CIS 455 – Homework #4

Due on myCourses by 11:59pm on Friday, April 13, 2018

Instructions: There are 5 problems worth 100 points total.

Submit on myCourses: solutions to the written parts, *all source code*, and screenshots of sample runs which thoroughly verify the correctness of your code. Please *rename and upload your source code as a **.txt file.***

Document and indent all your programs properly. You will be graded on both your solutions and your ability to show their correctness. Do **not** zip up your code, since I can't comment directly in your .doc/.pdf on myCourses.

Submitting modified versions of other people's work as your own (*or copying from websites*) is considered cheating and will result in an "F". 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 5-1. (10 points) Jones & Pevzner, Problem 6.6, page 212.
Find the number of different legal paths from *source* (0,0) to *sink* (n,m) in an n by m rectangular grid, and explain your answer.

Problem 5-2. (15 points) Jones & Pevzner, (*modified version* of Problem 6.14, page 213).
Two players play the following rock game with two piles of rocks of heights n and m.

At every turn a player must take two rocks from one pile (either the first pile or the second pile) **and** one rock from the other. The player who cannot complete their turn loses.

1. Who will win? Using dynamic programming, describe a general winning strategy for each n and m. (10 points)
2. Show the details of the winning strategy for n=m=6. (5 points)

Problem 5-3. (20 points) Jones & Pevzner, Problem 6.18, page 214.
What is the optimal global alignment for the string MOAT and the string BOAST?
Show all optimal alignments and the corresponding paths using the scoring matrix below and indel penalty -1.

	A	B	M	O	S	T
A	1	-1	-1	-2	-2	-3
B		1	-1	-1	-2	-2
M			2	-1	-1	-2
O				1	-1	-1
S					1	-1
T						2

(12 points for the correctly filled table; 5 points for the correct alignment; 3 points for the correct score)

Problem 5-4. (25 points) Jones & Pevzner, Problem 6.20, page 214.

Consider the sequence **v = TACGGGTAT** and **w = GGACGTACG**.
Assume that the match premium is **+1** and that the mismatch and indel penalties are **-1**.

1. (5 points) Fill out the dynamic programming table for a ***global*** alignment between **v** and **w**.
   a) (3 points) Draw arrows in the cells to store the backtrack information.
   b) (2 point) What's the score of the optimal global alignment & what alignment does this score correspond to?

2. (5 points) Fill out the dynamic programming table for a ***local*** alignment between **v** and **w**.
   a) (3 points) Draw arrows in the cells to store the backtrack information.
   b) (2 point) What's the score of the optimal *local* alignment and what alignment does this score correspond to?

3. (5 points) Suppose we use a gap penalty where it costs -20 to open a gap, and -1 to extend a gap. The scores of matches and mismatches are unchanged (still +1 and -1).
   What is the optimal ***global*** alignment in this case and what score does it achieve? (**Hint:** You do ***not*** need to refill in a dynamic programming table to solve this!*)*

Problem 5-5. (30 points)
Write a program that solves the *Local* Sequence Alignment problem using dynamic programming. (20 points)

The input includes strings ***v***, ***w*** and a scoring matrix $\boldsymbol{\delta}$.
The output is an alignment of ***v*** and ***w*** whose score (as defined by the matrix $\boldsymbol{\delta}$) is maximal among all possible alignments of ***v*** and ***w***.

1. Run your program to find the optimal ***local*** alignment for **1213434222** and **1343422421** under the match premium +1, mismatch penalty -1, and indel penalty -0.5.

2. (10 points) Submit your source code and screenshots of a sample run showing the correct output of your algorithm when given ***v*** = **1213434222** and ***w*** = **1343422421**

Submitting modified versions of other people's work as your own (*or copying from websites*) is considered cheating and will result in an "F". 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.*