Dimensionality Reduction:

Dataset →

No. of Rooms	Area	Locality	No. of doors	Price
			not so imp.	

1000

5×1000

→ 4×1000

$\times$ α ✓

PCA (principal component Analysis)

→ unsupervised technique.

→ linear transform

→ pattern recognition

→ It aims to find the direction of maximum variance in high dimensional data and projects it onto a new subspace with equal or fewer dimension than the original one.

Derivation :

$$\begin{bmatrix} 2 & 3 \\ 2 & 1 \end{bmatrix}_{2\times 2} \times \begin{bmatrix} 3 \\ 2 \end{bmatrix}_{2\times 1} = \begin{bmatrix} 12 \\ 8 \end{bmatrix}_{2\times 1} = 4 \times \begin{bmatrix} 3 \\ 2 \end{bmatrix}$$

$\begin{bmatrix} 2 & 3 \\ 2 & 1 \end{bmatrix}$ → A = input matrix

$\begin{bmatrix} 3 \\ 2 \end{bmatrix}$ → eigen vector (V)

4 → eigen value (λ)

$\begin{bmatrix} 3 \\ 2 \end{bmatrix}$ → eigenvector (v)

$$\boxed{A \cdot v = \lambda \cdot v}$$

Calculating eigen value :

$$A = \begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix}$$

$Av = v\lambda$

$Av - v\lambda = 0$

$\Rightarrow (A - \lambda) v = 0$

$\Rightarrow (A - \lambda) = 0$

$\Rightarrow (A - \lambda I) = 0$

Then, $|A - \lambda I|$

$$= \left| \begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right|$$

$$= \left| \begin{bmatrix} -\lambda & 1 \\ -2 & -3-\lambda \end{bmatrix} \right|$$

$$= (-\lambda \times (-3 - \lambda) \quad)$$

$$= \underline{\lambda^2 + 3\lambda + 2}$$

$$\lambda^2 + 3\lambda + 2 = 0.$$

$$\hookrightarrow \lambda_1 = -1 \quad , \quad \lambda_2 = -2$$

$$\boxed{\begin{aligned} \lambda_1 &= -1 \\ \lambda_2 &= -2 \end{aligned}} \longrightarrow \text{eigen values.}$$

calulating eigen vectors:

For λ_1, eigen vector is,

$$(A - \lambda_1 I) \cdot V_1 = 0.$$

$$\Rightarrow \begin{bmatrix} 1 & 1 \\ -2 & -2 \end{bmatrix} \begin{bmatrix} V_{1:1} \\ V_{1:2} \end{bmatrix} = 0.$$

$$\Rightarrow \quad V_{1:1} + V_{1:2} = 0.$$

and $\quad -2V_{1:1} - 2V_{1:2} = 0$

$$\hookrightarrow V_{1:1} = -V_{1:2}$$

$$V_1 = K_1 \begin{bmatrix} +1 \\ -1 \end{bmatrix}$$

$$V_2 = K_2 \begin{bmatrix} +1 \\ -2 \end{bmatrix}$$

Step-1: ① find co-variance matrix

variance:

$$s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$$

$$\boxed{cov(x,y) = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{n-1}}$$

co-variance matrix of 3 rows and 3 columns:

$$cov \cdot matrix = \begin{pmatrix} \underline{cov(x,x)}, & cov(x,y) & cov(x,z) \\ cov(y,x) & \underline{cov(y,y)} & cov(y,z) \\ cov(z,x) & cov(z,y) & \underline{cov(z,z)} \end{pmatrix}$$

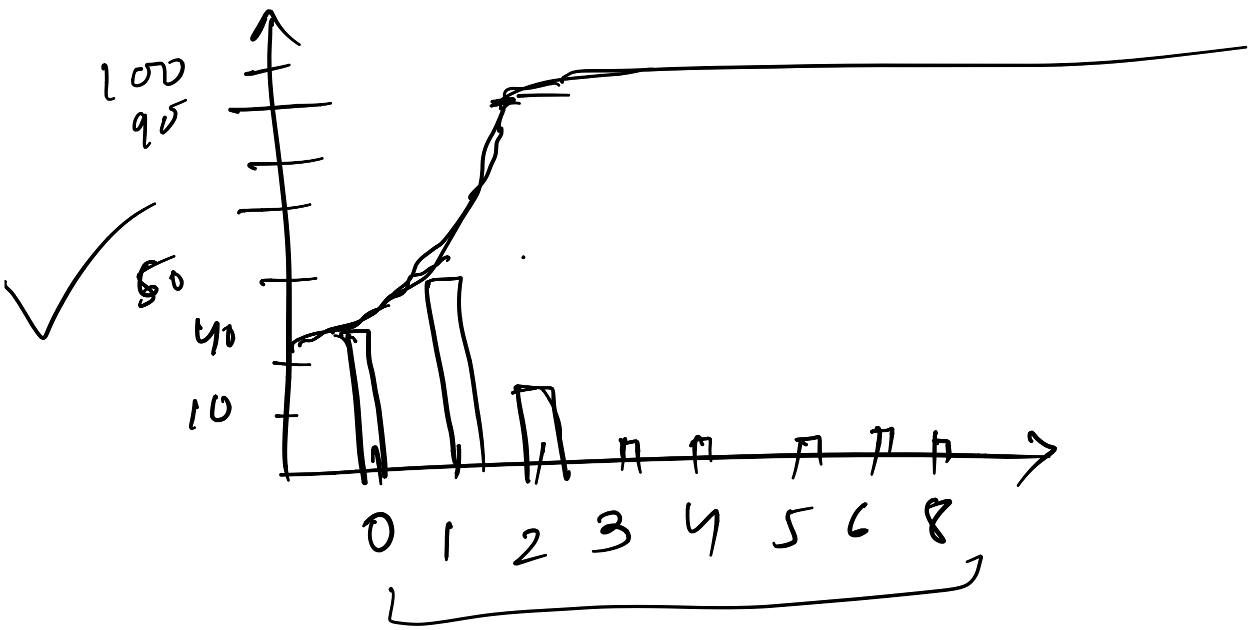
covariance of itself is the nothing but variance.

(2) calculate the eigenvectors of the co-variance matrix

(3.) select m eigenvectors that correspond to the largest m eigen values to be the new axis.

How to decide number of components?

Individual variance ratio.
Commulative variance ratio.



no. of pri. comp = 3