PCA

Vikash Tripathi

Feature Extraction

PCA (1)

(1) It will reduce curse of dimn.

(2) It reduce higher dimn to lower dimn

Benefit

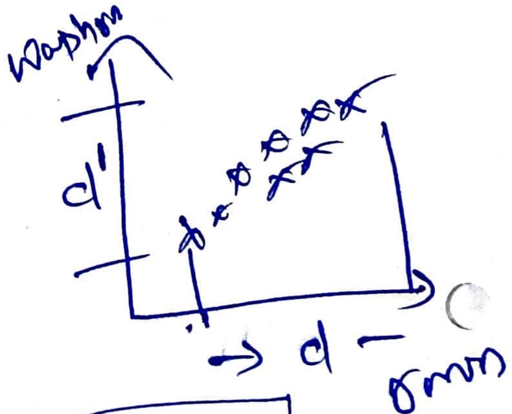
(1) faster execution of Algo

(2) It will help visualization
(because we can reduce Higher dimn to 2D also).

No of Room	# of grocery Shops	Price
3	2	60
4	0	130
5	6	170
2	10	90

d > d'
so Room have higher distribution as compair to d'

No of Room	No of Washroom	Price
3	2	20
4	3	50
5	5	100



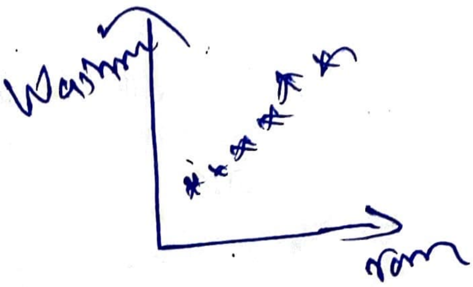
so we can not remove Washroom form feature

$\boxed{d \simeq d'}$

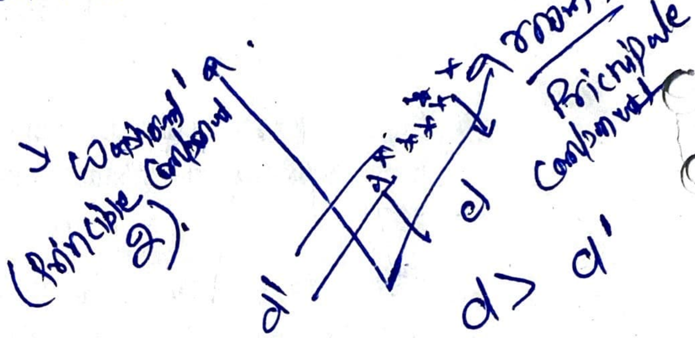
PCA ②

Now we can create New feature that is Size

Size	Price.



Note: **PCA** will create New feature from all existing feature



$\boxed{\#\text{ of PC} \le \text{no of feature}}$

Whey Variane is important

③ PCA

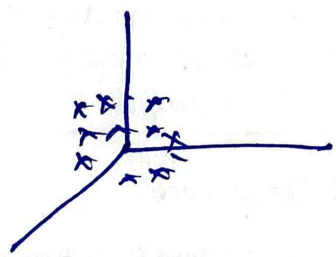
Variame

$$\frac{\sum (x_i - \bar{x})^2}{n}$$

Variane is proposnal to spread of data.

Example

f_1	f_2	f_3	y



Step I Mean Centric data convert.

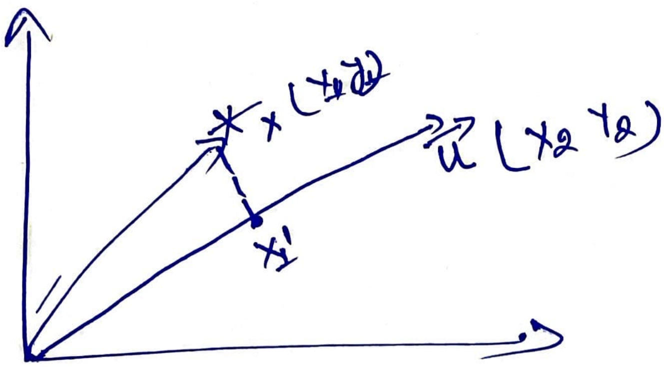
II. Find Co-variance Matrix

$$\begin{array}{c|ccc} & f_1 & f_2 & f_3 \\ \hline f_1 & V(f_1) & & \\ f_2 & C.V(f_1,f_2) & V(f_2) & \\ f_3 & C.V(f_1 f_3) & Cv(f_3, f_2) & V(f_3) \end{array}$$

III. Find the eign value & eign Vector of Co-variance in these Calc (3- eign valu, 3- Eign Vector)

Find the higest (Eign Vector, eign value)

Solved PCA 1



$$x' = \frac{\vec{u}\cdot\vec{x}}{|u|} = \vec{u}\cdot\vec{x} = u^T x$$

$$\boxed{|u| = 1}$$

Unit vector.

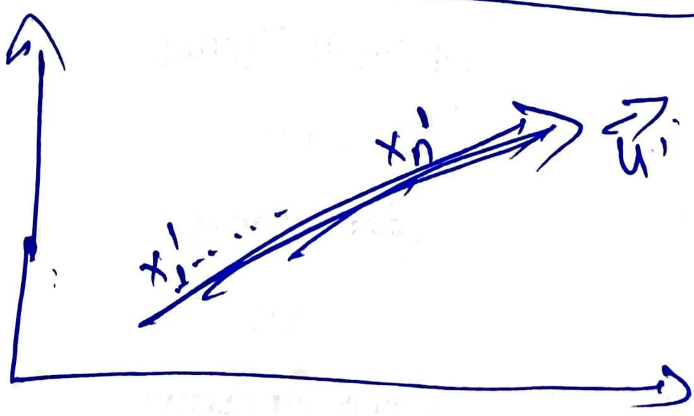
that:

$$\begin{array}{cc} x & u \\ [x_1\ y_1] & [x_2\ y_2] \end{array}$$

$$= X u^T$$

$$= \begin{bmatrix} x_1 & y_1 \end{bmatrix} \begin{bmatrix} x_2 \\ y_2 \end{bmatrix}$$

$$\boxed{x' = x_1 x_2 + y_1 y_2}$$



Variance $\frac{1}{n} \sum (x_i' - \bar{x})^2$

PCA 2

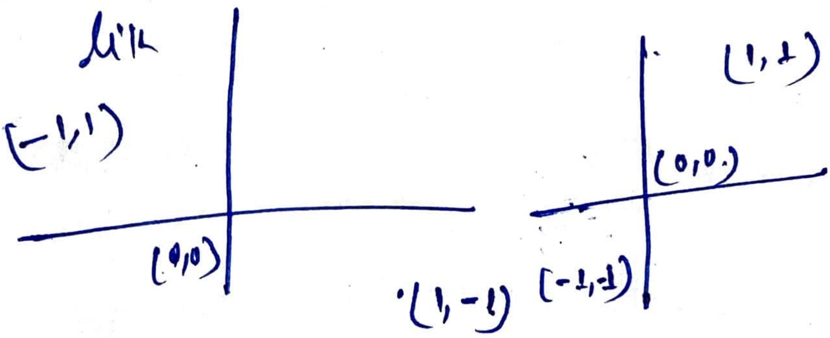
Variance = $\frac{1}{n}\sum_{i=1}^{n}(u^T x_i - u^T \bar{x})^2$

We have to Maximize the Variance

Co-variance & Matrix

Variance will not Provide exact answer

like



Variance $\frac{1+0+1}{3}$ $\frac{1+0+1}{3}$

both the same variance

but co-variance will

$$\frac{-1\times1 + 0\times0 + 1\times-1}{3} = \frac{1\times1 + 0\times0 + -1\times-1}{3}$$

direction $= -\frac{2}{3}$ → magnitude. $= \frac{2}{3}$

Here value same but direction is different

Solved P (1st) 3

Matrix will linear transform

Co-variance Matrix

	x_1	x_2

	x_1	x_2
x_1	$cov(x_1, x_1)$	$cov(x_2, x_1)$
x_2	$cov(x_1, x_2)$	$cov(x_2, x_2)$

$cov(x_1, x_1) = var(x_1)$

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

$\Downarrow$

Find eign dimcomposition
that is
Eign - Value - Eign - Vector

Eign - Vector (It will not change the direction) -
only change the Magnitue) .

It is eign Vector

$$\boxed{A\vec{v} = \underbrace{\lambda}_{\text{eign value}}\vec{v}}$$

Solved Example of PCA

Ex-1

Steps

1. Convert data into - Centering or Standardization

$$\boxed{X_i = \frac{x_i - \mu}{\sigma}}$$

where
μ - mean
σ - S.D

2. Use Co-variance Or Co-relation Matrix

3. Calculate Eigen Value & Eign Vector

4. Sort EignValue (descending Order)

5. Calculate - PC. Per. $PC_i = \frac{\lambda_i}{\sum \lambda_i} \times 100$

6. Select PC

7. Transform Data $PC_{New} = X_{centered} * \text{EignVector}$

Data

	C.G.PA	IQ
1	7	110
2	8.5	120
3	6	105
4	9	130
5	7.5	115

Solved ex PCA (2)

(2) to Standrizatn $X_i' = \frac{X_i - \mu}{\sigma}$

find Mean (μ) C.G.PA = 7.6
I.Q = 116.

$$\sigma = \sqrt{\frac{\sum (X_i - \mu)^2}{n}}$$

σ (C.G.A) = 1.06
σ (I.Q) = 8.6

Standrize -

Z =

	C.G.A	I.Q.
1	-.562	-0.697
2	.843	0.465
3	-1.499	-1.279
4	1.311	1.627
5	-0.094	-0.116

(3) Find - Co-variance Matrix

$$Cov = \frac{1}{n} Z^T_{2\times5} \cdot Z_{5\times2}$$

$$Z^T \cdot Z = \begin{bmatrix} 5 & 4.84 \\ 4.8 & 5 \end{bmatrix}$$

$$Cov = \frac{1}{5}\begin{bmatrix} 5 & 4.84 \\ 4.8 & 5 \end{bmatrix} = \begin{bmatrix} 1 & 0.96 \\ 0.96 & 1 \end{bmatrix}$$

Solved ex PCA-3

④ find Eignvalue

$$\det(\text{cov} - \lambda I) = 0$$

$$\det \begin{bmatrix} 1-\lambda & 0.96 \\ 0.96 & 1-\lambda \end{bmatrix} = 0$$

$$\lambda^2 - 2\lambda + 0.06 = 0$$

$$\lambda = \frac{2 \pm \sqrt{3.7552}}{2}$$

$$\lambda = \frac{2 \pm 1.93}{2}$$

$$\lambda_1 = 1.96 \qquad \lambda_2 = 0.03$$

⑤ find top P.C

$$\%PC1 = \left(1.96 / 1.96 + 0.03 \right) = 98.45\%$$

$$PC2 = \left(\frac{0.03}{1.96 + 0.03} \right) = 1.55\%$$

⑥ find Eijn ~~Valu~~ Vector.

$$\lambda_1 = 1.96$$

$$\begin{bmatrix} 1 - 1.96 & 0.96 \\ 0.96 & 1 - 1.96 \end{bmatrix} \begin{bmatrix} V_{1x} \\ V_{1y} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$-0.96 V_{1x} + 0.96 V_{1y} = 0$$

so take $\boxed{V_{1x} = 1}$ $\qquad V_{1x} = V_{1y}$

Solved ex PCA-4

$$||V_1|| = \sqrt{1^2+1^2} = \sqrt{2} = 1.41$$

Normalize PC_1

$$\boxed{e_1 = \begin{bmatrix} 1/1.41 \\ 1/1.41 \end{bmatrix} = \begin{bmatrix} 0.70 \\ 0.70 \end{bmatrix}}$$

for $\lambda_2 = 0.03$

$$(\text{Cov} - 0.03\, I)\, V_2 = 0$$

$$\begin{bmatrix} 0.96 & 0.96 \\ 0.96 & 0.96 \end{bmatrix} \begin{bmatrix} V_{2x} \\ V_{2y} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$0.96\, V_{2x} + 0.96\, V_{2y} = 0$$

$$\boxed{V_{2x} = -V_{2y}}$$

if $\boxed{\begin{matrix} V_{2x} = 1 \\ V_{2y} = -1 \end{matrix}}$

noralize value $= \sqrt{1^2+1^2} = 1.41$

$$\boxed{e_2 = \begin{bmatrix} 1/1.41 \\ -1/1.41 \end{bmatrix} = \begin{bmatrix} 0.70 \\ -0.70 \end{bmatrix}}$$

Solved ex PCA-S

⑦ Transform data

for PC1

PC1 = Stanz * e1
PC2 = Standz * e2

$e_1 = [0.7, 0.7]$
$e_2 = [0.7, -0.7]$

stand-data-1
PC1 = (-.562) * 0.70 + (-.697 * 0.70) = -.89
PC2 = (-5.62) * 0.70 + (-.697 * (-.70)) = 0.096

stand-data-2 {PC1, PC2}
(.843, 0.96)
(-1.49, -1.27)
stand-data-3 {PC1, PC2}
(1.31, 1.62)
stand-data 4 {PC1, PC2}

stand-data 5 {PC1, PC2}
(-0.09, -0.11)

Final Value

PC1	PC2
-.89	0.096
0.92	0.26
-1.96	-0.15
2.07	-0.224
-0.140	0.016