* PCA

→ It is not a ML Algo. It is Dimensionality Reduction.

- feature Selection ~~&~~ • Dimensionality Reduction

→ • Feature Selection

Column assumption is very high (i.e many clm)
(we have to reduce the feature due to many clm)

→ x_1 x_2 x_3 x_4 x_5 | x_6 ~~x_7~~ ~~x_8~~ x_9 x_{10}

Data → Subset Data

D → D'

This is called feature selection ⇒

Here we are focusing $\boxed{\text{Subset of feature}}$
from 10 clm we are taking only 5 of clm for my model building

(Here we are loosing some information, because we are discarding the feature based on conditions)

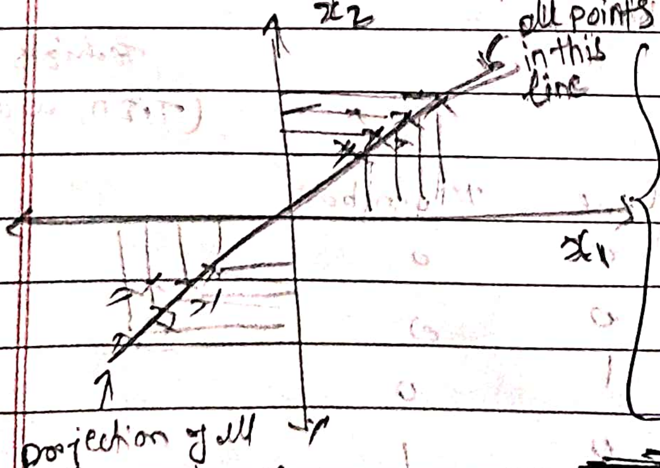
- Various Method of feature Selection

① feature Importance → (DT, Extra tree classifier)
② Statistical method
↳ Chi-square, Avova, P-value
③ Backward Elimination and forward Elimination
④ Laso regression
⑤ Correlation

use for elimination the feature

⇒ • Dimensionality Reduction

Interview



→ Here we can preserve information w.r.t both axis (i.e x_1 and x_2)
→ Means x_1 and x_2 we are converting into single axis
→ This is called Dimensionaly Reduction (or) feature Transformation

⇒ Here ~~we~~ we are combining two or more features into single feature

x_1	x_2
:	:

This data we are transforming into single axis
This is called Dimension reduction

Feature selection ⇒ we are Discarding the feature based on conditions.
Dimensionality Reduction ⇒ we are combining two feature into single feature.

→ Techniques of Dimensionality Reduction

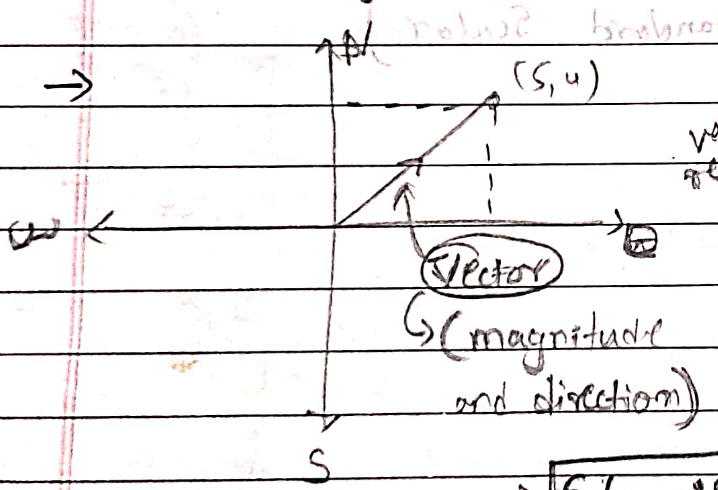
- (i) LDA → Linear Discriminant Analysis
- (ii) PCA → Principal Component Analysis ← (very old and famous technique)
- (iii) t-SNE → T-Stochastic Neighbor Embedding ← (State of Art Algo) (means new Algo)

Steps w.r.t PCA

→ ① Standardization of data
→ ② Covariance matrix
→ ③ Eigen value and Eigen vector
→ ④ Choose our PCA

PCA: Vector (Point) × Vector = Vector's
↑ This vector is called Eigen Vector
↑ This Eigen vector we generate from Eigen Value

- Vector × Eigen Vector = PCA



∴ Magnitude = $\sqrt{5^2+4^2}$

unit vector

$= \sqrt{25+16}$

vector repre ⇒ $\begin{bmatrix}5\\4\end{bmatrix}$ (or) $[5\ 4]$ row vector

$\sqrt{41}$

∴ magnitude of this vector = 6.40

Direction ⇒ North East

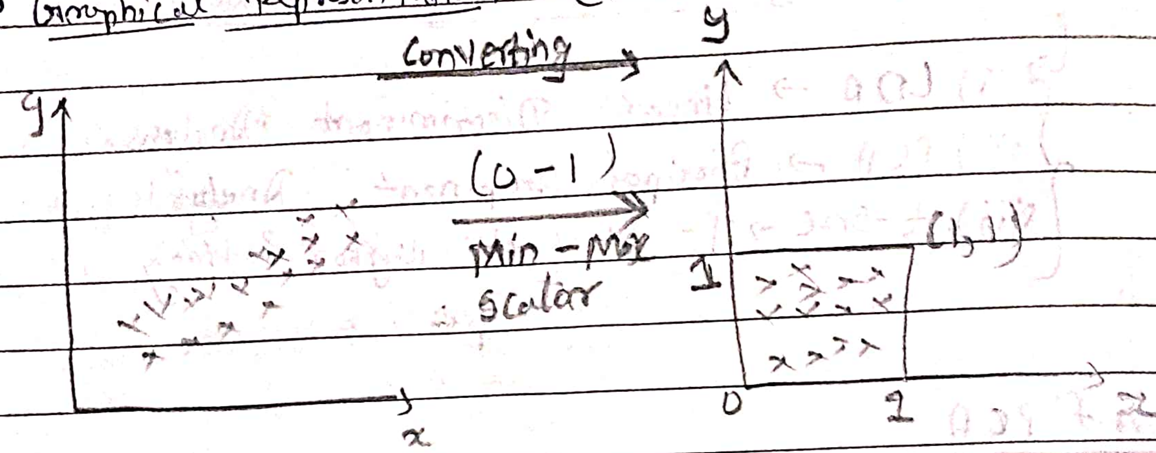
⇒ $\boxed{6.40\ NE}$

Step 1 :- Standardization of the Data
↳ Scaling of the data

Min-Max scaling :- Min Max
(0 to 1)

$$\boxed{\text{Min-Max} = \frac{X - X_{min}}{X_{max} - X_{min}}}$$

• Graphical Representation (Min - Max)



Note:- Scaling of data :- Converting data into certain range

methods are → (1) Min - Max scaling (range → 0 to 1) (converting data in minimum and maximum value)

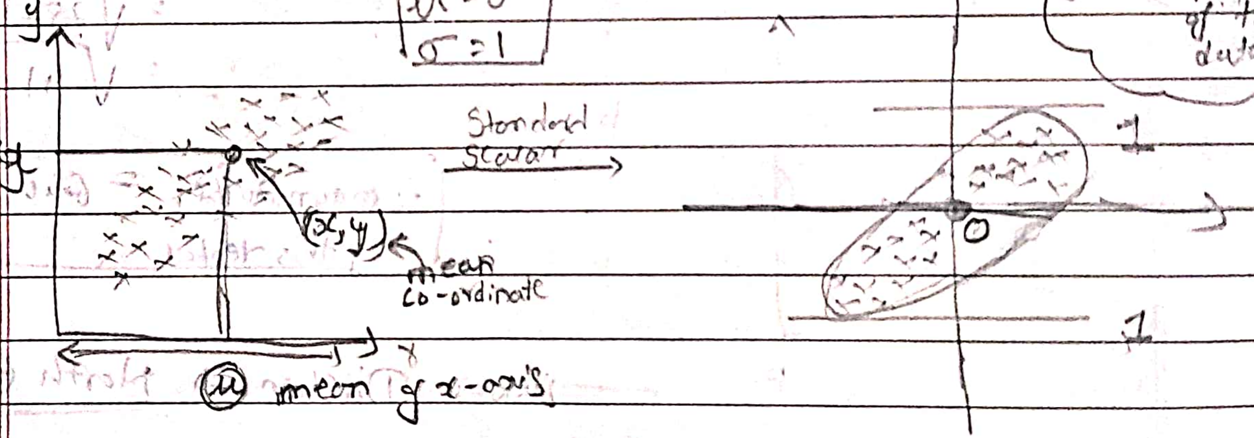
(2) Standarization (Standard scalar) {mean = 0 (μ), Standard deviation = 1(σ)}

• Representation (Standard Scalar)

μ = 0
σ = 1

μ = 0
σ (or) Spread = of the data

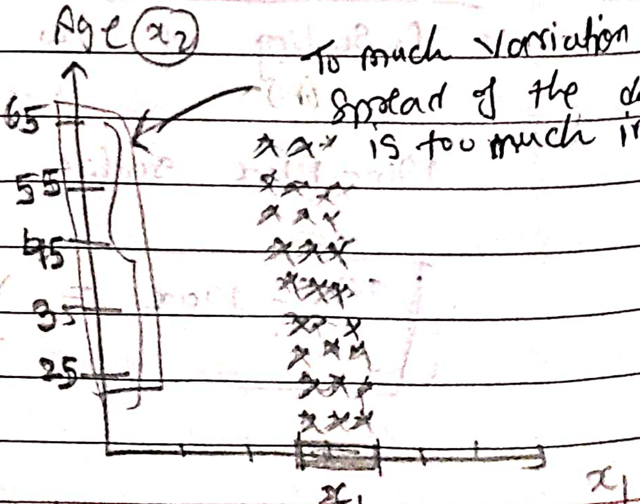
Assuming μ of y - mean y-axis



• Sprade of the data

Consider 2 clm { Person Age → x_2
Colour of Hair → x_1

Black to white
process is v.v. slow process



To much variation cor / Spread of the data is too much in x_2 axis

• x_2 axis ⇒ Too much Spread of Data

→ Concept of Co-variance (Refer Statics Notes)

$x_1\ x_2 \Rightarrow \frac{1}{n}\sum_{i=1}^{n} (x_{1i} - \bar{x}_1) \times (x_{2i} - \bar{x}_2)$ ⇐ between x_1 and x_2 the Covariance will be

$x_1\ x_1 \Rightarrow \frac{1}{n}\sum_{i=1}^{n} (x_1 - \bar{x})^2$ ⇐ between x_1 and x_1 the Covariance will be

variance from feature itself (from x_1 and x_1 features)

- Eigen value and Eigen Vector

$M \times V = \lambda \times V$

$MV - \lambda V = 0$

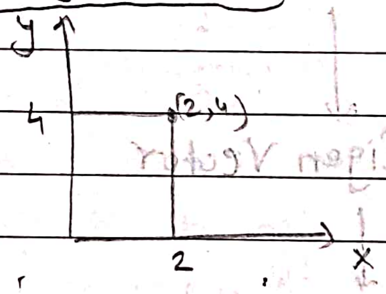
$\therefore \boxed{(M - \lambda I) V = 0}$ from here we have to calculate → PC

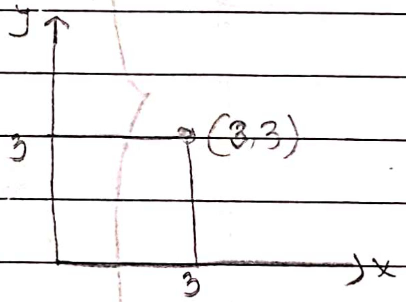
Identity matrix

Identical matrix $I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

Example • Eigen Value and Eigen Vector

$\begin{bmatrix} 2 \\ 4 \end{bmatrix}$ → Vector





$\begin{bmatrix} 3 \\ 3 \end{bmatrix}$ → Vector

⟹ Eigen Value and Eigen Vector we have to findout w.r.t matrix

\# $\therefore \boxed{AV = \lambda V}$ → Eigen Vector should fulfill this condition

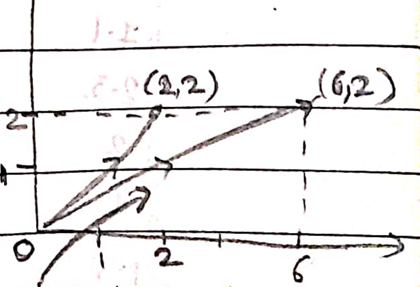
[ ] A-matrix, vector, scalar value, Vector

eg(1) $\begin{bmatrix} 1 & 2 \\ 1 & 0 \end{bmatrix} \times \begin{bmatrix} 2 \\ 2 \end{bmatrix} = \begin{bmatrix} (1\times2) + (2\times2) \\ (1\times2) + (0\times2) \end{bmatrix}$

matrix (A) Vector (V) multiply

$= \begin{bmatrix} 6 \\ 2 \end{bmatrix}$

Vector

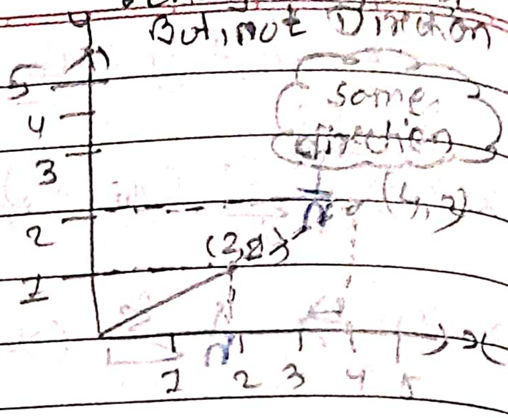


The direction is going to change of these vector

$\begin{bmatrix} 1 & 2 \\ 1 & 0 \end{bmatrix} \times \begin{bmatrix} 2 \\ 2 \end{bmatrix} = \begin{bmatrix} 6 \\ 2 \end{bmatrix}$

This is not Eigen vector of this matrix (A)

Eigen Vector → change magnitude But not Direction



② eg
$$\overset{\text{matrix}}{\begin{bmatrix} 1 & 2 \\ 1 & 0 \end{bmatrix}} \times \overset{\text{Vector}}{\begin{bmatrix} 2 \\ 1 \end{bmatrix}} = \begin{bmatrix} (1\times 2)+(2\times 1) \\ (1\times 2)+(0\times 1) \end{bmatrix}$$

Eigen vector of this matrix (A)

$$\rightarrow \begin{bmatrix} 1 & 2 \\ 1 & 0 \end{bmatrix} \times \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$$

$$= \underset{\text{Eigen value}}{2} \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$ //

$$\underset{\text{matrix}}{A} \times \underset{\text{vector}}{V} = \overset{\text{Eigen value}}{\lambda} \times \underset{\text{vector}}{V}$$

Here we are able to satisfyes the Equation

$$\boxed{A \times V = \lambda \times V}$$

Now this vector we have to multiply with our data → then we will be able to find out PC

\# Matrix A ⟶ $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$ × [data] ⟹ $\boxed{PC}$

Eigen vector w.r.t this matrix (A)

\# means converting Data into Single Dimension

• A is co-variance matrix, and this co-variance matrix we have generated from Data

Data
↓
A ← Co-Variance Matrix {we have generated from Data}
↓ from this Covariance matrix we have to find out Eigen Vector
Eigen Vector
↓ Then we are goin to multiply this (Eigen vector with our Data)
PC {then we getting PC}

Data

Ex:-

x	y
2.5	2.4
6.5	0.2
7.2	2.9
11.9	2.2
7.1	3.0
2.3	2.7
2	1.6
1	1.1
1.5	1.6
1.1	0.9

⟹ we have to convert this data into single dimension (or) we have to transform this data into single dimension.

Date: 4
Page:

Step :- ① Standrization of Data

(Assume already standardized Data)

$$Z = \frac{X - \mu}{\sigma}$$ After standerdization my → $\mu = 0$ and $\sigma = 1$

μ → mean, σ → SD

~~It~~ is also called Z-score

Step ② Co-variance Matrix

$$\begin{array}{c|c} & \begin{array}{cc} X & Y \end{array} \\ \begin{array}{c} X \\ Y \end{array} & \begin{bmatrix} cov(x,x) & cov(x,y) \\ cov(y,x) & cov(y,y) \end{bmatrix} \end{array}$$

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

{Co-variance between X and Y}

$$(x,y) = \frac{1}{n}\sum_{i=1}^{n}(x-\bar{x})(y-\bar{y}) \Rightarrow (X,Y) \rightarrow$$ { Showing relationship between feature (or) Data

{Co-variance between X and X}

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

$$(Y,Y) = \frac{1}{n}\sum_{i=1}^{n}(Y-\bar{Y})^2 \Rightarrow (Y,Y) \rightarrow$$ {Co-Variance between Y and Y

using the above co-variance formula we can find Co-variance

Co-Variance Matrix → Ans

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

{ cov(x,y) = cov(y,x)
Digonal will be always same

Step ③ Calculate Eigen Value and Eigen Vector

$\boxed{A - \lambda I = 0}$ ⟶ Eigen Value formula

A → matrix

$$= \begin{bmatrix} 0.6165 & 0.6154 \\ 0.6154 & 0.7165 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = 0$$

$$= \begin{bmatrix} 0.6165 & 0.6154 \\ 0.6154 & 0.7165 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} = 0$$

Determinants

$$[(0.6165-\lambda)\times(0.7165-\lambda)]-[(0.6154\times 0.6154)]=0$$

$\Rightarrow$ $\boxed{\lambda^2-1.33\lambda+0.630=0}$ ← Quadratic Equation

$\boxed{\lambda=?}$

Eigen Value $\Rightarrow$ $\boxed{\begin{matrix}\lambda_1 = 0.49 \\ \lambda_2 = 1.284\end{matrix}}$ } 2 Eigen value

- Eigen Vector

$\boxed{AV=\lambda V}$

Now we have to calculate x and y value

$$\begin{bmatrix}0.6165 & 0.6154 \\ 0.6154 & 0.7165\end{bmatrix}\begin{bmatrix}x_1 \\ y_1\end{bmatrix}=0.49\begin{bmatrix}x \\ y\end{bmatrix}$$

Vector

$$\Rightarrow \boxed{\begin{matrix}0.6165x_1+0.6154y_1=0.49x \\ 0.6154x_1+0.7165y_1=0.49y\end{matrix}}$$

$$x_1=-1.0845\,y_1$$

$\Rightarrow$ $\boxed{\lambda_1=0.49}$ $\dfrac{x_1}{y_1}=\dfrac{-1.0845}{1}$ $\begin{bmatrix}-1.0845 \\ 1\end{bmatrix}$ ← 1st Eigen Vector

$\Rightarrow$ $\boxed{\lambda_2=1.284}$ $\begin{bmatrix}0.6778 \\ 0.7351\end{bmatrix}$ ← 2nd Eigen vector

This 2 data we are rep in 1 data

$$\text{Data} \Rightarrow \begin{array}{cc} x & y \\ \end{array}$$

$$\begin{bmatrix}7.5 & 7.4 \\ 6.5 & 0.2 \\ 7.2 & 7.9 \\ 11.9 & 2.2 \\ \vdots & \vdots \\ \vdots & \vdots\end{bmatrix}\times\begin{bmatrix}-1.0845 \\ 1\end{bmatrix}\Rightarrow\boxed{PC_1}$$

Here we are getting 1st PC_1

$\Rightarrow$ Multiplying Data with 1st Eigen Vector

Data $\begin{bmatrix} 2.5 & 2.4 \\ 0.5 & 0.2 \\ \vdots & \vdots \end{bmatrix} \times \begin{bmatrix} 0.6718 \\ 0.7351 \end{bmatrix} \Rightarrow \boxed{PC_2}$ then we are getting 2^{nd} PC_2

This 2 Data we are Representing in 1 Data → 1 vector

one PC will show → High Variation (or) High spread of data

we are getting 2 PC

$\boxed{PC_1 \text{ and } PC_2}$

$\lambda_1 = 0.49$ $\lambda_2 = 1.284$

$\therefore 1.284 > 0.49 > \ldots$ which one is bigger in dese order

- If we have ⇒ $\boxed{2 \text{ features}}$

→ quadratic

↓

$\boxed{2 \text{ root}}$

↓

$\boxed{2 \text{ Eigen value}}$

↓

$\boxed{2 \text{ Eigen Vector}}$

↓

$\boxed{2 \text{ PC}}$

No. of feature = No. of PC