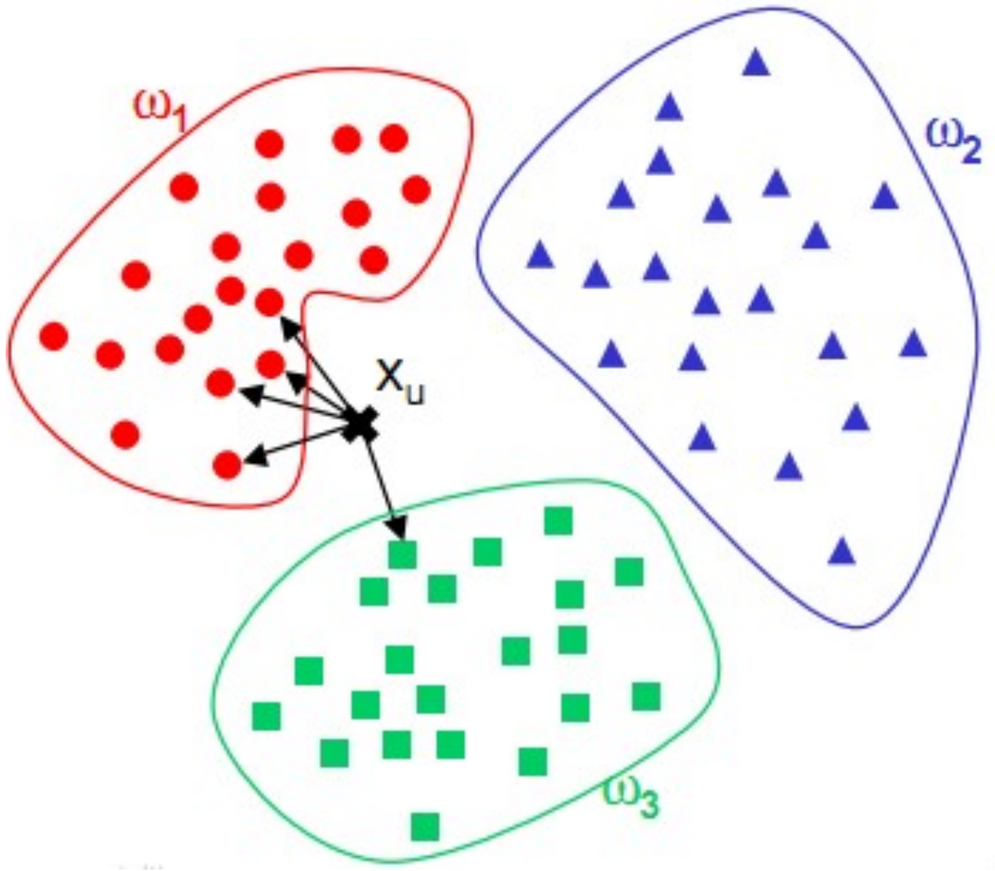
K-Nearest Neighbor

By:会飞的吴克

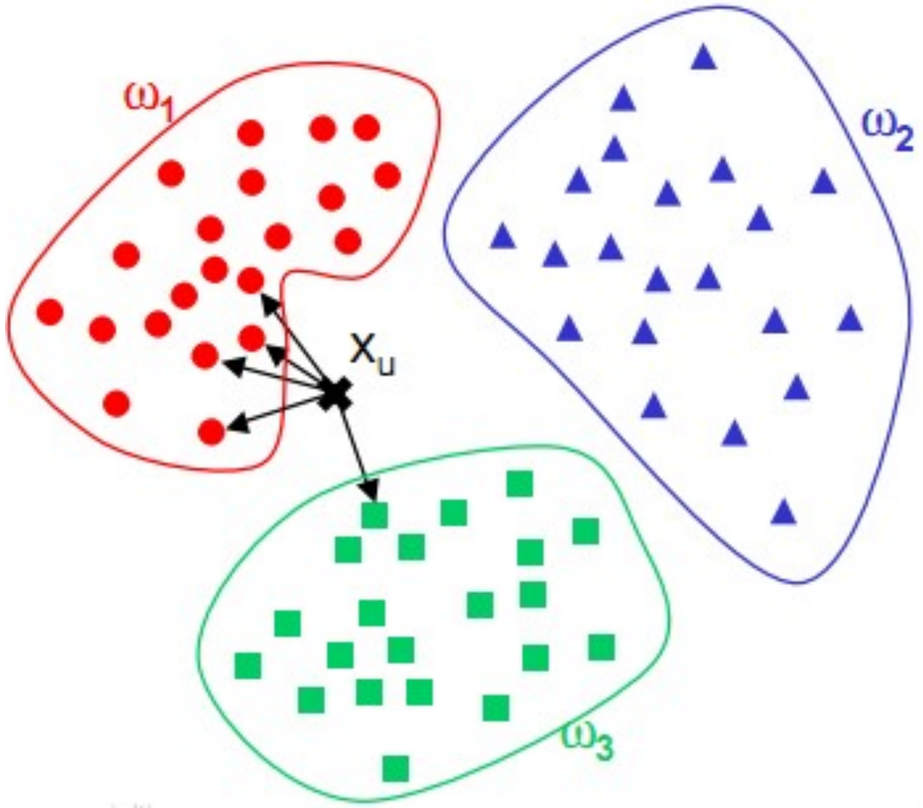
一、KNN分类算法

KNN分类算法



要解决的问题

已知M维空间中的N个点，对于一个新的点A，如何迅速找到N个点中距A最近的K个点



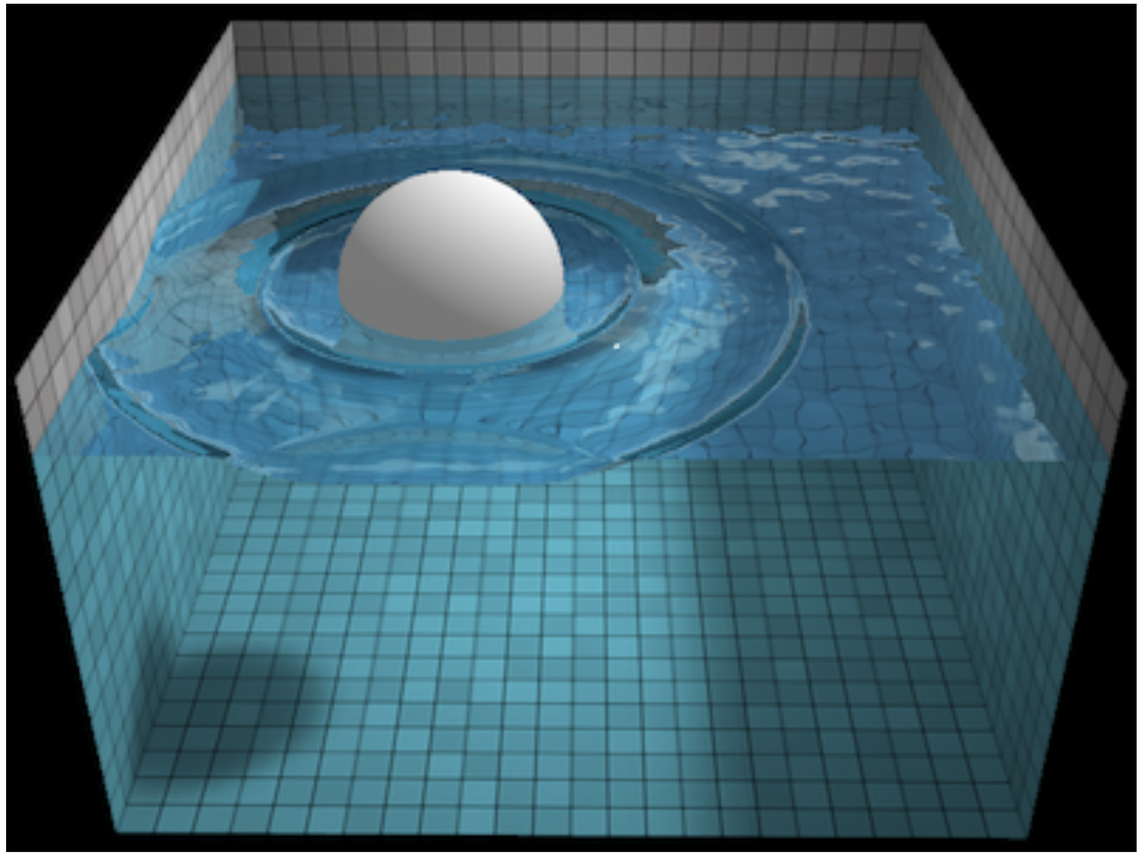
请找出以下数据中与0距离最近的数据:

0.0637, -0.1027, 0.4953, -0.668, 0.2641, -0.4978, 0.2596, -0.2779, 0.1106, -0.4123,
0.4672, -0.9177, 0.157, -0.276, 0.0728, -0.6118, 0.9371, -0.2979, 0.1902, -0.1842,
0.0888, -0.8097, 0.4618, -0.8971, 0.297, -0.4834, 0.7456, -0.8116, 0.3606, -0.1627,
0.9449, -0.7355, 0.9275, -0.6022, 0.4585, -0.8539, 0.8538, -0.9057, 0.8277, -0.0109,
0.6315, -0.2417, 0.8449, -0.7635, 0.4635, -0.4255, 0.3115, -0.8495, 0.228, -0.2333,
0.1318, -0.3214, 0.5186, -0.8322, 0.7893, -0.0771, 0.185, -0.9238, 0.5718, -0.3105,
0.0885, -0.3082, 0.689, -0.9794, 0.6801, -0.3416, 0.3894, -0.2797, 0.9114, -0.6186,
0.1856, -0.6707, 0.7603, -0.31, 0.4875, -0.9535, 0.2712, -0.7517, 0.3769, -0.617,
0.8399, -0.0924, 0.3908, -0.8972, 0.4542, -0.8141, 0.2736, -0.3282, 0.3829, -0.435,
0.0003, -0.995, 0.0717, -0.5736, 0.969, -0.4736, 0.3994, -0.9623, 0.4696, -0.4347

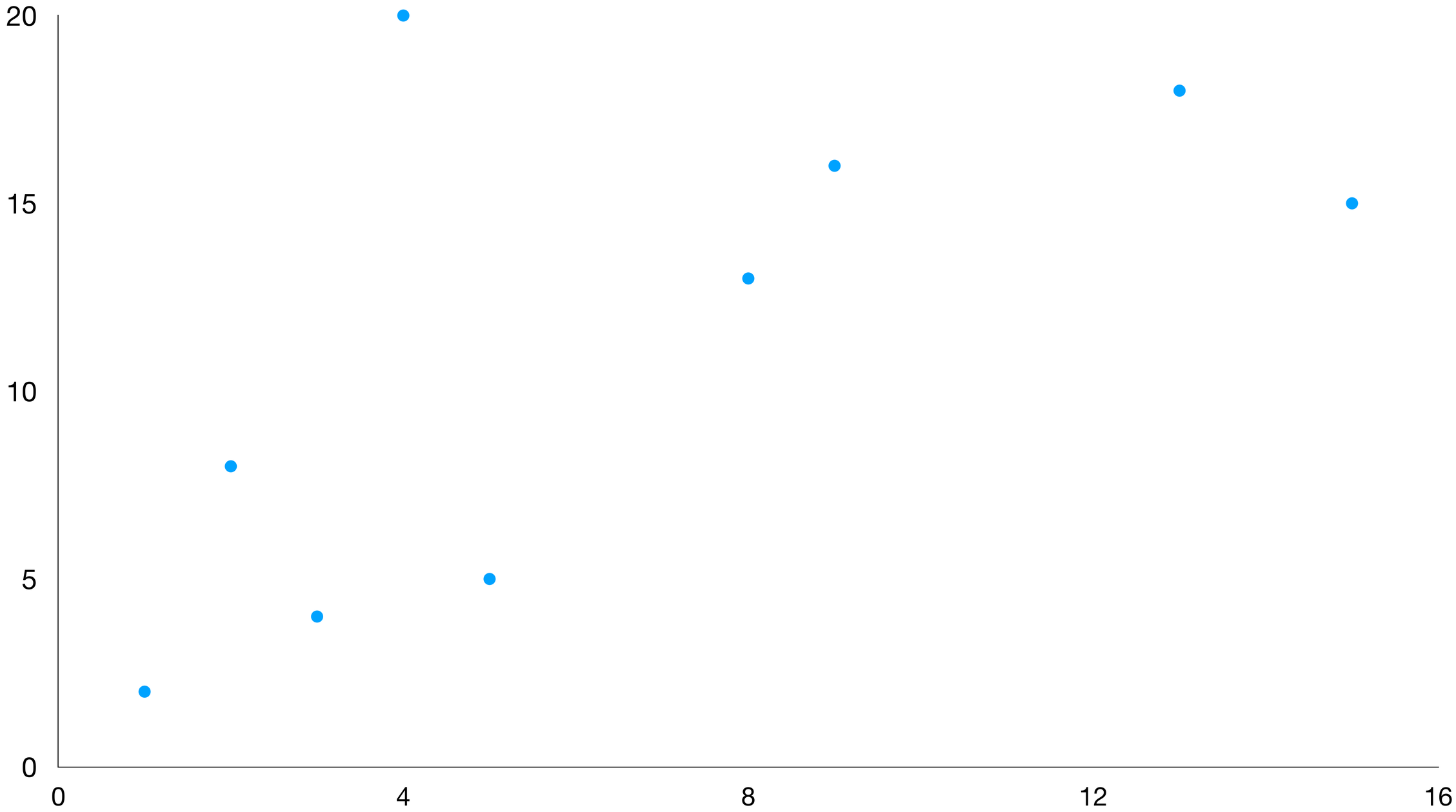
请找出以下数据中与0距离最近的数据:

-0.995, -0.9794, -0.9623, -0.9535, -0.9238, -0.9177, -0.9057, -0.8972, -0.8971, -0.8539, -0.8495, -0.8322, -0.8141, -0.8116, -0.8097, -0.7635, -0.7517, -0.7355, -0.6707, -0.668, -0.6186, -0.617, -0.6118, -0.6022, -0.5736, -0.4978, -0.4834, -0.4736, -0.435, -0.4347, -0.4255, -0.4123, -0.3416, -0.3282, -0.3214, -0.3105, -0.31, -0.3082, -0.2979, -0.2797, -0.2779, -0.276, -0.2417, -0.2333, -0.1842, -0.1627, -0.1027, -0.0924, -0.0771, -0.0109, 0.0003, 0.0637, 0.0717, 0.0728, 0.0885, 0.0888, 0.1106, 0.1318, 0.157, 0.185, 0.1856, 0.1902, 0.228, 0.2596, 0.2641, 0.2712, 0.2736, 0.297, 0.3115, 0.3606, 0.3769, 0.3829, 0.3894, 0.3908, 0.3994, 0.4542, 0.4585, 0.4618, 0.4635, 0.4672, 0.4696, 0.4875, 0.4953, 0.5186, 0.5718, 0.6315, 0.6801, 0.689, 0.7456, 0.7603, 0.7893, 0.8277, 0.8399, 0.8449, 0.8538, 0.9114, 0.9275, 0.9371, 0.9449, 0.969

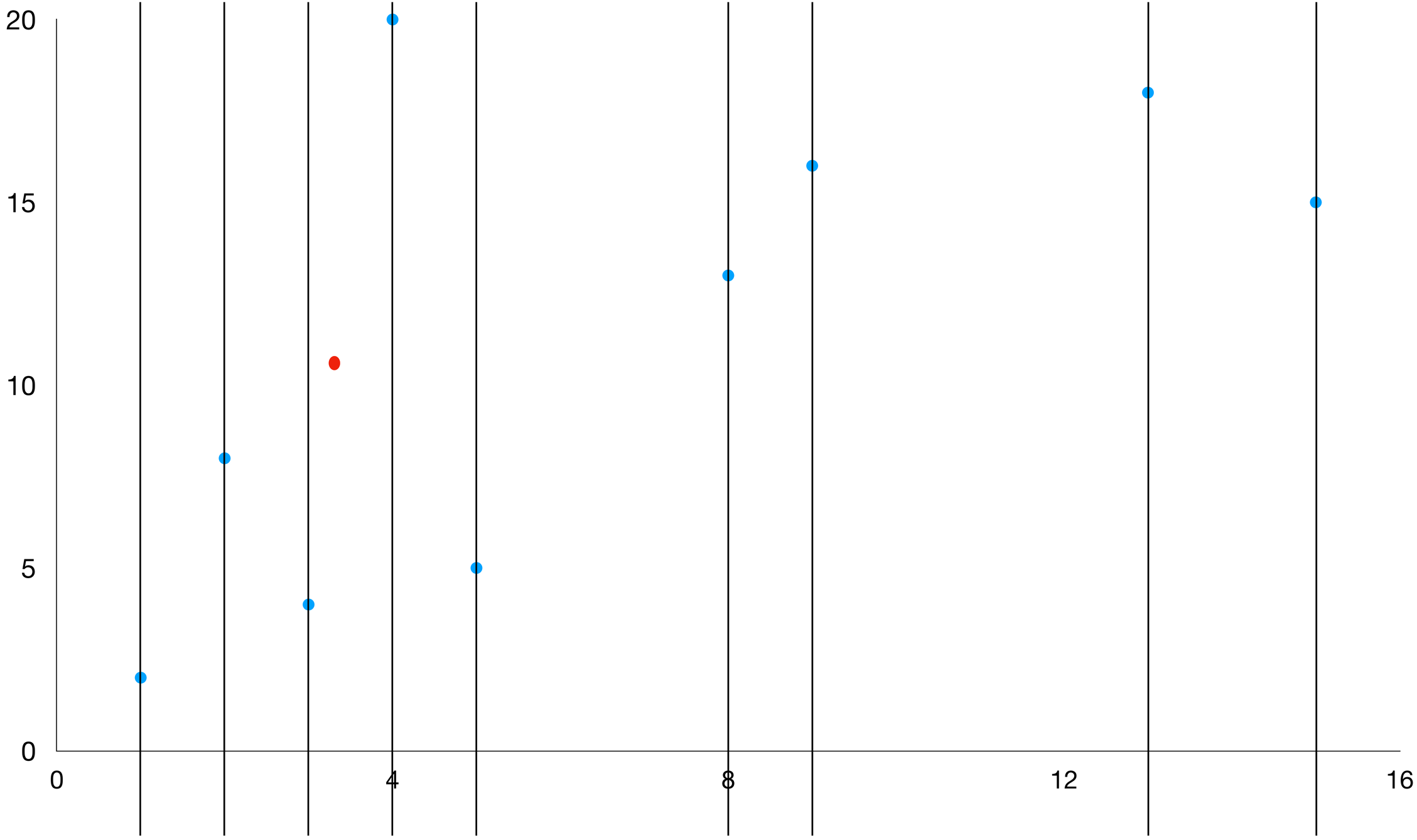
二、平衡KD树

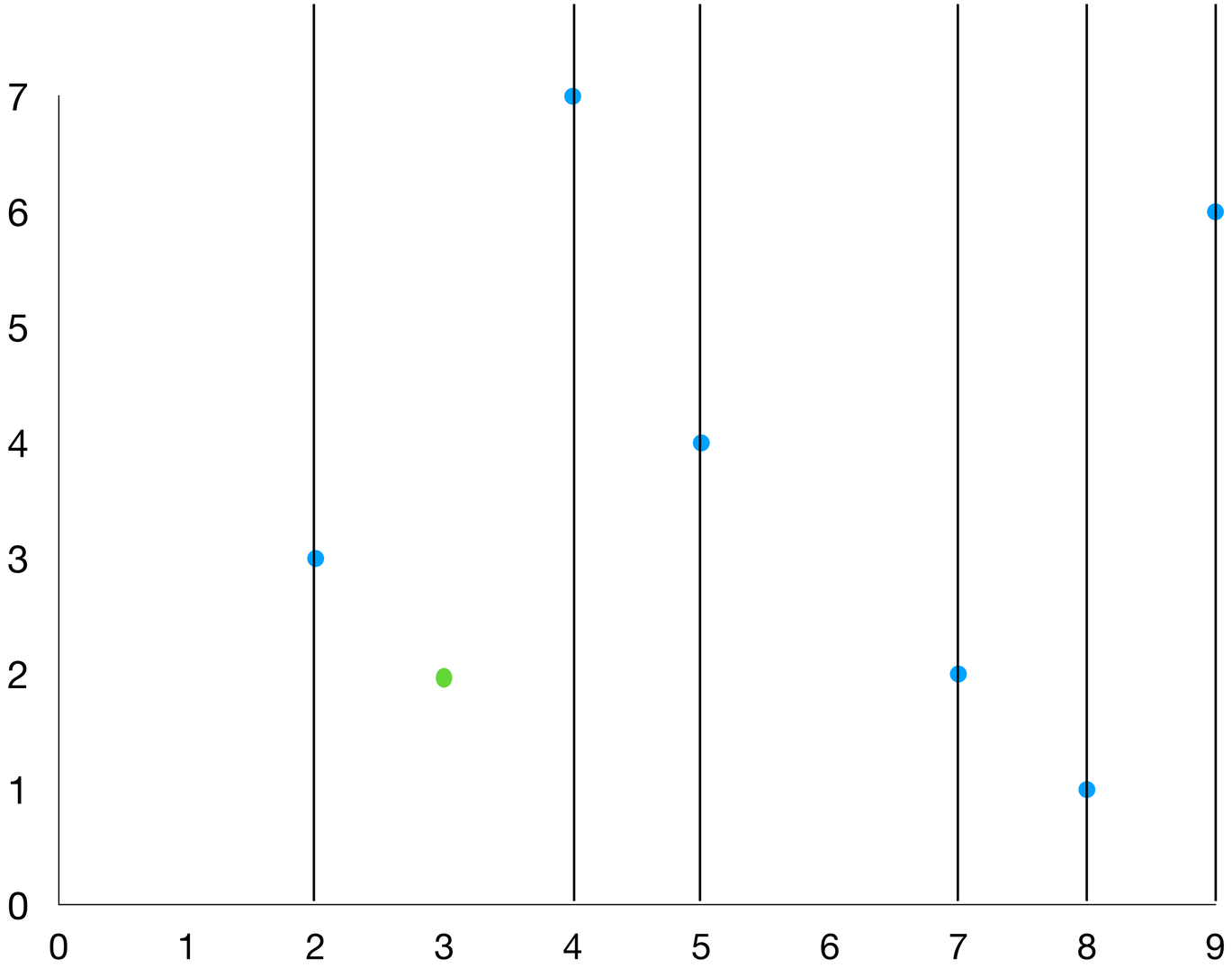


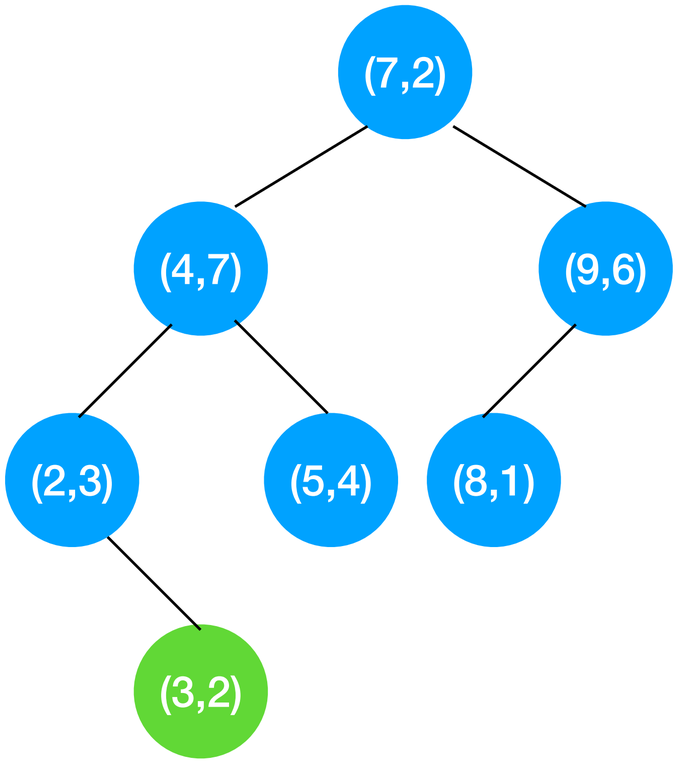
先考虑K=1的情况:


20
15
10
5
0
0
4
8
12
16

思路:

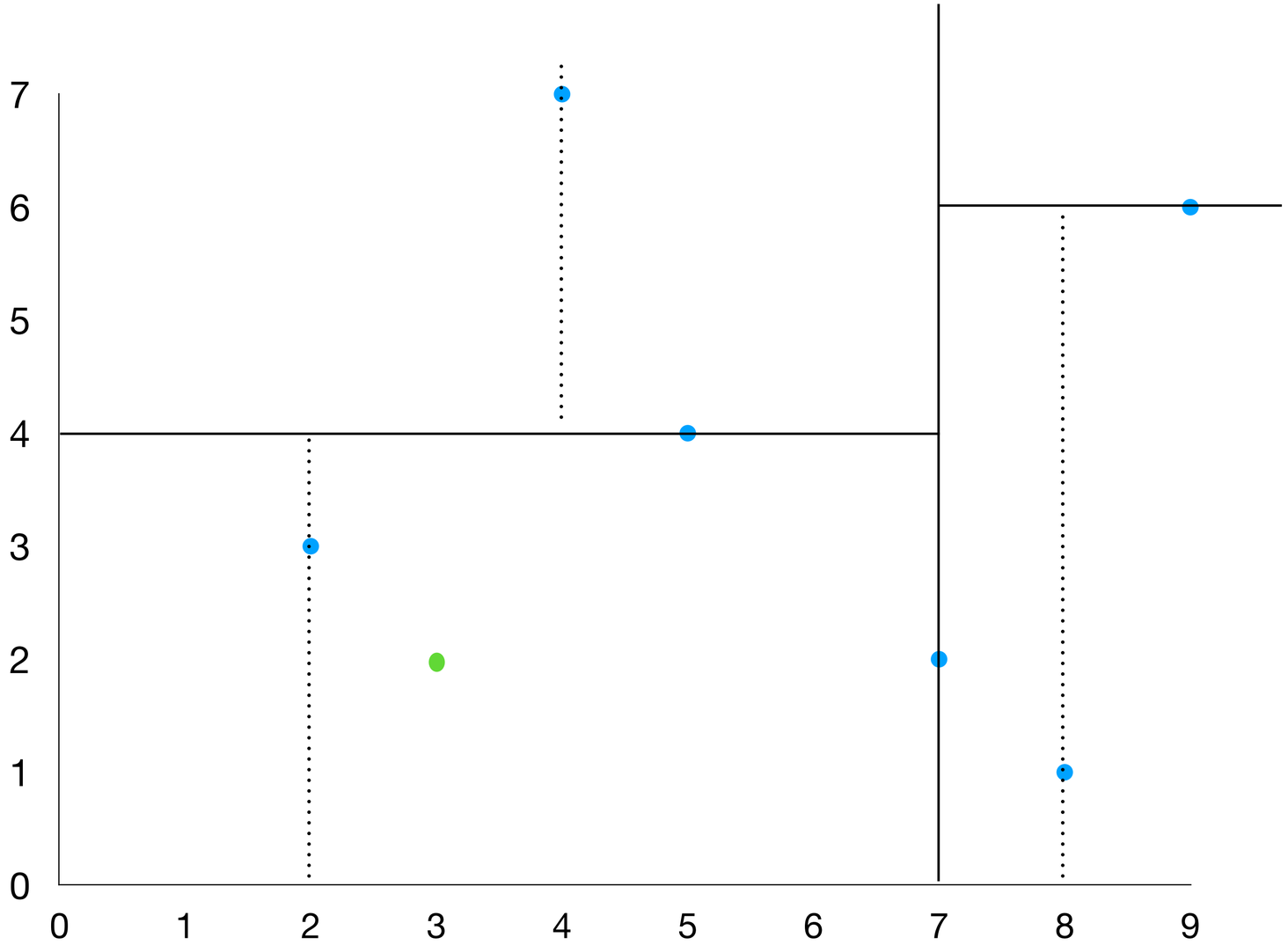


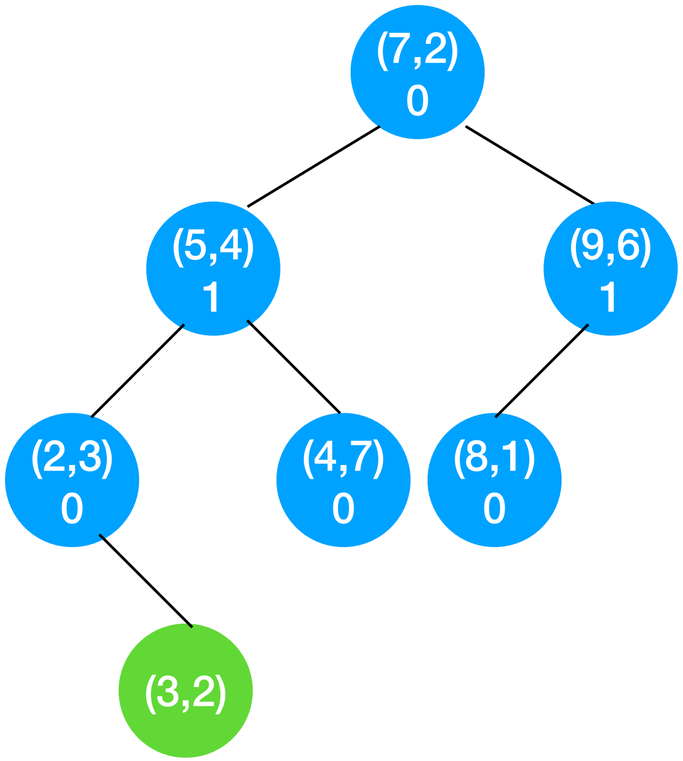




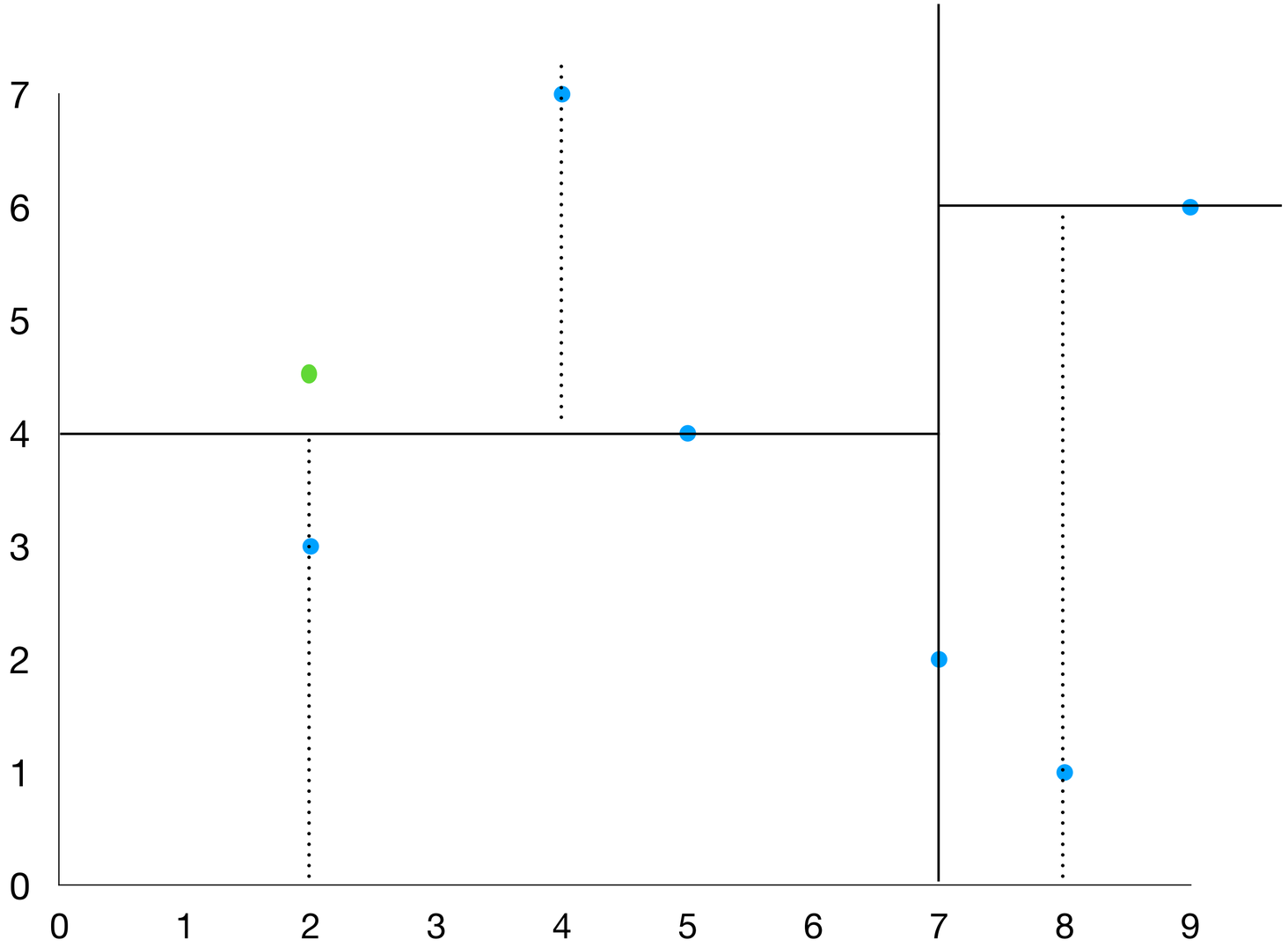
D = inf
P = null

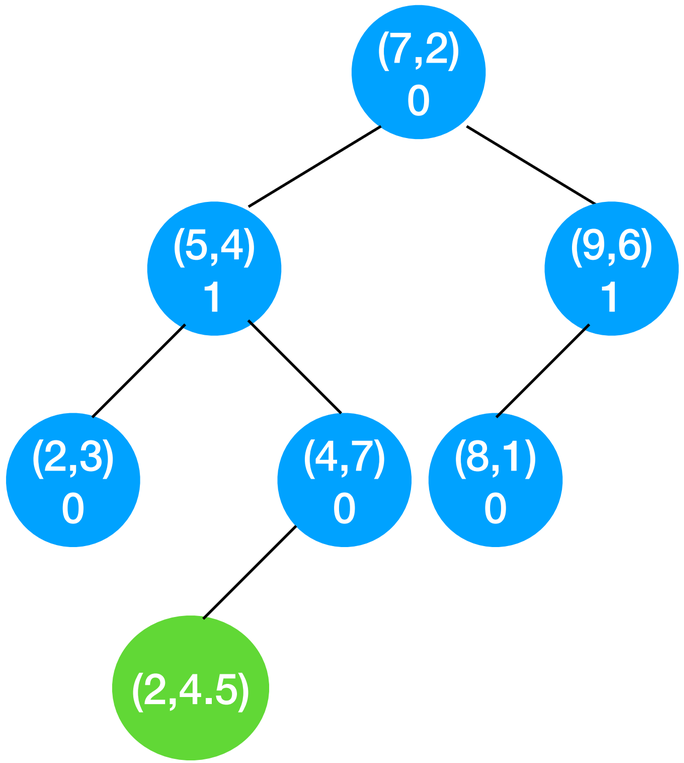
平衡KD树:





D = inf
P = null





D = inf
P = null

K > 1的情况

D1,D2,…,Dk
P1,P2,…,Pk