Dimensionality Reduction

- Variance & Covariance
- Eigen Values & Vectors

Principal Component Analysis (PCA)

- It uses covariance to compare features
- It uses Eigen vectors to find dimension of high co-variance

x_1	x_2	$(x_1-\mu_{x_1})$	$(x_2-\mu_{x_2})$	$(x_1-\mu_{x_1})^2$	$(x_2-\mu_{x_2})^2$	$(x_1-\mu_{x_1})(x_2-\mu_{x_2})$
1	2	-1	-1	1	1	1
2	3	0	0	0	0	0
3	4	1	1	1	1	1
				2/2	2/2	2/2
$\mu_{x_1}=2$	$\mu_{x_2}=3$			$=1$	$=1$	$=1$

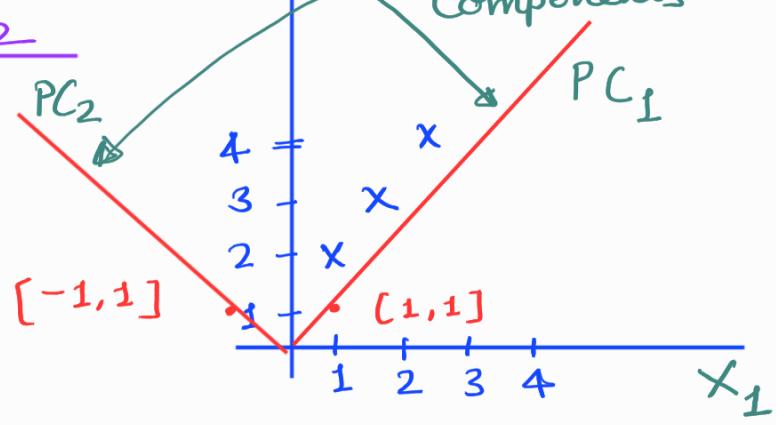
variance → $=1$; variance → $=1$; covariance → $=1$

Covariance

- $+ \Rightarrow$ Increasing/decreasing together
- $0 \Rightarrow$ Not related
- $- \Rightarrow$ One is increasing & another is decreasing

Covariance Matrix

	X_1	X_2
X_1	1	1
X_2	1	1



* Find Eigen Values & vectors of the co-variance matrix

λ	v
0	[-1, 1]
2	[1, 1]

plot here

Principal Components

- → Eigen Vectors of the covariance matrix
- → Principal components are orthogonal to each other
- → For an $N \times N$ matrix, there are N principal components.
- — Principal components points to the direction of high (co) variance
- → The eigen values decides the importance of a principal component.
- → Higher eigen values indicates better spread of data along that principal axis.
- → The principal components with low Eigen values, can be removed to reduce the number of dimensions.

Principal Component Analysis (PCA) steps

1. Remove mean from the data
2. Calculate co-variance matrix
3. Find Eigen values & Eigen vectors of the covariance matrix
4. Eigen vectors are the new principal components
5. Arrange Eigen vectors in descending order of Eigen values
6. Remove Eigen vector with low Eigen values to reduce dimensions.