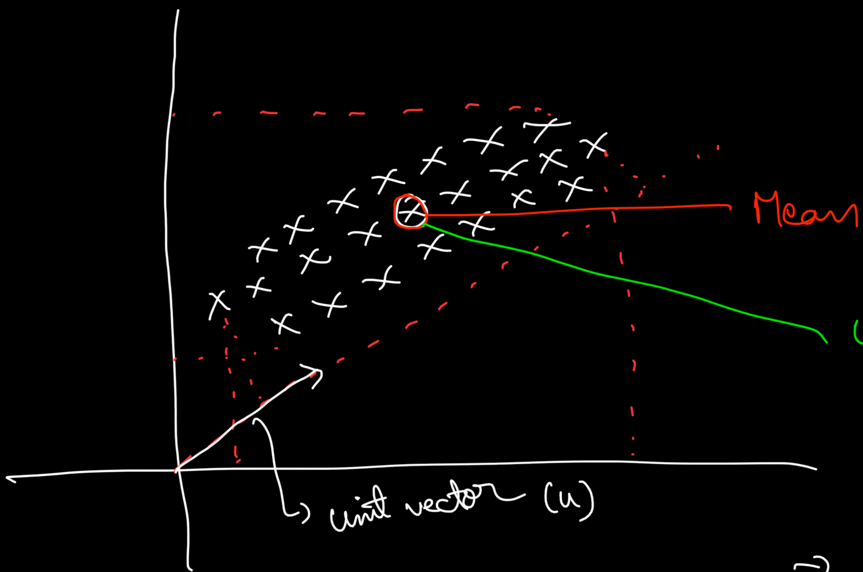
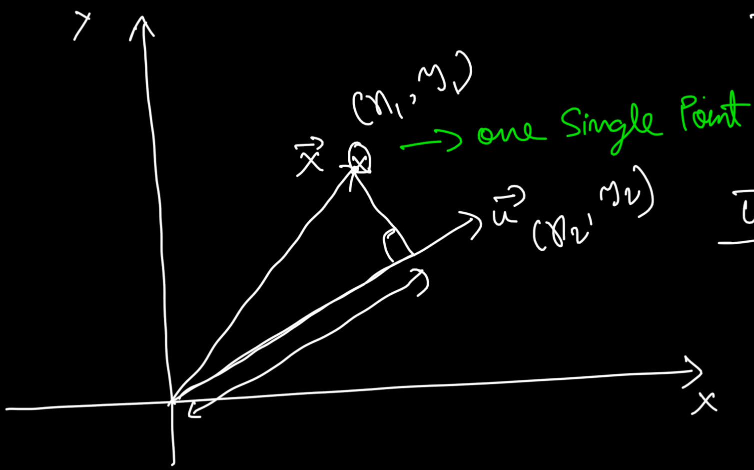
Principle Component Analysis (PCA) | Part 2 |

Problem Formulation and Step by Step Solution

Problem Formulation



$u^T \bar{x} \Rightarrow$ Mean of Projection



→ x is a vector and we have project x in another vector.

→ we have to find unit vector.

→ let say u is a unit vector

$$\frac{\vec{u}\cdot\vec{x}}{|u|} = \vec{u}\cdot\vec{x} = u^T x$$

$$\underset{x}{[x_1, y_1]} \quad \underset{u}{[x_2, y_2]}$$

$$[x_1, y_1]\begin{bmatrix} x_2 \\ y_2 \end{bmatrix}$$

$$= x_1 x_2 + y_1 y_2 = \text{Scalar}$$

(i) we are projecting every vector on a unit vector.

(ii) And for every individual vector, we get a Scalar quantity (length of projector of that vector on unit vector)

There can be $\boxed{\text{multiple unit vector}}$ but we have to $\boxed{\text{Select}}$ that $\boxed{\text{unit vector}}$ for which we get $\boxed{\text{maximum variance}}$.

$\Rightarrow$ $[u^T x_1]$ $[u^T x_2]$ . . . . $[u^T x_n]$

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

$$V = \left[\frac{\sum_{i=1}^{n} (u^T x_i - u^T \bar{x})^2}{n} \right] = \text{variance} \Rightarrow$$ This should be maximum

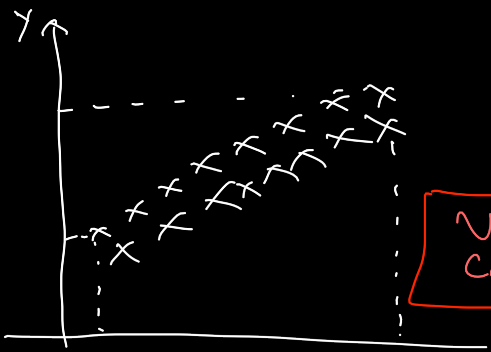
PCA has to find unit vector in such a way that V should be maximum.

Covariance and Covariance Matrix

Mean → central Tendancy (Mean can not help in finding Spread of data)

$\boxed{\text{that's why variance Come}}$

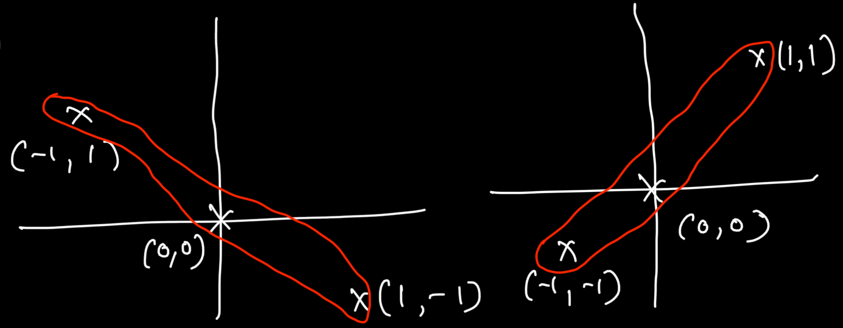
Problem



$\boxed{\text{Variance can be calculated in one axis.}}$

→ Variance one single metric

→ Variance new says anything about relationship between X and Y.



Corelatation
$\boxed{-1 \text{ to } 1}$

Acc to variance these two data are same but they are not.

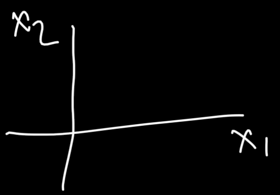
That's why

Covariance : Tells about relationship b/w x and y.

Covariance Matrix

⇓

x_1	x_2



Covariance Matrix

$$\begin{bmatrix} 2 \times 2 \end{bmatrix}$$

$$\begin{array}{c|c} & \begin{array}{cc} x_1 & x_2 \end{array} \\ \begin{array}{c} x_1 \\ x_2 \end{array} & \begin{bmatrix} Cov(x_1, x_1) & Cov(x_1, x_2) \\ Cov(x_2, x_1) & Cov(x_2, x_2) \end{bmatrix} \end{array}$$

$$\boxed{\begin{array}{c} Cov(x_1, x_1) \\ \Downarrow \\ Var(x_1) \end{array}}$$

$$\begin{bmatrix} Var(x_1) & Cov(x_1, x_2) \\ Cov(x_2, x_1) & Var(x_2) \end{bmatrix}$$

Square matrix
also Symmetric

In Covariance Matrix

- Diagonal Elements are variance of individual features
- Non-diagonal Elements are Covariance

Benefit of Covariance Matrix

- It tell about Spread for every axis (Variance)
- relationship between two pair of axis, if -ve then they are -ve correlated and vice versa.
- Covariance terms tells about orientation of data.

x	y	z

$$\begin{array}{c|c} & \begin{array}{ccc} x & y & z \end{array} \\ \begin{array}{c} x \\ y \\ z \end{array} & \begin{bmatrix} V_x & C_{x,y} & C_{x,z} \\ C_{y,x} & V_y & C_{y,z} \\ C_{z,x} & C_{z,y} & V_z \end{bmatrix} \end{array}$$

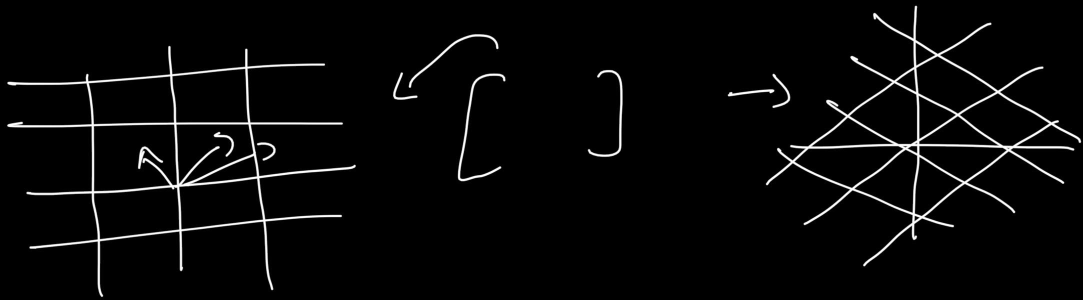
Square matrix
Symmetrical

Eigen decomposition of Covariance matrix

Linear Transformation, Eigen Vectors and Eigen values

→ **Matrices are linear transformation**

when we apply matrix transformation over any coordinates system, then it transforms coordinates systems. vectors direction and magnitude both changes. Basically vectors get knocked off from their span.



* with Identity matrix, No changes happen.
* Matrices are basically linear transformation that bring changes on coordinate system.

Eigen vectors : Special vector, only magnitude changes. Direction remain same, when linear transformation applied.

Eigen Values : How much eigen vector is stretching or srinking. factor by which eigen vector magnitude changes.

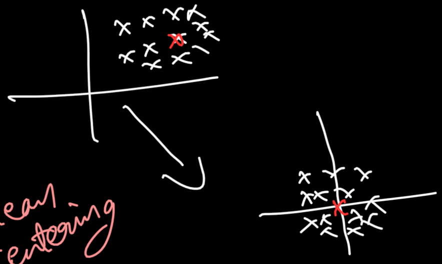
In 2D Space = 2 Eigen vector
In 3D Space = 3 Eigen vector
⋮

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

Matrix   Vector   Constant

⟹ If True then $\vec{v}$ is Eigen vector and λ is Eigen value

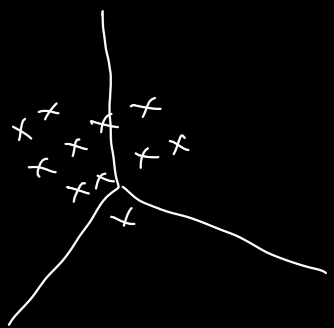
The largest eigenvector of the covariance matrix always points into the direction of the largest variance of the data, and the magnitude of this vector equals the corresponding eigenvalue.



Mean centering

Step by Step Solution

f_1	f_2	f_3	target



(i) Make data Mean centering (bring data to center), Not mandatory but if done then get good result.

(ii) find Covariance matrix

$$\begin{array}{c|ccc} & & & \\ f_1 & v(f_1) & c(f_1, f_2) & c(f_1, f_3) \\ f_2 & c(f_2, f_1) & v(f_2) & c(f_2, f_3) \\ f_3 & c(f_3, f_1) & c(f_3, f_2) & v(f_3) \\ & f_1 & f_2 & f_3 \end{array}$$

(3) find the Eigen value and Eigen vector for the Covariance matrix

3D $\rightarrow$ 3 vectors

λ_1 $\quad$ λ_2 $\quad$ λ_3

Greater eigen value $\leftarrow$ λ_1

PC 1 $\quad$ PC 2 $\quad$ PC3

PC 1 $\rightarrow$ 1D

PC 1, PC 2 $\rightarrow$ 2D

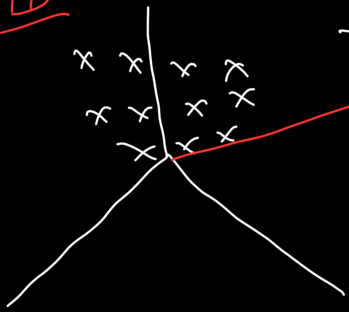
(4) After getting PCs, we have to bring actual data points from heigher dimension to lower dimension.

Data Points → 1000

How to transform Points?

3D to 1D

2D or 1D



→ PC1 (1, 3)

f_1	f_2	f_3	target

(1000, 4)

$U^T \cdot X = (1000, 3) \cdot (1, 3)^T$

$= \underbrace{(1000, 3) \cdot (3, 1)}_{(1000, 1)}$

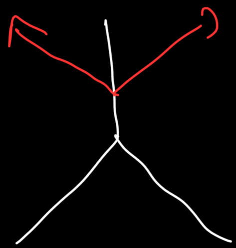
Add target →

PC1	tar

(1000, 2)

Shape ↗

3D to 2D



2 vector in 3D ←

$(1000, 3) \cdot (2, 3)^T$

$\underbrace{(1000, 3)(3, 2)}_{(1000, 2)}$

Add target →

PC1	PC2	tar

(1000, 3)

Shape ↗