

KNN - K-Nearest Neighbors

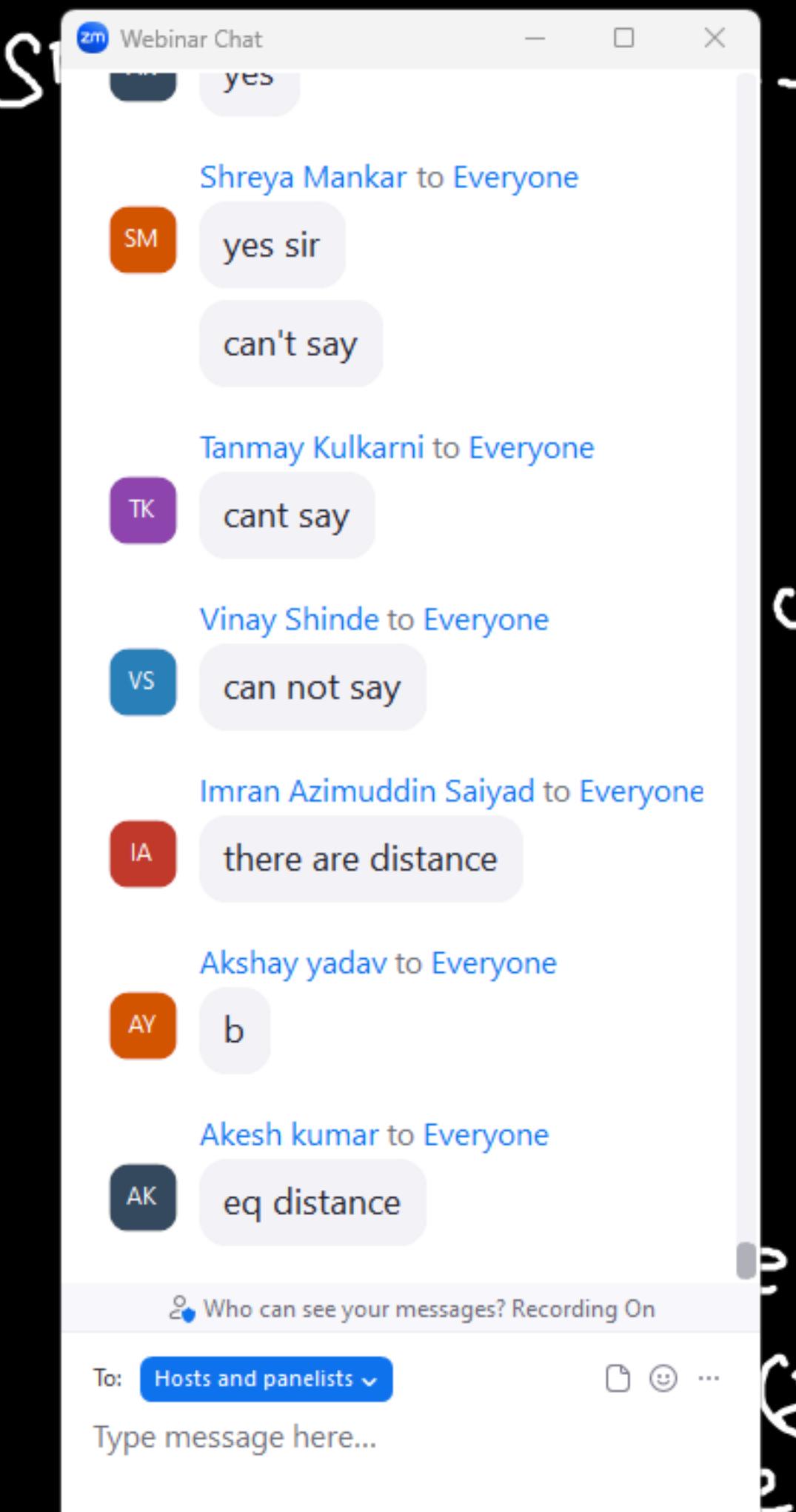
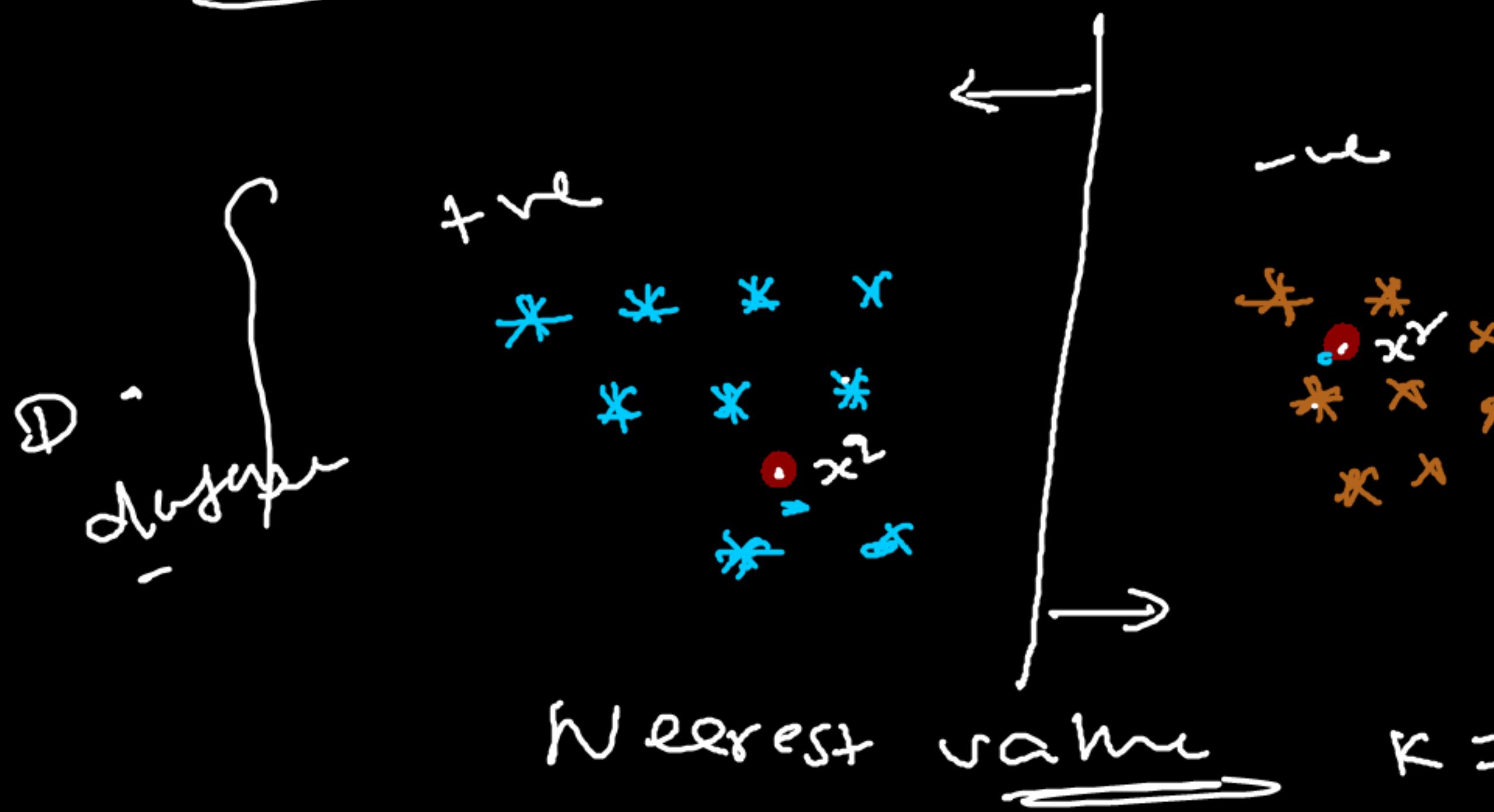
Supervised

ML

Domain

Regression → pls don't use
Classification ↗

Dataset - binary class { No
Yes }



el - closest

step

or (-)ve

or (-)ve

✓

nearest

can say x^1
 x^2 & x^3
the classes

Rule 1 :- Find the "K" nearest points to x^1 & x^2 in Dataset D.

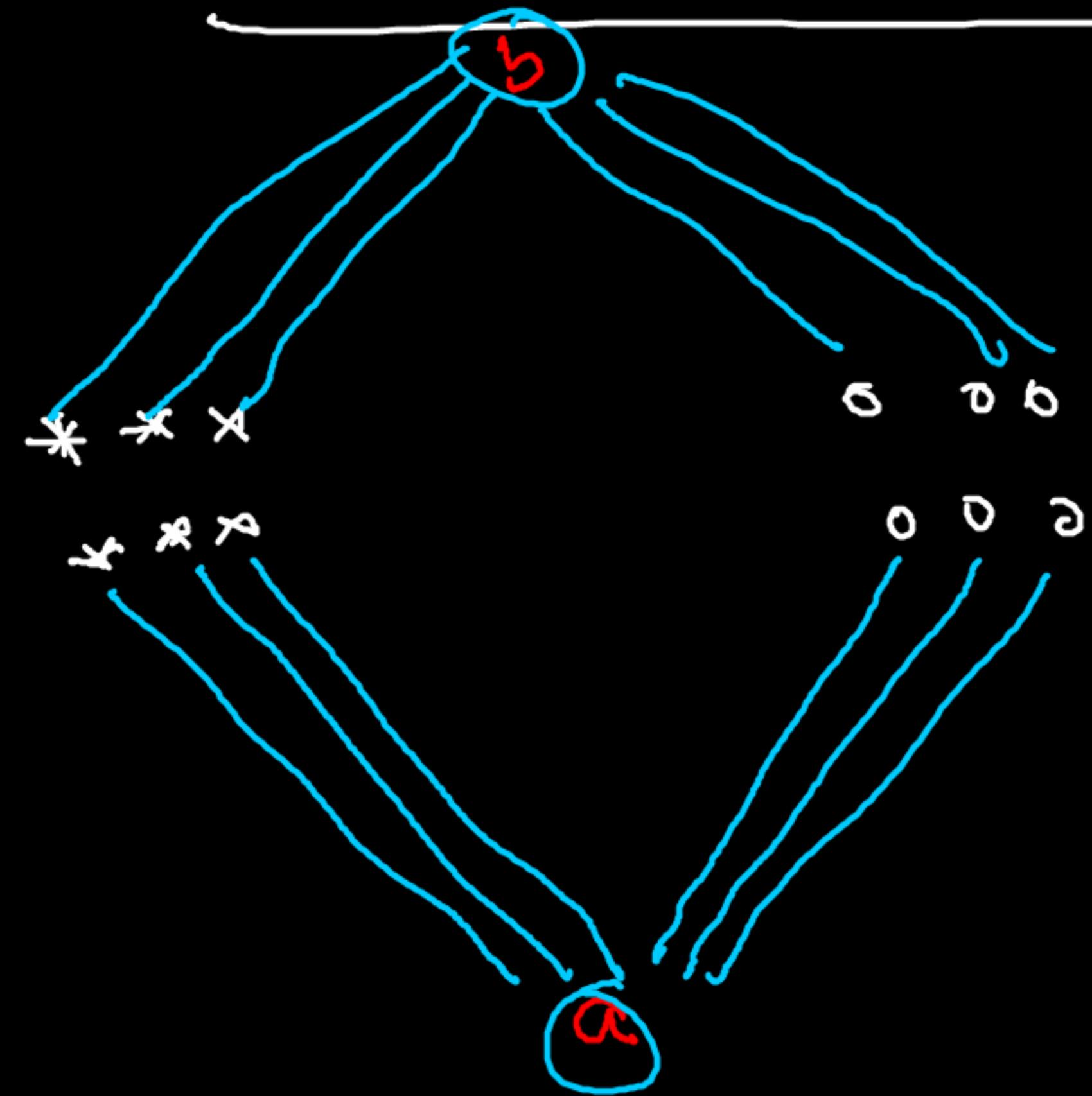
Rule 2 :- $\{y_1, y_2, y_3, \dots\}$ - Majority Vote

NOTE → Please ensure 'K' value always "Odd"

Even
 $K=4$ $\begin{cases} 2 \text{ Yes} \\ 2 \text{ No} \end{cases}$

Odd
 $K=5$ $\begin{cases} 4 = 5, 4, 3, 2, 1 \\ N=0, 1, 2, 13, 4, 5 \end{cases}$

failure cases of kNN



a & b are
case I :- far away from

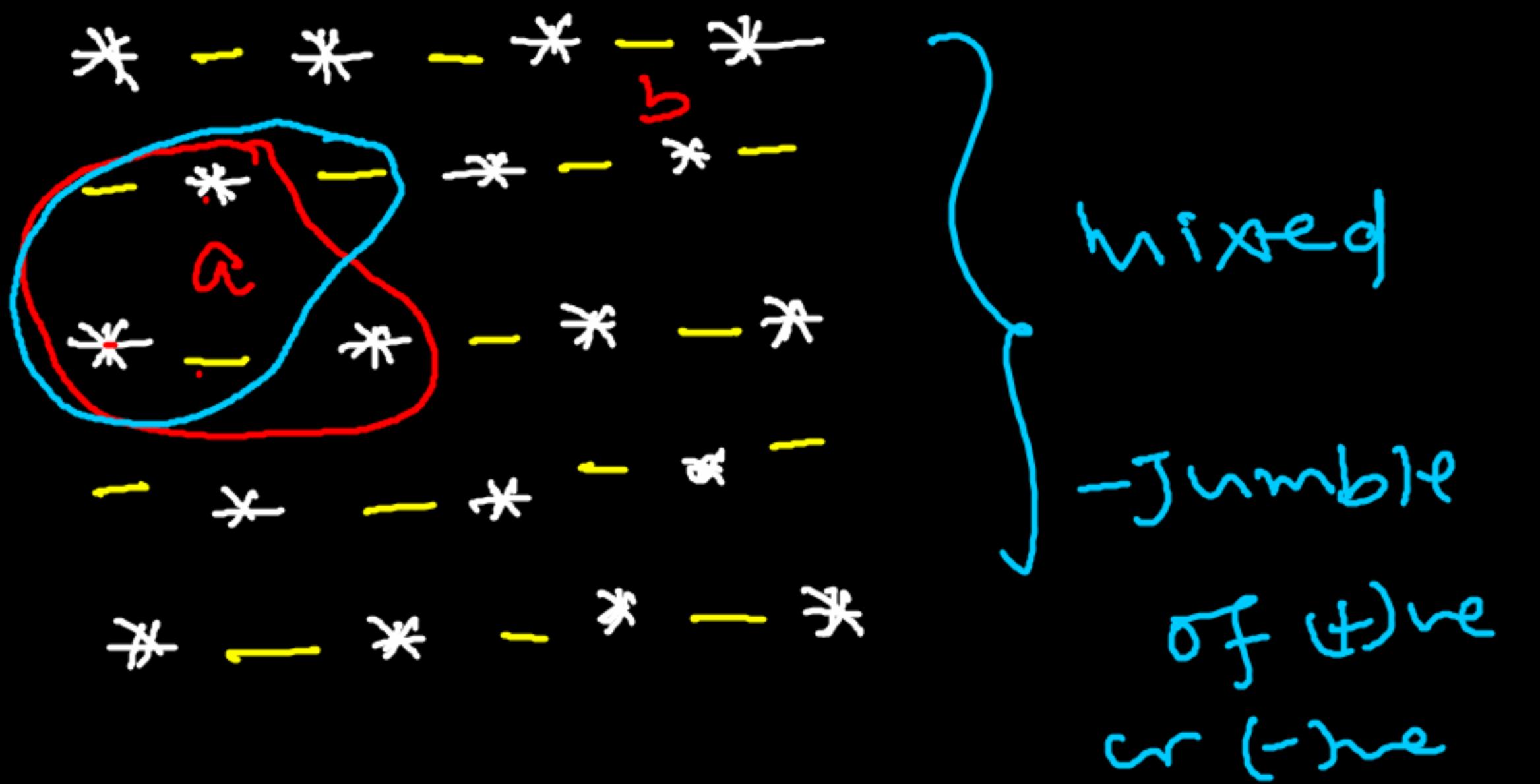
point in dataset D.

→ want - Nearest
Neighbour

You are screen sharing

Stop Share

Case II :- Not useful information



$$r = 5 \quad (+) = 3^2$$

$$a = (-) = 2^3$$

KNN - Distance matrix

by default

* Distance Measures : Euclidean distance (L2 Norm),

Manhattan Distance (L1 Norm)

by default Minkowski dist (L ϕ Norm)

Hammng dist

Chebyshev dist

String/Char

Cosine-dist, Mahalanobis dist

You are screen sharing

Euclidean dist (L2 Norm) $\|x\| = \text{length of } x$

$$\hookrightarrow \|x_1 - x_2\|_2$$

where

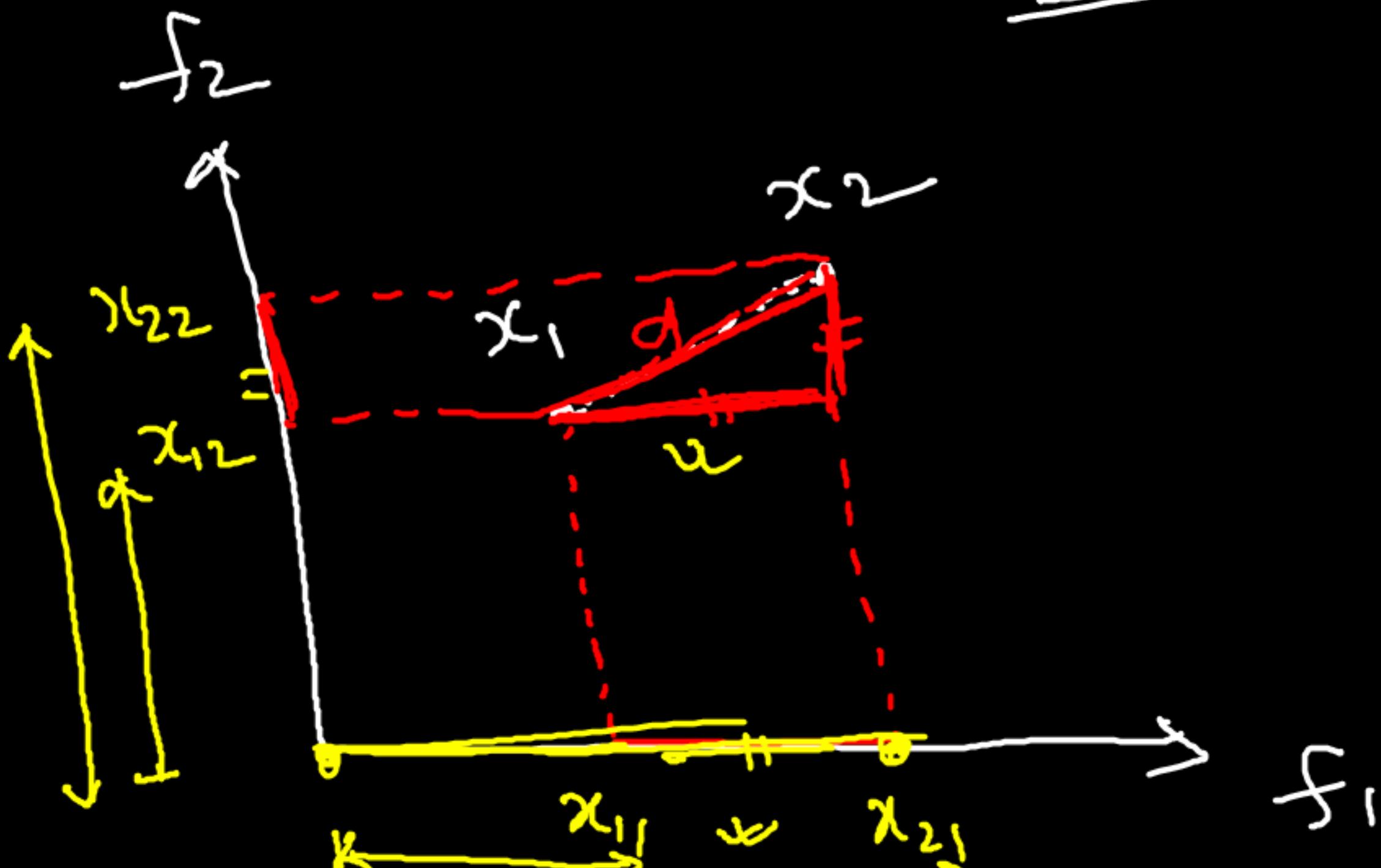
$$x \in \mathbb{R}^d$$

$$x_1 = (x_{11}, x_{12})$$

$$x_2 = (x_{21}, x_{22})$$

$d = \text{length of shortest line from}$

$$x_1 \& x_2$$



Euclidean dist = $\sqrt{(x_{21} - x_{11})^2 + (x_{22} - x_{12})^2 + \dots} = \|x_2 - x_1\|$

+
Euclidean dist : $\|x_2 - x_1\|_2 = \left(\sum_{i=1}^d (x_{2,i} - x_{1,i})^2 \right)^{\frac{1}{2}}$

$$\|x_1\|_2 = \left(\sum_{i=1}^d (x_{1,i})^2 \right)^{\frac{1}{2}} = \text{euclidean dist of } x_1$$

from origin.

You are screen sharing

	x_1	x_2	x_3	x_4	Label
A	100	10	1	2	Y
B	80	1	0	0	N
C	200	20	1	2	Y
D	150	15	1	2	Y
E	50	2	0	1	No
F	90	15	1	0	??

$$\begin{aligned} & \sqrt{10^2 + 5^2 + 0 + 2^2} = \sqrt{100 + 25 + 4} = \sqrt{129} \\ & \sqrt{10^2 + 14^2 + 1^2 + 0} = \sqrt{297} \\ & \sqrt{(110)^2 + 5^2 + 0 + 2^2} = \sqrt{12129} \\ & \sqrt{(60^2 + 0^2 + 0 + 2^2)} = \sqrt{3604} \\ & \sqrt{(40^2 + 13^2 + 1^2 + 1^2)} = \sqrt{1600 + 169 + 2} \\ & = \sqrt{1771} \end{aligned}$$

K-Nearest Neighbors

$K=1$ output = A \rightarrow "Yes"

$K=2$ output = B - "No"

$K=3$ output = E - "No"

F will also Set 2^{nd} = Y

$F < \underset{N=2}{\underline{\text{Y}=1}} \text{ "No"}$

$K=5 < \underset{N=2}{\underline{\text{Y}=3}}$

$F = \underline{\text{Y}=4}$

If you use RNN model apart from Clinical Sefer dataset, please apply only when your time less dataset.