**CS 238 – Spring 2023 – Homework 2 – Due May 24th at 11:59PM**

**NOTE: Do not use ChatGPT or other Large Language Models to try to solve this problem.**

Problem 1

Recall the computational problem of global alignment of two DNA sequences, A and B. In the most basic formulation, insertions/deletions and matches are allowed, each with a score contribution. The goal is the find the alignment between these two strings that maximizes the alignment score between A and B. In the dynamic programming solution discussed in class, the asymptotic runtime is O(m \* n) where m is the length of A and n is the length of B. We are attempting to reduce the time complexity of finding the solution by putting more constraints upon the solution. If we define that the maximum number of insertion/deletions is k, how does this change the asymptotic run time? To answer this, write the dynamic programming recurrence relationship and then describe the bounds on the time complexity given this new constraint.

**NOTE: You are allowed to use ChatGPT or other LLMs to write the code for this question but not answer the question. Please submit your code as part of this problem.**

Problem 2

Recall the DNA assembly problem given a set of shotgun reads from a genome. We would like to reconstruct the original genome by utilizing a Hamiltonian path approach to maximize the weight of the Hamiltonian path. Implement this approach to reconstruct the original genome sequence (hint: the length of the original genome is 1000 basepairs).

Download the reads from here:

<https://www.dropbox.com/s/dh0wn6inj8ms6gm/reads.fasta?dl=1>

**NOTE: You are allowed to use ChatGPT or other LLMs to write the code for this question but not answer the question. Please submit your code as part of this problem.**

Extra Credit Problem 3 – 5% of value of HW

Implement above the approach using the Eulerian path approach. How does the performance differ? Is this problem size big enough to matter?