CSC373 F24: Assignment 1

Due: September 30, by midnight

Guidelines: (read fully!!)

- Your assignment solution must be submitted as a typed PDF document. Scanned handwritten solutions, solutions in any other format, or unreadable solutions will not be accepted or marked. You are encouraged to learn the LATEX typesetting system and use it to type your solution. See the course website for LATEX resources.
- Your submission should be no more than 8 pages long, in a single column US Letter or A4 page format, using at least 9 rkus.teach.cs.toronto.edu/ To submit this assignm

gnment with at most one

ve the same mark on this

h partners in the group

ib -)problem, then the

 This is a group assignr other student. You are should work on and ar

assignment.

 Work on all problems to other student should p

- nsible for understanding all solutions to all prob You may not consult ar ass notes; your textbook and assigned readings with students other than your group partner, is a You may use any data ied in class, or in one of the prerequisites of this course, by just referring to it, and without describing it. This includes any
 - algorithm, or theorem we covered in lecture, in a tutorial, or in any of the assigned readings. Be sure to give a precise reference for the data structure/algorithm/result you are using.
- Unless stated otherwise, you should justify all your answers using rigorous arguments. Your solution will be marked based both on its completeness and correctness, and also on the clarity and precision of your explanation.

Question 1. (11 marks)

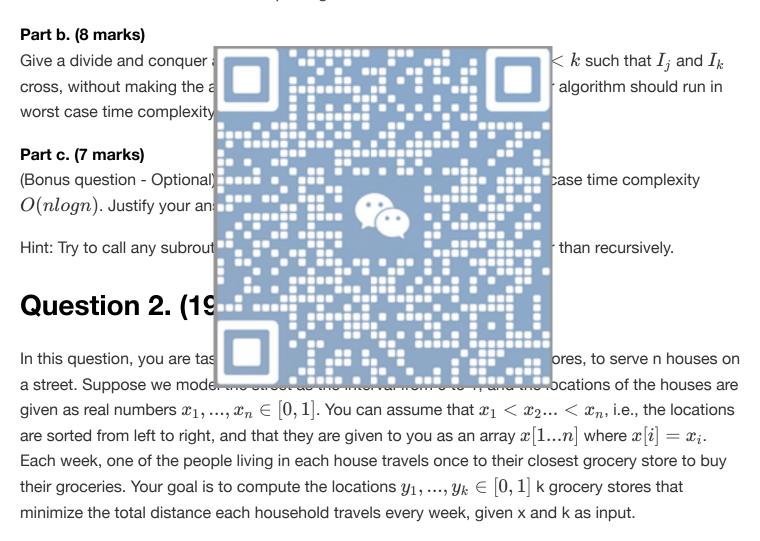
In this question you are given as input n intervals I1 = [a1, b1],..., In = [an, bn] on the real line. Each intervals I_j is specified by the two numbers a_j nd b_j . We assume that $a_j < b_j$ for all j. We also assume that no two intervals share any of their endpoints, i.e., all the numbers $a_1,...,a_n,b_1,...,b_n$ re distinct. The intervals are given in two arrays A[1...n] and B[1...n] where $A[j]=a_j$ and $B[j]=b_j$.

We will say that intervals I_j and I_k cross if $I_j \cap I_k \neq \emptyset$ but neither interval contains the other one. In other words, I_j and I_k cross if exactly one of the endpoints of I_k is contained in I_j .

Part a. (3 marks)

Suppose that there exists some number x so that $a_j < x$ for all j, and $b_j > x$ for all j. Give an algorithm running in worst case time complexity O(nlogn) to compute the number of pairs j < k such that I_j and I_k cross. Justify your answer.

Hint: use one of the divide and conquer algorithms from class.



Part a. (4 marks)

Prove that if k=1, then the optimal location of the single grocery store is the median of $x_1,...,x_n$. I.e., the total distance each household travels to the grocery store at location y $|x_1-y|+...+|x_n-y|$ is minimized by choosing y to be the median of $x_1,...,x_n$. Here we define the median as follows: recalling that $x_1 < ... < x_n$ are sorted, the median is $x_{\lceil (n+1)/2 \rceil}$. So, the median of $x_1 < x_2$

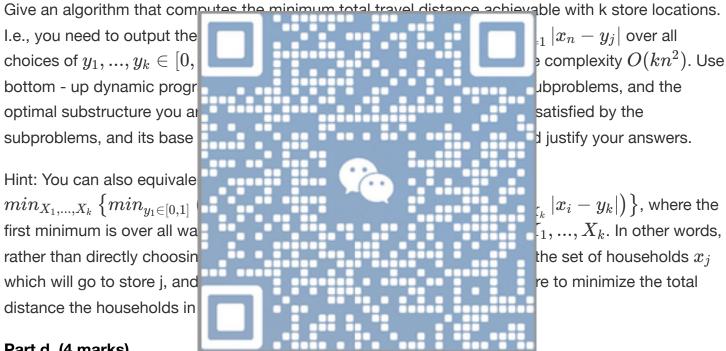
is x_2 , and the median of $x_1 < x_2 < x_3$ is x2 as well. (This may not be how you have seen medians defined for n even; any of the standard definitions would work, but please stick to the one above.)

Hint: There are many ways to prove this. One possibility is to show that, for $x_1 \leq ... \leq x_n$, $|x_1-y|+...+|x_n-y| \geq |x_n-x_1|+|x_{n-1}-x_2|+...+|x_{\lceil (n+1)/2\rceil}-x_{\lfloor (n+1)/2\rfloor}|$, and that the two sides of the inequality are equal when y is the median.

Part b. (5 marks)

For $1 \leq i \leq j \leq n$, let M[i,j] equal $|x_i-y|+...+|x_j-y|$ where y is the median of $x_i,...,x_j$. Give an algorithm that computes all values of M[i,j] for all $1 \leq i \leq j \leq n$ worst case time complexity $O(n^2)$ (i.e., constant time per pair of i and j). Justify your answer.

Part c. (6 marks)



Part d. (4 marks)

Modify your algorithm from the previous problem to also output the choice of store locations $y_1, ..., y_k$ that minimizes the total distance travelled. Your algorithm should still run in worst case time complexity $O(n^2k)$. Justify your answer.

Question 3. (14 marks)

The Galactic Research Society (GRS) has decided to organize an interstellar expedition. The president insists that exactly k members, including herself, join the expedition, and all selected members are required to participate.

There are n GRS members, and they are organized in a strict hierarchical structure (i.e., a tree), with the president at the root. Every society member is represented as a node in the tree. Every member except the president has a supervisor (their parent in the tree), and members supervise at most two other members (their children in the tree).

The society's HR office has determined a "tension coefficient" between each member and their supervisor. This is a real number, and represents the degree to which there is tension between the member and their supervisor: a large positive number means their relationship is quite tense and there is potential for conflict, and a negative number with large absolute value means they get along very well.

Your task is to use a dynamic programming algorithm to choose exactly k members to join the expedition such that the total tension is minimized. You can assume that $k (1 \le k \le n)$ is given to you, and also that the GRS hierarchy is given to you as a rooted binary tree T=(V,E), and each edge $e=(u,v)\in E$ of the tree is labeled by the tension coefficient $t_e=t_{u,v}.$ The total tension of a set $S \subseteq V$ of k GRS mem ${
m I}$ where u and v are both in s. I.e., we add up the tensic r and their supervisor who are both selected for the ex Part a. (7 marks) Define the subproblem stru atisfied by each subproblem, as well as the hould allow you to compute the value of a sub ssuming the values of "smaller" subproblems hav Part b. (3 marks) Using your recurrence from mic programming algorithm to compute the s S members that includes the president (root of the tre ll in the total number of GRS members n. Give pseudocode for your algorithm, and analyze its worst case running time.

Part c. (4 marks)

Modify your algorithm from the previous subproblem to also output a set S of k GRS members, including the president, that minimizes the total tension. Analyze the modified algorithm's worst case running time.