Feedback — Homework 4

You submitted this homework on **Sat 7 Nov 2015 3:43 AM PST**. You got a score of **100.00** out of **100.00**.

In this Homework, we will consider several interesting problems that can be solved using dynamic programming. For the main part of the Homework, we will consider the problem of computing both global and local alignments of two sequences using dynamic programming.

You may wish to review these notes on the problem of sequence alignment before you begin the Homework.

Question 1

Questions 1-3 will test your understanding of the feasibility conditions of global pairwise sequence alignment (that is, what makes an alignment). Let $X = AC$ and $Y = CA$ be two DNA sequences. Which of the following are global alignments of X and Y?

Your Answer		Score	Explanation
☑ $X' = AC$ and $Y' = CA$.	✔	1.00	
☑ $X' = AC--$ and $Y' = --CA$.	✔	1.00	
☐ $X' = AC$ and $Y' = --CA$.	✔	1.00	
☐ $X' = CA$ and $Y' = CA$.	✔	1.00	
☐ $X' = CA-$ and $Y' = -AC$.	✔	1.00	
Total		5.00 / 5.00	

Question Explanation

Remember that deleting the dashes from the global alignments should yield the original sequences.

Question 2

How many possible global alignments are there for two sequences X and Y if $|X| = |Y| = 1$?

How many possible global alignments are there for two sequences X and Y if $|X| = |Y| = 2$?

Give your answer below as two numbers separated by a space (e.g., 17 100)

You entered:

3 13

Your Answer		Score	Explanation
3	✔	2.50	Correct. There is one global alignment of length one and two global alignments of length two.
13	✔	2.50	Correct. There is one global alignment of length two, six global alignments of length three, and six global alignments of length four.
Total		5.00 / 5.00	

Question Explanation

Try working out the number of possible global alignments for each possible length of the alignment.

Question 3

Recall from the feasibility conditions of a sequence alignment that for X' and Y' to be an alignment of two sequences X and Y, X' and Y' must satisfy this condition (among other conditions):

- There does not exist an i such that $X'_i = Y'_i = -$.

What happens if we remove this condition; that is, allow $X'_i = Y'_i = -$ for some values of i?

Your Answer	Score	Explanation

○ The number of possible alignments of a pair of sequences

doubles (as compared to the number of alignments that satisfy the condition).

○ The number of possible alignments of a pair of sequences does not change (as compared to the number of alignments that satisfy the condition).

● The number of possible alignments of a pair of sequences becomes infinite.

✔ 5.00   Correct. The resulting alignment could contain as many pairs of matching dashes as we wish.

○ The number of possible alignments of a pair of sequences becomes smaller (as compared to the number of alignments that satisfy the condition).

○ The number of possible alignments of a pair of sequences increases exponentially in the lengths of X and Y (as compared to the number of alignments that satisfy the condition).

Total   5.00 /
                                 5.00

Question 4

Questions 4-5 will test your understanding of the optimality of <u>global</u> pairwise sequence alignment (that is, what the score of an alignment is and what makes an alignment optimal).

Consider the scoring matrix M for alphabet $\Sigma = \{A, C, T, G\}$ with the following entries:

- $M_{\sigma,\sigma} = 5$ for every $\sigma \in \Sigma$.
- $M_{\sigma,\sigma'} = 2$ for every $\sigma, \sigma' \in \Sigma$ and $\sigma \neq \sigma'$.
- $M_{\sigma,-} = -2$ for every $\sigma \in \Sigma$.
- $M_{-,\sigma} = -4$ for every $\sigma \in \Sigma$.

What is the score of the following alignment:

X' = A C - - C T

Y' = T A C G G T

Enter your answer below as a single number.

You entered:

3

Your Answer		Score	Explanation
3	✔	5.00	Correct.
Total		5.00 / 5.00	

Question Explanation

Remember to sum the scores for the pairs of corresponding elements in the two sequences.

Question 5

Let $X = AC$ and $Y = GG$ be two DNA sequences ($\Sigma = \{A, C, T, G\}$), and consider the scoring matrix M given by:

- $M_{\sigma,\sigma} = 6$ for every $\sigma \in \Sigma$.
- $M_{\sigma,\sigma'} = 2$ for every $\sigma, \sigma' \in \Sigma$ and $\sigma \neq \sigma'$.
- $M_{\sigma,-} = M_{-,\sigma} = -4$ for every $\sigma \in \Sigma$.

Under this scoring matrix, which pair of sequences below is the optimal global alignment of X and Y?

Your Answer		Score	Explanation
○ $X' = A - C -$ and $Y' = -G - G$.			
○ $X' = AC - -$ and $Y' = - - GG$.			
○ $X' = AC -$ and $Y' = -GG$.			
◉ $X' = AC$ and $Y' = GG$.	✔	5.00	Correct. This alignment has maximal score of 4.

○ $X' = -AC-$ and
$Y' = -GG-$.

Total	5.00 / 5.00

Question Explanation

Compute the score of each possible global alignment and choose the one that yields the maximal score.

Question 6

Let $B(m,n)$ denote the number of global alignments of two sequences X and Y of lengths m and n, respectively. A recursive formula for $B(m,n)$ is:

Your Answer		Score	Explanation
○ $B(m,n) = B(m-1,n) + B(m,n-1)$.			
◉ $B(m,n) = B(m-1,n) + B(m-1,n-1) + B(m,n-1)$.	✔	5.00	Correct.
○ $B(m,n) = n \cdot B(m,n-1)$.			
○ $B(m,n) = m \cdot B(m-1,n)$.			
○ $B(m,n) = m \cdot B(m-1,n) + n \cdot B(m,n-1)$.			
Total		5.00 / 5.00	

Question Explanation

Consider how adding one element to the end of one or both sequences increases the number of possible global alignments.

Question 7

Overview

Questions 7-10 develop a dynamic programming (DP) algorithm for solving the Global Pairwise Alignment Problem. Algorithm **GlobalAlignment** below calls two helper algorithms to compute a global alignment.

Algorithm 1: GlobalAlignment.

Input: Sequences X and Y, and scoring matrix M.
Output: Optimal global pairwise alignment X' and Y' of sequences X and Y.

1 $S \leftarrow \mathbf{ComputeGlobalAlignmentScores}(X, Y, M)$;
2 $(X', Y') \leftarrow \mathbf{ComputeAlignment}(X, Y, M, S)$;
3 **return** (X', Y');

These two helper functions do the following:

- Algorithm **ComputeGlobalAlignmentScores** computes a *global alignment matrix* S whose entries $S[i, j]$ are the maximum scores over all possible global alignments for the pair of sequences $X[0 \ldots i-1]$ and $Y[0 \ldots j-1]$.
- Algorithm **ComputeAlignment** takes this alignment matrix and traces backwards from the entry $S[m, n]$ to the entry $S[0, 0]$ to compute the actual optimal global alignment of X and Y.

If you find it easier to refer to, you can open this figure in another window with this link: figure.

Question

Since the entries $S[i, j]$ of the global alignment matrix are the maximum scores for all possible global alignments for the pair of sequences $X[0 \ldots i-1]$ and $Y[0 \ldots j-1]$, which values of S will this dynamic programming method use in computing $S[i, j]$ when $i > 0$ and $j > 0$?

Your Answer		Score	Explanation
⊙ $S[i-1, j-1]$, $S[i-1, j]$, and $S[i, j-1]$.	✔	10.00	Correct. These three terms model the three possible ways an alignment can be extended.
○ $S[i-1, j-1]$.			
○ $S[i-1, j]$, and $S[i, j-1]$.			
○ $S[i-1, j-1]$, $S[i-1, j]$, $S[i, j-1]$, and $S[i, j]$.			
Total		10.00 / 10.00	

Question Explanation

Consider the various ways that a global alignment can be generated one pair of elements at a time.

Question 8

In this question, we will complete the partial pseudo-code for Algorithm

ComputeGlobalAlignmentScores that computes the global alignment matrix S. In particular, the entry $S[i, j]$ of this matrix should contain the maximum score over every possible global alignment of the pair of sequences $X[0 \ldots i-1]$ and $Y[0 \ldots j-1]$.

Algorithm 2: ComputeGlobalAlignmentScores.

Input: Sequences X and Y, and scoring matrix M.
Output: The dynamic programming (DP) table S.

$m \leftarrow |X|; \quad n \leftarrow |Y|;$
$S[0, 0] \leftarrow 0;$
for $i \leftarrow 1$ *to* m **do**
1   $S[i, 0] \leftarrow \ldots;$
for $j \leftarrow 1$ *to* n **do**
2   $S[0, j] \leftarrow \ldots;$
for $i \leftarrow 1$ *to* m **do**
  for $j \leftarrow 1$ *to* n **do**
3     $S[i, j] \leftarrow \ldots;$

return S;

If you find it easier to refer to, you can open this figure in another window with this link: figure.

The pseudo-code is missing details on Lines 1, 2, and 3. Which of the following options completes the algorithm so that it correctly computes the global alignment matrix S?

Your Answer	Score	Explanation

○
- Line 1: $S[i-1, 0]$
- Line 2: $S[0, j-1]$
- Line 3: $\max \begin{cases} S[i-1, j-1] + M_{X_{i-1}, Y_{j-1}} \\ S[i-1, j] + M_{X_{i-1}, -} \\ S[i, j-1] + M_{-, Y_{j-1}} \end{cases}$

○
- Line 1: $S[i-1, 0] + M_{X_{i-1}, -}$
- Line 2: $S[0, j-1] + M_{-, Y_{j-1}}$

- Line 3: $\max \begin{cases} S[i-1,j-1] \\ S[i-1,j] \\ S[i,j-1] \end{cases}$

⊙   ✔   10.00   Correct.

- Line 1: $S[i-1,0] + M_{X_{i-1},-}$
- Line 2: $S[0,j-1] + M_{-,Y_{j-1}}$
- Line 3: $\max \begin{cases} S[i-1,j-1] + M_{X_{i-1},Y_{j-1}} \\ S[i-1,j] + M_{X_{i-1},-} \\ S[i,j-1] + M_{-,Y_{j-1}} \end{cases}$

○

- Line 1: $S[i-1,0]$
- Line 2: $S[0,j-1]$
- Line 3: $\max \begin{cases} S[i-1,j-1] + M_{X_{i-1},Y_{j-1}} \\ S[i-1,j] \\ S[i,j-1] \end{cases}$

○

- Line 1: $M_{X_{i-1},-}$
- Line 2: $M_{-,Y_{j-1}}$
- Line 3: $\max \begin{cases} S[i-1,j-1] + M_{X_{i-1},Y_{j-1}} \\ S[i-1,j] + M_{X_{i-1},-} \\ S[i,j-1] + M_{-,Y_{j-1}} \end{cases}$

○

- Line 1: $M_{X_{i-1},-}$
- Line 2: $M_{-,Y_{j-1}}$
- Line 3: $\max \begin{cases} S[i-1,j-1] + M_{X_{i-1},Y_{j-1}} \\ S[i-1,j] \\ S[i,j-1] \end{cases}$

Total   10.00 / 10.00

Question 9

Here is the pseudo-code of an incomplete **ComputeAlignment** algorithm:

Algorithm 3: ComputeAlignment.

Input: Sequences X and Y, scoring matrix M, and a dynamic programming table S.

Output: A global pairwise alignment of X and Y using the DP table S and scoring matrix M.

$i \leftarrow |X|;\quad j \leftarrow |Y|;$
$X' \leftarrow \varepsilon;\quad Y' \leftarrow \varepsilon;$
while $i \neq 0$ *and* $j \neq 0$ **do**
    if $S[i,j] = S[i-1,j-1] + M_{X_{i-1},Y_{j-1}}$ **then**
    1 $X' \leftarrow ...;$
    2 $Y' \leftarrow ...;$
        $i \leftarrow i - 1; j \leftarrow j - 1;$
    else
        if $S[i,j] = S[i-1,j] + M_{X_{i-1},-}$ **then**
    3 $X' \leftarrow ...;$
    4 $Y' \leftarrow ...;$
            $i \leftarrow i - 1;$
        else
    5 $X' \leftarrow ...;$
    6 $Y' \leftarrow ...;$
            $j \leftarrow j - 1;$

while $i \neq 0$ **do**
    $X' \leftarrow X_{i-1} + X';\quad Y' \leftarrow \text{``}-\text{''} + Y';$
    $i \leftarrow i - 1;$
while $j \neq 0$ **do**
    $X' \leftarrow \text{``}-\text{''} + X';\quad Y' \leftarrow Y_{j-1} + Y';$
    $j \leftarrow j - 1;$
return (X', Y')

If you find it easier to refer to, you can open this figure in another window with this link: figure.

The pseudo-code above is missing details on Lines 1-6. Which of the following options completes the algorithm so that it correctly computes an optimal global alignment using the global alignment matrix S that was computed using Algorithm **ComputeGlobalAlignmentScores**?

Your Answer	Score	Explanation

○
- Line 1: $X_{i-1} + X'$
- Line 2: $Y_{j-1} + Y'$
- Line 3: $"-" + X'$
- Line 4: $"-" + Y'$
- Line 5: $"-" + X'$
- Line 6: $"-" + Y'$

○
- Line 1: $X_{i-1} + X'$
- Line 2: $Y' + Y_{j-1} + Y'$
- Line 3: $X_{i-1} + X'$

- Line 4: $\varepsilon + Y'$
- Line 5: $\varepsilon + X'$
- Line 6: $Y_{j-1} + Y'$

○

- Line 1: $\varepsilon + X'$
- Line 2: $\varepsilon + Y'$
- Line 3: $X_{i-1} + X'$
- Line 4: $" - " + Y'$
- Line 5: $" - " + X'$
- Line 6: $Y_{j-1} + Y'$

○

- Line 1: $\varepsilon + X'$
- Line 2: $\varepsilon + Y'$
- Line 3: $X_{i-1} + X'$
- Line 4: $\varepsilon + Y'$
- Line 5: $\varepsilon + X'$
- Line 6: $Y_{j-1} + Y'$

◉   ✔   10.00   Correct.

- Line 1: $X_{i-1} + X'$
- Line 2: $Y_{j-1} + Y'$
- Line 3: $X_{i-1} + X'$
- Line 4: $" - " + Y'$
- Line 5: $" - " + X'$
- Line 6: $Y_{j-1} + Y'$

○

- Line 1: $X' + X_{i-1}$
- Line 2: $Y' + Y_{j-1}$
- Line 3: $X' + X_{i-1}$
- Line 4: $Y' + " - "$
- Line 5: $X' + " - "$
- Line 6: $Y' + Y_{j-1}$

Total   10.00 / 10.00

Question Explanation

Use your answer to Question 7 to guide your solution here.

Question 10

Given two strings X and Y of lengths m and n, respectively, which of the following gives the tightest worst-case running time of Algorithm **GlobalAlignment** as given by the pseudo-code in Question 7?

Your Answer		Score	Explanation
⦿ $O(m\,n)$	✔	5.00	Correct.
○ $O(m^2 n^2)$			
○ $O(m+n)$			
○ $O(2^{m+n})$			
Total		5.00 / 5.00	

Question 11

Recall that given a string x, the string v is a *substring* of x if there exist strings u and w such that $x = u + v + w$. Given a string x of length n, how many substrings v of x are there?

You may assume that each character of x is distinct if necessary to clarify the problem. Enter your answer below as a mathematical expression in n.

You entered:

(n*(n+1)/2)+1

[Preview]   Help

Your Answer		Score	Explanation
(n*(n+1)/2)+1	✔	5.00	Correct. The expression is $\frac{1}{2}(n+1)(n+2) - n$.
Total		5.00 / 5.00	

Question Explanation

If $x = u + v + w$, let $|u| = i$, $|v| = j$, and $|w| = k$. Now, construct an expression that counts the number of tuples of the form (i, j, k) where $n = i + j + k$. Then adjust this expression appropriately based on the observation that the tuples $(i, 0, k)$ all correspond to the empty string.

Question 12

In Questions 12-15, we will make simple modifications to **ComputeGlobalAlignmentScores** and **ComputeAlignment** that yield an efficient algorithm for the Local Pairwise Alignment Problem.

In this question, we will focus on modifying **ComputeGlobalAlignmentScores** to compute a matrix of local alignment scores. Our modification is as follows: Whenever Algorithm **ComputeGlobalAlignmentScores** (in Question 8) computes a value to assign to $S[i, j]$, if the computed value is negative, the algorithm instead assigns 0 to $S[i, j]$. The result of this computation is the *local alignment matrix* for the two sequences. No other modification is done.

As an example, consider two strings $X = AA$ and $Y = TAAT$ over the alphabet $\Sigma = \{A, C, T, G\}$ and the scoring matrix M given by:

- $M_{\sigma, \sigma} = 10$ for every $\sigma \in \Sigma$.
- $M_{\sigma, \sigma'} = 4$ for every $\sigma, \sigma' \in \Sigma$ and $\sigma \neq \sigma'$.
- $M_{\sigma, -} = M_{-, \sigma} = -6$ for every $\sigma \in \Sigma$.

Given the two sequences X and Y and the scoring matrix M, what values would the modified algorithm assign to the entries $S[0, 2]$, $S[2, 0]$ and $S[2, 2]$ of the local alignment matrix S?

Enter your answer below as three numbers separated by spaces.

You entered:

0 0 14

Your Answer		Score	Explanation
0	✔	1.67	Correct. The scores in row zero and column zero are all zero.
0	✔	1.67	Correct. The scores in row zero and column zero are all zero.

| 14 | ✔ | 1.67 | Correct. The score at $S[2,2]$ is $4 + 10$. |

| Total | | 5.00 /
5.00 | |

Question Explanation

Remember that the entries in the DP table S can never be negative.

Question 13

During the computation of a global alignment, the traceback in the global alignment matrix S starts at the bottom right corner of the matrix ($S[m,n]$) and traces to the upper left corner ($S[0,0]$). Given the local alignment matrix S computed in Question 12, Algorithm **ComputeAlignment** (in Question 9) can be modified to generate a local alignment of two sequences as follows:

Start the traceback from the entry in S that has the maximum value over the entire matrix and trace backwards using exactly the same technique as in **ComputeAlignment**. Stop the traceback when the first entry with value 0 is encountered. If the local alignment matrix has more than one entry that has the maximum value, any entry with maximum value may be used as the starting entry.

As an concrete example of this process, what is the maximum value in an entry in the local alignment matrix S that you computed in Question 12? Enter your answer below as a single number.

You entered:

20

Your Answer		Score	Explanation
20	✔	5.00	Correct.

| Total | | 5.00 / 5.00 | |

Question Explanation

Remember that the entries in the local alignment matrix have a minimum value of zero.

Question 14

At what entry does the modified **ComputeAlignment** (in Question 13) start the traceback and at what entry does it end the traceback?

If the traceback starts at entry $S[i, j]$ and ends at entry $S[k, l]$, what are the values for the starting entry (i, j) and ending entry (k, l). Enter your answer below as four individual numbers i, j, k, l, separated by spaces. For example, if the traceback starts at entry $S[15, 20]$ and ends at entry $S[2, 4]$, you should enter 15 20 2 4 below.

You entered:

> 2 3 0 1

Your Answer		Score	Explanation
2	✔	1.25	
3	✔	1.25	
0	✔	1.25	
1	✔	1.25	
Total		5.00 / 5.00	

Question Explanation

Remember that the entries of the sequence X correspond to the rows of the local alignment matrix S and the entries of the sequence Y correspond to the columns of the local alignment matrix S.

Question 15

Using the modified **ComputeGlobalAlignmentScores** and **ComputeAlignment** algorithms in Questions 12 and 13, what is the local alignment they compute on the sequences X and Y

using the scoring matrix in Question 12?

Your Answer	Score	Explanation
⦿ $X' = AA$ and $Y' = AA$	✔ 5.00	Correct.
○ $X' = AA-$ and $Y' = AAT$		
○ $X' = -AA$ and $Y' = TAA$		
○ $X' = A$ and $Y' = A$		
○ $X' = -AA-$ and $Y' = TAAT$		
Total	5.00 / 5.00	

Question 16

Given two strings X and Y of lengths m and n, respectively, which of the following gives the tightest worst-case running time of computing a local alignment of the two strings using the modified algorithms in Questions 12 and 13?

Your Answer	Score	Explanation
○ $O(m+n)$		
○ $O(m^2 n^2)$		
○ $O(2^{m+n})$		
⦿ $O(m\,n)$	✔ 5.00	Correct. The worst case running time is proportional to the size of the alignment matrix.
Total	5.00 / 5.00	

Question 17

Is this statement true or false?

If all entries in a scoring matrix M are non-negative, then the score of an optimal local alignment and an optimal global alignment of two sequences X and Y using M are identical.

Your Answer		Score	Explanation
◉ True	✔	5.00	Correct. The largest value in the local alignment matrix will lie in $S[m,n]$ while the smallest value (zero) will lie in $S[0,0]$.
○ False			
Total		5.00 / 5.00	

Question Explanation

Try again. :)