☆ PCA - Principal Component Analysis - is used to reduce dimensions

☆ Why to reduce dimesions? (Dimensionality Reduction)

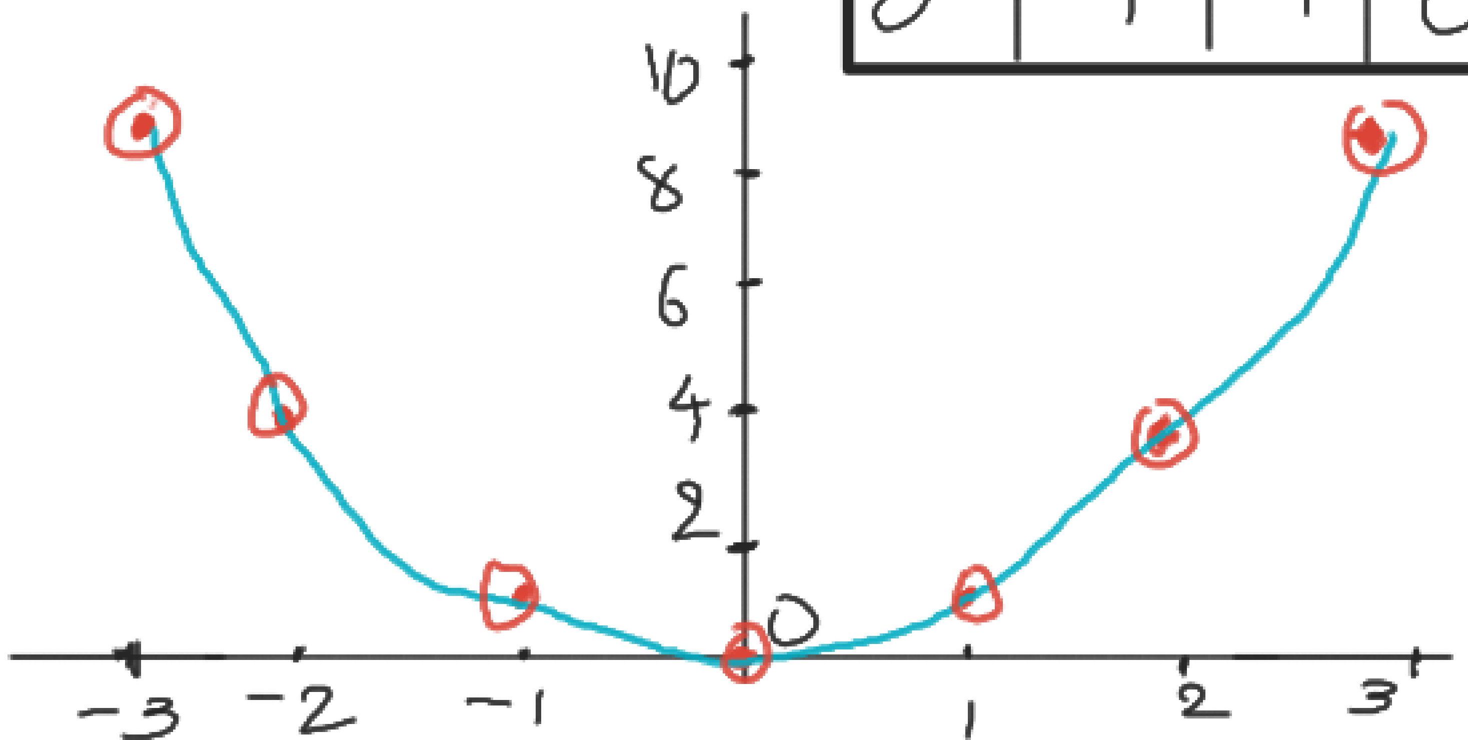
→ Reduce time

→ Calculation becomes simple

→ curse of dimensionality

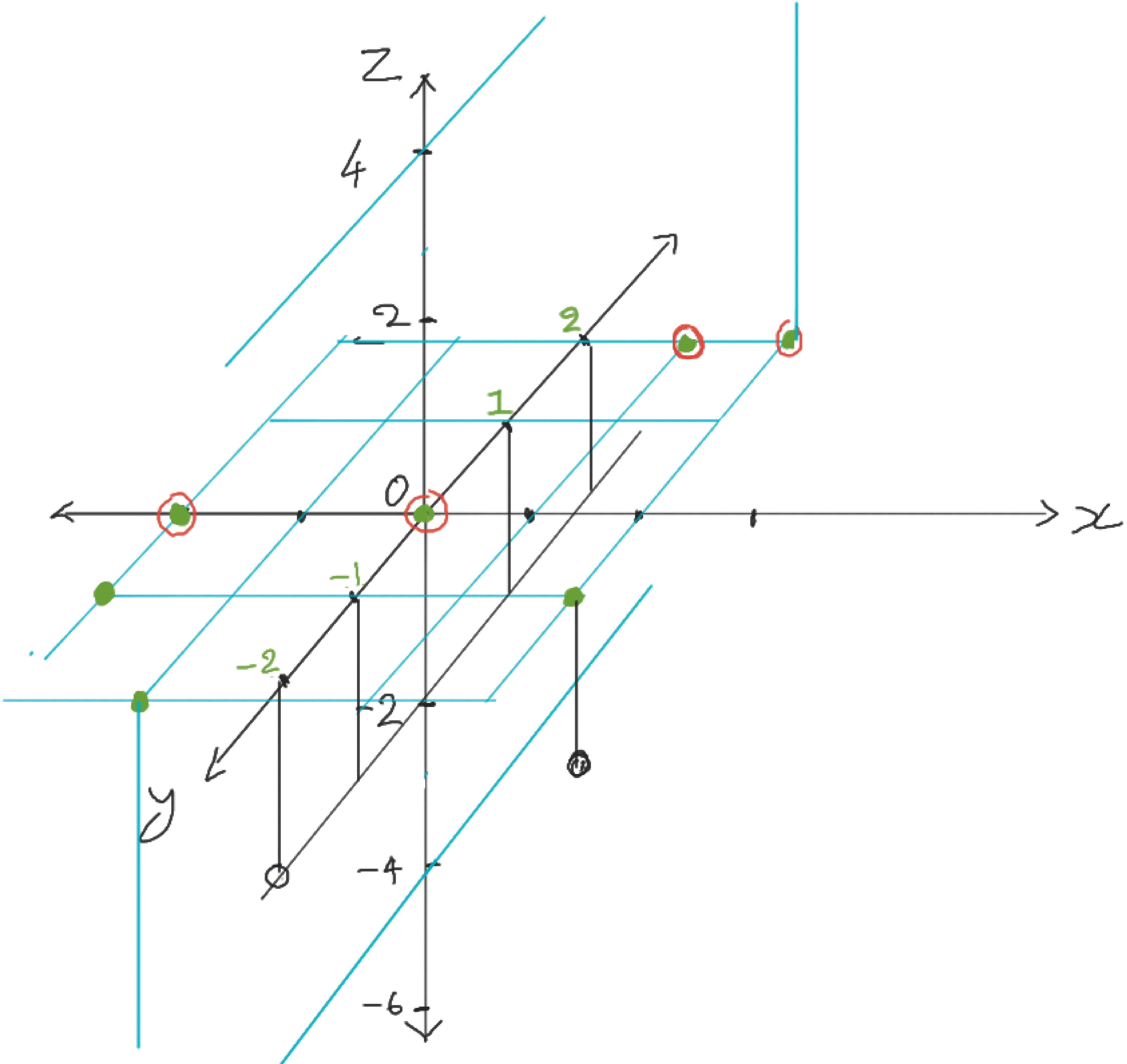
$f(x) = x^2$

x	-2	-1	0	1	2
y	4	1	0	1	4



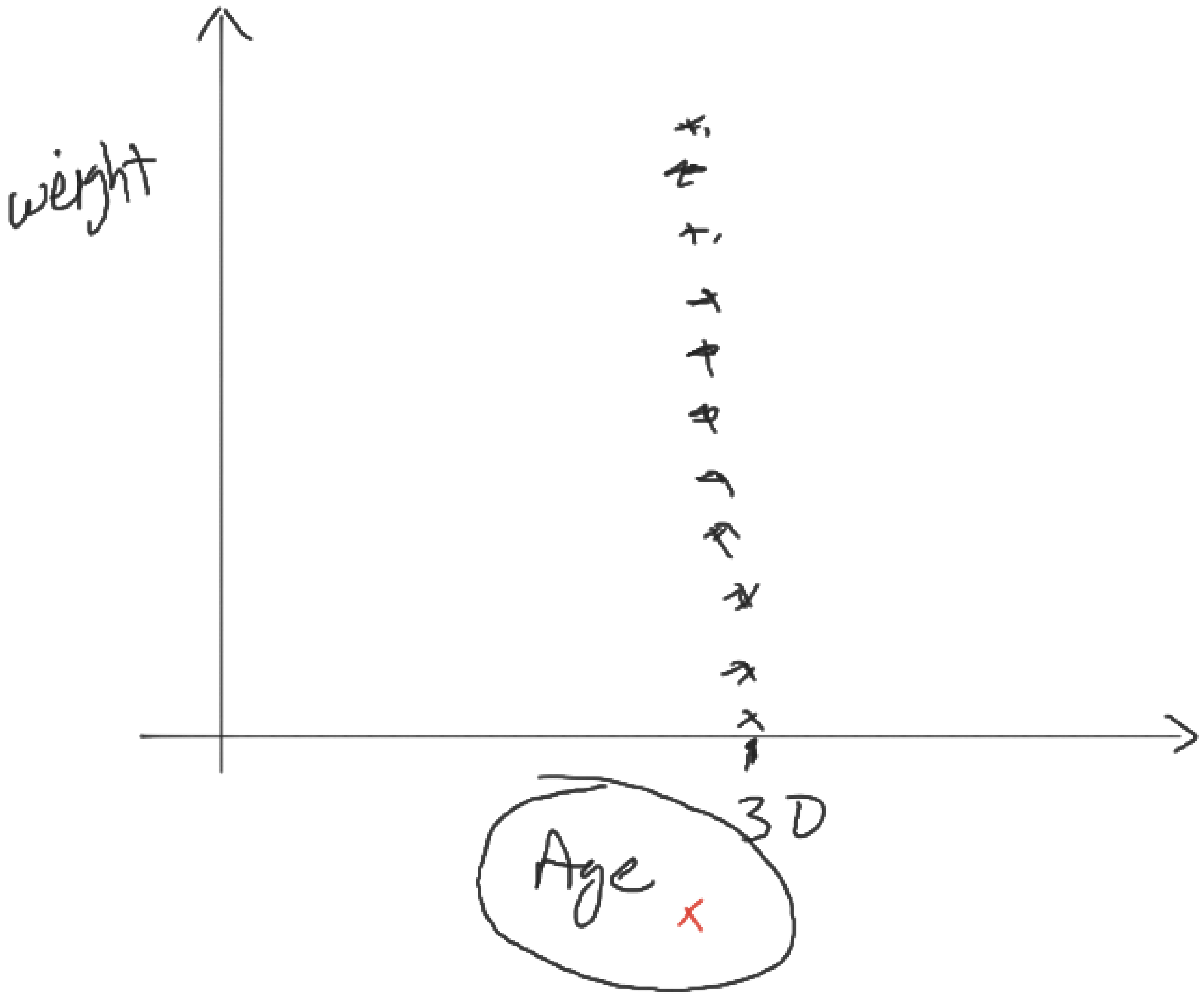
$g(x) = x^2y - xy$

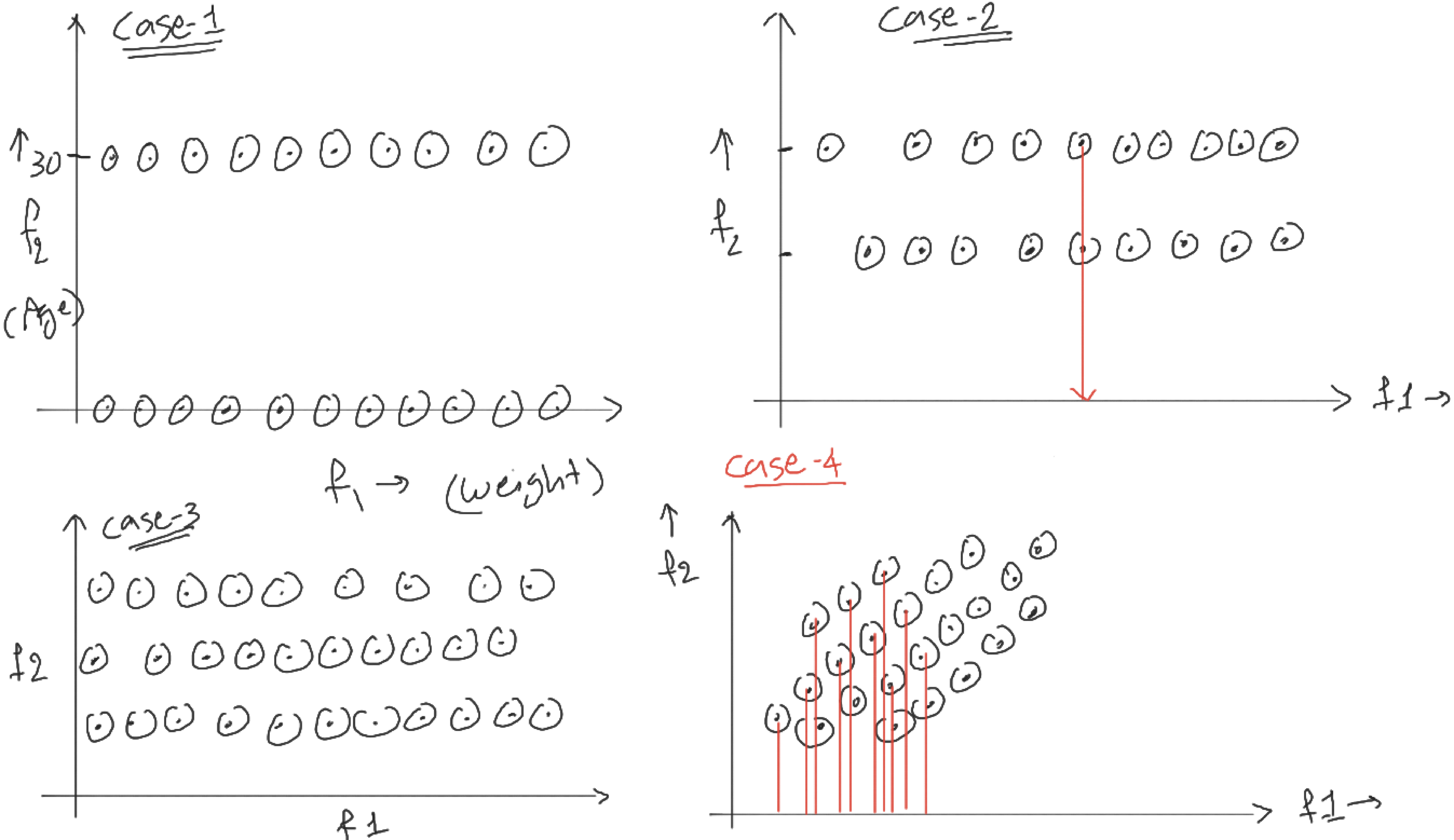
x	-2	-1	0	1	2	-2	2
y	-1	-2	0	2	2	0	-1
z	-6	-4	0	0	4	0	-2



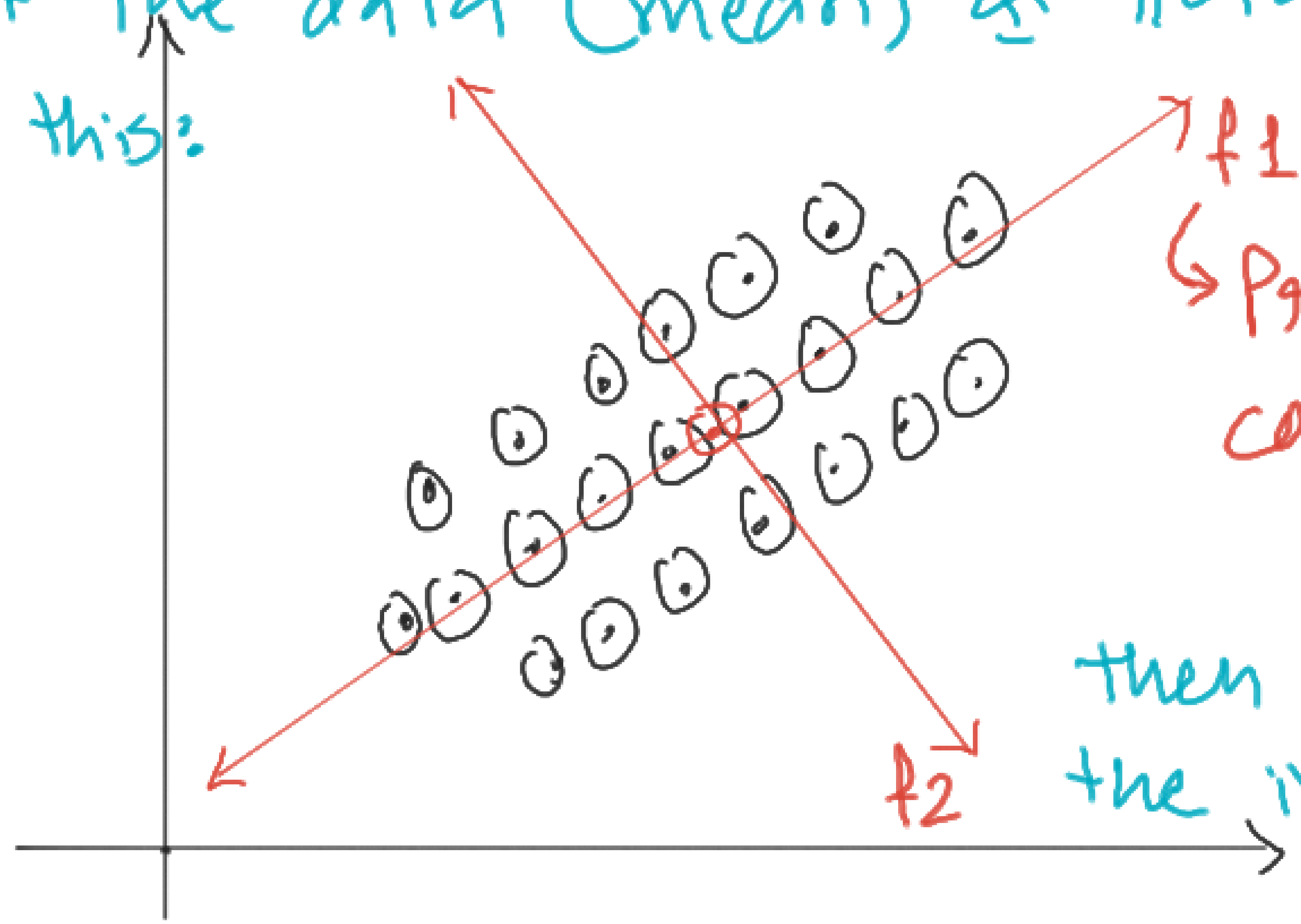
z
4
2
-2
-4
-6
x
y
0
1
2
-1
-2

Weight	Age	diabetes
80	30	1
60	30	0
55	30	1
65	30	0
70	30	1
68	30	1
72	30	0
66	30	0
79	30	0




case-1
30
f_2
(Age)
$f_1 \rightarrow$ (weight)
case-2
f_2
$f1 \rightarrow$
case-3
$f2$
$f1$
case-4
$f2$
$f1 \rightarrow$

If we shift our origin on the center of the data (mean) & rotate the axis as this:



then we can minimize the information loss.

$$= \frac{x - \bar{x}}{\sigma}$$

20,000

$\mu =$

z

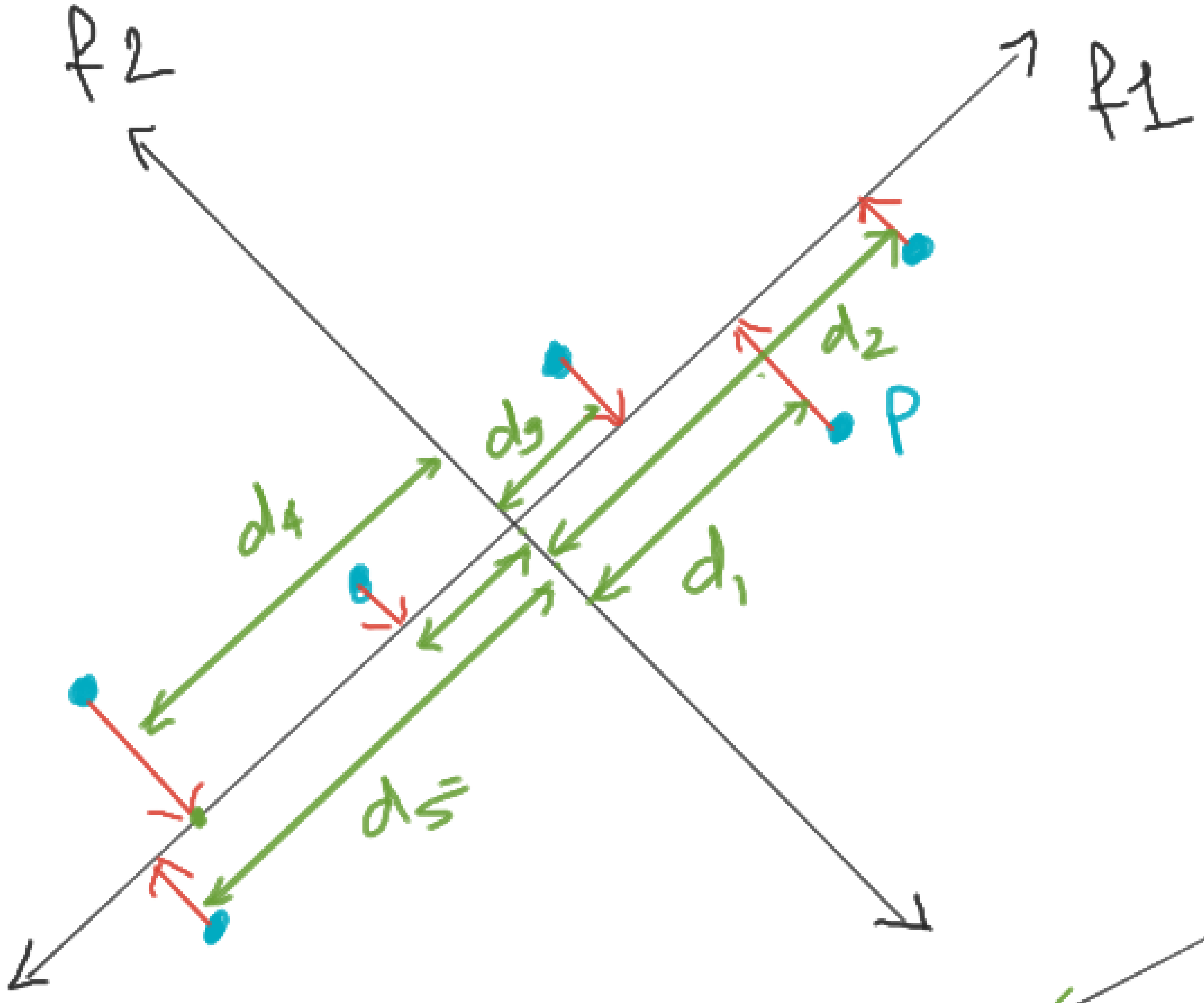
0 1 2

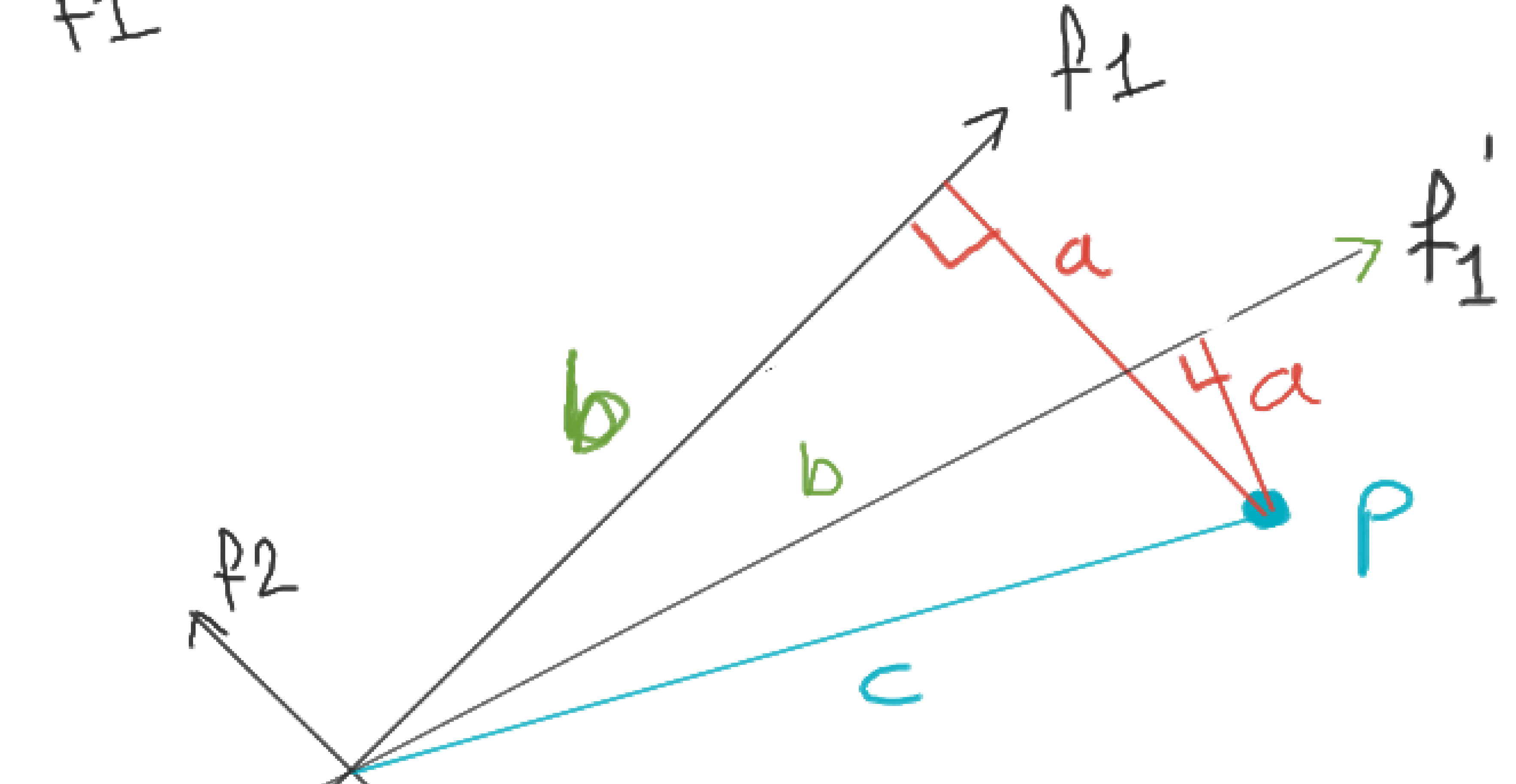
x

25000
24000
18000
16000
20000

How to shift origin?

Ans = normalization | standardization





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

↓ ↓ ↑

? dec fix

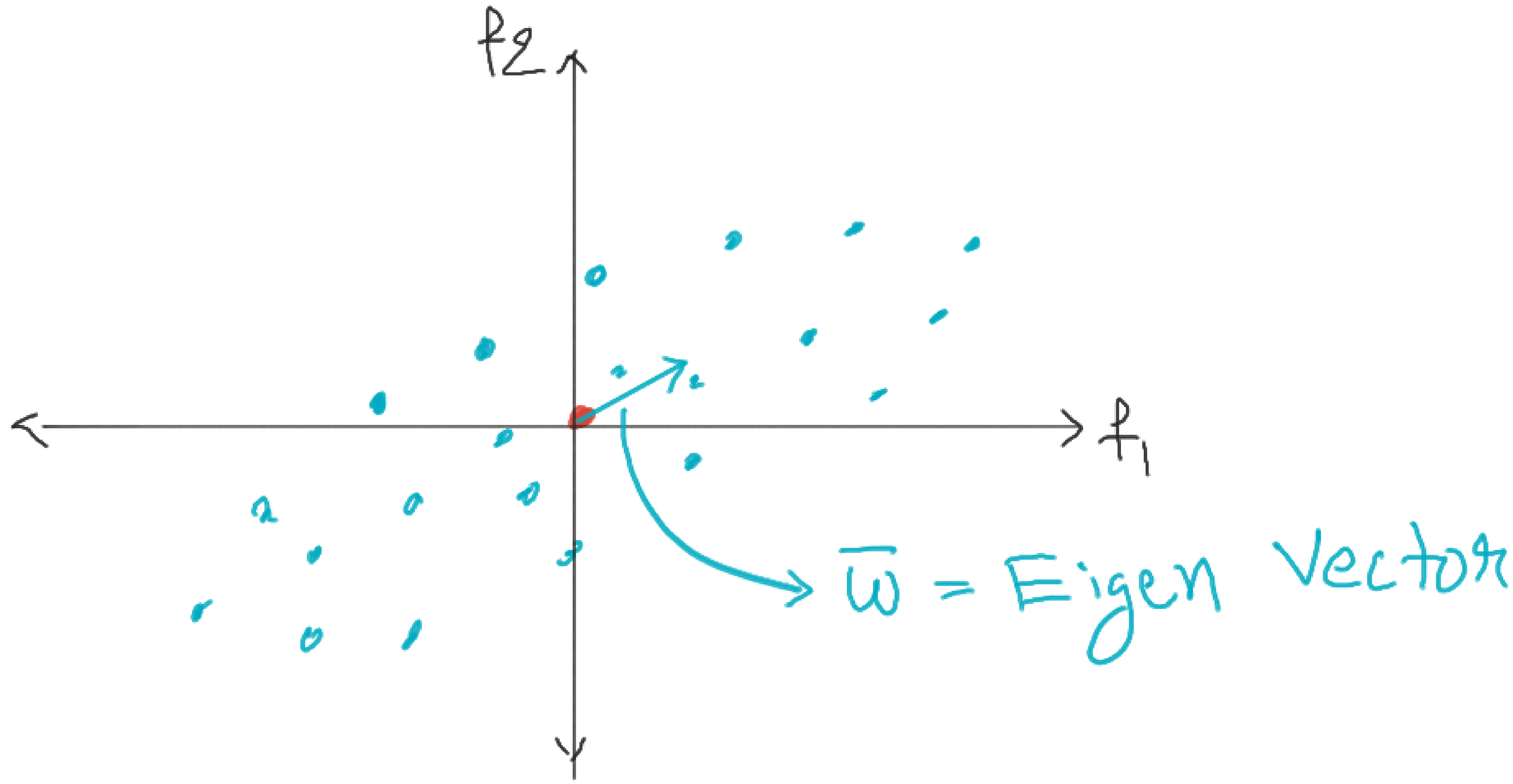
$$(\ )^2 + (4)^2 = (5)^2$$

$$(\ )^2 + (3)^2 = 25$$

$$d_1^2 + d_2^2 + d_3^2 + \dots + d_n^2 = ss(\text{distance})$$

(sum of square distance)

(Eigen Value)



f2
f1
$\overline{w}$ = Eigen Vector

d_1^2 = distance of x_1 from mean $(\bar{x})$ = $(x_1-\bar{x})^2$

d_2^2 = " " x_2 " " = $(x_2-\bar{x})^2$

⋮

d_n^2 = " " x_n " " = $(x_n-\bar{x})^2$

$$\sum \frac{(x_i-\bar{x})^2}{d.o.f.} = \text{variance} = \frac{ss(\text{distance})}{d.o.f.} = \frac{ss(\text{distance})}{(n-1)}$$

★ So, the procedure is : (1) Standardize the data (as above)

In general, $\dfrac{x-\mu}{\sigma}$

(2) Finding a line $(\bar{\omega})$ s.t. variance along that direction is maximum.

converting into mathematical form:

variance $$= \frac{1}{(n-1)} \sum (x_i - \mu)^2$$

$$x_i = x_i \cdot \hat{\omega}$$

$$x_i = \frac{x_i \cdot \bar{\omega}}{\|\bar{\omega}\|}$$

$$\underset{\bar{\omega}}{\operatorname{argmax}} \frac{1}{(n-1)} \sum \left(\frac{x_i \cdot \bar{\omega}}{\|\bar{\omega}\|} - \mu \right)^2$$

Here, μ will become 0 as we have standardize the data. hence, if we take $\|\bar{\omega}\| = 1$ then our calculation will become simple

$$\therefore \underset{\bar{\omega}}{\operatorname{argmax}} \frac{1}{(n-1)} \sum (x_i, \bar{\omega})^2 \text{ s.t. } \|\bar{\omega}\| - 1 = 0$$

Converting it into an uncontrained problem using Lagrange's multiplier:

$$\underset{\bar{\omega}}{\operatorname{argmax}} \frac{1}{(n-1)} \sum (x_i \cdot \bar{\omega})^2 + \lambda (\|\bar{\omega}\| - 1)$$

★ We can also write our constraint $\|\bar{\omega}\| = 1$ as $\|\bar{\omega}\|^2 = 1$

∴ The formula will become:

$$\boxed{\bar{\omega} = \underset{\bar{\omega}}{\operatorname{argmax}} \frac{1}{(n-1)} \sum (\bar{x}_i \cdot \bar{\omega})^2 + \lambda \left(\|\bar{\omega}\|^2 - 1\right)}$$

→ The objective of P.C.A.

Explanation of the previous formula:

$$\begin{bmatrix} \cdots & x_1 & \cdots \\ \cdots & x_2 & \cdots \\ \cdots & x_3 & \cdots \\ & \vdots & \\ \cdots & x_n & \cdots \end{bmatrix} \cdot \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ \vdots \\ w_d \end{bmatrix}$$

$$n \times d \qquad d \times 1 \quad = n \times 1$$

☆ Finding "best $\bar{w}$" using gradient descent:

As we know (from the last pdf): $p^2 = p^T \cdot p$

As well as usually $n-1 \approx n$ as no. of datapoints will be large

$$\left[\frac{1}{n}\sum(x_i \cdot w)^T \cdot (x_i \cdot w)\right] + \lambda(w^T \cdot w - 1)$$

Now, differentiating w.r.t. $\bar{w}$: (Also taking $\bar{x_i}$ as X)

$$\frac{1}{n}\sum \frac{\partial}{\partial w}\left[w^T \cdot X^T \cdot X \cdot w\right] + \lambda \frac{\partial}{\partial w}(w^T \cdot w - 1)$$

A fact: $(a \cdot b)^T = b^T \cdot a^T$

☆ __2nd fact:__ $f(w) = w^T \cdot S \cdot w \Rightarrow \frac{\partial}{\partial w} f(w) = (S + S^T) \cdot \bar{w}$

$x^T \cdot x \cdot w = \lambda' \cdot w$ where $\lambda' = -n\lambda$

$\lambda' \rightarrow$ Eigen value

$w \rightarrow$ Eigen vector