Instructions.

1. You could submit a pdf of the homework, either printed or hand-written and scanned, as long as it is easily readable.

2. If your solution is not clearly written, it might not be graded.

3. Prove the correctness of your answer. A correct answer without a proof might not be accepted.

4. If you have discussed the solution with other students, mention their names clearly on the homework. These discussions are not forbidden and are actually encouraged. However, you must write your whole solution yourself.

5. All questions have same weight.

6. You could refer to a data structure studied in class, and just mention briefly their guaranties. For example “It is known that a Red-Black tree could support the insert/delete/find operations on a set of $n$ elements in time $O(\log n)$. ”

7. If your answer uses one of the data structures or algorithms that were studied in class, you could refer to it without having to repeat details studied in class. If your answer requires only modifications of one of the algorithm, it is enough to mention the required modifications, and what’s the effect (if any) on the running time and on other operations that the algorithm performs.

8. In general, a complete answer should contain the following parts:

   (a) High level description of the data structures (if needed). E.g. We use a binary balanced search tree. Each node contains, a key and pointers to its children. We augment the tree so each node also contains a field...

   (b) High level description of the algorithms

   (c) Proof of correctness (why your algorithm provides what is required).

   (d) A claim about the running time, and a proof showing this claim.

9. When discussing trees, do not confuse the children of a node with the descendants of the node.
1. Your friend (same one from hw1) has this time dropped you at a point $p$ in the desert. You wish to walk to the nearest road, from which you could hitchhike back home. However, you have no idea in which direction you should head.

Let $q$ be the nearest road point, and let $d$ be the distance from $p$ to $q$. Explain how you could walk a distance of $c \cdot d$ and reach a road point (not necessarily $q$). Here $c$ is a constant (does not depend on $d$) that you have to compute.

For simplicity, assume that every road that crosses the desert is an infinite straight line. Each road is either horizontal (oriented East-West) or vertical (South-North).

Comments:

(a) Assume that the desert is flat, so every path is possible. Also assume that you are equipped with an accurate GPS, showing you location as well as the direction to North. You have no access to maps.

(b) For simplicity, assume that you need to be physically on the road to see it. So ignore range-of-visibility issues.

(c) Obviously, roads could cross each other, and if they do, they are orthogonal to each other.

(d) If your solution is iterative, explain what exactly is executed at the $i$’th iteration (for every $i$)

Figure 1: The roads network. Note that roads are not equally-spaced
2. Due to reasons not to be disclosed here, but related to the very same friend from the previous question, you have found yourself at a point \( p \) in a lake. Crocodiles are habituating the lake, strongly suggesting that you’d swim to the nearest point on the lake’s shore before becoming a crocodile treat. As before, you have a GPS, but no map of the lake. Suggest your strategy, so the distance you have to swim is at most \( c \cdot d \). Here \( d \) is the distance from your starting position \( p \) to the nearest point \( q \) of the shore of the lake, and \( c \) is a constant (not depending on \( d \)).

Note that the lake is not convex, but there are no island in the lake, so its shore is continuous. Show the best value of \( c \). You do not have a map, and you don’t know the crocodiles’ locations.

Note also that you do not have to find \( q \) itself. As usual, if your solution is iterative, describe what exactly happens at the \( i \)th iteration (for an arbitrary \( i \)).

You could not use techniques proposed at [www.youtube.com/watch?v=kmHOPP_zAKo](http://www.youtube.com/watch?v=kmHOPP_zAKo). (Thanks to Jeremy M. for the insight).

Figure 2: The lake. You have landed at \( p \), and wish to swim to the shore. The nearest point of the shore is \( q \).

(credit: ClipPanda.com)

3. You are inserting \( n \) different keys into a SkipList \( L \). \( L \) was an empty Skiplist before the insertions took place. What is the probability that after all insertions took place, all the keys appear only on level 1, but none of them appears in level 2?

What is the probability that all keys appear in level 2, but none of them appears in level 3?

4. In this question, we are discussing a SkipList \( L \) created by inserting a set \( S \) of \( n \) keys into an (initially empty) Skiplist. As you know, in a skip list, a key that appears in level \( i \) appears also in level \( i + 1 \) with probability \( p = 0.5 \) (that is, a key has a 50% chance to be promoted to the next level).

Now assume that we re-create a Skiplist by inserting the same keys, in the same order, but this time the probability \( p \) is 0.01. Will the time to perform \( \text{find}(x) \) operation increase or decrease, compared to time for the same operation in the original SkipList?

Repeat the question, but now assume \( p = 0.9 \).

5. Prove that in the ‘perfect’ SkipList discussed in class, the worst-case time for a query is \( O(\log n) \) where \( n \) is the number of keys. You can refer to the slides at [www.cs.arizona.edu/classes/cs445/spring16/SL.pdf](http://www.cs.arizona.edu/classes/cs445/spring16/SL.pdf).
6. Let $S = \{p_1 \ldots p_n\}$ be a set of $n$ points in the plane. For example, in Figure 3, $S = \{p_1 \ldots p_4\}$. Each point $p_i$ is given by its $x$ and $y$ coordinates $(x_i, y_i)$. Assume that if $i \neq j$ then $x_i \neq x_j$.

(a) Suggest a data structure that supports the following operations:

i. **Insert** $(x_i, y_i)$. Insert a new point whose coordinates are $(x_i, y_i)$ into $S$.

ii. **Delete** $(x_i, y_i)$. Remove the point whose coordinates are $(x_i, y_i)$ from $S$.

iii. **Sum** ($q$) returns the sum of the $y$-coordinates of all the points of $S$ whose $x$ coordinates are strictly smaller than $q$. That is, the sum of $y$-coordinates of the points of $S$ which are to the left of the vertical line $x = q$. If there is no such point, return zero.

The time for each operation should be $O(\log n)$. 

![Figure 3: $S = \{p_1 \ldots p_4\}$](image-url)
**Example:** For the set of points in Figure 3

i. Sum(2) returns 3

ii. Sum(4) returns 3 + 1 = 4. The line \( x = 4 \) is highlighted, and the points to its left are \( p_1 \) and \( p_2 \).

iii. Sum(6) returns 3 + 1 + 2 = 6,

iv. Sum(8) returns 3 + 1 + 2 + 4 = 10

![Figure 4:](image)

(b) Next, in addition to the insert and sum operations (and without jeopardizing their running time) also support the operation \( \text{highest}(a, b) \) that should return, in \( O(\log n) \) time, what is the point with the maximum \( y \) value, among the points whose \( x \) coordinates are between \( a \) and \( b \). That is \( (x_i, y_i) \in S \) should be considered only if \( a \leq x_i \leq b \). There are the points “sandwiched” between the line \( x = a \) and the line \( x = b \). This region is depicted in light gray in Figure 3, for the lines \( x = 1.1 \) and \( x = 4 \).

Example: In Figure 4, \( \text{highest}(1.1, 4) \) returns \( p_3 \), while \( \text{highest}(2, 6) \) returns \( p_7 \).

7. Suggest a data structure that supports the following operations on the grades that a student at the university receives. Assume that the total number \( n \) of exams that the student might take is large. The data structure should support the following operations: (each in \( O(\log n) \) time, where \( n \) is the number of exams taken).

   (a) \text{insert}(\text{exam date}, \text{exam grade}). \) This operation informs the data structure that the student received at date \( \text{exam date} \) the grade \( \text{exam grade} \).

   (b) \text{average}(\text{date}_1, \text{date}_2). \) As a response to this operation the data structure should answer what is the average of all grades that are received between the dates \( \text{date}_1 \) and \( \text{date}_2 \). Assume again that all dates are different, and for simplicity, assume that no exam took place in \( \text{date}_1 \) nor in \( \text{date}_2 \).

Hint: Start by solving the simpler task of being able to report the number of exams that took place between \( \text{date}_1 \) to \( \text{date}_2 \).

8. Suggest a data structure that helps Tucson keep track of its population. The operations that are needed from the data structure are
(a) **Joined**(*x, date*). This operation informs the data structure that a person whose Social Security Number is *x* has moved to Tucson on data *date*. Note that this date might be in the past.

(b) **Left**(*x, date*). This operation informs the data structure that a person whose Social Security Number is *x* has left Tucson on data *date*. Note that this date might be in the past.

(c) **Population**(*date*). The data structure needs to report *how many* people were living in Tucson on data *date*. That is, what was Tucson population at this date.

Before any operation took place, assume that Tucson is empty. Note that these operations could be given in any order (not necessarily chronologically). Each operation should be performed in \(O(\log n)\) time, where *n* is the total number of people ever lived in Tucson. Once a person has left Tucson, he/she could not move back.

9. What is the running time (in terms of \(O(\cdot)\) and \(\Omega(\cdot)\)) of the median selection algorithm studied in class, if during the partition algorithm, the pivot is always selected as one of the 5 smallest keys in the array? Assume that *n* is an arbitrary large number.