CENG 465
Introduction to Bioinformatics
Fall 2015-2016

Assignment #1
Written Assignment on Sequence Alignment

1. Let “end alignment” be an alignment between a suffix of a sequence \( v = v_1 \ldots v_n \) and a prefix of a sequence \( w = w_1 \ldots w_n \). For example, if \( v = ACACAC \) and \( w = CCCTTT \), then a (not necessarily optimal) end alignment between \( v \) and \( w \) is:

\[
\begin{align*}
CAC \\
CCC
\end{align*}
\]

Optimal end alignment is an alignment that maximizes the global alignment score between \( v_i \ldots v_n \) and \( w_j \ldots w_n \), where the maximum is taken over all suffixes \( v_i \ldots v_n \) of \( v \) and all prefixes \( w_1 \ldots w_j \) of \( w \).

Give an algorithm which computes the optimal end alignment, and runs in time \( O(nm) \).

2. Suppose that we have sequences \( v = v_1 \ldots v_n \) and \( w = w_1 \ldots w_n \), where \( v \) is longer than \( w \). We wish to find a substring of \( v \) which best matches all of \( w \). Global alignment will not work because it would try to align all of \( v \). Local alignment also will not work because it may not align all of \( w \). Let us call this alignment the Fitting alignment. Fitting a sequence \( w \) into a sequence \( v \) is a problem of finding a substring \( v' \) of \( v \) that maximizes the score of global alignment \( s(v', w) \) among all substrings of \( v \). For example, if \( v = GTAGGCTTAAGGTTA \) and \( w = TAGATA \), the best fitting alignment might be:

\[
\begin{align*}
TAGGCTTA \\
TAGA--TA
\end{align*}
\]

Give an algorithm which computes the optimal fitting alignment. Explain how to fill in the first row and column of the dynamic programming table and give a recurrence to fill in the rest of the table. Give a method to find the best alignment once the table is filled in. The algorithm should run in time \( O(nm) \).

Submission

Submit your solutions as a single PDF/DOC/ODT document (scanned copies of handwritten solutions are accepted as well) via ODTU-Class before the deadline. Late submission is -20 pts per day.

Note: Although it is possible to find solutions of these problems on the internet, try to solve these problems on your own or at least understand the solutions fully as there will be questions in the midterm exam related to these problems.