CIS 455 – Homework #1

Due on myCourses by 11:59pm on Sunday, Sept. 20, 2015

Instructions: There are 6 problems worth a total of 100 points.

Submit on myCourses: solutions to the written parts, hard copies of any source code and hard copies of sample runs which thoroughly verify the correctness of your code. Other than for Problem 1-6, you may write your code in whichever programming language you prefer (so long as you can demonstrate that your code runs correctly).

Document and indent your programs properly. You will be graded on both your solutions and your ability to show their correctness.

If you feel it would help, you are encouraged to work together on homework. But remember that you must submit your own work, as the point of the homework is to learn the material. *If you do work with others on homework, you must write the names of those you worked with on your homework.*

Late Homework will be penalized!

Problem 1-1. (15 points) Jones & Pevzner, Problem 2.1, page 54.
Write an algorithm that, given a list of n numbers, returns the largest and smallest numbers in the list.
Estimate the running time of the algorithm.
Can you design an algorithm that performs only $3n/2$ comparisons to find the smallest and largest numbers in the list?

Problem 1-2. (20 points) Jones & Pevzner, Problem 2.2, page 54.
Write two algorithms that iterate over every index from $(0, 0, \ldots, 0)$ to $(n_1, n_2, \ldots, n_d)$.
Make one algorithm recursive, and the other iterative.

Problem 1-3. (15 points) Jones & Pevzner, Problem 2.3, page 54.
Is $\log n = O(n)$?
Is $\log n = \Omega(n)$?
Is $\log n = \Theta(n)$?

Problem 1-4. (20 points) Adopted from: Jones & Pevzner, Problem 2.17, page 56.
There are n bacteria and 1 virus in a Petri dish. Within the first minute, the virus kills one bacterium and produces another copy of itself, and all of the remaining bacteria reproduce, making 2 viruses and $2 \cdot (n-1)$ bacteria. In the second minute, each of the viruses kills a bacterium and produces a new copy of itself (resulting in 4 viruses and $2(2(n-1)-2) = 4n-8$ bacteria; again, the remaining bacteria reproduce. This process continues every minute.

Will the viruses eventually kill all the bacteria?
If so, design an algorithm that computes how many steps it will take *and submit sample runs showing the output of your code*.
How does the running time of your algorithm depend on n?

Problem 1-5. (10 points) Adopted from: http://www.bioalgorithms.info/problems/00_asmt.htm.

Searching a Nucleotide Sequence Database

1. Go to the following web page:
http://www.ncbi.nlm.nih.gov/Class/MLACourse/Modules/BLAST/q_jurassicparkDNA.html
2. Copy the DNA sequence from the book Jurassic Park.
3. Go the NCBI Blast home page at http://www.ncbi.nlm.nih.gov/BLAST/. Go to the link that says "nucleotide BLAST".
4. Copy and paste the DinoDNA DNA sequence into the text box, make sure the "nucleotide collection (nr/nt)" database is selected and hit the "BLAST" button at the bottom.
5. Sort the results by "Total score".
    1. What is the highest score?
    2. What is the description of the result with the highest score?
    3. What is the query coverage of this result?
    4. Is the DNA sequence in Jurassic park fictional (i.e. made up/random) or "borrowed" (i.e. copied from real DNA)?

Problem 1-6. (20 points)
Write a Matlab program that does something fun.
This is your choice entirely -- it can be original or from http://www.mathworks.com/products/bioinfo/ or some other website. You won't be graded on how much fun I find it.
Describe the program in a sentence or two, give a sample input and output, and include a snapshot/screenshot of it in action.
Matlab is available in Dion 303 and on the Virtual Computing Lab: http://www.umassd.edu/cits/vcl