CS 222A winter 2011, Homework 1, due tuesday Jan. 11.

1. Solve the following recurrence by unwrapping and by use of the Master method (see the notes on the class website if you don't know these methods):

$$T(n) = 7T(n/4) + n^2$$
$$T(1) = 1$$

2. Review Big-Oh, Ω and Θ notation. Then answer the following questions:

Sometimes people say something like "Algorithm X takes at least $O(f(n))$ time in worst case", where $f(n)$ is some function of input size n, for example $O(n^2)$. Explain why that statement is vacuosly true, and therefore useless. Of course, the people who made that statement did not intend to say something vacuosly true. What is the correct statement they intended?

This week, when Gov. Cuomo of NY was sworn in, he said something like "one cannot underestimate the seriousness of the budget deficit". What does this statement have to do with the above question?

3. Solve Problem 4 on page 247 of the book.

4. This problem is couched in terms of liars and truth tellers, but it has real applications in identifying which componants of a complex system are good (functioning correctly) and which are faulty.

Assume we have a community of n people and *we know* an integer number $t < n/2$, which has the property that most t of the n people are liars. This does not say that there actually are t liars, but only that there are at most t liars. We will identify the liars in the community by successively picking pairs of people, (X, Y) say, and asking X: Is Y a liar?. The response is either "yes" or "no"; we assume that the truth-tellers are always truthful and correct, but in responding to any question, a liar may or may not be telling the truth. If we know for sure that person X is a truth-teller, then we can find all the liars by simply asking person X his opinion about everyone else. The problem is that in general we don't know whether a given person X is a liar or not, and so interpreting his responses can be tricky.

Our goal as computer scientists is to find an algorithm that chooses pairs of people X, Y (where X gives an opinion of Y) so that we can identify the liars with as few questions as possible. Our problem as mathematicians is to state an upper bound on the number of questions that are needed in worst case to identify the liars in a n person community that contains at most $t < n/2$ liars. Note that in this problem, the algorithm can be *adaptive*, i.e. it does not select all the pairs first and then ask the questions, but rather it chooses the first pair, asks a question and gets a response, then chooses the second pair, gets a response etc. At every iteration, the algorithm can use all of the responses obtained so far in deciding which next pair to pick. Later we will look at a non-adaptive version of this problem.

Below we will give an algorithm that will always identify the liars with at most $n + 2t - 1$ questions. It turns out that there are methods that do better, never using more than $n + t - 1$. Those methods are more complex and may be included in future homeworks.

We introduce the ideas behind the algorithm. You fill in the details and explain why they are correct.

The algorithm will use the responses to its questions to build up two disjoint sets of people, A and B, which will have the following properties: at any point in the algorithm, set A will have a distinguished member, denoted y, such that if y is a liar then everybody in A is a liar; set B will have the property that at least half the members of B are liars. Hence if A and B grow to a point where $|A| + |B|/2 > t$, then we know that y, the distinguished member of A, must be a truth-teller; if he was instead a liar, then every member of A is a liar, and since at least half the members of B are liars, there would be more than t liars, a contradiction. Once y is identified as a truth-teller, the algorithm uses it to identify all the liars and truth-tellers.

The algorithm chooses pairs so that $|A|+|B| = 2$ after the first question, and $|A|+|B|$ grows by at least one with each additional question and response. Hence after at most $2t$ questions $|A| + |B|/2$ will be greater than t (give a clean explain of this), and the distinguished member of A will be identified as a truth-teller. After an additional $n-1$ questions, all the n people will be properly classified, hence the method takes at most $n + 2t - 1$ questions.

Devise an algorithm that has the above properties and that solves the liars problem. Explain why it is correct. As a hint, I am thinking of an algorithm where, both X and Y are put into set B (and X is moved from A if it is in there) if X says Y is a liar.

Preview question for next homework, due in two weeks: Try to devise an algorithm that uses only $n + t - 1$ queries.