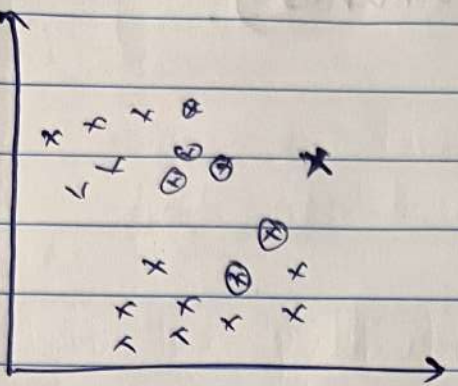
KNN (K-Nearest Neighbour)



f_1	f_2	y (0/1)
-	-	

Steps:

$K=5$

1) Determine the K-value

~~$K>0$~~ ~~$0<K$~~ $0 < K \leq \infty$

2) Calculate the nearest neighbour of the test data (No. of neighbour will depend on K).

3) Count the no. of points which belong to 0 category & 1 category.
Which ever is greater, the test data will belong to that category.

Q How to calculate K-nearest nearest neighbour?

1) ~~Eu~~ Euclidean distance
2) ~~Ham~~ Manhattan distance

Euclidean formula (Mostly used in airlines).

(x_1, y_1) (x_2, y_2)

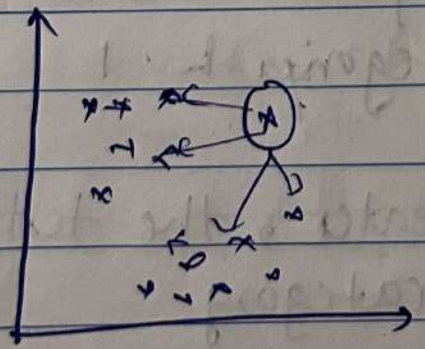
$$\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

Manhattan distance:

$$|x_1 - x_2| + |y_1 - y_2|$$

KNN - Classification we have seen right now.

KNN Regression:



step: we will compute the distance from test data. (nearest)

- Take the average of these points

KNN Optimization:

- Inorder to find the nearest neighbour, we don't need to find the distance from all the points. (Time complexity - O(n)).

We can optimize this by using:
- KD Tree
- Ball Tree.

KD Tree:

It splits the data recursively according the axes (creates a BST).
Then, the new test data is searched.
In KD Tree, we have to backtrack to the datpoints.

Ball Tree:

- It creates group of nearest points.
- This process is repeated till all the datapoints are covered.
- If whenever the new data comes, the group is checked and the nearest

neighbours are then selected.

- This doesn't require back-tracking.