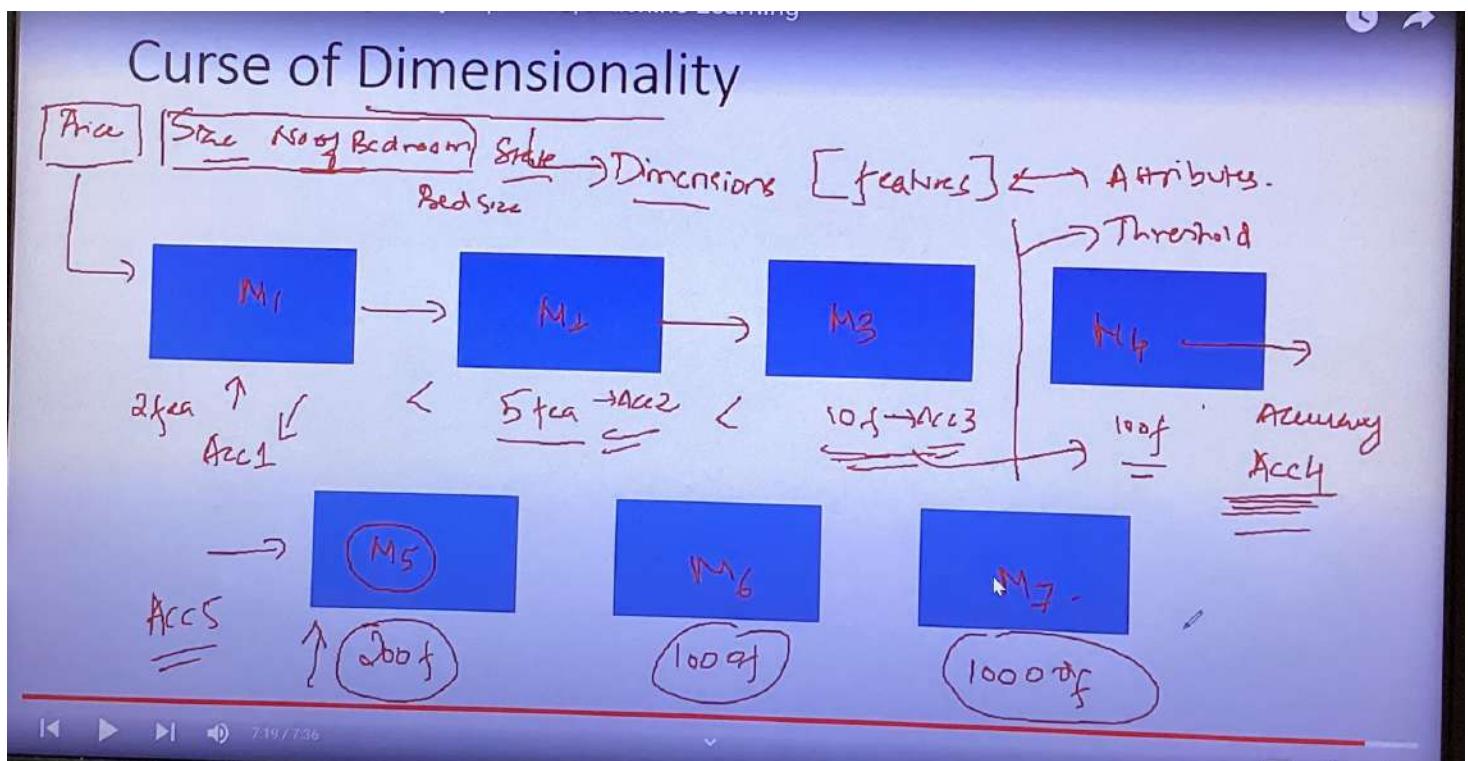


Curse of Dimensionality:-

AGENDA:

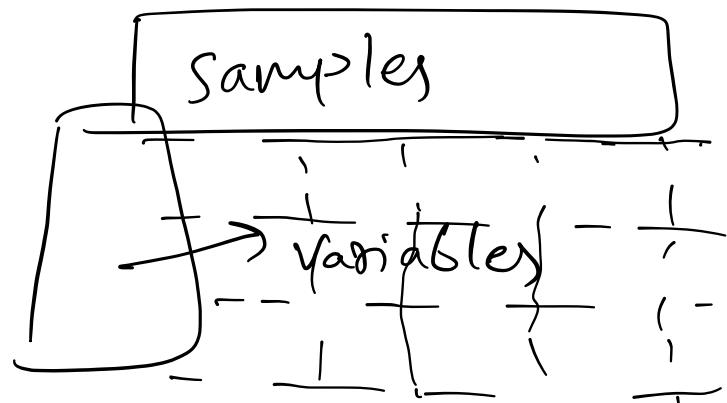
→ picking only useful features (independent)



→ there limitations (or) abrupt change in accuracy depending on no. of features is called as **Curse of Dimensionality**.

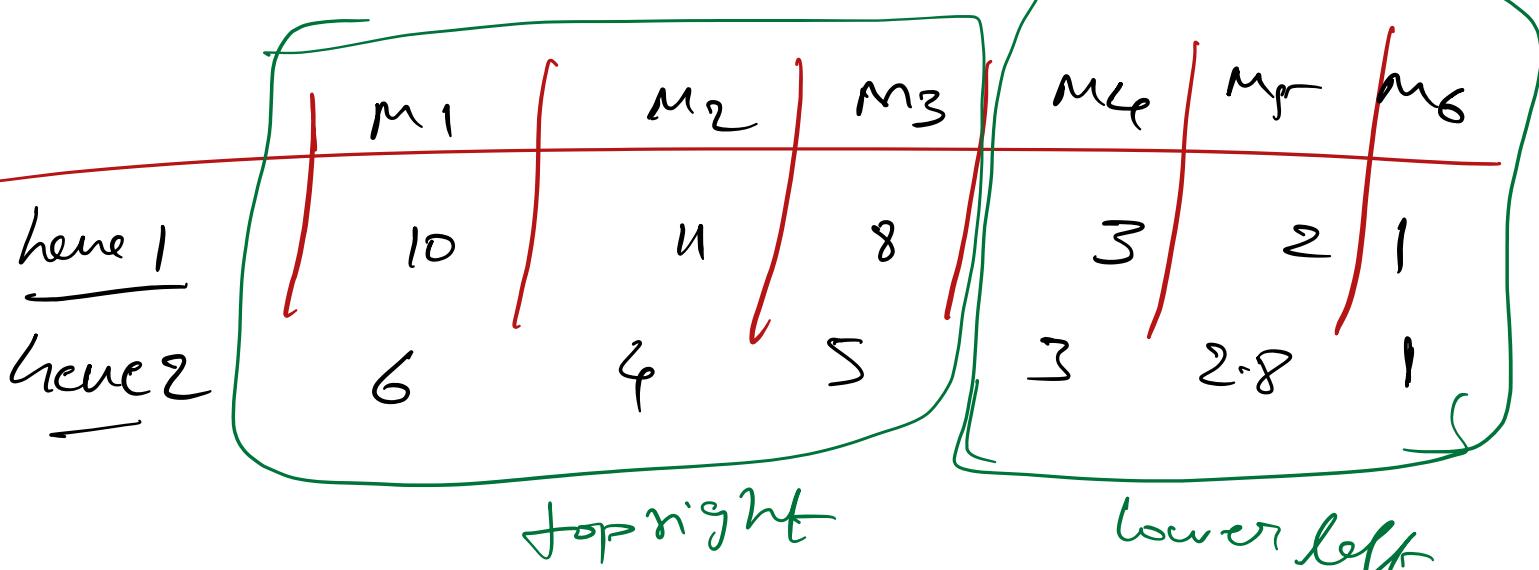
Principal Component Analysis:

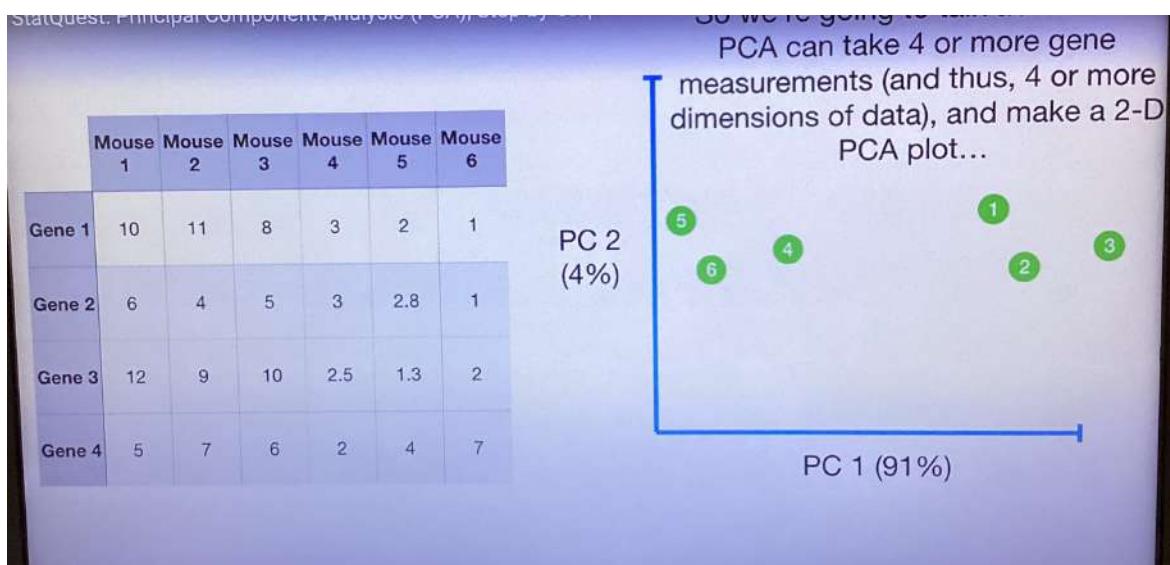
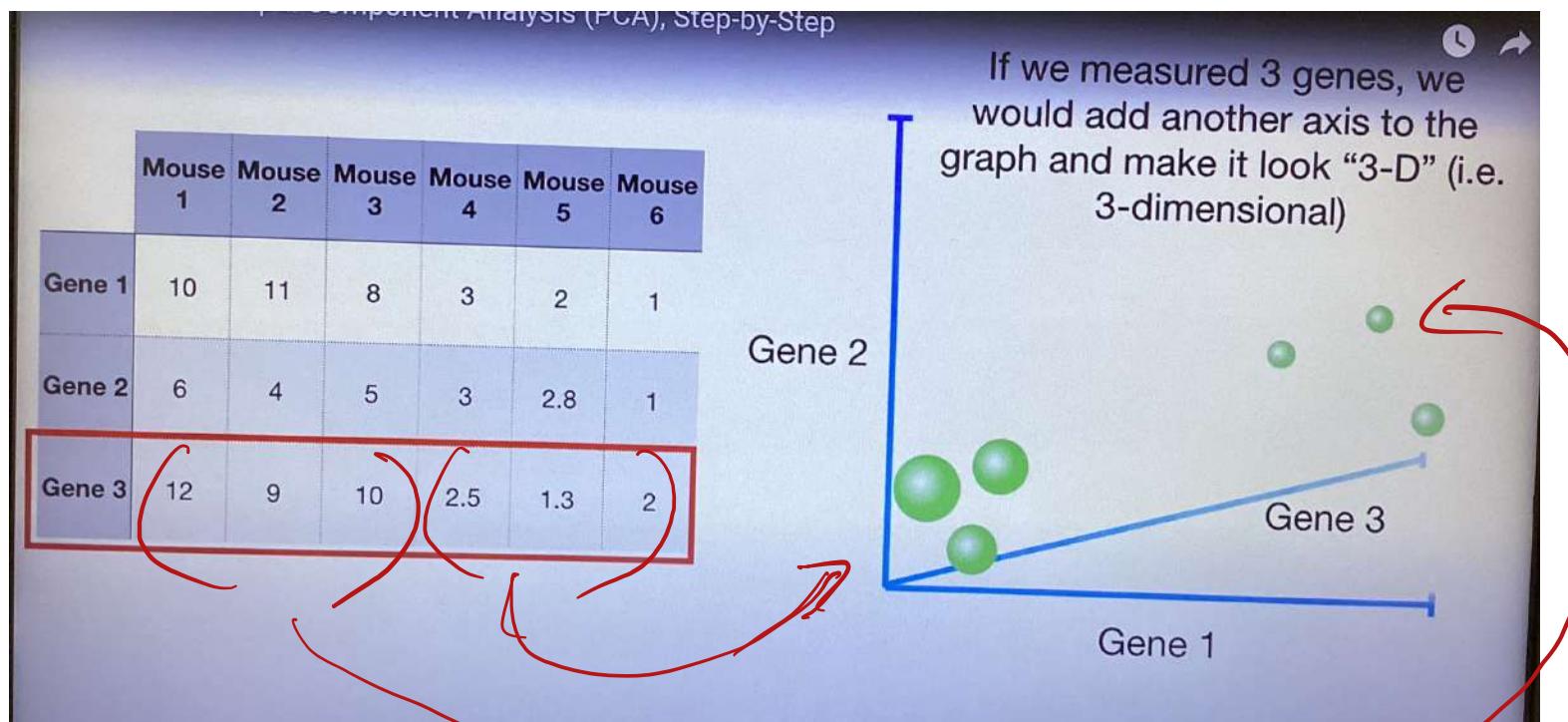
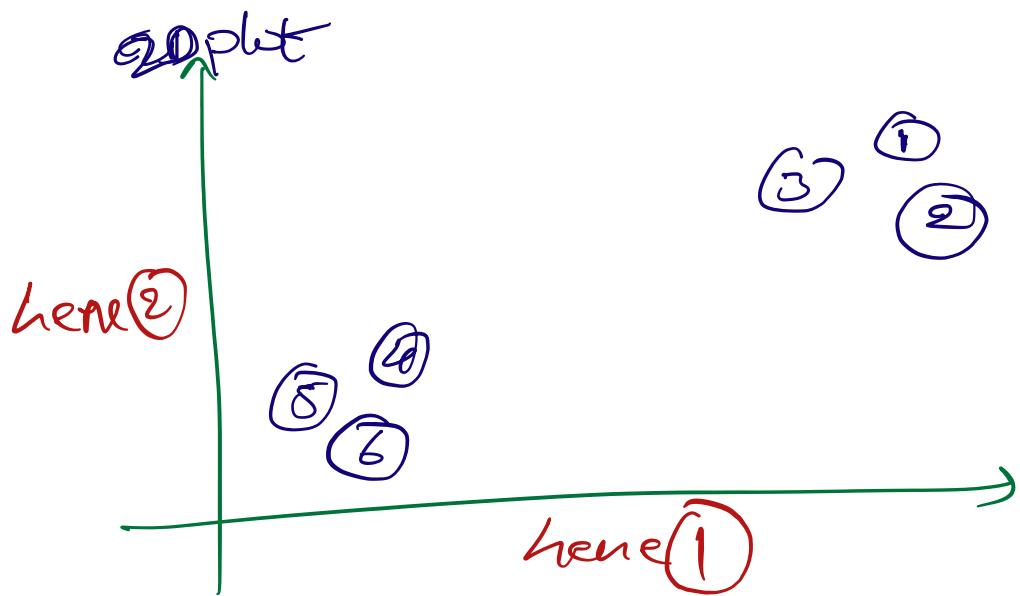
→ as the no. of dimensions increase, it is called a curse. (as the accuracy is affected)



ex:-

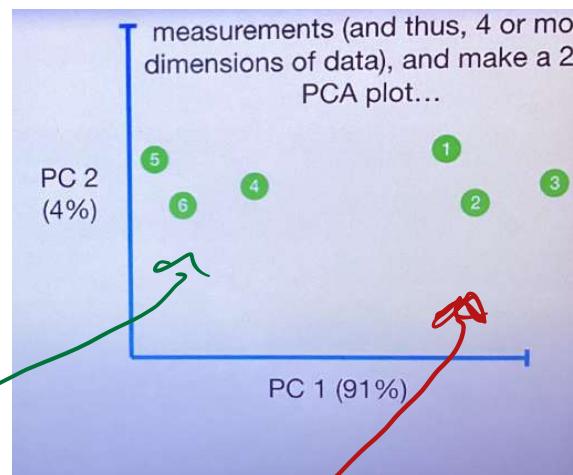
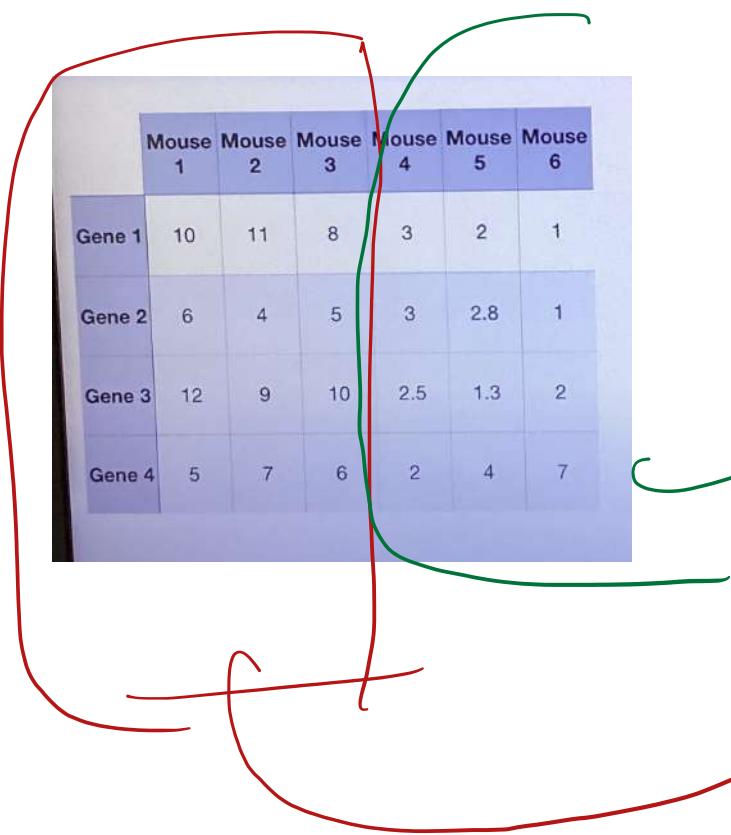
	m_1	m_2	m_3	m_4	m_5	m_6
hence 1	10	11	8	3	2	1





If we measured 4 genes, however,
we can no longer plot the data
-4 genes require 2 dimensions

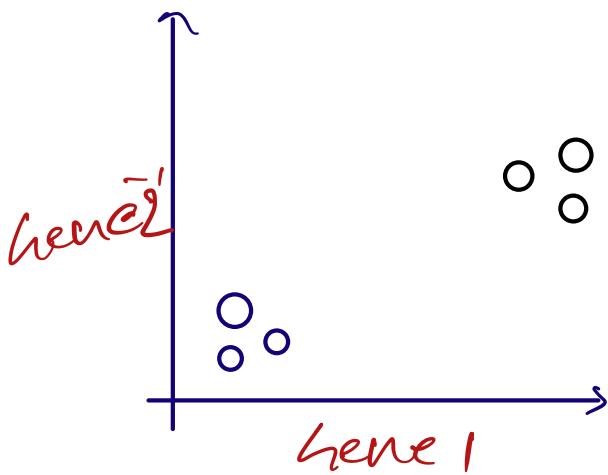
→ So, we're going to talk about how
PCA can take 2 or more gene
measurements (and thus, 4 or more
dimensions of data) and make a 2D
PCA plot.



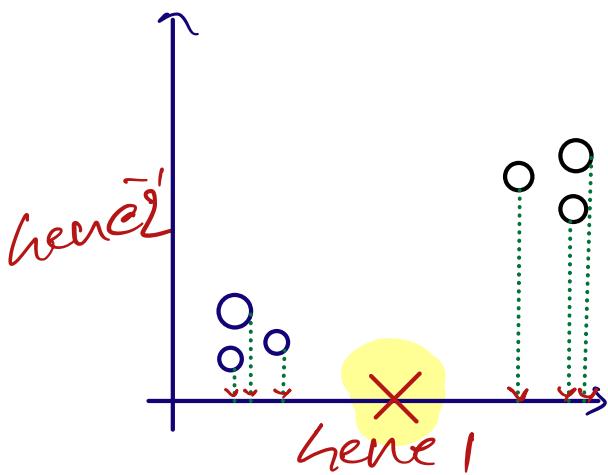
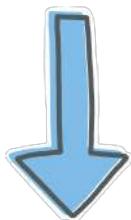
Finally, we'll talk about how PCA can

tell us how accurate the ~~2D~~ graph
is.

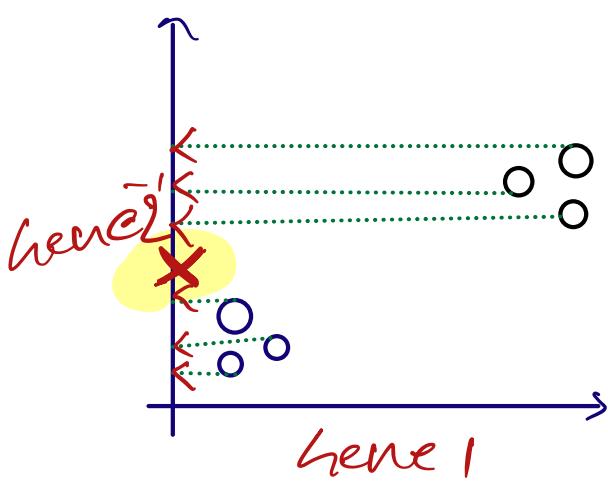
lets go back to the data of 2-gene:



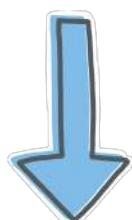
→ after plotting the data as shown.

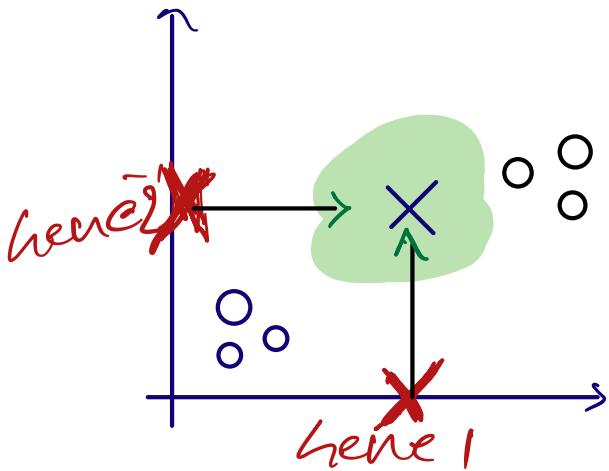


→ we take the avg measurement for gene 1

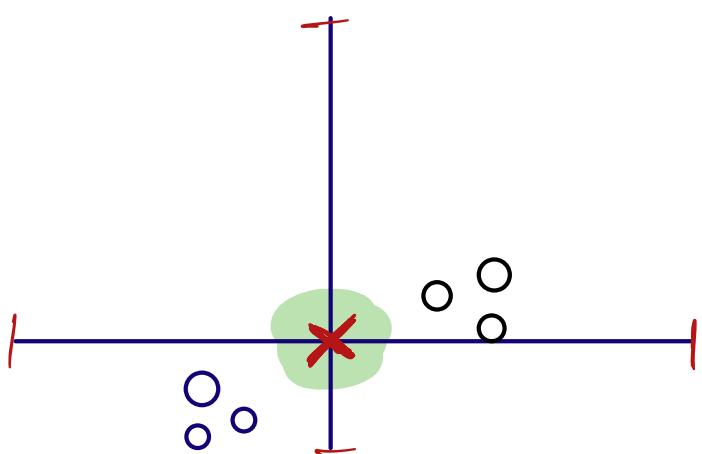


→ and for gene 2





→ with the avg values
we can calculate the
centre of the data



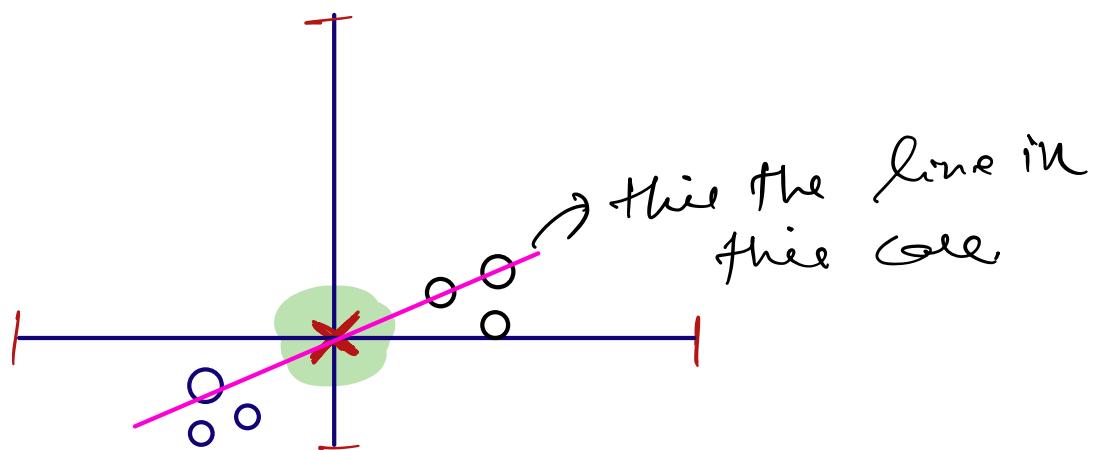
→ Now we'll shift the
data so that the
center is on top of
the origin $(0,0)$ in
the graph.

→ now, we try to find a line which
passes through the origin where our
centre of data is located

→ And our Aim is to, find a line in a
way that has minimum distance after
projection from deal points.

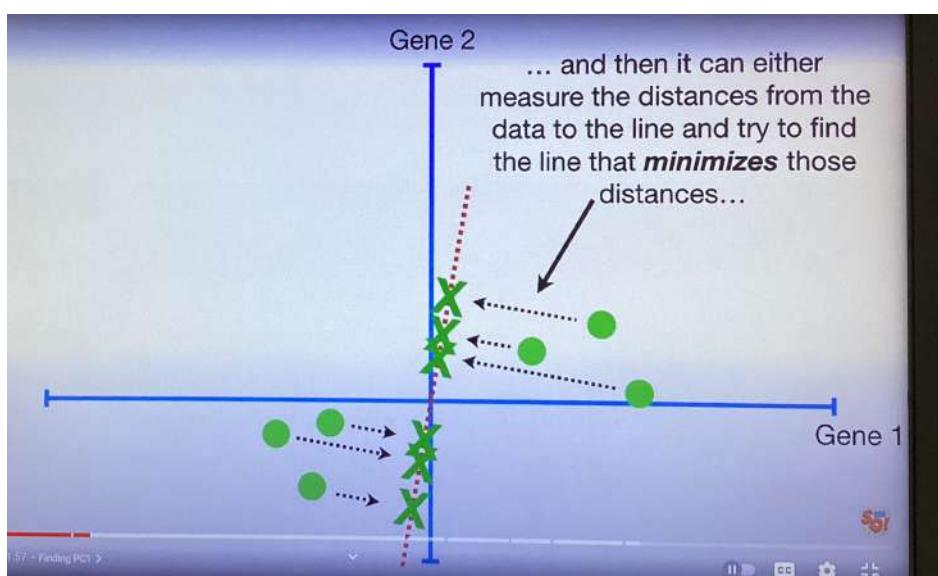
→ we can say it like, either decrease the distance of real value and projected value (or)

increase the distance of projected points from origin.

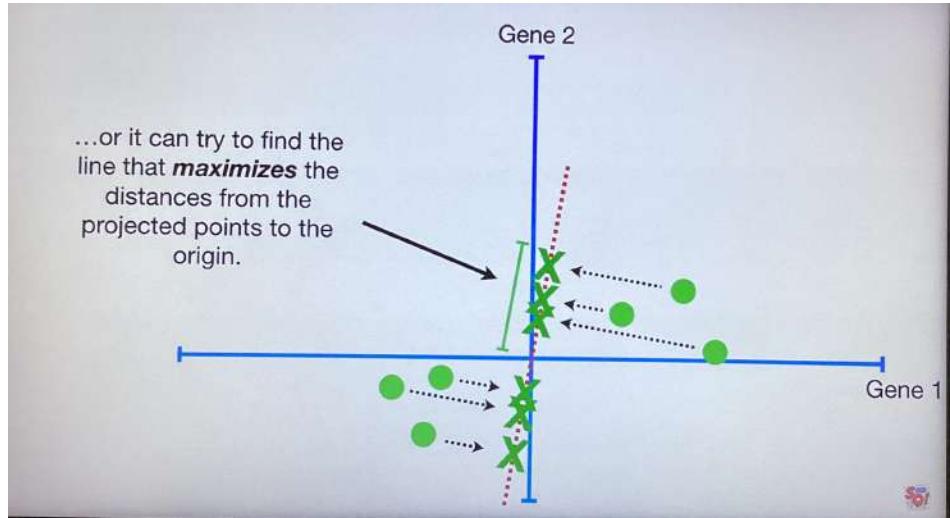


how PCA decides this is the best fit line?

①

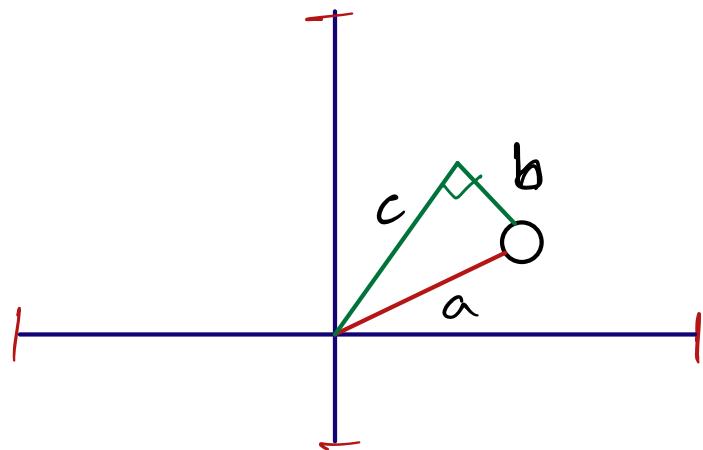


②



Intuition:

- ① These distances get larger when the line fits better



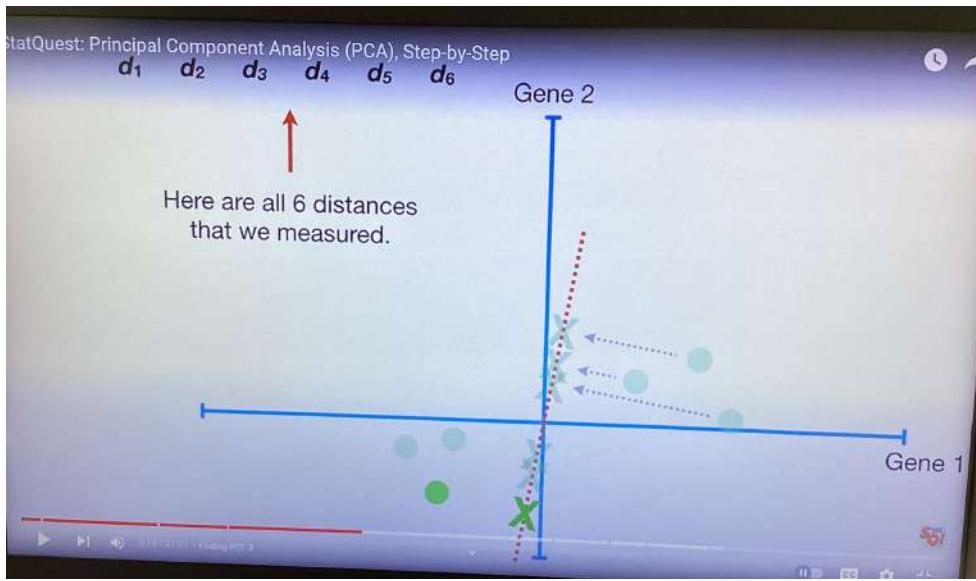
$$a^2 = b^2 + c^2$$

if $b \uparrow$

$c \uparrow$

\Rightarrow then $c \downarrow$

\Rightarrow then $b \downarrow$



→ now, we will take square of distance,
and sum them. \downarrow (SS distance)

Our dist can be ve.

→ our Aim is it, to find a line
where new

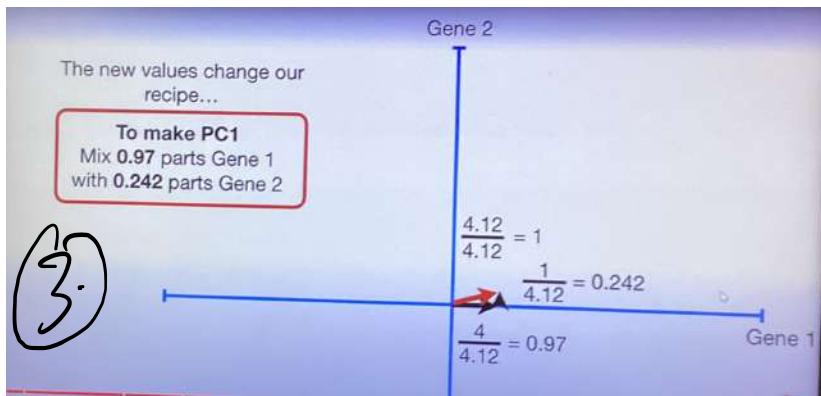
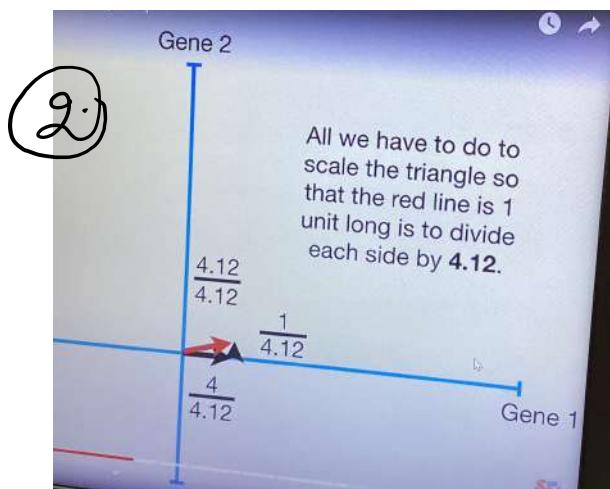
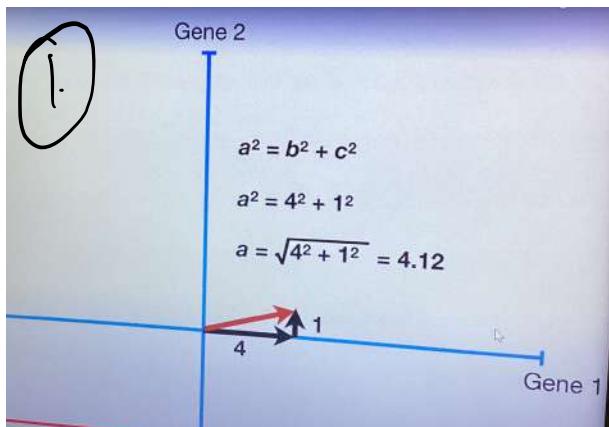
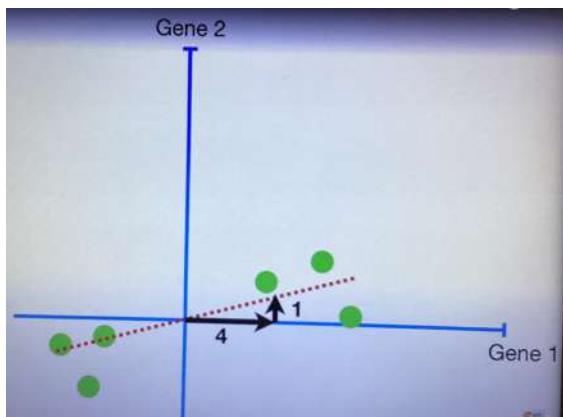
(SS dist) is largest

⇒ This is called principal Component 1
(PC1 for short)

⇒ PC1 is a linear combination of
Variables

In this ex slope of line is 0.25

means 4 parts Gene 1
and 1 part Gene 2



So, this unit vector will guide us, and we can fit the different line... -

Note: This 1 unit long vector, consisting of 0.97 parts Gene 1 and 0.242 parts Gene 2, is called the "Singular Vector" (in the yellow bubble) and the "EigenVector" (in the yellow bubble) for PC1

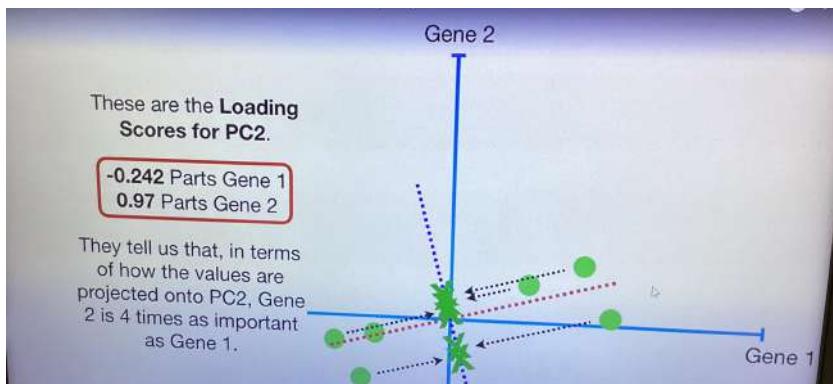
↳ and the proportions of each gene are called "Loading Scores" (in the green bubble)

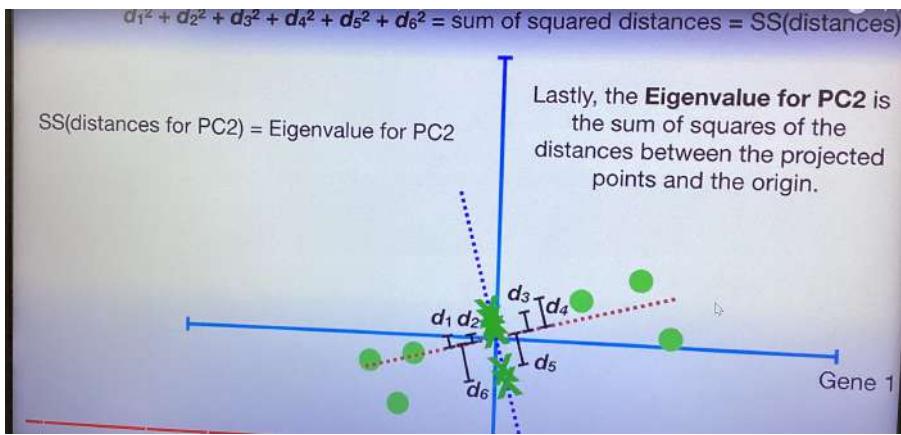
$$SS(\text{diff for PC1}) = \text{Eigenvalue for PC1}$$

$$\rightarrow \text{Eigenvalue for PC1} = \text{Singular Value for PC1}$$

PC2:

a 1st line to PC_1 passing through $(0,0)$

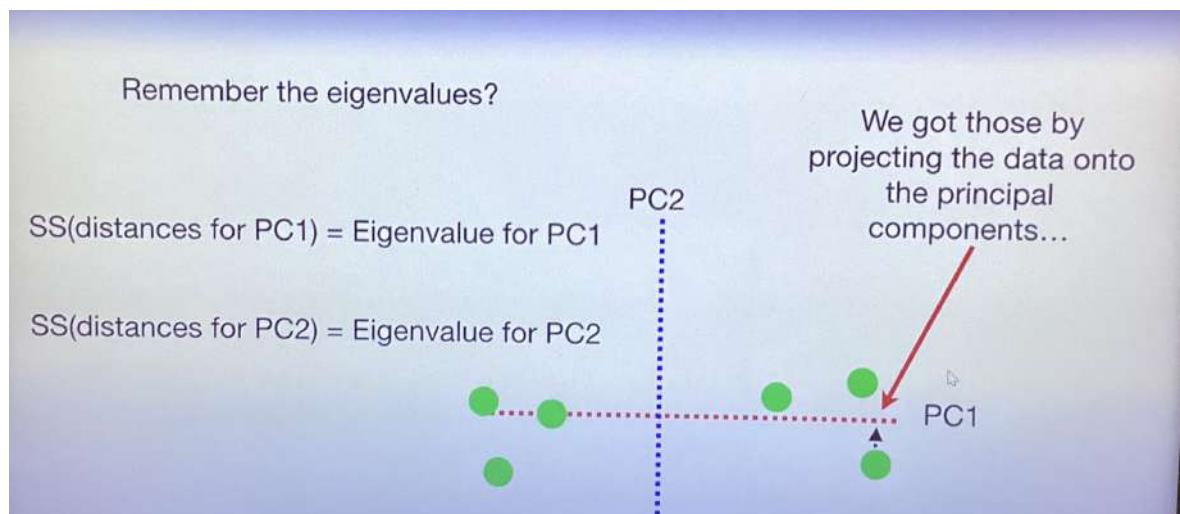




→ this is our PC_2

DON'T FORGET

- ① we simply rotate everything so that PC_1 is horizontal
- ② then we use the projected points to find where the samples go in the PCA plot.



Calculating percent variation for each PC and scree plot?

$SS(\text{distance for } \text{PC}_1) = \text{Eigenvalue for } \text{PC}_1$

$SS(\text{distance for } \text{PC}_2) = \text{Eigenvalue for } \text{PC}_2$

→ we convert them into variation around the origin (0,0) by dividing by the sample size minus 1 (ie $n-1$)

$\frac{SS(\text{dist for } \text{PC}_1)}{n-1} = \text{Variation for } \text{PC}_1$

$\frac{SS(\text{dist for } \text{PC}_2)}{n-1} = \text{Variation for } \text{PC}_2$

let, in the above ex;

$$\begin{aligned} \text{Variation of } \text{PC}_1 &= 15 \\ \text{or} \quad \text{Variation of } \text{PC}_2 &= 3 \end{aligned} \quad \left\{ \begin{array}{l} \text{PC}_1 + \text{PC}_2 \\ = 18 \end{array} \right.$$

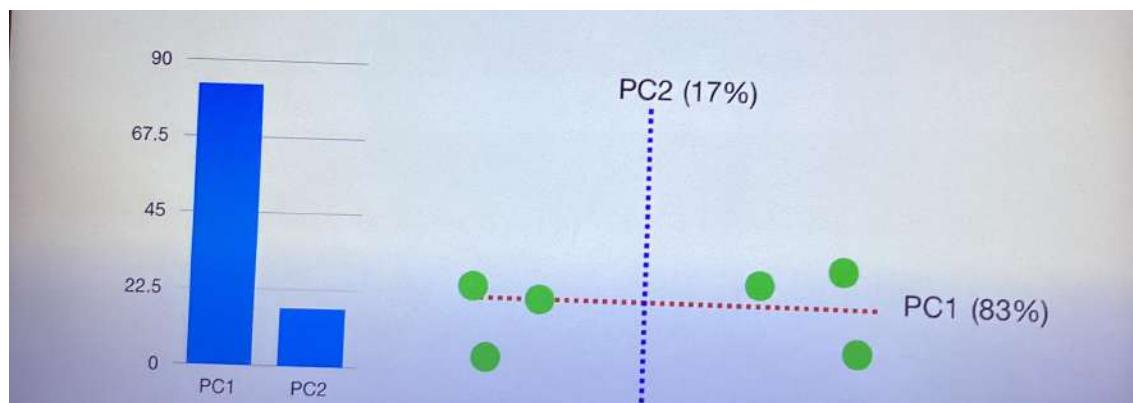
$$\Rightarrow \text{PC}_1 \text{ accounts for } 15/18 = 0.83 \\ = 83\% \quad \checkmark$$

the total variation around the PCs.

$$\Rightarrow \text{PC}_2 \Rightarrow 3/18 = 0.17 = 17\%$$

Scree plot!

\rightarrow it is a graphical representation of the percentages of variation that each PC accounts for.



\rightarrow it is a line plot of the eigenvalues of factors (or) Principal Components in an analysis.

→ The scree plot is used to determine the no. of factors to retain in an exploratory factor analysis (FA) or principal components to keep in a.

a. PCA

In short:

- ① Standardize the range of continuous initial variables
- ② Compute the covariance matrix to identify correlations
- ③ Compute the eigen vectors and eigen values of the covariance matrix to identify the principle components.
- ④ Create a feature vector to decide which principle components to keep
- ⑤ Recast the data along the principle component axes.

Standardization

Value - mean

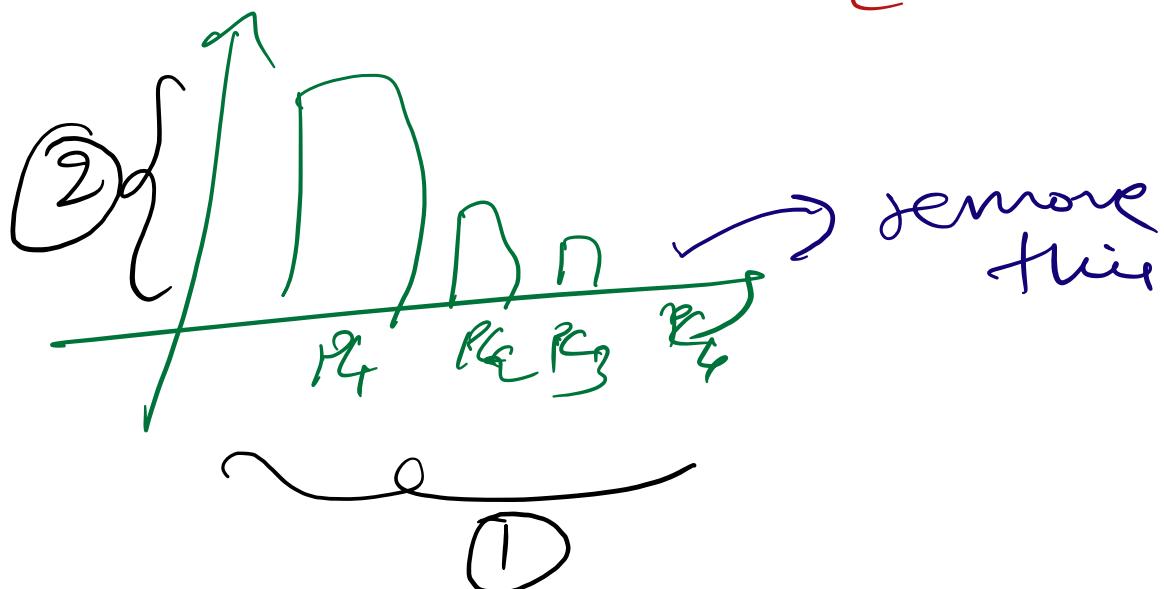
Standard deviation

Cov matrix

(x, x)	(x, y)	(x, z)
(y, x)	(y, y)	(y, z)
(z, x)	(z, y)	(z, z)

① explained Variance

② cumulative explained variance



Final!

$$\underline{\text{Data set}} = \underline{\text{Feature Vector}}^T \times \underline{\text{Standardized Original Dataset}}^T$$

Ro ight

Linear Discriminant Analysis!

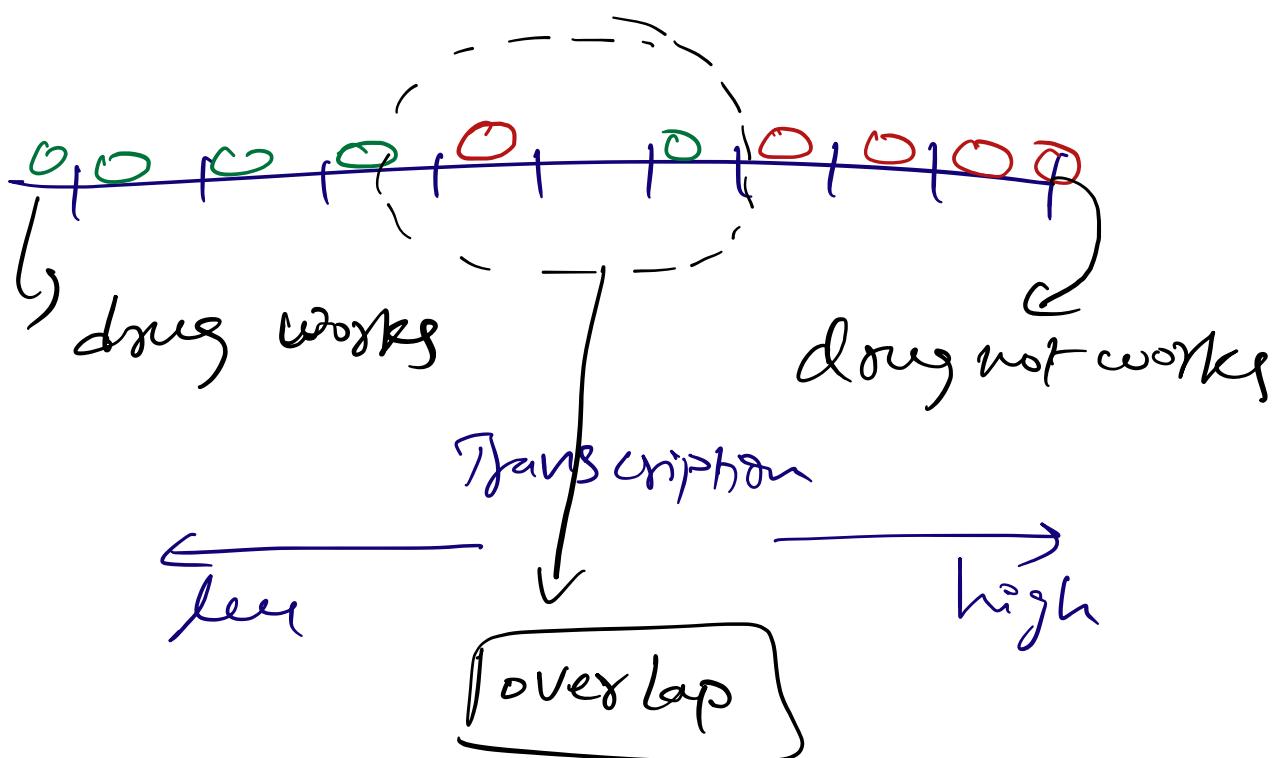
Ex! we have a Cancer drug

- It works great for some people
- but not for others.

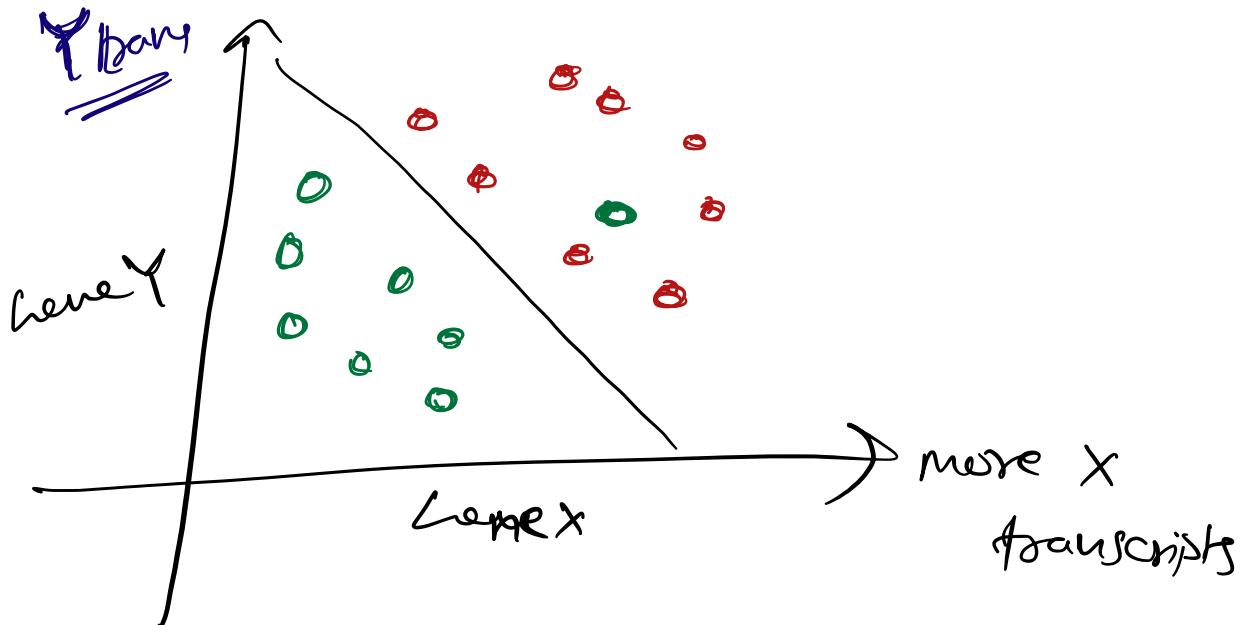
Now, we decide whom to give the drug?

- Gene expression can help us decide.

⇒ let's take a gene and find,



If we take 2 gene



If we have 3-D then it will be more complicated to understand the distribution.

diff b/w PCA and CDA:

- PCA is useful for plotting data with a lot of dimensions (or lot of genes) onto a simple X/Y plot
- However, in this case we're not super interested in the gene with the most variation.
- Instead, we're interested maximizing the separability b/w the two

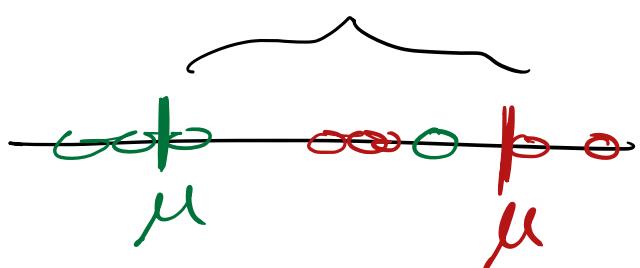
groups so we can make the best decision.

⇒ LDA is like PCA but it focuses on maximizing the separability among known categories

let

- we have a data in 2D, and we try to convert it to 1D, so, that we get the maximum separability
- if we just neglect Gene X (or)
Gene Y It will create loss of data
- LDA uses both genes to create a new axis and projects the data onto this new axis in a way to maximize the separation of the two categories.
- The new axis is created according to two Criteria (simultaneously):

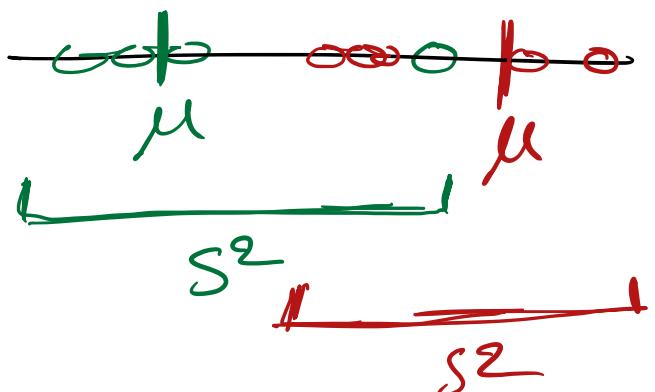
(1) Maximize the dist b/w means



(2) Minimize the Variation (which LDA calls "scatter" and is represented by s^2) within each category

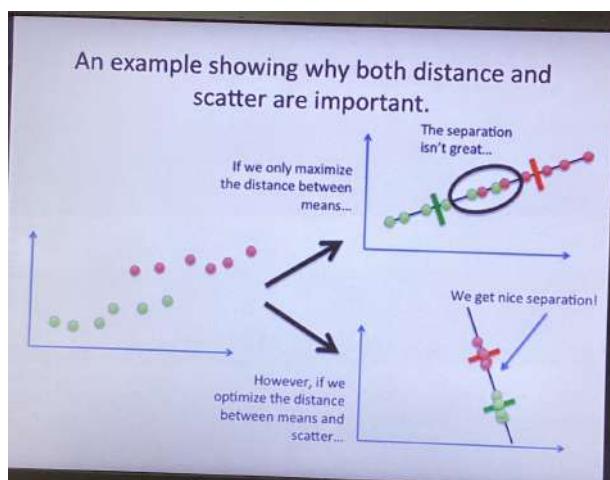
(simultaneously)

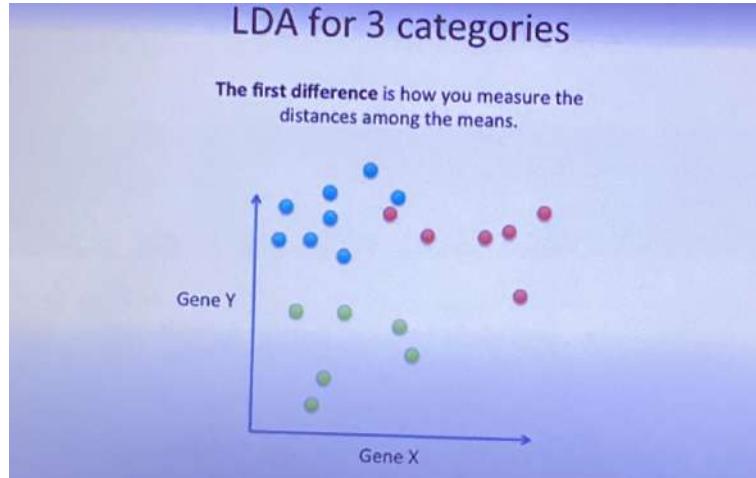
$$\frac{(\mu_1 - \mu_2)^2}{s^2 + s^2}$$



S₂: Cuz any diff can be big and others less.

$$\text{let } (\mu_1 - \mu_2) = d$$



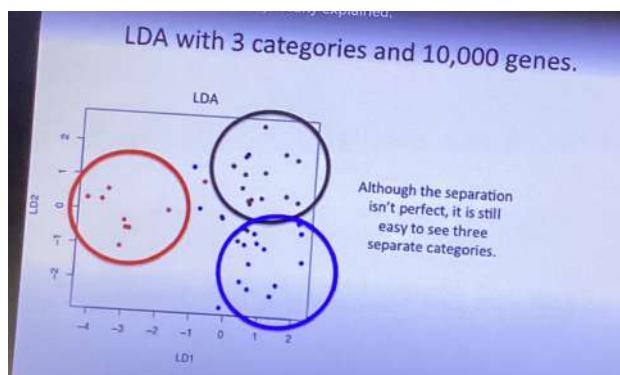


→ '2' genes but '3' Category

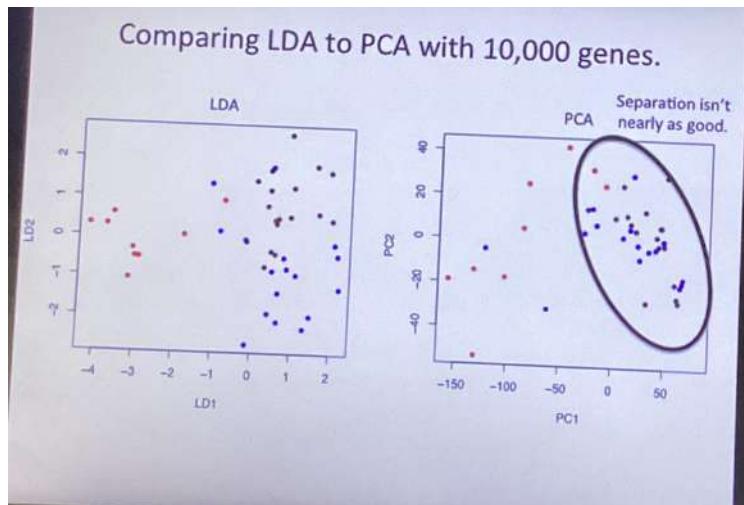
- first find centre for all points
- find distance from each Cat centre to main centre

$$= \frac{d_g^2 + d_r^2 + d_b^2}{S_g^2 + S_r^2 + S_b^2}$$

- we take two lines finally to construct a linear line.



D



diff b/w LDA and PCA

Factor Analysis:

- ① what is factor analysis
- ② latent variables
- ③ Assumptions in factor analysis
- ④ purpose of factor analysis
- ⑤ types of factor analysis
- ⑥ Example with factor analysis
- ⑦ Basic logic of factor analysis

Def.

⇒ It is a technique that is used to reduce a large no. of variables into fewer no. of factors.

⇒ It is a way of condense the data in many variables into just few variables

⇒ It is an example of latent variable model.

Ex:

If you want to go to a hotel,
the factors which you consider are:

- ① waiting time
- ② cleanliness
- ③ staff behaviour
- ④ taste of food
- ⑤ food freshness
- ⑥ food temperature

This might be little diff to choose any hotel.

Let, { ① Service } these are called
 ② Food Quality } all latent variables.

Latent Variables!

- ⇒ they are variables that are not directly observed but are inferred from other variables.
- ⇒ mathematical models that aim to explain observed variables in terms of latent variables are called latent variable models.

e.g. moral, happiness } in real world.

Assumptions in factor analysis:

- ① There are **no outliers** in the data
 - ② **Sample size** is supposed to be **greater than** the factor.
 - ③ Variables must be **interrelated**
(Barrett test)
 - ④ metric variables are expected
(interval data)
 - ⑤ multivariate normality not required
- to analyze the correlation.

interval of numbers in particular.

Type of factor Analysis:

- ① Exploratory factor analysis (EFA)
 - ↳ used to discover underlying structure.
- ② confirmatory factor analysis (CFA)

Used to test if the data fit a priori expectation for data structure

→ used structural equations modelling

EFA is divided into:

- ① PCA
- ② Common factor Analysis
- ③ Image factoring
- ④ Maximum likelihood Analysis
- ⑤ Alpha factoring and weight squar.

Issues with factor Analysis:

- ① use PCA or FA
 - ② How to interpret?
 - ③ How many factors?
- if you want to find latent variable for many variables then use FA,
if you want to eliminate some variables with high variance use PCA.

② Loading:

Factor loading is the correlation coefficient for the variable and factor.

↳ 10 var \rightarrow 3 var

2. How to Interpret?			
	Factor 1	Factor 2	Factor 3
Variable 1	0.7	0.2	0.1
Variable 2	0.4	0.4	0.2
Variable 3	0.7	0.1	0.2
Variable 4	0.1	0.1	0.8
Variable 5	0.6	0.2	0.2
Variable 6	0.5	0.3	0.2
Variable 7	0.7	0.3	0
Variable 8	0.1	0.8	0.1
Variable 9	0.1	0.3	0.6
Variable 10	0.3	0.4	0.3

\Rightarrow variance is denoted.

① Communality (h^2) (1st row)

$$= (0.7)^2 + (0.2)^2 + (0.1)^2$$

② Eigen value (for 1st column)

/

$$\rightarrow (0.7)^2 + (0.4)^2 + (0.2)^2 = \dots$$

⇒ Some times for a particular variable it shows high correlation for more than one factor called as

Cross loading.

in this scenario, Variable rotation should be performed.

③ Sample size

5 observations / variable

e.g.: 10 variables / 50 observations.

No. of factors:

100 variables: 5, 8, 10, ? ... factors

make a scree plot: Bend in plot

Latent Root Criterion:

EigenValues > 1

Basic logic of factor Analysis:

- ① Items you want to reduce
- ② Create mathematical combination to find PC
- ③ New Combination from residual variance -
 Σ nd PC
- ④ Select minimal no. of factors
- ⑤ Interpret the factors

(

SVD: [Singular Value Decomposition]

→ Every $m \times n$ matrix A factors into

$$A_{m \times n} = U_{m \times m} \cdot \Sigma_{m \times n} \cdot V^T_{n \times n}$$

U is a matrix

U is an orthogonal matrix
 Σ is a diagonal matrix
 V^T is again an orthogonal matrix

Operations:

eigen value

$$\Rightarrow A^T \cdot A = (V \Sigma^T U^T) U \Sigma V^T$$

$$\Rightarrow A^T \cdot A = V \cdot \Sigma^T \cdot \Sigma \cdot V^T$$

↑
 ↗ of $A^T A$
 is same as
 σ^2 for A

$$\Rightarrow A \cdot A^T = (U \Sigma V^T) \cdot (V \Sigma^T U^T)$$

$$\Rightarrow A \cdot A^T = U \Sigma \Sigma^T U^T$$

$$A = [U_1 \ U_2] \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix} \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix}$$

↙
this is of $m \times n$ so
it can be a rectangular
matrix as well.

(c)

instead of eigen values
we call them singular
values.

left singular
vectors

right singular vectors ↘

Two geometrical ops in SVD!

- ① Rotation
- ② Stretching

$U = \text{Rotation}$
 $\Sigma = \text{Stretching}$
 $V = \text{Rotation}$

$$A = (R)(S)(R)$$

ex!

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$



in SVD we call it
unitary transformation

ex1

U Σ V^T

$$A = \begin{bmatrix} 2 & 2 \\ 1 & 1 \end{bmatrix} = \sqrt{2} \begin{bmatrix} 2 & 1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} \sqrt{10} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

↓

Orthogonal matrix

$\frac{\sqrt{5}}{\sqrt{2}}$

ex1 People

gene	1	2	3
1	1	0	0
2	0	1	0
3	0	0	1
4	0	0	0

all to Statistics:

→ Values having greatest variance in this table.

all to Normal Eng!

→ the most information

will be achieved from the special matrix $U\Sigma V^T$

whose,

Rank = 1

$U_1 = \text{Comb of people}$

$\Sigma = \text{Biggest number possible}$

$V_1 = \text{Comb of genes}$

Uses: (Linear Algebra)

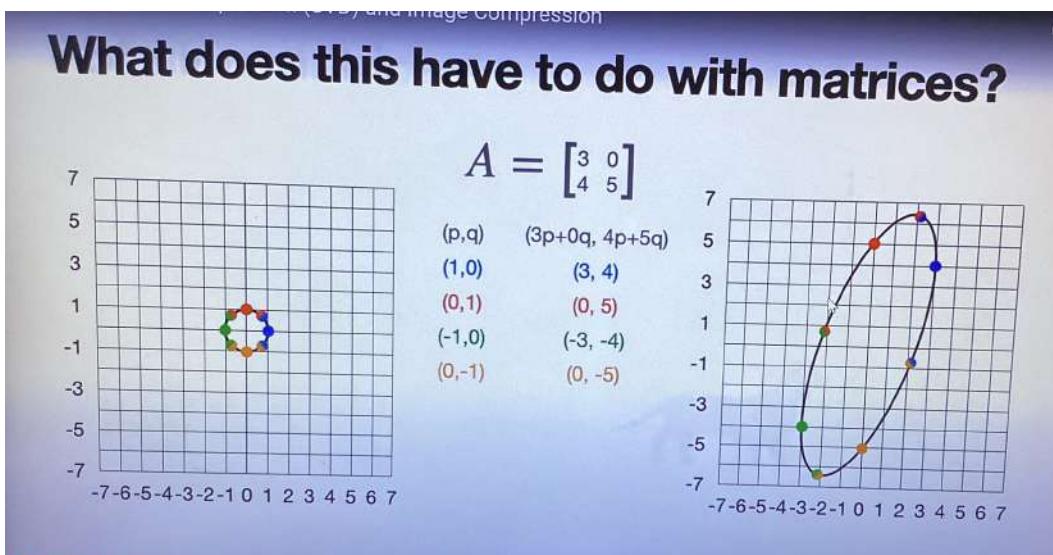
(1) Solve $Ax = b$ for non square A

↳ Linear regression

(2) basis of PCA

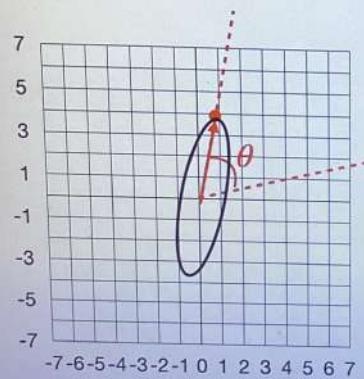
(3) Google, Facebook, Netflix etc---

①



②

Rotation matrices

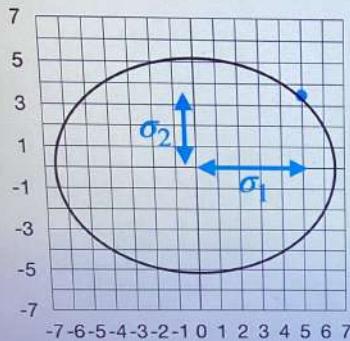


$$\begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$$



3.

Stretching matrices

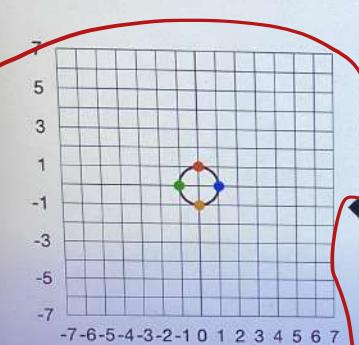


$$\begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix}$$



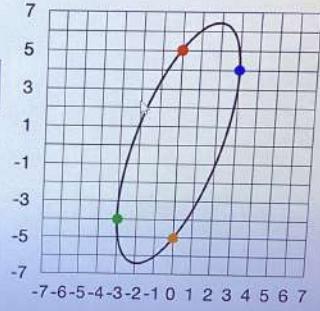
4.

What does this have to do with matrices?



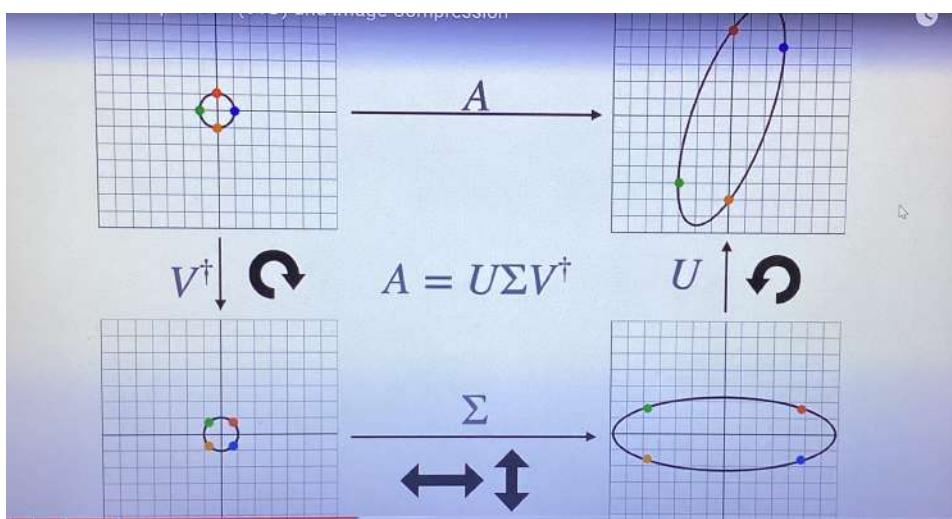
$$A = \begin{bmatrix} 3 & 0 \\ 4 & 5 \end{bmatrix}$$

$$\begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix} \leftrightarrow \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix} \leftrightarrow \begin{bmatrix} \cos(\phi) & \sin(\phi) \\ -\sin(\phi) & \cos(\phi) \end{bmatrix}$$



$$U \cdot \Sigma \cdot V^T = A$$

5.



ex:

$$A = \begin{bmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{bmatrix}_{2 \times 3}$$

$$A^T = \begin{bmatrix} 3 & -1 \\ 1 & 3 \end{bmatrix}_{3 \times 2}$$

$$A \cdot A^T = \begin{bmatrix} 11 & 1 \\ 1 & 11 \end{bmatrix}_{2 \times 2}$$

$$(A - \lambda I) = 0 \quad \text{Characteristic eqn to find ev and } \vec{ev}$$

$$\begin{bmatrix} (11-\lambda) & 1 \\ 1 & (11-\lambda) \end{bmatrix} = 0$$

$$(11-\lambda)^2 - (1)^2 = 0$$

$$(11-\lambda+1)(11-\lambda-1) = 0$$

$$(12-\lambda) = 0 \quad | \quad 10-\lambda = 0$$

$\lambda=12$ $\lambda=10$

These are eigenvalues

→ to find eigen vectors.

$$\lambda_1 = 12 \Rightarrow \begin{pmatrix} (11-\lambda) & 1 \\ 1 & (11-\lambda) \end{pmatrix}$$

$$\Rightarrow \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} \text{ these are the } \underline{\text{ev}}$$

we need to perform orthogonalisation,
with the help of,

Gram-Schmidt Orthogonalization

process.

note:

from the formula $A = U \Sigma V^T$

U and V^T need to be orthogonal matrix,
to get them we are performing
orthogonalisation.

$$\begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} \rightarrow \sqrt{(-1)^2 + (1)^2} = \sqrt{2}$$
$$\begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} \rightarrow \sqrt{(1)^2 + (-1)^2} = \sqrt{2}$$

$$\begin{pmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix} = U$$

note:

$$A^T \cdot A = V$$

$$A \cdot A^T = U$$

$$A^T \cdot A = \begin{pmatrix} 10 & 0 & 2 \\ 0 & 10 & 4 \\ 2 & 4 & 2 \end{pmatrix}$$

$$(A - \lambda I) X = 0$$

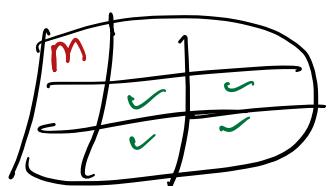
$$\begin{pmatrix} (10-\lambda) & 0 & (2) \\ 0 & (10-\lambda) & 4 \\ 2 & 4 & (2-\lambda) \end{pmatrix} = 0$$

$$\lambda^3 - s_1 \lambda^2 + s_2 \lambda - s_3 = 0$$

$$\Rightarrow S_1 = \text{trace}(A) \Rightarrow 10 + 10 + 2 = 22$$

$$\Rightarrow S_3 = \det(A) \Rightarrow 10(4) + 2(-20) = 0$$

$$\Rightarrow S_2 = \text{minor of } \begin{matrix} & \\ \swarrow & \text{diagonals} \end{matrix} \Rightarrow 4 + 16 + 100 = 120$$



$$\lambda^3 - S_1\lambda^2 + S_2\lambda - S_3$$

$$\Rightarrow \lambda^3 - 22\lambda^2 + 120\lambda = 0$$

$$\lambda(\lambda^2 - 22\lambda + 120) = 0$$

$$\lambda(\lambda^2 - 10\lambda - 12\lambda + 120) = 0$$

$$\lambda(\lambda(\lambda - 10) - 12(\lambda - 10)) = 0$$

$$\lambda(\lambda - 10)(\lambda - 12) = 0$$

$$\Rightarrow \lambda = 0, 10, 12$$

$$U = \{U_1, U_2, U_3\}$$

\Rightarrow Row matrix

$$V^T = \begin{bmatrix} V_1^T \\ V_2^T \\ V_3^T \end{bmatrix} \Rightarrow \text{Column matrix}$$

DON'T
FORGET

Coeff: $\lambda_1 = 12$

$$\Rightarrow \begin{bmatrix} -2 & 0 & 2 \\ 0 & -2 & 4 \\ 2 & 4 & -10 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 0$$

Crammers rule

$$\frac{x_1}{\begin{vmatrix} -2 & 4 \\ 4 & -10 \end{vmatrix}} = \frac{-x_2}{-8} = \frac{x_3}{4}$$

$$\Rightarrow \frac{x_1}{1} = \frac{-x_2}{2} = \frac{x_3}{1}$$

$$\Rightarrow V^2 \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}$$

Cole 2

$$\lambda = 10$$

$$\begin{pmatrix} 0 & 0 & 2 \\ 0 & 0 & 4 \\ 2 & 4 & -8 \end{pmatrix}$$

$$\frac{x_1}{-16} = \frac{-x_2}{-8} = \frac{x_3}{0}$$

$$\Rightarrow \frac{x_1}{2} = \frac{-x_2}{1} = \frac{x_3}{0}$$

$$V^2 \begin{pmatrix} 1 & 2 \\ 2 & 1 \\ 1 & 0 \end{pmatrix}$$

Cole 2 $\lambda = 0$

$$\begin{pmatrix} 10 & 0 & 2 \\ 0 & 10 & 4 \\ 2 & 4 & 2 \end{pmatrix}$$

$$\frac{x_1}{4} = \frac{x_2}{-8} = \frac{x_3}{-20}$$

$$\frac{x_1}{1} = \frac{-x_2}{\pm 2} = \frac{x_3}{-5}$$

$$V = \begin{pmatrix} 1 & 2 & 1 \\ 2 & 1 & 2 \\ 1 & 0 & -5 \end{pmatrix}$$

$$\Rightarrow V^T = \begin{pmatrix} 1 & 2 & 1 \\ 2 & 1 & 0 \\ 1 & 2 & -5 \end{pmatrix}$$

now Gram-Schmidt process we calculate orthogonal matrix.

$$V^T = \left(\begin{array}{ccc} \sqrt{1^2 + 2^2 + 1^2} = \sqrt{6} & & \\ & \sqrt{2^2 + 1^2 + 0^2} = \sqrt{5} & \\ & & \sqrt{1^2 + 2^2 + (-5)^2} = \sqrt{30} \end{array} \right)$$

$$= \begin{pmatrix} \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} & 0 \\ \frac{1}{\sqrt{30}} & \frac{2}{\sqrt{30}} & \frac{-5}{\sqrt{30}} \end{pmatrix}$$

$$\text{and, } A = [2 \times 3] \Rightarrow \Sigma = [2 \times 3]$$

observing, I^{U} and N we can see $(12, 10)$

are the most repeated values.



higher value
lower value