COMP 550.002: Fall 2023
Assignment 4

Announced: October 23, 2023

Due Date: November 6, 2023

All problems are collaborative. You can form a group of at most four students to collaborate. You MUST mention the names of your collaborators, and cite any material you took help from (including discussions on canvas by other students) except the textbook and provided slides.

CLRS refers Cormen et al. textbook.

Problem 1 (15 Points)

A sorting algorithm is stable if equal elements are in the same relative order in the sorted array as in the original array. Which of Insertion Sort, Selection Sort, and Merge Sort are stable, and which are unstable? For each algorithm that is unstable, give an example input array and corresponding output array that shows different relative order of equal elements in the original and sorted array. Keep your example array as small as possible (there should be an array of size 3–4 that would show this). For Insertion Sort and Merge Sort, consider the pseudo-code taught in class. For Selection Sort, refer to the pseudo-code in HW1.

Note that a sorting algorithm being unstable is often a pseudo-code specific issue. For example, we have seen that Counting Sort is stable, but it is possible to change the pseudo-code to make it unstable without altering the key ideas. Also, any unstable sorting algorithm can be made stable with simple modifications. Can you figure out how? (You're not required to answer this.)

Problem 2 (5 Points)

[**This problem is based on CLRS Problem 8.1-1**] What is the smallest possible depth of a leaf in a decision tree for a comparison sort? Justify your answer.

Problem 3 (8 Points)

Illustrate the operation of Counting Sort, according to the pseudo-code given in CLRS, on the array $A = \langle 4, 0, 0, 1, 3, 4, 1 \rangle$. You are only required to show the arrays A, B, and C after each iteration of the **for** loop of lines 11–14.

Problem 4 (7 Points)

Using Fig. 8.3 in CLRS as a model, illustrate the operation of RADIX-SORT on the following numbers: 1259, 48, 513, 6321, 197, 982.

Problem 5 (15 + 15 = 30 Points)

Suppose there are n houses along a long road L. The road L can be viewed as a line segment. The location of each house is given by a real number a_i, which is the distance between the house and the left end-point of the road. The locations of all houses are given in an array $A = \langle a_1, a_2, \cdots, a_n \rangle$.

A wireless service provider wants to set up base stations to cover all houses in A. Each base station can cover a maximum of d miles in each direction from it. The provider has a set of m candidate locations $B = \langle b_1, b_2, \cdots, b_m \rangle$ to place its base stations. The provider wants to minimize its operating costs by setting up as few base stations as possible. You come up with the following two greedy choices to solve this problem.

(i) Select the farthest base station that is within d miles from the left-most house. Remove all houses covered by the selected base station and repeat the process for the remaining houses.

(ii) Select the base station that covers the maximum number of houses. (You can break ties in any way you want.) Remove all houses covered by the selected base station and repeat the process for the remaining houses.

Now, answer the following questions.

(a) Prove or disprove that the greedy choice in (i) is always part of some optimal solution.

(b) Prove or disprove that the greedy choice in (ii) is always part of some optimal solution.

If a greedy choice is always part of some optimal solution, then you should provide a statement as in Theorem 15.1 in CLRS and prove it. Otherwise, you should provide an example (by drawing houses and base station candidate locations along a road) that shows that the greedy choice does not work.

Problem 6 (9 + 6 = 15 Points)

Answer the following questions regarding Huffman encoding.

(a) Suppose we have a text containing three characters a, b, c with frequencies f_a, f_b, and f_c. In each of the following cases (i)–(iii), mention whether such codewords can be produced by huffman encoding for some values of f_a, f_b, and f_c. If such a codeword is possible, then give an example of frequencies f_a, f_b, and f_c that demonstrates the possibility. Otherwise, explain why such a codeword is not possible.

(i) $a = 0$, $b = 10$, $c = 11$.

(ii) $a = 01$, $b = 10$, $c = 11$.

(iii) $a = 0$, $b = 01$, $c = 1$.

(b) Under Huffman encoding, what is the length of the longest codeword for a set of n characters with frequencies $f_1, f_2, \cdots, f_n$? Give an example set of frequencies that would produce this case.

Problem 7 (20 Points)

Implement the Huffman encoding algorithm to determine an optimal encoding of a text file. Details to be provided in a separate document.

Problem 8 (Extra Credit: 15 Points)

[**This problem is based on CLRS 15.1-4**] You are given a set of activities to schedule among a large number of lecture halls, where any activity can take place in any lecture hall. You wish to schedule all the activities using as few lecture halls as possible. Give an efficient greedy algorithm to determine which activity should use which lecture hall.