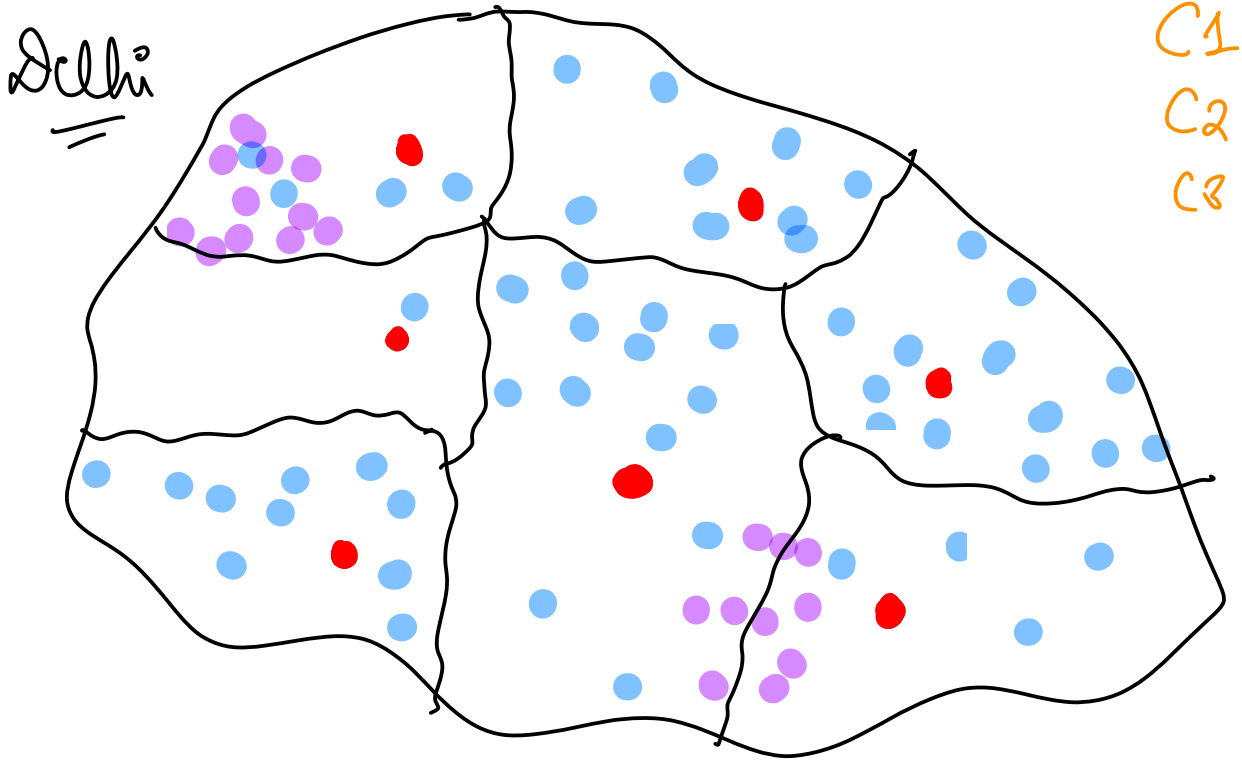
KNN-1

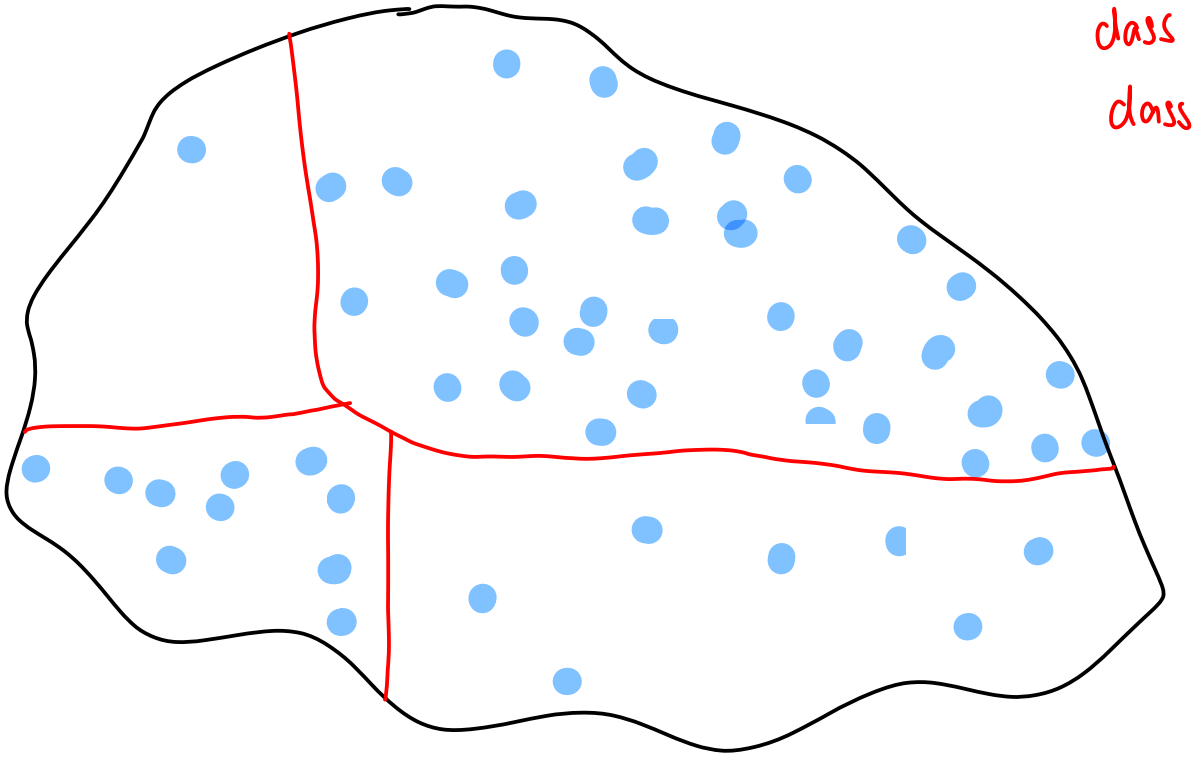
① Notes

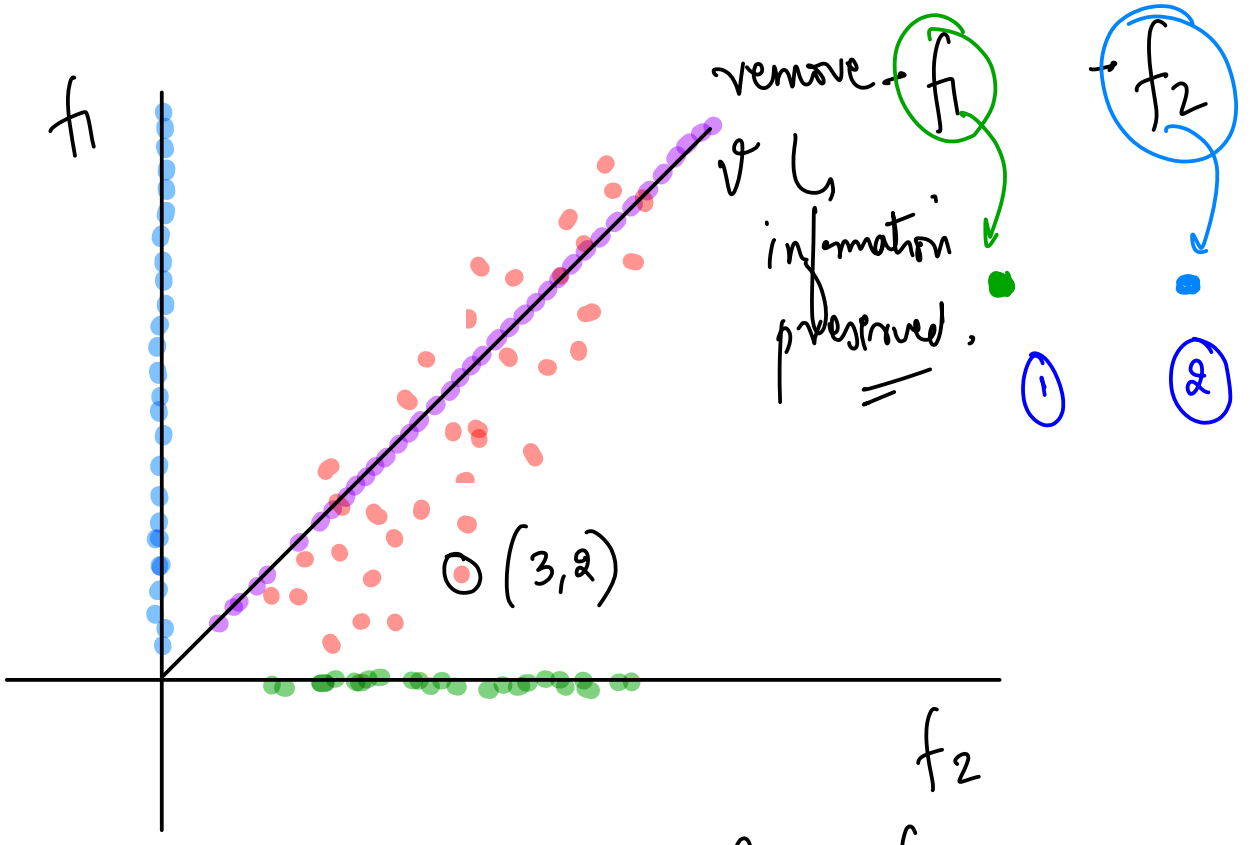
② Classroom:- T1 → 1st → 2nd → (if req) 3rd

③ Doubts : nikhil.sanghi@Scaler.com

KNN → K Nearest Neighbours



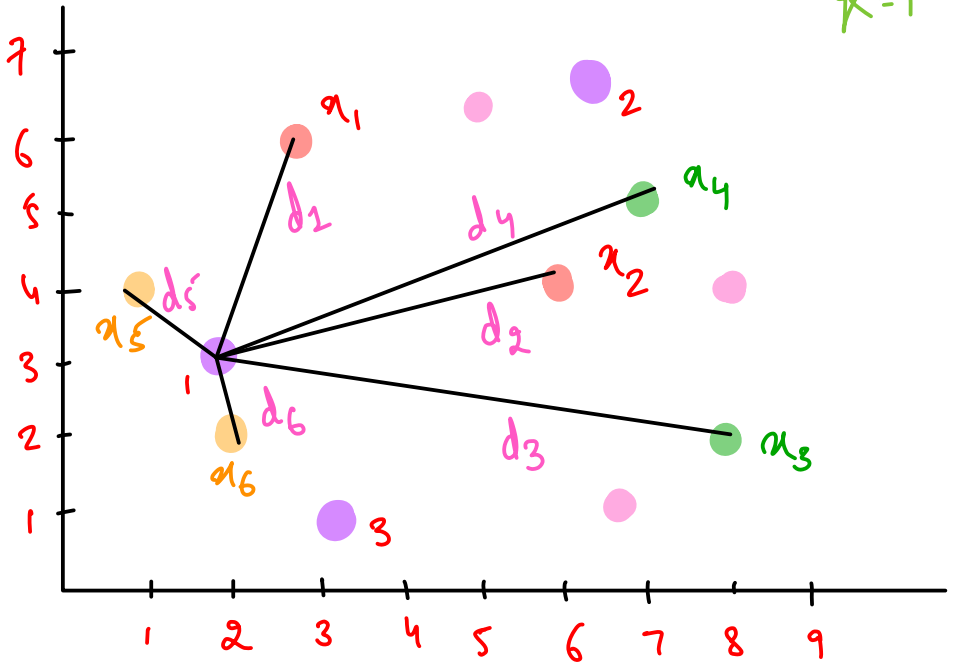

High → class 1
class 2
class 3



$$v = \alpha f_1 + \beta f_2$$

$k=1$

	f_1	f_2	y	
x_1	3	6	●	1
x_2	6	4	●	1
x_3	8	2	●	2
x_4	7	5	●	2
x_5	1	4	●	3
x_6	2	2	●	3
x_q	2	3	●	



$d_6 < d_5 < d_1 < d_2 < d_4 < d_3$

3 3 1 1 2 2

K ⇒ hyperparameter

We try to avoid even values of K

We try to choose odd values of K

K = 1 x_q ⇒ 3

K = 2 x_q ⇒ 3

K = 3 x_q ⇒ 3

K = 4 x_q ⇒ 3/1

K = 5 x_q ⇒ 3/1

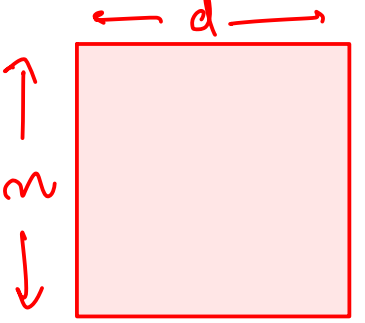
K = 6 x_q ⇒ 3/1/2

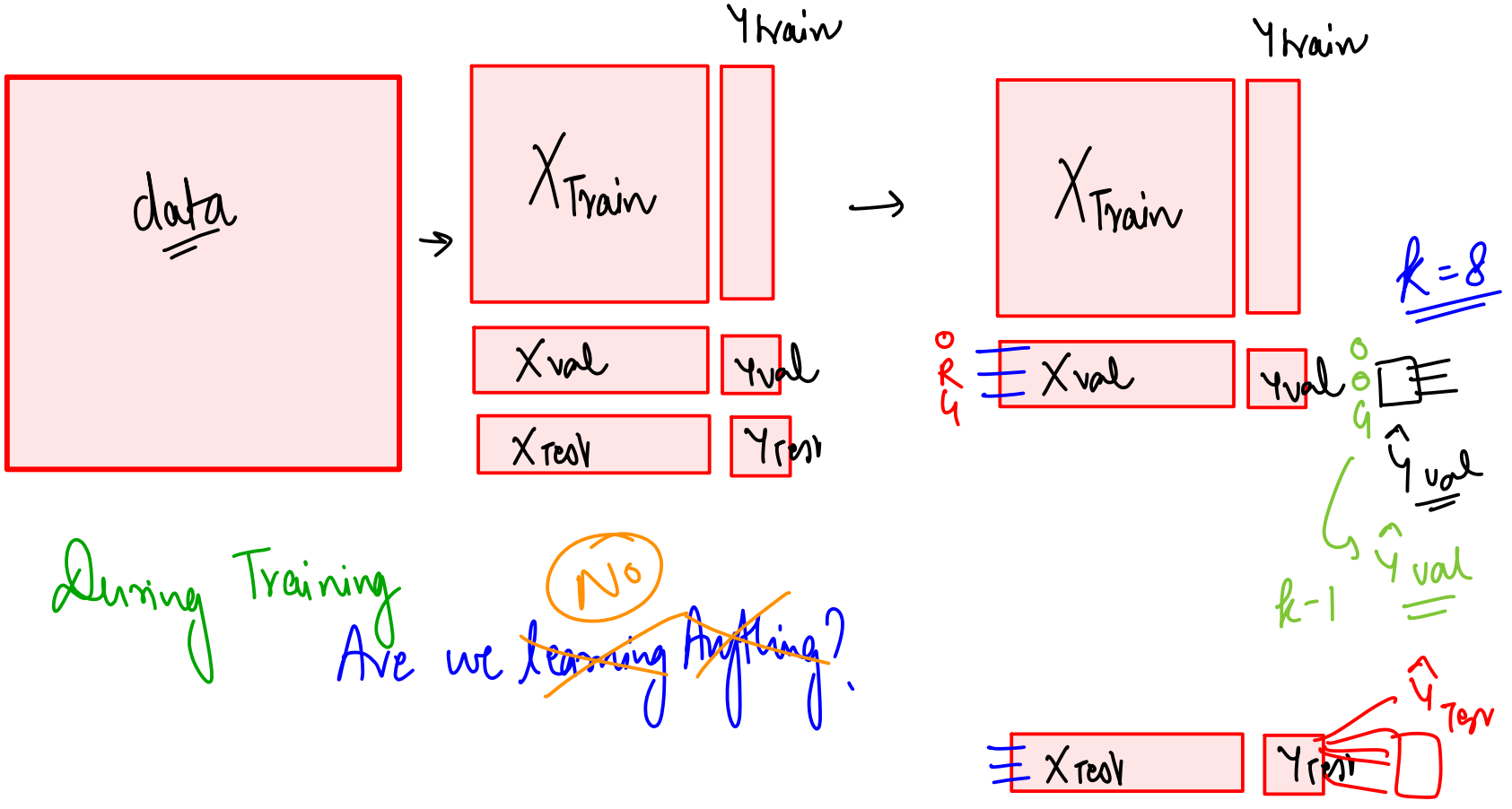
100M datapoints x_q

1. Calculate distances of x_q with all points. $O(n \times d)$
2. Sort distances in ascending order. $O(n \log n)$
3. Pick top k values (odd) $O(k)$ ↓↓↓
4. Majority Voting. Time Complexity = $O(nd + n\log n)$

$$\sqrt{(x_1^d - x_q^d)^2 + (x_1^{d-1} - x_q^{d-1})^2 \cdots\cdots}$$

d calculations




data
y_{train}
X_{Train}
X_{val}
y_{val}
X_{test}
y_{test}
y_{train}
X_{Train}
$k=8$
OR
X_{val}
y_{val}
OR
y_{val}
$\hat{y}_{val}$
$k-1$
$\hat{y}_{test}$
X_{test}
y_{test}
During Training
NO
Are we learning Anything?

During Training → fast (1)

slow (2)

At inference → fast (1)

slow ✓ (2)

1 Billion Training

In testing

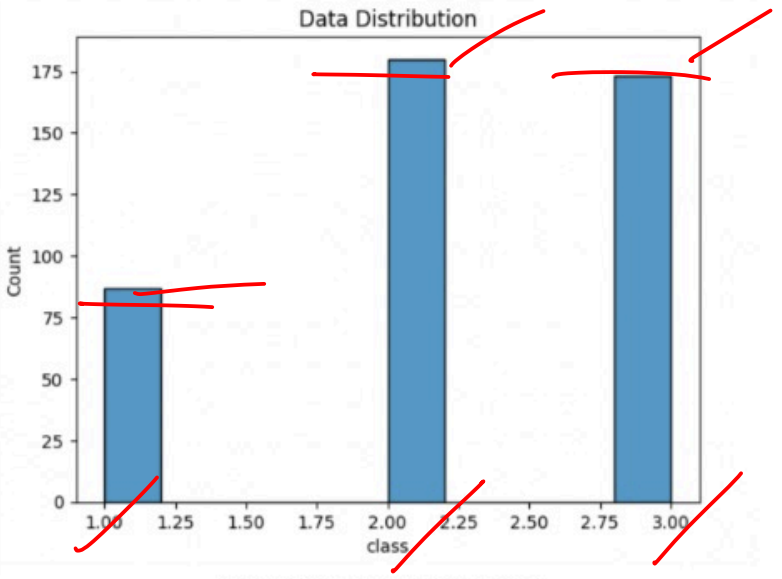
lot of data ⇒ ~~KNN~~

Advantage:

→ Non linear separable data

→ Multi class classification

→ Simple & intuitive



What can be said about the data ?

45 users have participated

A	Multi-class balanced data	9%
B ✓	Multi-class imbalanced data	87%
C	Binary-class imbalanced data	4%
D	Binary-class balanced data	0%

Quiz time!

Quiz Ended!

How will Logistic Regression handle non-linear, multi-class data ?

4[illegible] users have participated

A	Polynomial, OneVsRest	59%
B	Linear, OneVsRest	18%
C	OneVsRest, Polynomial	11%
D	OneVsRest, Linear	11%

Quiz time!

Quiz Ended!

Arrange the statements in correct order based on kNN algo

s1- find majority vote

s2- perform euclidean distance

s3- sort and select k datapoints

s4- give class to x_q datapoint

33 users have participated

A	s2,s1,s3,s4	9%
B	s2,s3,s4,s3	15%
✓ C	s2,s3,s1,s4	64%
D	s2,s4,s3,s1	12%

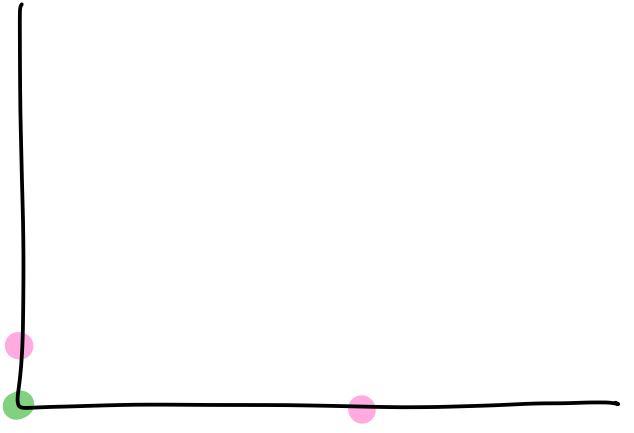
Quiz time!

Quiz Ended!

If x1 at (4,0) , x2 at (0,1) and x3 (5,0) and xq at (0,0). Which is nearest point to xq ?

46 users have participated

	A	x1	4%
✓	B	x2	93%
	C	x3	2%



Quiz time!

Quiz Ended!

how kNN is better than logistic regression. Select the correct option