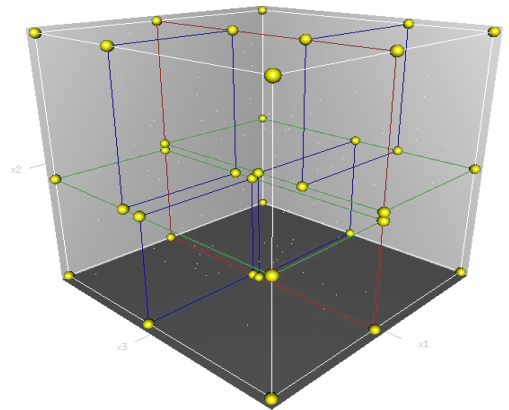
Kd-trees

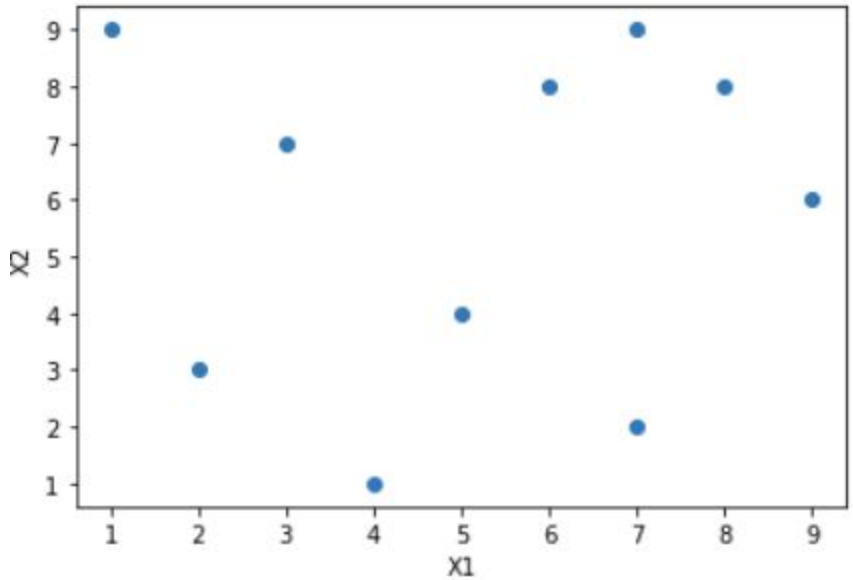
- Invented in 1970s by Jon Bentley
- Binary data algorithm for storing data structures in computer science.
- Name originally meant “3d-trees, 4d-trees, etc” where k was the # of dimensions
- Now, people say “kd-tree of dimension d”
- Idea: Each level of the tree compares against 1 dimension.



k-d tree example

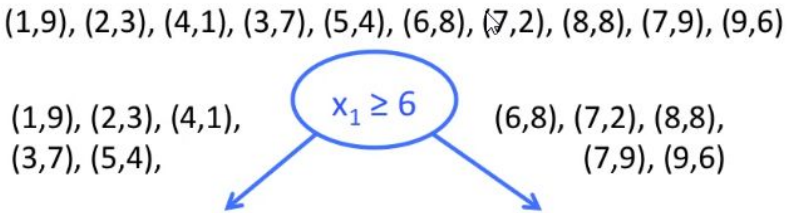
Dataset

X1	X2
1	9
2	3
4	1
3	7
5	4
6	8
7	2
8	8
7	9
9	6

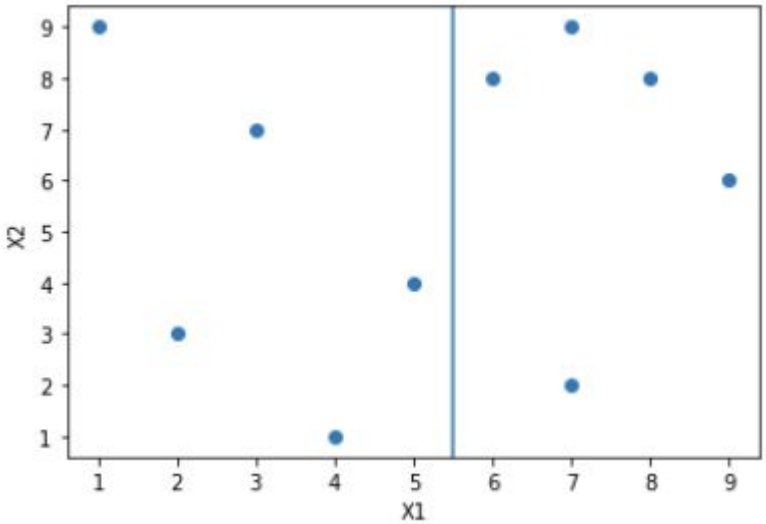


k-d tree algorithm organizes the dataset as tree

1) Pick a feature at random, say X1
2) Find Median of X1 and split dataset based on the median of X1

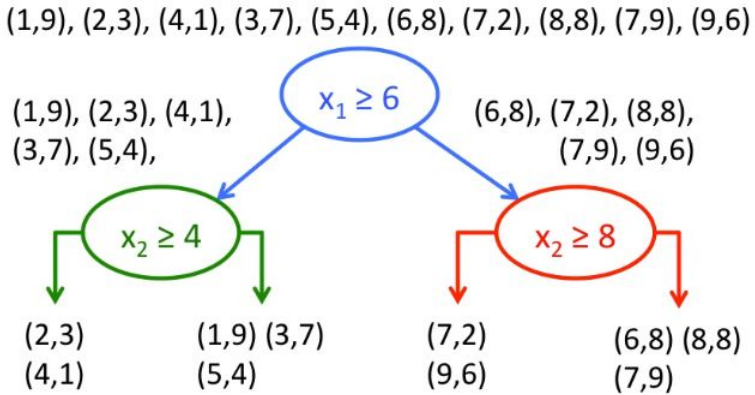


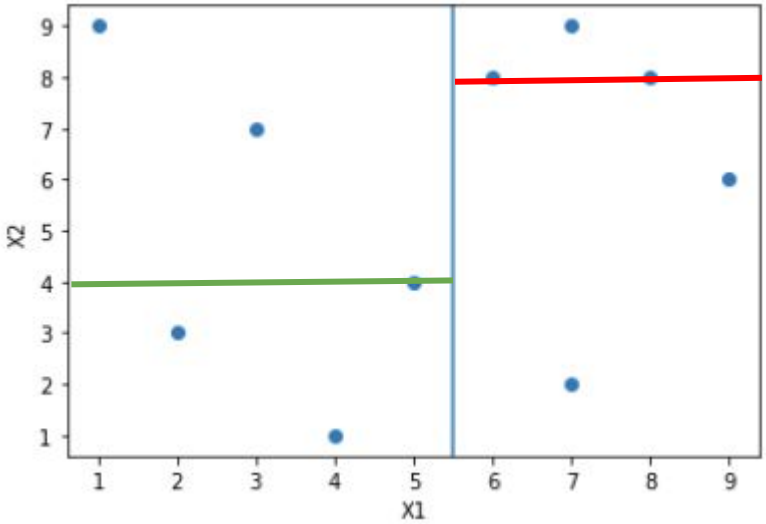
Half the dataset end up on each side



k-d tree algorithm cont..

3) Take the median of X2 and split data again





Finding KNN for new data point z (X1=7, X2=5)

- Search logic will be via matching the new Z point to the tree logic
- Pick the best Neighbour based on euclidean distance
- This process can repeated N times to find N neighbours
- Note k-d tree doesn't always find optimum NN but its fast