CSCI 538 – Artificial Intelligence
Assignment 2 - Sequence Alignment
Spring 2019

Due Date: Tuesday April 2nd 2019
Total Points: 100

The sequence alignment problem takes two DNA sequences *X[1..m]* and *Y[1..n]* and returns the optimal alignment cost. This is the same problem as the spelling correction problem using Minimum Edit Distance that we went through in class. In the assignment page you will find two sequences that can be used as input to your program. The first DNA sequence represents the Human species, while the other gene is from the mice species. In this assignment you are going to write a menu driven program with the following properties:

a) The program allows the user to choose between two strategies for calculating insertion, deletion and substitution costs. The first one is a generic strategy and follows what we did in class where the deletion and insertion cost is 1 and substitution is 2. The second strategy also has the deletion and insertion costs as 1 but assumes that during a mismatch base *A* will most likely pair with base *G,* while *T* will most likely pair with *C*. The substitution cost for this strategy is shown below:

	A	T	C	G
A	0	5	5	1
T	5	0	1	5
C	5	1	0	5
G	1	5	5	0

b) The program allows the user to choose between either:
   a. calculating the sequence alignment of the entire human DNA with the entire mice DNA
   b. calculating the sequence alignment of a certain part of the human DNA with the same part for the mice DNA. The program allows the user to pick from the following percentage: 10%, 20%, 30%, 40%, 50% and 75%. Any percentage chosen will always start from the beginning of the DNA sequence. For example, 105 means the first 10% of the DNA sequence.

c) The program will display (as an output of an alignment) the optimal alignment cost. The actual alignment will not be printed out on the screen but will be saved in a text file called *alignment.txt*. Your program needs to trace back through the alignment matrix in order to construct the sequence alignment.

What to hand in

Submit your project electronically through D2L. Please hand your Python program.