05 - K Nearest Neighboor (KNN)

1> Classification

2> Regression

1) Classification

$K = 5$

$0 = 2$

$1 = 3$
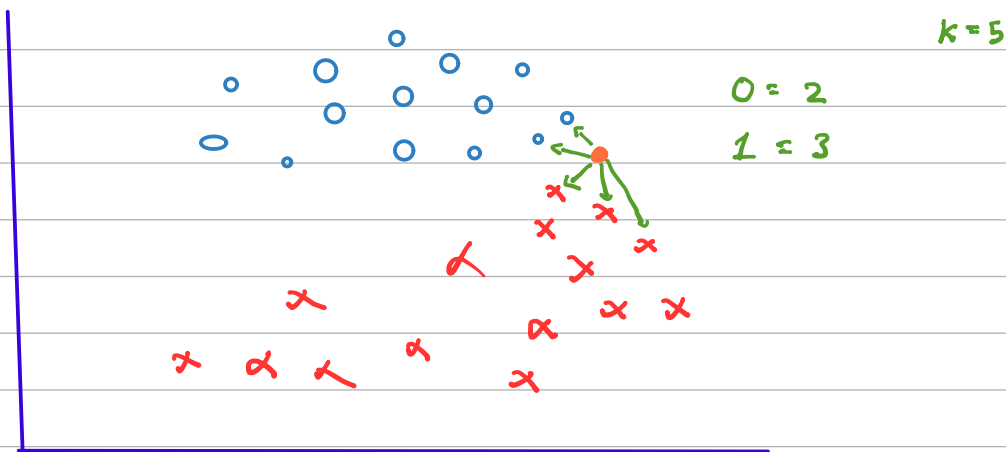
Steps to solve

1> We have to initialize the K value
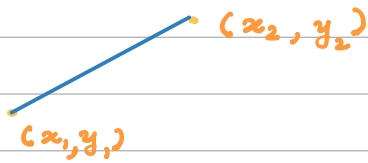
$K > 0$

$k = 1, 2, 3, 4, 5, 6 \Rightarrow$ Hyper parameter

2> Find the K nearest Neighbour from the test data

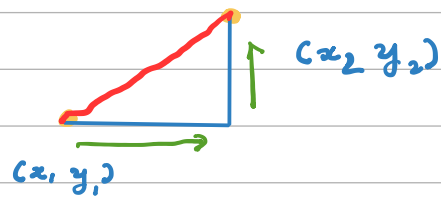3> From those $K = n$ how many neighboors belongs to 0 category and 1 Category

Distance Metric

① Euclidian Distance

(x_2, y_2)

(x_1, y_1)

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

$\text{Distance} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}$ (3D)

② Manhattan Distance

(x_2, y_2)

(x_1, y_1)

$|x_1 - x_2| + |Y_1 - Y_2|$

$|x_1 - x_2| + |Y_1 - Y_2| + |z_1 - z_2|$

2} Regression

$K = 5$

→ Average / median
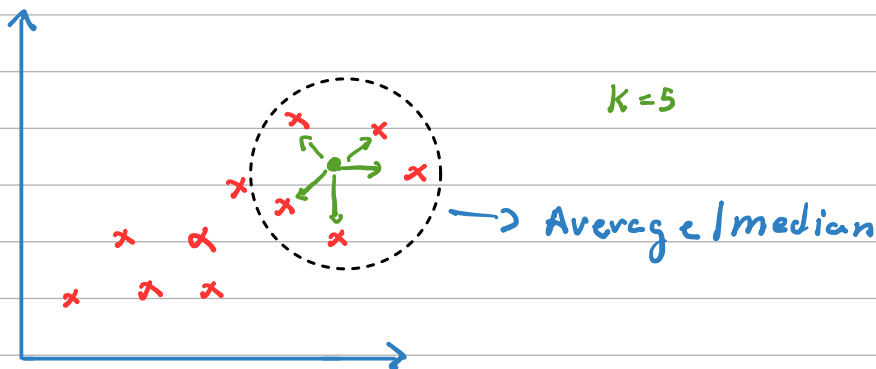
Take the no of K's and make the average of the distance of the K points.

Limitations of KNN.

1) Huge dataset

2) Outliers ← Sensitive to Outliers
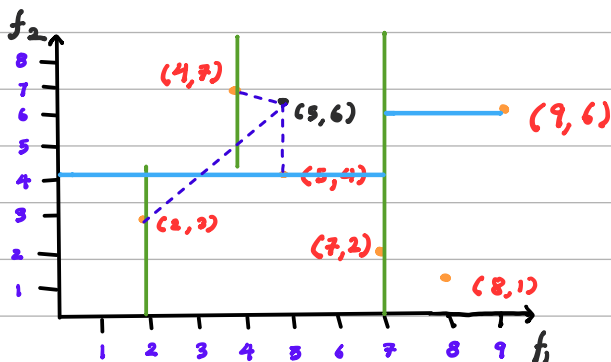
3) Sensitive to missing Values

Varient of KNN

→ Time Complexity $O(N)$ ↑ ↑

① KD Tree ⎫ Binary Tree
② Ball Tree ⎭

⇓

Time Complexity ↓↓

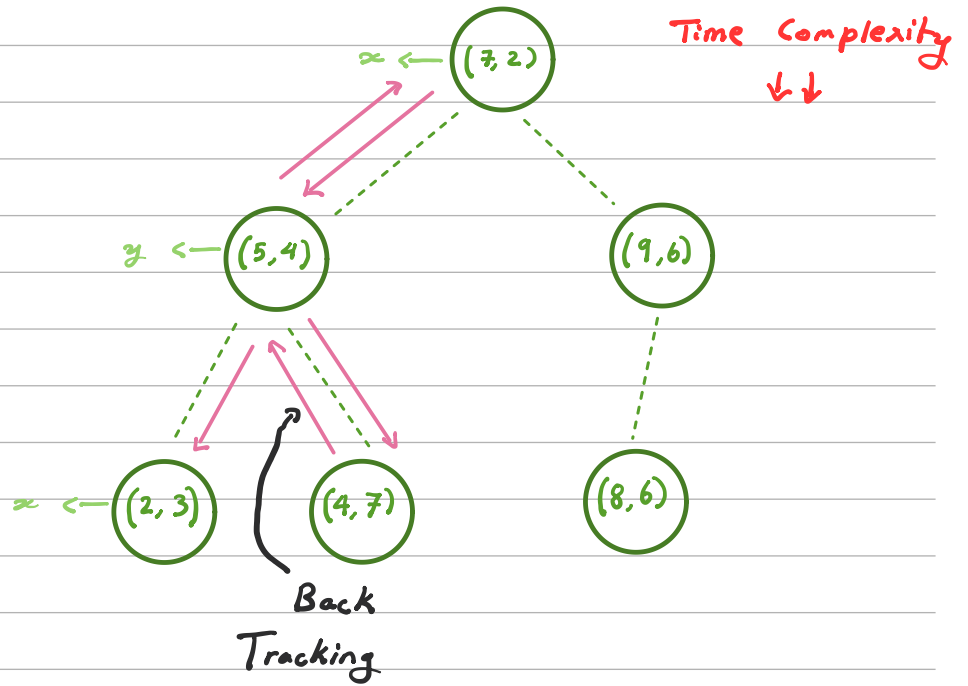* By default KNN uses Brute Search (Calculating each and every point)

K-D Tree ⇒ K Dimention Tree

① Median of x coordinates

2, 4, 5,7, 8, 9

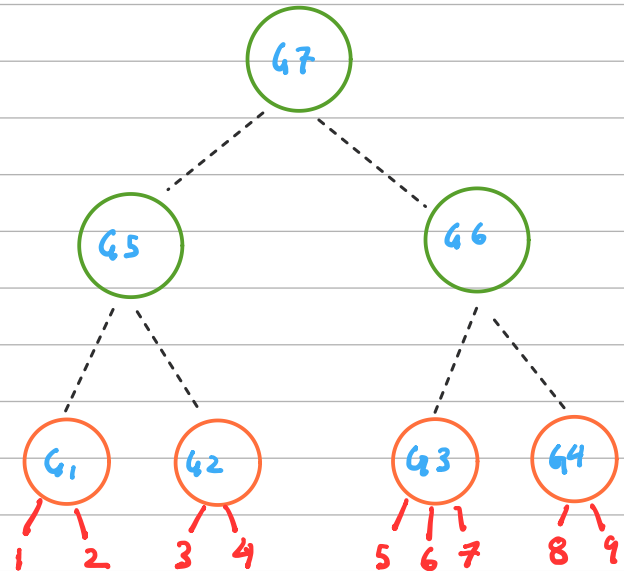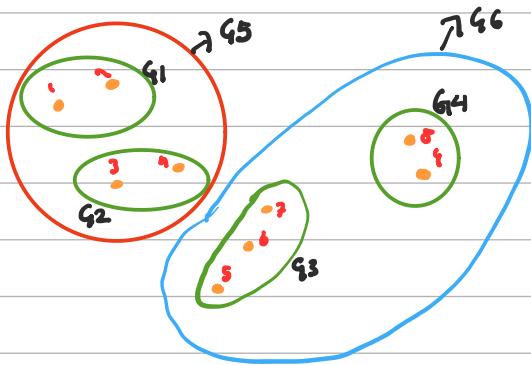$$\frac{5+7}{2} = 6.5 \Rightarrow 7$$

② Median of y axis

1, 2, 3,4, 6, 7

Time Complexity ↓↓

x ← (7, 2)

y ← (5, 4)   (9, 6)

x ← (2, 3)   (4, 7)   (8, 6)

Back
Tracking

23 Ball Tree

↗ G5   ↗ G6

G1   G4
G2
          G3

G7

G5   G6

G1   G2   G3   G4

1   2   3   4   5 6 7   8 9

Grouping each and every clusters individually then grouping
together.

* Back propagation is not nessecarry in Ball Tree