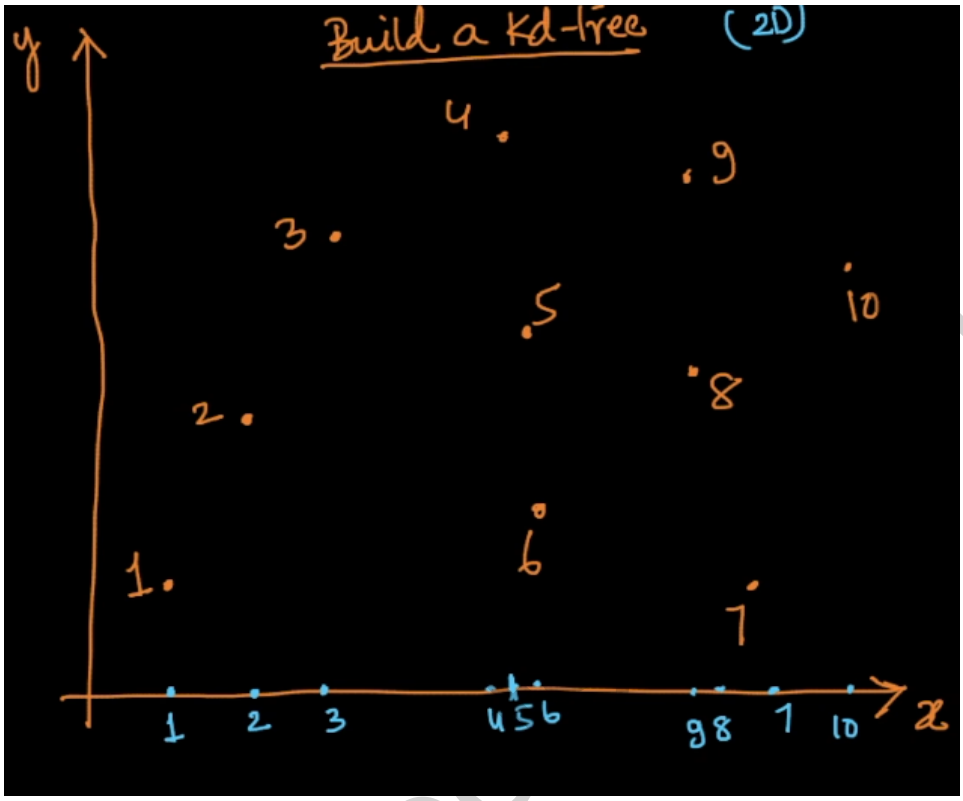
29.22 How to build a KD-Tree

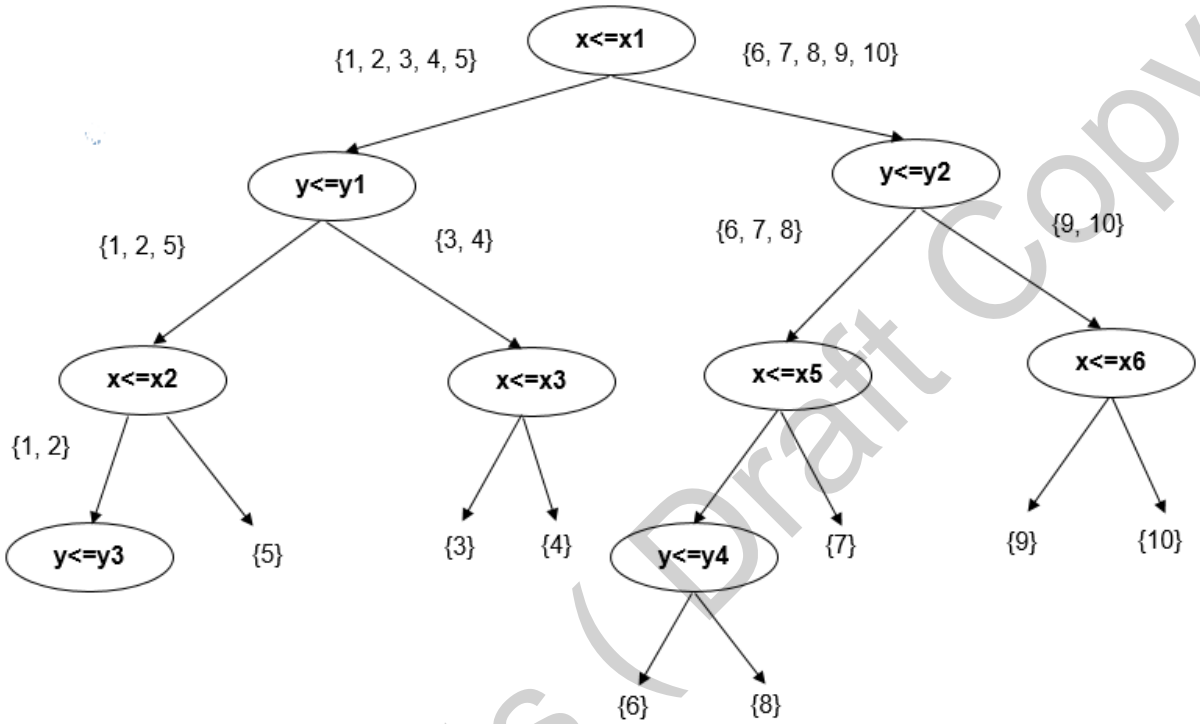
Procedure to construct a KD-Tree

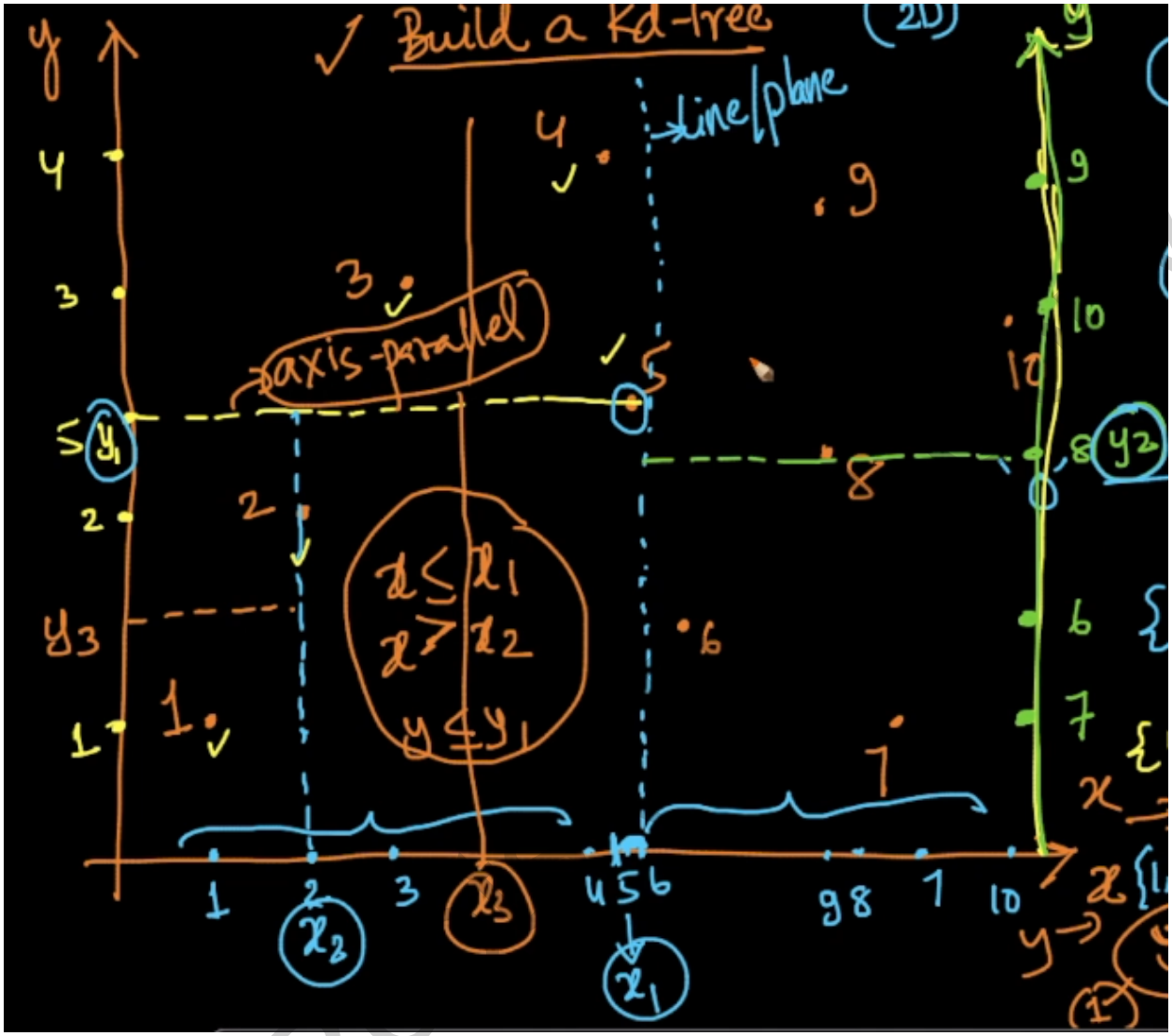
Let the given points be projected in a 2D space as shown below



1) Pick the 'X' axis, project all the points on the 'X' axis. Compute the median. Split the data on the basis of the median 'x_1'.
   All the values that are less than or equal to 'x_1' should go into the left subtree(ie., 1,2,3,4,5), and those values which are greater than 'x_1'(ie., 6,7,8,9,10) should go into the right subtree. 'x_1' would be the root of the tree.
2) Pick the 'Y' axis, project all these points on the 'Y' axis. Compute the median. Split the data on the basis of the median 'y_1'. Now 'y_1' is the root node of the left subtree.
   All the values that are less than or equal to 'y_1'(ie.,1,2,5) should go into the left subtree, and those values which are greater than 'y_1'(ie., 3,4) should go into the right subtree.
   Compute the median of the values {6,7,8,9,10} and it should be the root node of the right subtree. Let it be 'y_2'. All the values that are less than or equal to 'y_2' will now go into the left subtree and the remaining values should now go into the right subtree.

This way we have to keep alternating between the axes. The final KD-Tree looks like below





Note:

In KD-Tree, we are breaking the space using axis-parallel lines/planes into rectangles(in 2-D)/ cuboids(in 3-D)/ hyper cuboids (in n-D).

We have to switch the axes alternatively till we reach the leaf nodes. We can build the KD-Tree not only with the 'median' statistic, but also with the 'mean' statistic.' But if we go with the 'mean' statistic, we should make sure, there are no outliers in our data.