Design and Analysis of Algorithms 2024
Programming Assignment

- Programs should be written in C and should compile with gcc.
- Your submission should be a zip archive with the programs, compilation instructions and sample input/output.
- Mail your submission to debrup.otro@gmail.com within June 5, 2024. The subject line of the email should be *DAA 2024: Assignment*

1. Write a program to multiply two arbitrary long integers using the Karatsuba multiplication. The following conventions are to be followed:

   (i) The input would be provided in a file with two integers along with their length in bytes. An example input can be as:

   ```
   0x6b8b4567327b23c6643c986966334873 16
   0x74b0dc5119495cff2ae8944a625558ec 16
   ```

   Where two integers of 16 bytes are provided in hexadecimal representation.

   (ii) The output integer should be expressed in hexadecimal and written in a file.

   (iii) Your program should take the input and output file names as command line arguments. No other user input should be taken by your program.

2. You are given a set of points in $\mathbb{R}^n$ and an integer k. You need to find k clusters in the data using MST as discussed in class. The following steps should be followed:

   (i) Find the distances between each pair of the given points.

   (ii) Construct a complete weighted graph where the vertices are the given points and the edge weights are the distances computed in step (i).

   (ii) Use Kruskal's algorithm to find the minimal spanning tree of the graph constructed in step (ii). In the implementation of the Kruskal's algorithm, you should use randomized quicksort as the sorting procedure and for detecting cycles you should use the disjoint set data structure with union by rank and path compression.

   Your program should assume the following:

   (a) The input would be provided in a file whose first line would have the number of data points, the second line would contain the dimensionality of the data, the third line should contain the value of k (i.e., the number of clusters) and the following lines would contain the data. The structure of the input file with 20 points in $\mathbb{R}^4$ for which we need to find three clusters would be as:

```
20
4
3
5.1 3.5 1.4 0.2
7.0 3.2 4.7 1.4
6.3 3.3 6.0 2.5
4.9 3.0 1.4 0.2
6.4 3.2 4.5 1.5
5.8 2.7 5.1 1.9
4.7 3.2 1.3 0.2
6.9 3.1 4.9 1.5
7.1 3.0 5.9 2.1
4.6 3.1 1.5 0.2
5.5 2.3 4.0 1.3
6.3 2.9 5.6 1.8
5.0 3.6 1.4 0.2
6.5 2.8 4.6 1.5
6.5 3.0 5.8 2.2
5.4 3.9 1.7 0.4
5.7 2.8 4.5 1.3
7.6 3.0 6.6 2.1
4.6 3.4 1.4 0.3
6.3 3.3 4.7 1.6
```

(b) Your output should be written in a file. The output file will have the data points along with the cluster name (named 1 to k) in the last column. The output file corresponding to the input file displayed in (a) would be as:

```
5.1 3.5 1.4 0.2 1
7.0 3.2 4.7 1.4 2
6.3 3.3 6.0 2.5 3
4.9 3.0 1.4 0.2 1
6.4 3.2 4.5 1.5 2
5.8 2.7 5.1 1.9 3
4.7 3.2 1.3 0.2 1
6.9 3.1 4.9 1.5 2
7.1 3.0 5.9 2.1 3
4.6 3.1 1.5 0.2 1
5.5 2.3 4.0 1.3 2
6.3 2.9 5.6 1.8 3
5.0 3.6 1.4 0.2 1
6.5 2.8 4.6 1.5 2
6.5 3.0 5.8 2.2 3
5.4 3.9 1.7 0.4 1
5.7 2.8 4.5 1.3 2
7.6 3.0 6.6 2.1 3
4.6 3.4 1.4 0.3 1
6.3 3.3 4.7 1.6 2
```

(c) Your program should take the input and output file names as command line arguments. No other user input should be taken by your program