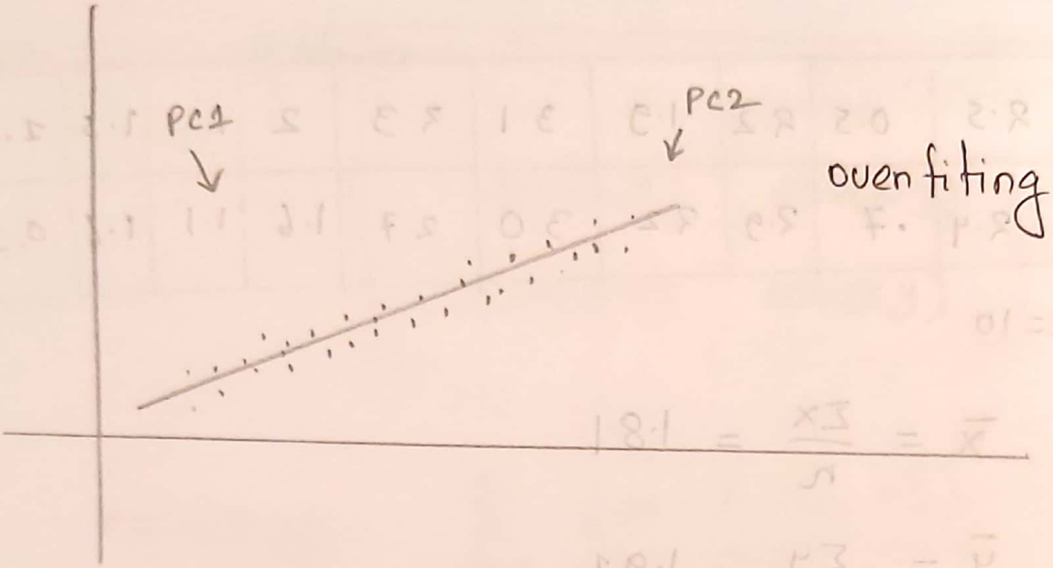
Principle Component Analysis

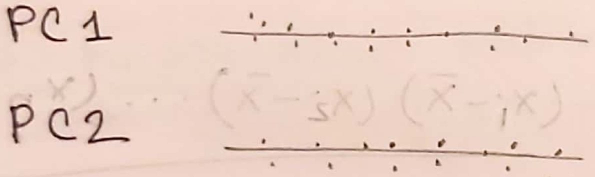
→ solves overfitting problem

→ It is a statistical technique used for simplifying and summarizing the complexity in high dimensional data whitle retaining its essential feature.

→ helps to indentify and emphasize most important pattenns and nelationship within a dataset reducing it to a smaller set of variable called principle component



Principle component



x	2.5	0.5	2.2	1.9	3.1	2.3	2	1	1.5	1.1
y	2.4	.7	2.9	2.2	3.0	2.7	1.6	1.1	1.6	0.9

$n = 10$

$$\bar{x} = \frac{\Sigma x}{n} = 1.81$$

$$\bar{y} = \frac{\Sigma y}{N} = 1.91$$

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

$i = 1 - 10$

x	$x - \bar{x}$	$(x - \bar{x})(x - \bar{x})$
2.5	$2.5 - 1.81 = .69$	$.69^2 = 0.476$
0.5	-1.31	1.7161
⋮	⋮	⋮
1.1	$-.71$	.5041

Sum = 5.5490

$n-1=9$

$\therefore n = 10$

$n=10$

$$\text{cov}(x, x) = \frac{5.5490}{9} = .6165$$

features x, y = 2 dimmesion (2×2)

if 3 feature = 3 " (3×3)

Covariance Matrix

$$= \begin{bmatrix} cov(x,x) & cov(x,y) \\ cov(y,x) & cov(y,y) \end{bmatrix}$$

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

Formula 1.2

step-1:- Find Covariance Matrix

$$cov(y, x) = cov(x, y)$$

i	x	y	$x-\bar{x}$	$y-\bar{y}$	$(x-\bar{x})(y-\bar{y})$
1	2.5	2.4	0.69	0.49	.338
2	0.5	0.7	-1.31	-1.21	1.5851
⋮	⋮				
10	1.1	0.9	...	...	...

sum = 5.5393

n - 1 = 9

$$\therefore cov(x, y) = \frac{5.5393}{9} = .6154$$

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

i	y	$y-\bar{y}$	$(y-\bar{y})(y-\bar{y})$
1	2.4	$2.4-1.9 = .49$	$.49^2 = .2401$
2	0.7	-1.21	1.4641
⋮	⋮	⋮	⋮
10	0.9		1.91881

sum = 6.449

$n-1 = 10-1 = 9$

$$\text{conv}(y,y) = \frac{6.449}{9}$$

$$= 0.7165$$

$$C = \begin{bmatrix} cov(x,x) & cov(x,y) \\ cov(y,x) & cov(y,y) \end{bmatrix}$$

$$= \begin{bmatrix} 0.6165 & 0.6154 \\ 0.6154 & 0.7165 \end{bmatrix}$$

Step 2:- Find eigen value

Now: $\boxed{C - \lambda I = 0}$ Formula - 2

λ = eigan vecton

$$\begin{bmatrix} .6165 & .6154 \\ .6154 & .7165 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = 0$$

$$\Rightarrow \lambda^2 - 1.333\lambda + 0.0630 = 0$$

$$\therefore \lambda_1 = 0.0490$$

$$\therefore \lambda_2 = 1.2840$$

$\boxed{\text{Step-3 :- Find eigen Vectors}}$

Formula 1.3 $\boxed{C \cdot V = \lambda \cdot V}$

λ = eigan Value
V = eigen Vector

$\lambda = \lambda_1$

$$\begin{bmatrix} .6165 & .6154 \\ .6154 & .7165 \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = 0.0490 \begin{bmatrix} x_1 \\ y_1 \end{bmatrix}$$

$\Rightarrow 0.5674\, x_1 = -0.6154\, y_1$

$\Rightarrow 0.6154\, x_1 = -0.6674\, y_1$

choose any

$x_1 = -1.0845\, y_1$

assume $y_1 = 1$

$$\begin{bmatrix} -1.0845 \\ 1 \end{bmatrix}$$

$x = \sqrt{-1.0845^2 + 1^2}$
$= 1.4751$

$\therefore$ eigan vector for $x = \lambda_1$ $= \begin{bmatrix} -\frac{1.0845}{1.4751} \\ +\frac{1}{1.4751} \end{bmatrix} = \begin{bmatrix} -0.7351 \\ 0.6778 \end{bmatrix}$

same Do fon

$\lambda = \lambda_2$

$x_2 = .92194\, y_2$

$\therefore$ eiganvecton for λ_2 $= \begin{bmatrix} 0.6778 \\ 0.7351 \end{bmatrix}$

— o —

PCA finds a new set of Dimmensions (on a set of views) such that all dimmensions are onthogonal (linearly Independent)

Goals :-

1. Find linearly independent dimmension

How does PCA works

① Calculate the covariance Matrix

$$\text{cov matrix} = \begin{bmatrix} c(x,x) & c(x,y) \\ c(xy,x) & c(y,y) \end{bmatrix}$$

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

② Calculate eigen vectors and corresponding eigen values

$$C - \lambda I = 0 \Rightarrow (i)$$
$$CV = \lambda V \Rightarrow (ii)$$

③ sort the eigen vectors according to eigen values in descending order

④ choose first k eigen vectors that will be new k dimmension

⑤ Transform the original n dimmension data points to k dimmension solving a problem using Sir's format

Q Reduce the dimmesion from 2 to 1 using PCA

Feature				
x_1	4	8	13	7
x_2	11	4	5	14

Step - 1

$$\bar{x}_1 = \frac{1}{4}(4+8+13+7) = 8$$

$$\bar{x}_2 = \frac{1}{4}(11+4+5+14) = 8.5$$

Step - 2

$$S = \begin{bmatrix} cov(x_1, x_1) & cov(x_1, x_2) \\ cov(x_2, x_1) & cov(x_2, x_2) \end{bmatrix}$$

$$cov(x_1, x_1) = \sum_{i=1}^{n} \frac{(x - x_1)(x - x_1)}{n-1}$$

$$= 14$$

$cov(x_1, x_2) = -11$

$cov(x_2, x_1) = -11$

$cov(x_2, x_2) = 23$

$$S = \begin{bmatrix} 14 & -11 \\ -11 & 23 \end{bmatrix}$$

Step-3

$$\boxed{S - \lambda I = 0}$$

$$\begin{bmatrix} 14 & -11 \\ -11 & 23 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = 0$$

$\Rightarrow \lambda^2 - 37\lambda - 201 = 0$

$\lambda_1 = 30.3849$

$\lambda_2 = 6.6151$

Step-4 assume

eigan vecton $U = \begin{bmatrix} U_1 \\ U_2 \end{bmatrix}$

$$\boxed{(S - \lambda I)\, U = \begin{bmatrix} 0 \\ 0 \end{bmatrix}}$$

$$\Rightarrow \begin{bmatrix} 14-\lambda & -11 \\ -11 & 23-\lambda \end{bmatrix} \begin{bmatrix} U_1 \\ U_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} (14-\lambda)U_1 - 11U_2 \\ -11U_1 + (23-\lambda)U_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

(i) $\rightarrow (14-\lambda)U_1 - 11U_2 = 0$

(ii) $\rightarrow -11U_1 + (23-\lambda)U_2 = 0$

$$\frac{U_1}{11} = \frac{U_2}{14-\lambda} = t$$

$$\therefore\; U = \begin{bmatrix} 11 \\ 14-\lambda \end{bmatrix}$$

eigan vecton 1

$$U_1 = \begin{bmatrix} 11 \\ 14-\lambda_1 \end{bmatrix}$$

$$||U_1|| = \sqrt{(11)^2 + (14-\lambda_1)^2}$$

$$\lambda_1 = 30.3849$$

$$\therefore ||U_1|| = 19.734$$

eigan vecton-2

$$U_2 = \begin{bmatrix} 11 \\ 14-\lambda_2 \end{bmatrix}$$

$$||U_2|| = \sqrt{11^2 + (14-\lambda_2)^2}$$

$$\lambda_2 = ?$$

$$\therefore ||U_2|| = ?$$

$\boxed{\text{step-5}}$ computing Principle component 1

$K=1$

PCA components No

$$e_1 = \begin{bmatrix} \frac{11}{110{,}11} \\ \frac{14-\lambda_1}{110{,}11} \end{bmatrix} = \begin{bmatrix} 0.5574 \\ -.8303 \end{bmatrix}$$

Transpose of e1

$$PC = e_1^T \begin{bmatrix} \overset{a}{x_{1k}} - \bar{x}_1 \\ \underset{b}{x_{2k}} - \bar{x}_2 \end{bmatrix}$$

$$= \begin{bmatrix} 0.5574 & -.8303 \end{bmatrix} \begin{bmatrix} 4 - 8 \\ 11 - 8.5 \end{bmatrix}$$

$$= -4.30535$$

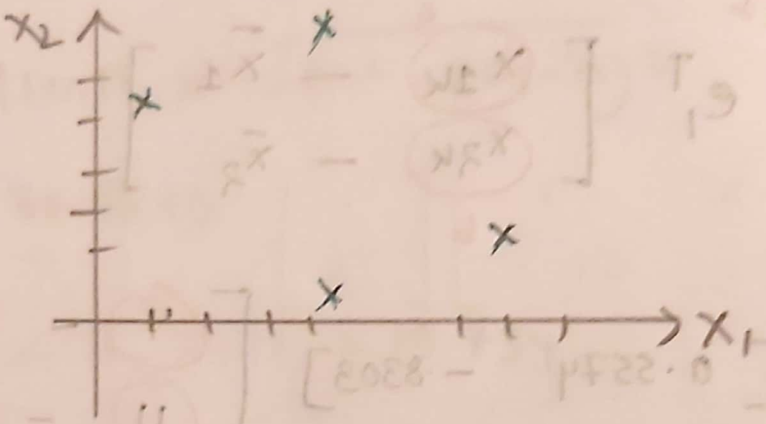
$\therefore PC_{i=1} = -4.3035$

Same find PC 2, PC 3, PC 4

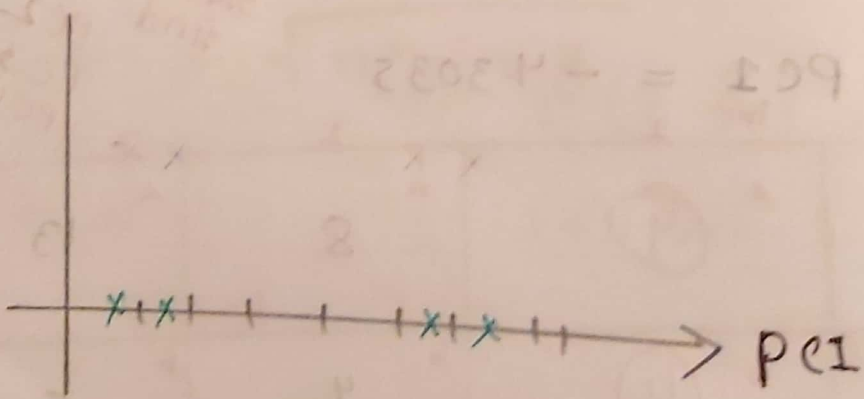
	1	2	3	4
x_1	a 4	a 8	a 13	a 7
x_2	b 11	b 4	b 5	b
PC1	PC = -4.3053	PC = 3.7361	PC = 5.6928	PC = -5.128

Given Data points projected to 2d to 1D

Before PCA



After PCA   1 Dimmension oNly.



PC1 :- -4.3052, 3.7361, 5.69, -5.1238