Data Structure & Algorithms
CS210A
Semester II, 2015-16, CSE, IIT Kanpur

Theogramming Assignment

Deadline : 11:55 PM on 8th April

Note: This assignment has the flavor of a programming assignment as well a theoretical assignment. That is why, it is named as a *theogramming assignment*. Being the last assignment, it has to be truly inspiring and memorable. You are given the choice of two problems. Attempt exactly one problem. In case you face any problem in understanding these problems or their solution, you may contact the instructor on or before 3rd April. I

Inter-group collaboration is **strictly prohibited** for this assignment.

1 Implementing the median finding algorithm

(*Marks : 60*)
Implement the linear time algorithm for finding median of a set of n numbers that we discussed in the class. You need to find experimentally the size of the *small* arrays for which we compute median in a brute force manner. In particular, which of 3,5,7,9, proves to be the best size ? In addition you need to demonstrate that the running time is indeed $O(n)$.

2 Thinking beyond the boundaries

(*Marks: 100*)
Though it is known that we can compute median in $O(n)$ time, it is natural to ask for the minimum number of comparisons needed to compute median of n numbers. Researchers pursued this question and proved that this number is surely greater than $2n$ for any algorithm. In other words, for every algorithm that computes median of a given set using comparison as the basic operation, there exists an input of n numbers on which the algorithm must perform more than $2n$ comparisons. Does that mean we should stop searching for an algorithm that performs close to n comparisons ? Think over this question for at least 30 minutes and in total silence...

Sampling is a very powerful technique. In real life also, we have seen the power of this technique. When we go to a grocery store with many varieties of rice lying in big and open sacks, how do you find the best quality rice ? Well, we take a handful of sample from each sack and judge their quality. The underlying intuition is that a sample, picked uniformly, will reflect the true nature of the original population. This idea, which we imbibed during our childhood, has played a crucial role in many applications, including the design of efficient algorithms.

You need to design and implement an algorithm that makes use of this sampling to break the barrier of $2n$. Basically, for any input, this algorithm is supposed to perform much less than $2n$ comparisons *almost always*. You have to make use of the following hint. Did you get inspired by the linear time algorithm for median that we discussed in the class ? If yes, could you realize that the key idea was to prune away a large chunk of numbers so that we can avoid comparing them. You have to use the tool of sampling to achieve essentially the same objective. You have to demonstrate experimentally that you are able to break the barrier of $2n$ with a huge margin. In addition, you have to make a sincere attempt to analyze this algorithm theoretically.

At this stage, some of you may start feeling that this assignment is beyond the scope of this course. But, rest assured, the tools needed to achieve all the goals of this assignment are very well covered in this course. All, that is needed to solve this problem, is creativity combined with a little bit of analytical skills. Of coure, each one of you is equipped with these tools.

Note: The description of the second problem is indeed open ended, and it is done so intentionally. Moreover, the expectations of the instructor are also very high for this assignment. Therefore, if you are not confident about it, you are advised to attempt the first problem. It is also known that many students resort to internet whenever they get stuck on a problem. If you are desparate for marks, even at the cost of such unfair means, go ahead and take such short cuts (searching on the internet or seeking hint from another classmate). But those students, who might consider research as their future plan with non-zero probability, should refrain from it. This assignment is a golden opportunity for them to get a flavor of what research is about. Such students should confine themselves to their group partner and meditate like a true yogi over the problem.

Grading: You have to upload the code for this assignment on the website. The input output format of this assignment will be announced by Tuesday (29th March). You will be assigned a time slot of 15 minutes for demo and viva starting from 9th April.