Big| \quad \mu_{x_2} = 3.375$$

② Calculate covariance matrix

$$\begin{array}{c|cc} & x_1 & x_2 \\ \hline x_1 & \nearrow & \triangleleft \\ x_2 & \triangleleft & \nearrow \end{array}$$

Covariance of x_1, x_2

Variance of x_1

Variance of x_2

$$Var(x_i) = \frac{\sum (x_i - \mu_{x_i})^2}{(n-1)}$$

$$\text{Covariance}(X_i, X_j) = \frac{\sum (x_i - \mu_{x_i})(x_j - \mu_{x_j})}{(n-1)}$$

Covariance Matrix

$$A = \begin{bmatrix} 3.43 & 2.64 \\ 2.64 & 2.84 \end{bmatrix}$$

③ Calculate Eigen values & vector of the covariance matrix

$$A - \lambda I = 0$$

$$\begin{bmatrix} 3.43 & 2.64 \\ 2.64 & 2.84 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} = 0$$

$$\therefore \begin{bmatrix} 3.43 - \lambda & 2.64 \\ 2.64 & 2.84 - \lambda \end{bmatrix} = 0$$

$$\therefore (3.43 - \lambda)(2.84 - \lambda) - (2.64)^2 = 0$$

$$\therefore \lambda^2 - 6.27\lambda + 2.78 = 0$$

$$\therefore \boxed{\lambda_1 = 5.78 \qquad \lambda_2 = 0.48}$$

Eigen vector V_1 for λ_1

$$A V = \lambda V$$

Covariance matrix

Eigen Value

Eigen Vector

$$\begin{bmatrix} 3.43 & 2.64 \\ 2.64 & 2.84 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = 5.78 \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$$

$$3.43 v_1 + 2.64 v_2 = 5.78 v_1$$

$$v_2 = 0.89 v_1$$

$$V_1 = \begin{bmatrix} 1 \\ 0.89 \end{bmatrix}$$

Principal Component I

Eigen vector V_2 for λ_2

$$\begin{bmatrix} 3.43 & 2.64 \\ 2.64 & 2.84 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = 0.48 \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$$

$$3.43 v_1 + 2.64 v_2 = 0.48 v_1$$

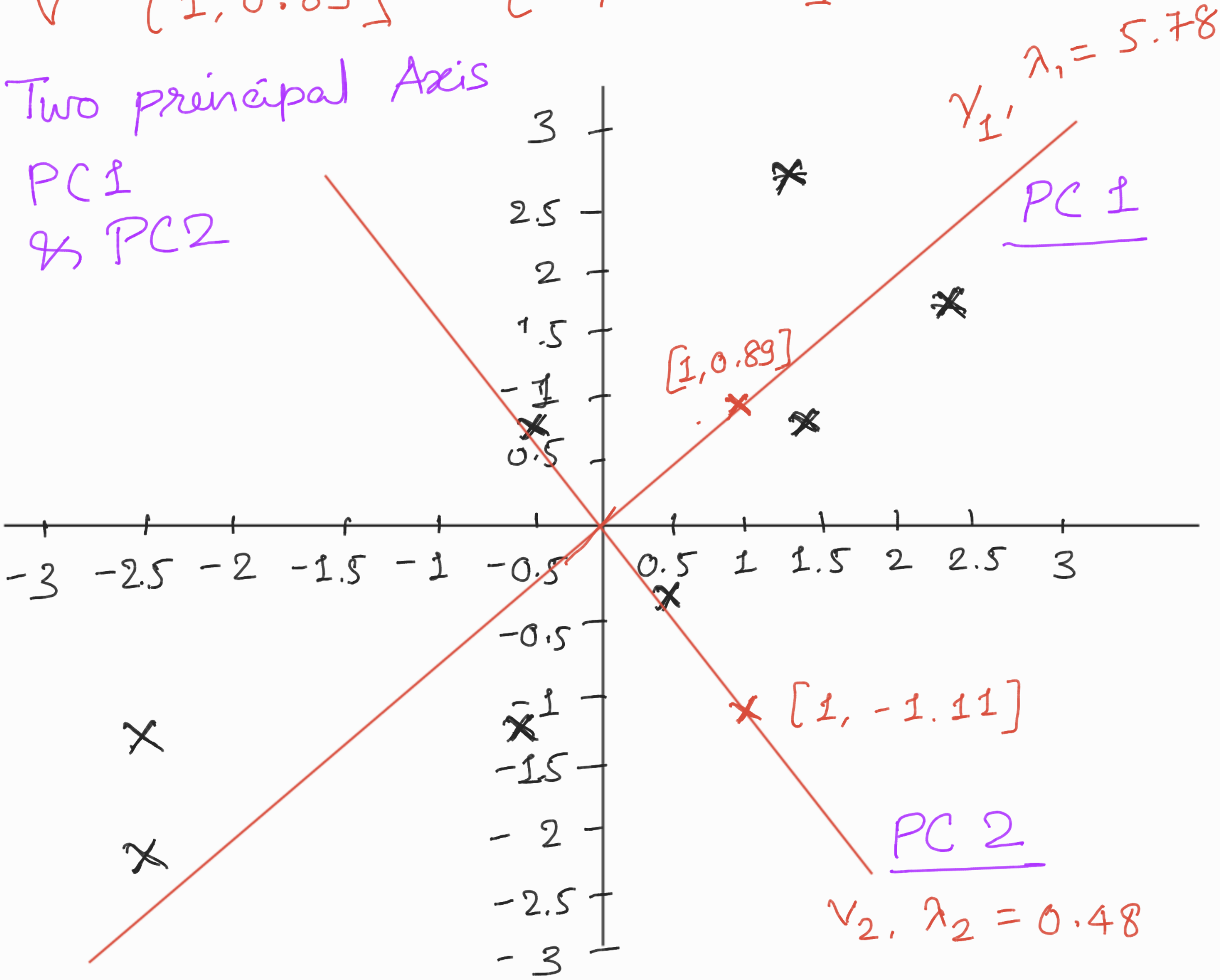
$$v_2 = -1.11 v_1$$

$$V_2 = \begin{bmatrix} 1 \\ -1.11 \end{bmatrix} \longleftarrow \text{Principal Component 2}$$

4. Arrange in descending order of Eigen values.

	PC1	PC2
λ	5.78	0.48
V	[1, 0.89]	[1, -1.11]

Two principal Axis PC1 & PC2



④ Remove Eigen vector with low Eigen value,

- Remove PC_2 because Eigen value is low.

- PC 1 [1, 0.89]

Original Features X_1, X_2

New Features, PC_1, $\cancel{PC_2}$

New Data points = New Feature Vector × Old data points

Principal component

old data

$$\underline{[1 \times 2]} \times \underline{[2 \times n]}$$

New data points

$$\overset{1 \times n}{\begin{bmatrix} 1.89 \\ 2.8 \\ \vdots \end{bmatrix}} = \overset{1 \times 2}{\begin{bmatrix} 1 & 0.89 \end{bmatrix}} \overset{2 \times n}{\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 3 & 2 \\ \vdots & \vdots \end{bmatrix}}$$