Homework 4

11/17 points (64%)

Quiz, 17 questions

✖ Required to pass: 70% or higher

You can retake this quiz up to 3 times every 8 hours.

Back to Week 3

Retake

✖ 0 / 1
points

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?

☐ $X' = AC$ and $Y' = CA$.

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

☐ $X' = AC - -$ and $Y' = - - CA$.

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

☐ $X' = AC$ and $Y' = - - CA$.

Un-selected is correct

☐ $X' = CA$ and $Y' = CA$.

Un-selected is correct

☐ $X' = CA -$ and $Y' = -AC$.

This should not be selected

Homework 4

11/17 points (64%)

Quiz, 17 questions

✔ 1 / 1
points

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)

    3 13

▲

Correct Response

Correct. There is one global alignment of length one and two global alignments of length two. There is one global alignment of length two, six global alignments of length three, and six global alignments of length four.

✔ 1 / 1
points

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?

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

○

Homework 4

Quiz, 17 questions

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).

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

Correct

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

✔ 　1 / 1
　points

4.

Questions 4-5 will test your understanding of the optimality of global 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'=AC--CT

Y'=TACGGT

Enter your answer below as a single number.

> 3

Correct Response
Correct.

Homework 4

Quiz, 17 questions

11/17 points (64%)

✔ 1 / 1 points

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?

- ○ $X' = -AC-$ and $Y' = -GG-$.

- ○ $X' = A - C -$ and $Y' = -G - G$.

- ◉ $X' = AC$ and $Y' = GG$.

Correct
Correct. This alignment has maximal score of 4.

- ○ $X' = AC -$ and $Y' = -GG$.

- ○ $X' = AC - -$ and $Y' = - - GG$.

✖ 0 / 1 points

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:

- ○ $B(m, n) = B(m - 1, n) + B(m, n - 1)$.

- ○ $B(m, n) = B(m - 1, n) + B(m - 1, n - 1) + B(m, n - 1)$.

- ○ $B(m, n) = m \cdot B(m - 1, n)$.

- ◉ $B(m, n) = m \cdot B(m - 1, n) + n \cdot B(m, n - 1)$.

This should not be selected
Incorrect. This recurrence doesn't account for the recursive case when we have two sequences of length $m - 1$ and $n - 1$.

Homework 4 ○ $B(m, n) = n \cdot B(m, n - 1).$ **11/17 points (64%)**

Quiz, 17 questions

✗   0 / 1
          points

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 $j > 0$?

○ $S[i - 1, j - 1], S[i - 1, j]$, and $S[i, j - 1]$.

○ $S[i - 1, j - 1]$.

○ $S[i - 1, j]$, and $S[i, j - 1]$.

This should not be selected

Incorrect. Remember that global alignments include matches from corresponding pairs of elements in each sequence.

Homework 4

11/17 points (64%)

Quiz, 17 questions

○ $S[i-1,j-1]$, $S[i-1,j]$, $S[i,j-1]$, and $S[i,j]$.

✔   1 / 1
points

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 $X[0 \dots i-1]$ and $Y[0 \dots j-1]$.

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    $S[i,0] \leftarrow \dots;$
for $j \leftarrow 1$ *to* n **do**
2    $S[0,j] \leftarrow \dots;$
for $i \leftarrow 1$ *to* m **do**
   for $j \leftarrow 1$ *to* n **do**
3      $S[i,j] \leftarrow \dots;$
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?

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

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

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

Correct
Correct.

○
- Line 1: $M_{X_{i-1},-}$

- Line 2: $M_{-,Y_{j-1}}$

Homework 4

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

11/17 points (64%)

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

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

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

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

○
- Line 1: $S[i-1,0]$

- Line 2: $S[0,j-1]$

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

✓   1 / 1
points

9.

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

Homework 4

Quiz, 17 questions

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

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

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**?

- ○
  - 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: $Y_{j-1} + Y'$

Correct
Correct.

- ○
  - 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: $X_{i-1} + X'$

- Line 2: $Y_{j-1} + Y'$

- Line 3: $" - " + X'$

- Line 4: $" - " + Y'$

- Line 5: $" - " + X'$

- Line 6: $" - " + Y'$

11/17 points (64%)

○
- Line 1: $\varepsilon + X'$

- Line 2: $\varepsilon + Y'$

- Line 3: $X_{i-1} + X'$

- Line 4: $\varepsilon + Y'$

- Line 5: $\varepsilon + X'$

- Line 6: $Y_{j-1} + Y'$

○
- 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: $Y' + Y_{j-1}$

○
- Line 1: $\varepsilon + X'$

- Line 2: $\varepsilon + Y'$

- Line 3: $X_{i-1} + X'$

- Line 4: $" - " + Y'$

- Line 5: $" - " + X'$

- Line 6: $Y_{j-1} + Y'$

✖   0 / 1
         points

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?

○ $O(m + n)$

Homework 4

Quiz, 17 questions

11/17 points (64%)

○ $O(m^2 n^2)$

This should not be selected

Incorrect. The correct running time is smaller.

○ $O(2^{m+n})$

○ $O(m\,n)$

✗   0 / 1 points

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.

Preview unable to update: Error parsing submission

(n-2)(n-1)/2

Incorrect Response

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.

Reveal correct answer

✓   1 / 1 points

12.

Homework 4

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

11/17 points (64%)

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.

> 0 0 14

Correct Response
Correct. The scores in row zero and column zero are all zero. The
score at $S[2, 2]$ is $4 + 10$.

✓   1 / 1
points

13.

Homework 4

Quiz, 17 questions

11/17 points (64%)

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.

20

Correct Response
Correct.

✓　　1 / 1
points

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.
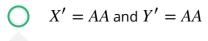
2 3 0 1

Correct Response

✓　　1 / 1
points

Homework 4

Quiz, 17 questions

11/17 points (64%)

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?

○ $X' = AA$ and $Y' = AA$

Correct
Correct.

○ $X' = -AA$ and $Y' = TAA$

○ $X' = -AA-$ and $Y' = TAAT$

○ $X' = A$ and $Y' = A$

○ $X' = AA-$ and $Y' = AAT$

✖   0 / 1
         points

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?

○ $O(m + n)$

○ $O(m^2 n^2)$

This should not be selected
Incorrect. The worst case running time is smaller.

○ $O(2^{m+n})$

○ $O(m\, n)$

✔   1 / 1
         points

17.

Is this statement true or false?

Homework 4

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.

11/17 points (64%)

○ True

Correct
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