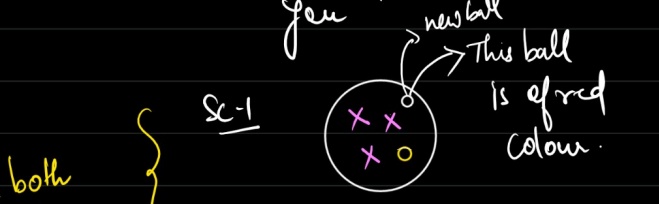
K-Nearest Neighbour (KNN)

① KNN Classifier
② KNN Regressor

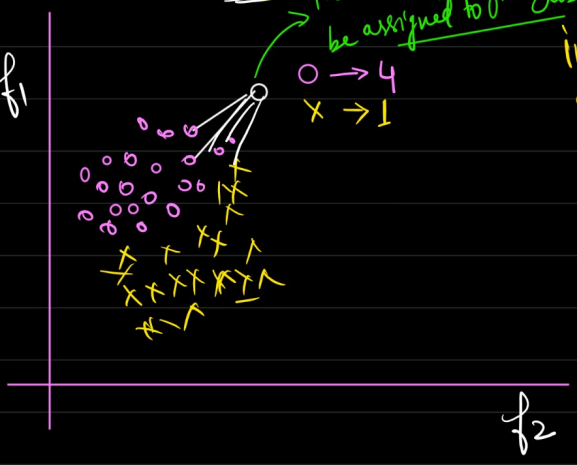
① KNN Classifier

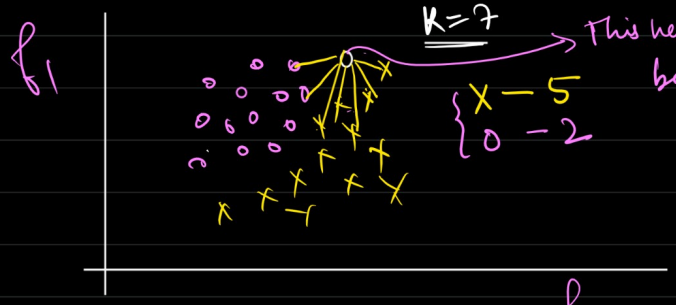
idea:- You will be like people surrounding you



in both of the cases you have considered the majority colour ball in the bag as deciding factor.

f_1	f_2	y
—	—	0
—	—	1
—	—	0
—	—	0
—	—	1





→ As K changes the class of new dp might also change.

⇓

K is a hyperparameter

① plot the datapoints in n-d space

② Initialise the K value (No of Neighbours you want to consider)

$K \in 1 \to \infty$

$K > 0$

(generally taken as $K > 3$)

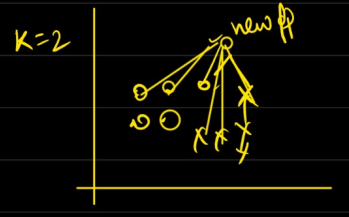
→ Calculate the distance of new datapoint wrt to all dp's.

→ Sort the distance.

→ based on K find the class of that K nearest datapoint.

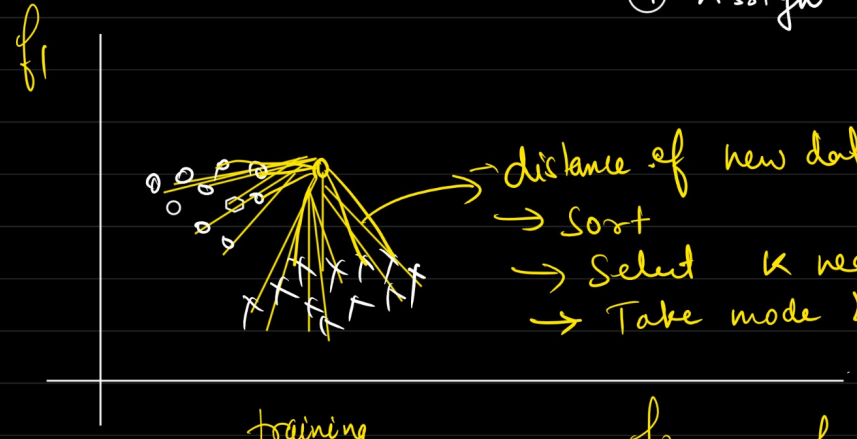
③ Find the mode of the class

Corner Case



K → odd → 1, 3, 5, 7 ⇒

④ Assign the class.



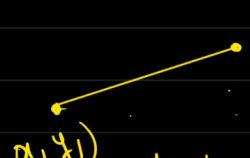
→ distance of new datapoint from all the dp's
→ Sort
→ Select K nearest distance dp
→ Take mode & assign the class.

K=3

f_1	f_2		dist		majority Predicted class = 1
		0		40	0
		1	58	58	1
		0	60	60	1
		1	60		
		0	40		

To calculate distance

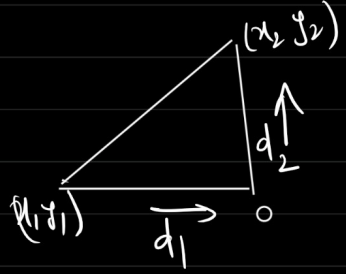
① Euclidean distance



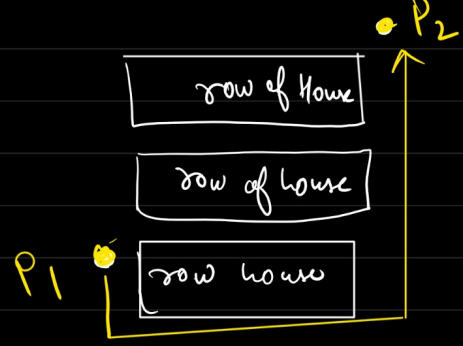
$$\text{distance} = \sqrt{(x_2-x_1)^2+(y_2-y_1)^2}$$

A ——— B by Aeroplane → Euclidean distance.

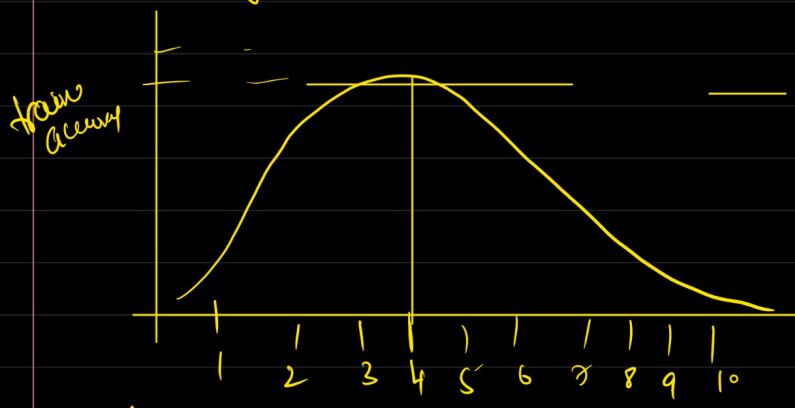
② Manhattan distance



$$\text{distance} = d_1 + d_2$$

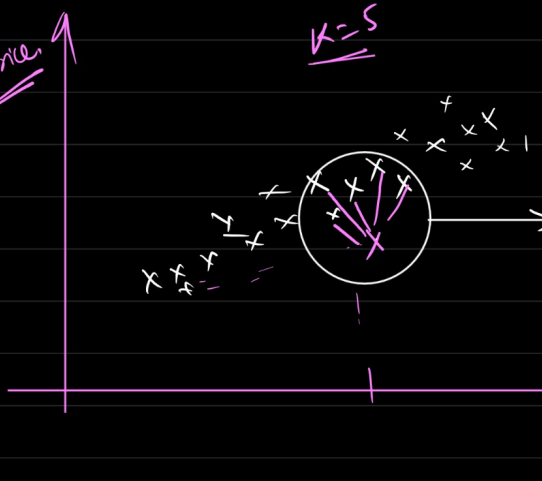


* K – hyperparameter



* for different K, you keep track of train/test accuracy.

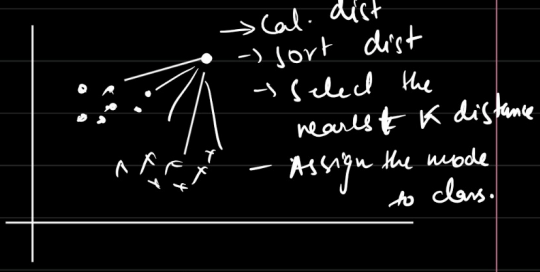
② Regression



→ Prediction for new dp will be avg y (price) of nearest 5 dp's

→ In case of outlier, instead of Avg you can take median.

Conclusion of Classification



* **Advantage**

→ Easy to Understand | very intuitive.

→ Performance of model in terms of evaluation metric is good

* **disadvantage.**

→ Lazy learner

↳ At the real time the distance of test data from each of dp is Calculated ⇒ Brute force.

→ n dp → n-1 distance you will be calculating

→ Computationally expensive

Time complexity O(N)

ML models

- Eager learner (All the model parameters are calculated while training and used for prediction)
- Lazy learner ⇓ you are calculating Parameters of model in real time

* **Variants of KN**

① KD Tree
② Ball tree

} optimizers