CS 482/682, 2022, Assignment 1.

Due date: Feb. 1, 11:59pm.

Assignment 1. Fit Alignment

Fit alignment: Given two sequences S and T, the fit alignment asks to find a substring T' of T, such that the global sequence alignment score between S and T' is maximized.

Write a program to "fit align" two sequences. The program reads from a FASTA file that contains exactly two DNA sequences; and prints the alignment score and the alignment to an output file. We will call your program by the following command line:

assn1-part1 in.fasta out.txt

Note that the file names may be different when we call.

To make things simpler, we will use a simplified FASTA format, where each entry in the file takes exactly two lines: the first starts with a '>' sign and is the "header line" or "annotation line". The second is the actual sequence. But keep in mind that the sequences can be very long.

The output file should consist of three lines:

- The first line is the alignment score;
- The second and third lines correspond to the first and second sequences in the input, with dash inserted to form the alignment.

If there are multiple alignments that can give the same optimal score, your program only needs to output any one of them.

We use the following score scheme and linear gap penalty: match = 1, mismatch = -1, indel=-1.

For example, the following input:

```
>seq1
AACCCCTAG
>seq2
TTAATCCCCAGGGTCGTTT
```

Should produce the following output (or possibly another solution with equal score to the following):

```
6
AA-CCCCTAG
AATCCCC-AG
```

Your program needs to be reasonably efficient (quadratic time complexity and at most quadratic space complexity) to get full marks.

What to submit:

Your program and a readme.pdf file. Your program can contain multiple source files. For this program, only standard libraries for the programming language are allowed (calling a bioinformatics library to read fasta file or conducting alignment are not allowed). The readme.pdf file contains your information (names, student number, etc.), a brief description of the code (general structure), and how to run it. Keep the readme.pdf file short. Further details about the submission procedure will be announced later.

Programming language:

Python.