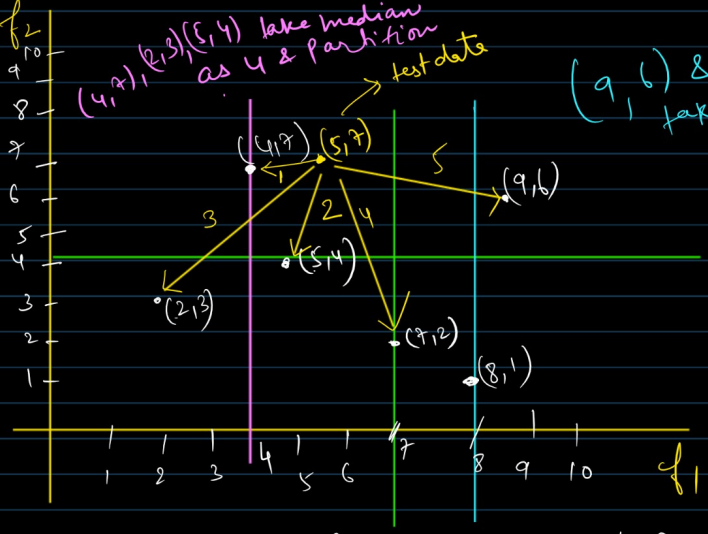
Variants of KNN

① KD Tree (K-dimensional tree)
② Ball Tree

* In KNN, Distance calculation test data is done with all the other datapoints in the dataframe. => computationally expensive

① KD tree

f_1	f_2
7	2
5	4
→ 9	6
2	3
4	7
8	1



(9,6) & (8,1) take median as 8 and partitioned

① $f_1 \Rightarrow 2, 4, 5, 7, 8, 9$, $f_2 = 1, 2, 3, 4, 6, 7$

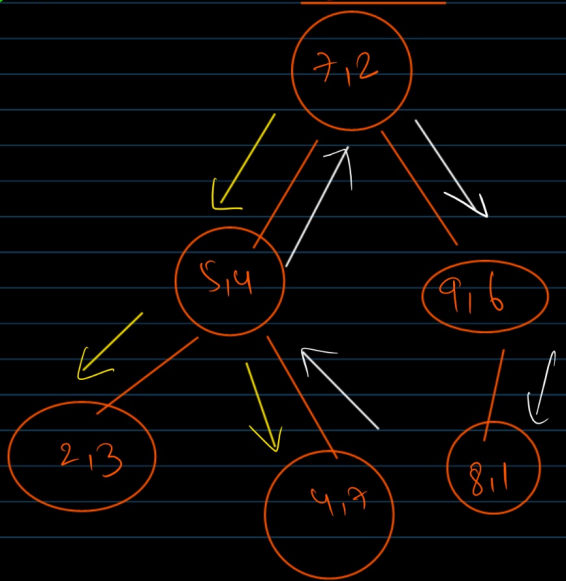
② median = $\frac{5+7}{2} = 6.5$ take 4 as median

Just take 7 as median (any among 5 & 7) → reason – 6.5 doesn't exist in data, so take the closest value that exit.

③ Partition the data recursively (Check each sub partition & Calculate median)

Eg: (9,6) ,(8,1)
Take 8 as median.

KD tree



KD tree

idea :- Partition the data into a binary tree.
In order to implement the idea follow steps below:-

Step 1 :→ Sort the feature f_1 & f_2

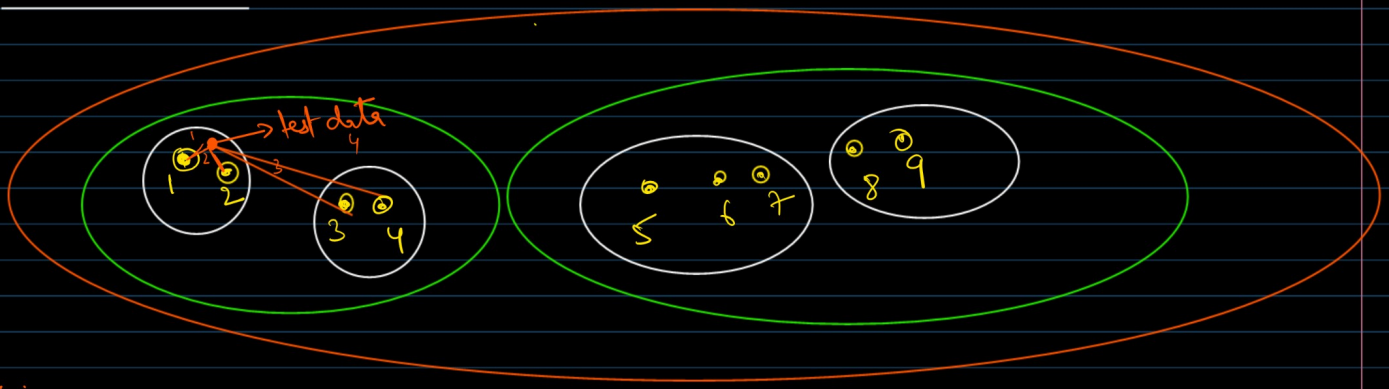
Step 2 :→ Calculate the median f_1 & f_2

Step 3 Partion the data based on median recursively (keep partitioning data in each subpartion based on median)

Step 4 KD tree (Binary tree ↓ two child node)

Advantage
→ you don't need to calculate all the distances.

② Ball tree



Steps:

① Each dp is a ball/group in itself

② Combine nearest ball recursively.
Until it becomes 1 ball.

③ for a test data → only need to calculate the distance its own ball/group element or the nearest ball/group element