

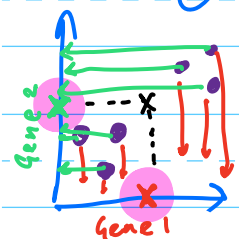
Principal Component Analysis :

Steps:

① Standardize the data

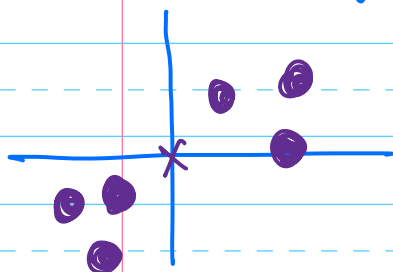
- all features have mean = 0
and variance = 1

to ensure each feature contributes equally to analysis.



Average measurement of each q1 & q2 avg value → centre of data

• shift data to the origin (centre of data)



② Calculate Covariance Matrix

cv b/w 2 variables tells how they change together

relation b/w different variables

features = $n \times n$ matrix

$$\text{matrix} = \begin{bmatrix} \text{cov}(q_1, q_1) & \text{cov}(q_1, q_2) \\ \text{cov}(q_2, q_1) & \text{cov}(q_2, q_2) \end{bmatrix}$$

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

③ Calculate eigen value & eigen vectors

find $\lambda_1, \lambda_2, v_1, v_2$

dir → max variance
mag in the dir of max variance

$$\det(A - \lambda I)v = 0$$

$$\det\left(\begin{bmatrix} 3 & 1 \\ 2 & 8 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}\right) = 0 \Rightarrow \det\left(\begin{bmatrix} 3-\lambda & 1 \\ 2 & 8-\lambda \end{bmatrix}\right) = 0$$

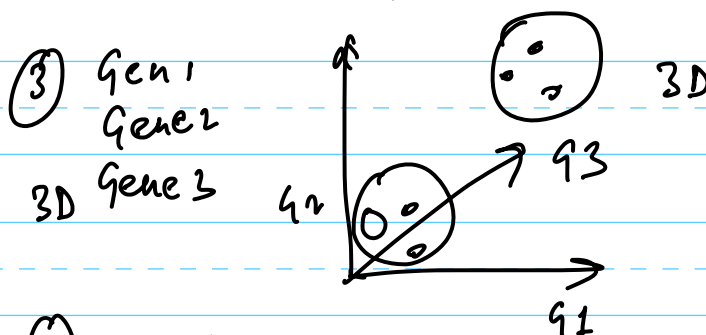
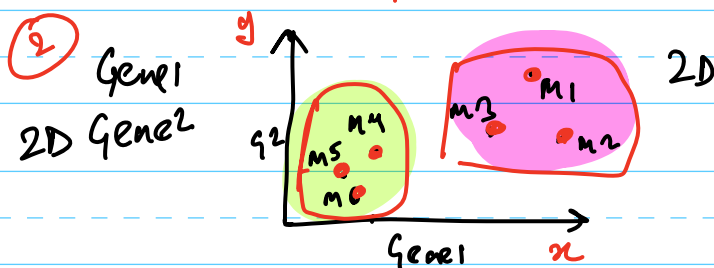
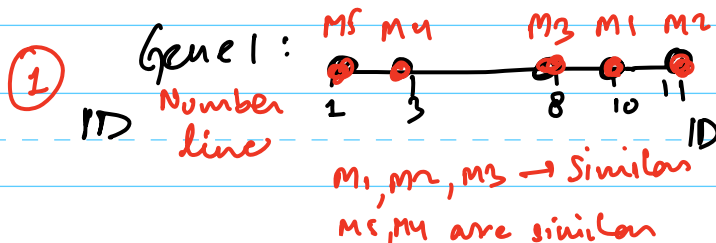
λ_1, λ_2 from here

variables

samples

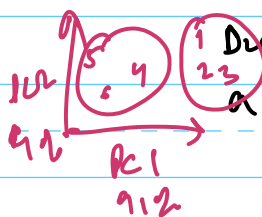
	1	2	3	4	5	...
Math	95	85	93	75
Reading	96	79	98	84

	M1	M2	M3	M4	M5	...
Gene1	10	11	8	3	1	...
Gene2	6	4	5	3	2	...



④ Gene1 4D 3 we can't plot data

DCA → can take 4 or more
Dim data and can make
a 2D PCA plot.



eigen vector: v

$$Av = \lambda v$$

eigen value: λ

A = matrix (square)

for any square matrix A , if there is a vector v , such that

$$Av = \lambda v \rightarrow \text{eigen vector} \\ \hookrightarrow \text{eigen value}$$

then

$$(A - \lambda I) v = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$(A - \lambda I) \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

PC1 \Rightarrow highest eigen value

PC2 \rightarrow smaller eigen value