Lesson B05:

Principal Component Analysis

PCA : identify the pattern in data and expressing the data in more compact way!

Step 1: Get some data

x	y
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	1.1
1.5	1.6
1	0.9

Mean value.

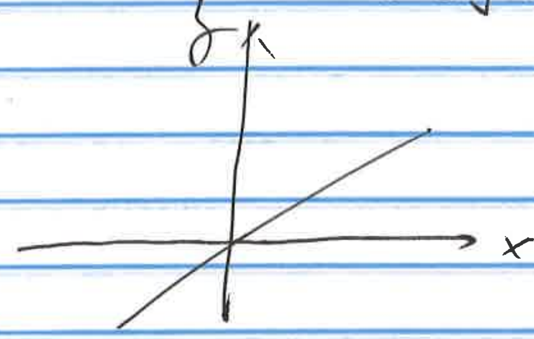
$\bar{x} = 1.81$

$\bar{y} = 1.91$

Step 2: substract the mean to make data pass through the origin (Data Adjust)

X	Y
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

the mean of $\begin{bmatrix} X \\ Y \end{bmatrix}$ will be $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$.



Step 3: Calculate the Covariance Matrix.

→ Covariance is a measure between 2 dimensions. It shows how Two variables vary together

$$\text{Cov}(x, y) = \sum_{i=1}^{n} \frac{(x-\bar{x})(y-\bar{y})}{n-1}$$

for example.

x	2.1	2.5	3.6	4.0
y	8	10	12	14

$$\left\{ \begin{array}{l} \bar{x} = 3.1 \\ \bar{y} = 11 \end{array} \right. \rightarrow \text{Cov}(x, y) = \frac{[(2.1-3.1)(8-11) + (2.5-3.1)(10-11)}{+ (3.6-3.1)(12-11) + (4-3.1)(14-11)]} \div 3$$

$$= 2.26$$

How about Covariance Matrix?

	x	y
x	Cov(x, x)	Cov(x, y)
y	Cov(y, x)	Cov(y, y)

3

Step 4: Check Covariance Matrix

The Covariance Matrix for the given data set is:

$$Cov = \begin{bmatrix} 0.616 & 0.6154 \\ 0.6154 & 0.1165 \end{bmatrix}.$$

Since the non-diagonal element in Covariance Matrix are positive $\Rightarrow$ x. y increase together.

Step I: Calculate the eigen value and eigen vector for the Covariance Matrix.

Eigen values: $\lambda_1 = 0.4908$.

$\lambda_2 = 1.25402$.

Eigen Vector: $\vec{v}_1 = \begin{bmatrix} -0.735 \\ 0.677 \end{bmatrix}$.

$\vec{v}_2 = \begin{bmatrix} -0.678 \\ -0.73 \end{bmatrix}$.

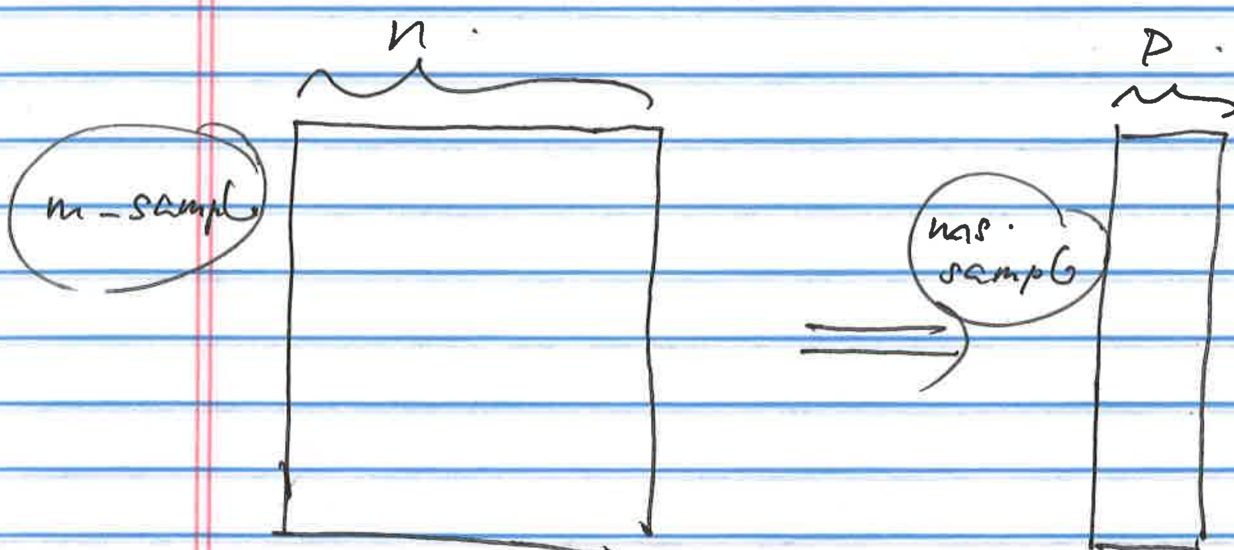
The most important (Principal) Eigen-vector would be the direction in which the variables strongly correlate.

Step 6: The Eigen Vectors with highest Eigen-Value will be chosen as PCA.

n _ dimension of variable (feature)

$\Downarrow$

Choose P eigenvectors. $(p < n)$ with largest eigenvalue.



☆ The original data had Two axes — x & y; the data are formulated in term of x & y.

☆ The newly formulated are in term of Primary Eigenvectors.