

# Project: Find All Duplicates

CS3334, 100 Pts

November 10, 2023

## 1 Overview

In this individual project, you are required to explore different methods (at least three) to find out all duplicates in a given list of numbers.

**It is not allowed to use any C++ STL containers (e.g., `std::vector`, `std::set`, `std::map`, and others).**

## 2 Detailed Requirements

In this project, you are asked to implement and compare different methods of finding out all duplicates in a given list of numbers. In particular, you need to provide at least **three** (i.e.,  $\geq 3$ ) different methods and compare their performance with time complexity analysis. For example, to find all duplicates in the list  $\{7, 4, 5, 9, 5, 8, 3, 3\}$ , you can try using array based brute force approach or binary search tree to find out the targeted output  $\{5, 3\}$ .

In your submitted report, you need to explicitly explain:

- the algorithms and data structures of the methods you have tried;
- the time complexity of each method (average time complexity, worst-case time complexity, etc.);
- the input scenario causing the worst case for each method.

Note that you are NOT required to provide formal time complexity analysis, but some intuitive or informal discussion. The challenging part would be designing some input data that trigger the worst-case running time of each method.

### Experimental Evaluations

In this project, you need to design and generate the test data for each method. Then you are required to evaluate and compare the differences in efficiency among different methods over different generated inputs. Figures that intuitively show how the input data affects the selection among different methods are appreciated. Since theoretical analysis is too sophisticated, experiments would be of great importance.

Note that, the compiling command for compiling your programs (written in C++) should be clearly stated in the report. Some available compiling options are `-std=c++11`, `-lm`, `-O2`.

If you **cannot pass the OJ test cases**, the maximum marks you can get is 30. If your codes **containing the direct usage of C++ STL containers** for one of the methods, you will get 0 mark in this project.

### 3 OJ Input and Output

You can check this material in Question 897 on the OJ system.

#### Input

The input contains multiple test cases. For each test case, it contains a line of integers (the number of integers is less than  $10^4$  while each integer is  $n \in (-10^8, 10^8)$ ).

#### Output

Each line of the output is the list of all duplicates. If one integer has duplicates, it is printed out once (the first time it appears; so the order of the output matters).

Sample Input	Sample Output
-1 1 -1 8	-1
1 2 3 3 3 4 4 5	3 4
2 3 1 5 4 3 2 1	2 3 1

### 4 Due Date and Submission

- The project is due at **23:59 on Dec 7, 2023**.
- Test your solutions by submitting your codes to the Online Judge system (Question 897).
- Submit a zip file (size  $\leq 5\text{MB}$ ) onto Canvas, containing all your materials, i.e., 1) codes (including all codes submitted to the OJ), 2) generated data (including files containing specially constructed data and indicate in your report how to use them), and 3) the final report.