

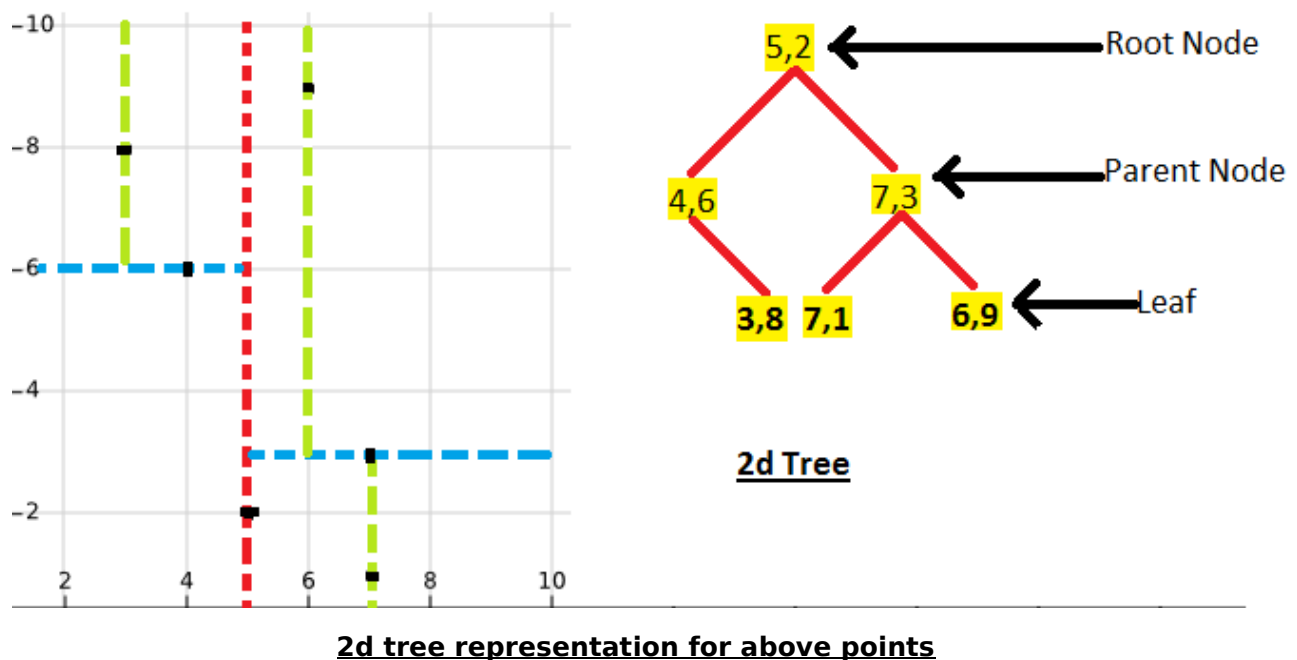
Review for: An efficient k-Means Clustering Algorithm: Analysis and Implementation

By Tapas Kanugo, David M. Mount, Nathan S. Netanyahu, Christine D. Piatko, Ruth Silverman and Angela Y. Wu

K-d Tree:

K-d Tree is used to find the K nearest neighbors of a specific point or location. For example, let $K=2$, i.e. a 2 dimensional tree to be formed with a set of coordinates as below-

$X = \{(5,2), (7,3), (4,6), (6,9), (3,8), (7,1)\}$



- Let's take (5,2) as root node.
- Next point is (7,3) the X dimension value is bigger than the same as root node ($7 > 5$). Therefore, it is placed at the right branch of the root node.
- Next point is (4,6). As $4 < 5$, it is placed at the left of the root node.
- Next point is (6,9). As ($6 > 5$) it moves to the right branch but this place is already taken by (7,3). Therefore, now dimensional component is compared and as ($9 > 3$) the point takes place at the right branch of Parent node (7,3)

In this manner the process goes on until there is no points left to place in the tree.

Now for the graphical representation:

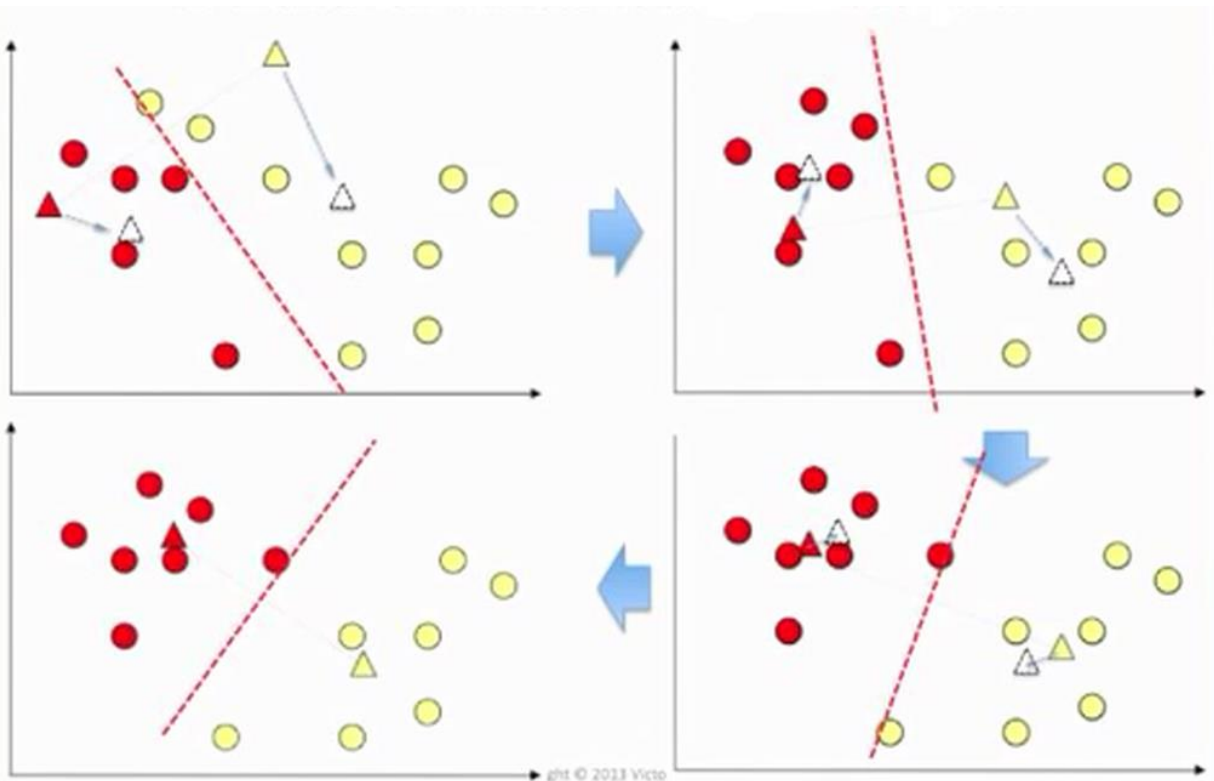
At the root point (5,2) a divider is drawn as $X=5$ which divides the surface in two subplots, we did in the tree. Next $Y=6$ and $Y=3$ make the rest of subplot in another two section. And the process carries on according to the tree.

For more than 2 dimension after X and Y dimension the remaining dimensions are considered and followed accordingly.

Filtering Algorithm:

It is based on Lloyd's Algorithm which states that:

- K is input finite number which represents the number of cluster
- X is a set of points (x_1, x_2, \dots, x_n)
- Now we intersect two dimension using kd tree, and assign random centroid to the points in each intersection.
- Next we calculate the mean value of the distance from that centroid in that cluster and move the previous centroid to new centroid achieved.
- This process keeps going until we have no centroid to be moved thus optimizing the cluster centroid to its best position.



Filtering Algorithm Moving of centroids to new optimal centroids