Advanced Database Systems (CS60113)

Assignment 2. NN Search on R-Tree (Due Date: September 20 2020)

This is a follow up assignment for Assignment # 1.

Write a C program for performing k-NN queries on R-Trees. Use the same R-tree code written by you in Assignment 1. However, consider point data in n-dimensions, i.e., at the leaf level the entries are n-dimensional points. So nearest neighbor will be with respect to the leaf level entries with no need to go to the data in tables for finding NN distances. The coordinates of the data points will vary between 0 and 30 for each dimension. Determine M and m based on the corresponding data. Page size is 4 KB.

Consider 50 random query points. The range of values for each dimension of the query points should be between 10 and 40. Determine the query time and number of nodes visited as average for 50 query points.

Report your results in a table as follows:

For k = 1 (i.e., get the nearest neighbor: 1-NN)

Srl. No.	Number of Dimensions (n)	Average Time taken (msec) (T)	Average no. of nodes/pages accessed (V)	No. of data points (N)	M, m	Linear Search number of pages accessed= N/4096 (L)	L/V
1	2			5,000,000			
2	5			5,000,000			
3	10			5,000,000			
4	50			5,000,000			
5	100			5,000,000			

For k = 5

Srl. No.	Number of Dimensions (n)	Average Time taken (msec) (T)	Average no. of nodes/pages accessed (V)	No. of data points (N)	M, m	Linear Search number of pages accessed= N/4096 (L)	L/V
1	2			5,000,000			
2	5			5,000,000			
3	10			5,000,000			
4	50			5,000,000			
5	100			5,000,000			

For k = 20

Srl. No.	Number of Dimensions (n)	Average Time taken (msec) (T)	Average no. of nodes/pages accessed (V)	No. of data points (N)	M, m	Linear Search number of pages accessed= N/4096 (L)	L/V
1	2			5,000,000			
2	5			5,000,000			
3	10			5,000,000			
4	50			5,000,000			
5	100			5,000,000			

Submit your C program and a pdf file containing the result in the above format.