PCA (Principal Component Analysis)
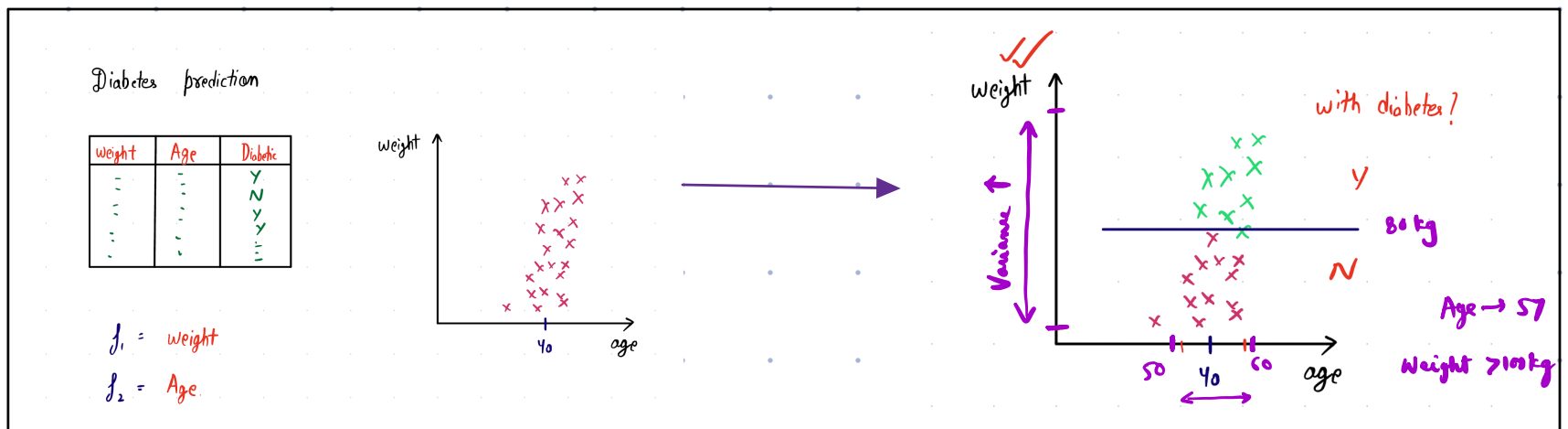
1. Recap of Dimensionality Reduction
2. Properties of PC
3. Mathematical Intuition
4. Code Implementation
5. Summary

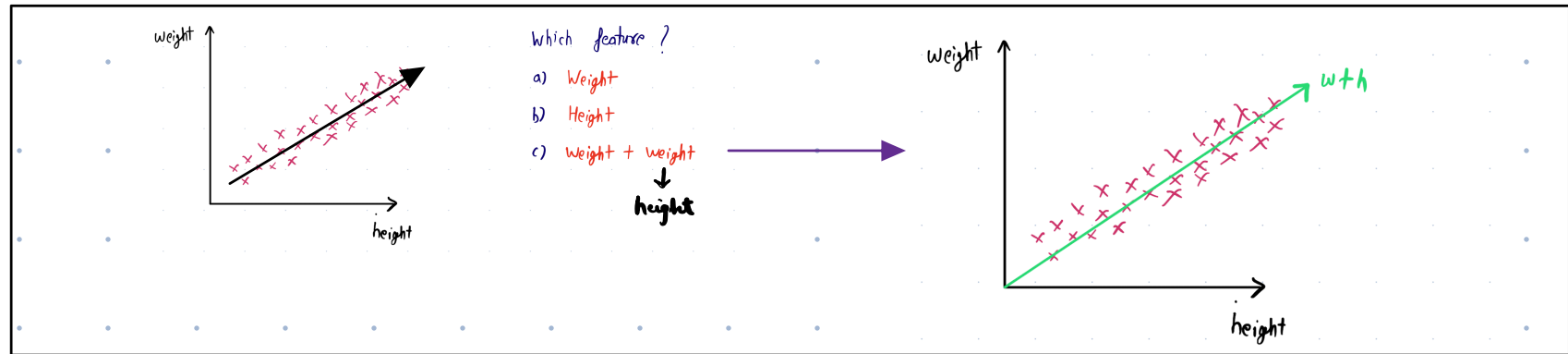## Data of Patient's 'Weight' and 'Age'

Example-1.
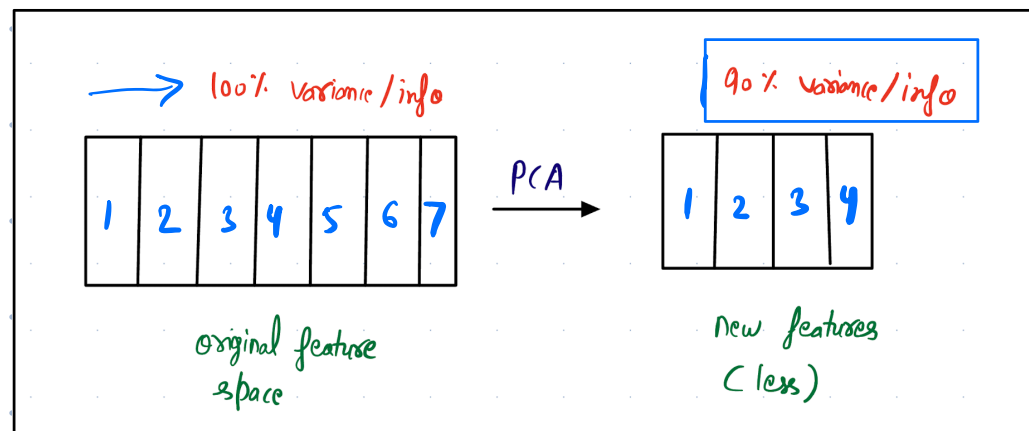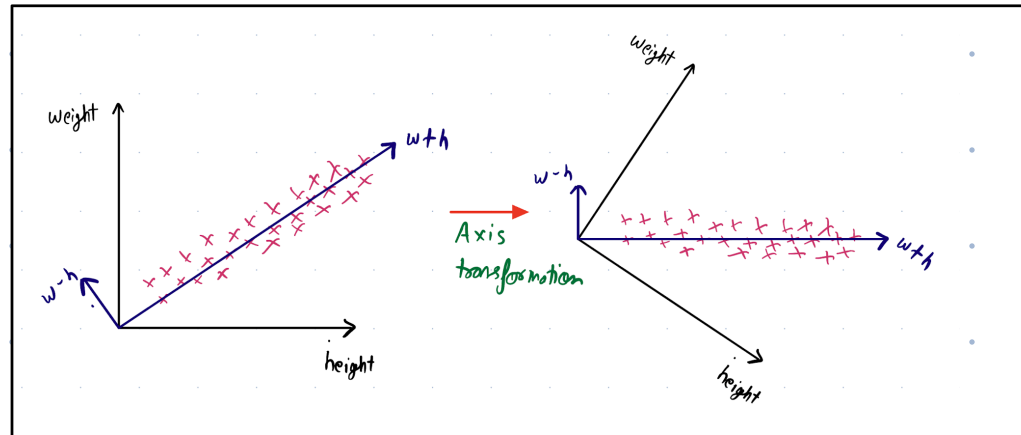

'Weight' and 'Height'

↓

Which feature?
a) Weight
b) Height
c) Weight + weight → height

•AXIS TRANSFORMATION

Axis transformation

100% variance/info

90% variance/info

PCA

original feature space

new features (less)
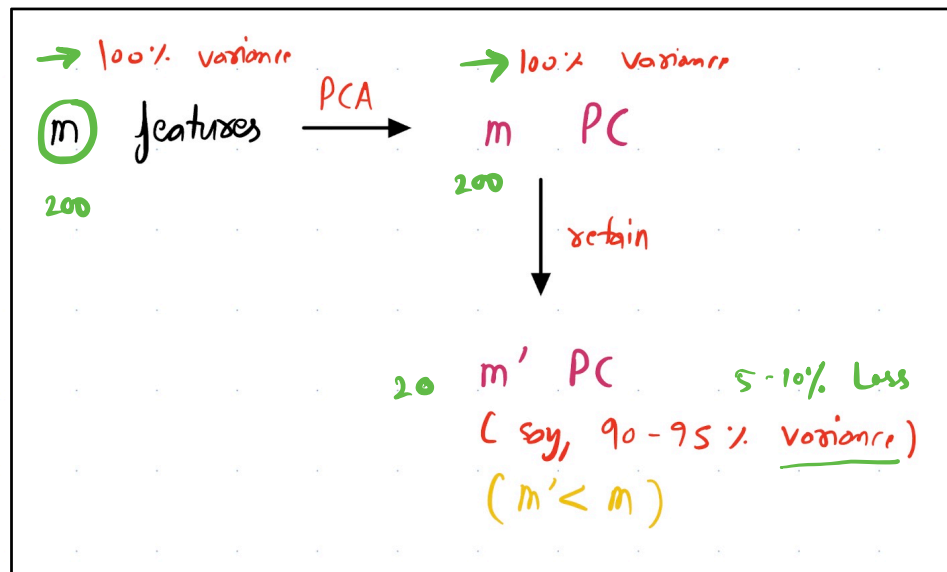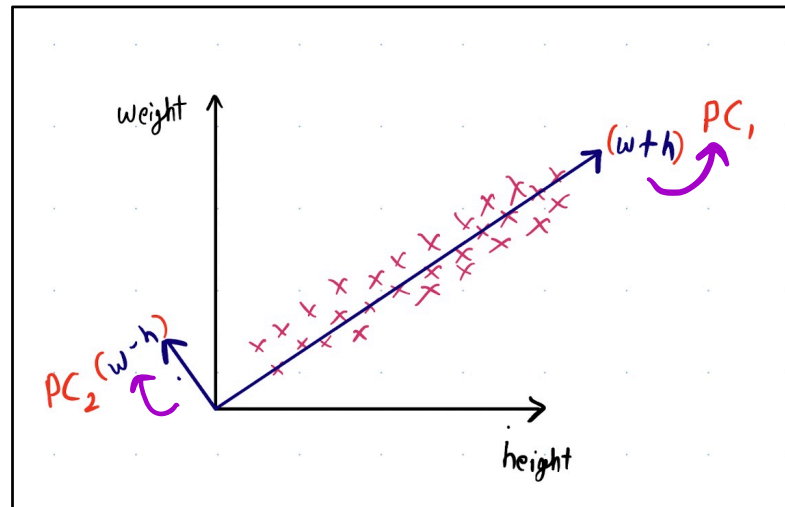
Against a small trade-off of information loss:

① we get reduced number of features

② these new less features are called as 'Principal Components'

- Properties of PC :

  ① They are perpendicular / orthogonal to each other.

  ② Number of PCs is equal to Number of features present.

- Mathematical Intution of PC :

  - Eigen Value

  - Eigen Vectors

Note : Our direction of max. variance can be
found out by calculating the 'eigen vector'

$$A = \begin{bmatrix} -5 & 2 \\ -7 & 4 \end{bmatrix}_{2 \times 2}$$

$$\boxed{\det(A - \lambda I) = 0}$$

$$+2 \rightarrow \begin{bmatrix} 2 \\ 7 \end{bmatrix}$$

$$7x - 2y = 0$$

$$\det\left(\begin{bmatrix} -5 & 2 \\ -7 & 4 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}\right) = 0$$

$$\boxed{A x = \lambda x}$$

Eigen values $\longrightarrow \lambda_1 = +2$, $\lambda_2 = -3$

$$\boxed{\begin{array}{c}(\lambda I - A) x = 0 \\ x = \begin{bmatrix} x \\ y \end{bmatrix}\end{array}}$$

- Information loss in PCA

Original Data

5 dim $\xrightarrow{PCA}$ 5 PCs

(5 features)

$\rightarrow \underline{10}$
$\rightarrow \underline{7}$
$\rightarrow \underline{6}$ } eigen values
$\rightarrow \underline{3}$
$\rightarrow \underline{1}$

$$\left(\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} - \begin{bmatrix} -5 & 2 \\ -7 & 4 \end{bmatrix}\right)\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} 7 & -2 \\ 7 & -2 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} 7x - 2y \\ 7x - 2y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$
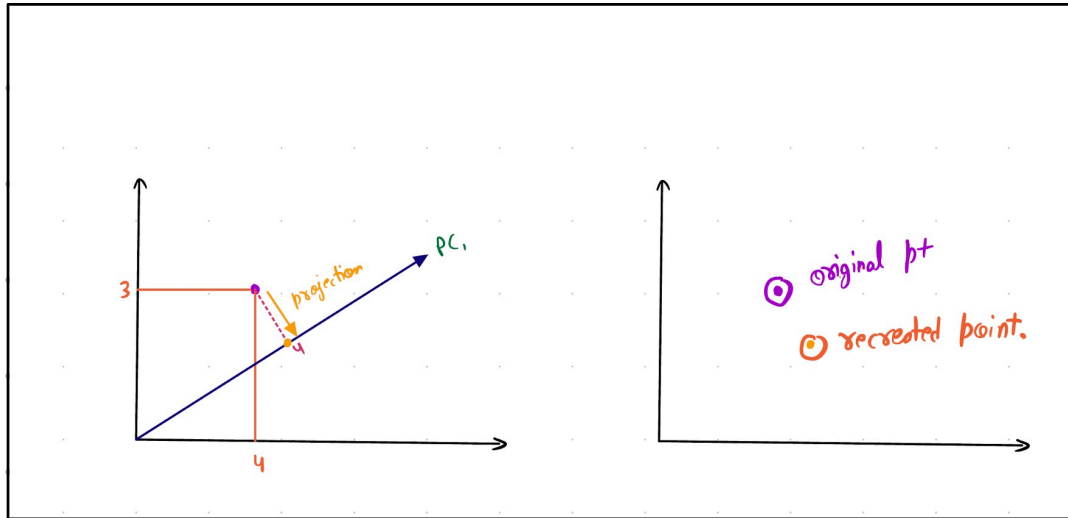
$$\rightarrow 7x - 2y = 0$$

$$y = \frac{7x}{2} \leftarrow$$

$$\begin{bmatrix} 1 \\ \frac{7}{2} \end{bmatrix} - \quad x = 1$$

$$\hookrightarrow \begin{bmatrix} 2 \\ 7 \end{bmatrix}$$

PC₁

projection

3

4

PC₁

⦿ original pt

⊙ recreated point.

$$\text{explained variance ratio} = \frac{10+7+6}{10+7+6+3+1} = \frac{23}{27} \simeq 0.85$$

$$= 85\%$$

$\Rightarrow$ 85% of variance is preserved by top 3 PCs.

5 PCs $\longrightarrow$ 3 PCs

(Reduced the Dimension)