

The Sequence Alignment Problem

Classes introduced:

- **Gene:** Enum to map the gene to its place in the **Scoring Matrix**
- **Scoring Matrix:** Add function `getScore` to easily retrieve score of two genes

Approach:

Tabulation dynamic programming was used to solve each subproblem bottom-up. By initializing an array of size (sequence 1 length) x (sequence 2 length), where the first row/column maps a match with gap, we can save the results of the optimal solution for each subsequence.

In this example, for the presented sequences the optimal solution would be:

TACGC-A
-ATGCC-

Timing Analysis:

By plotting (sequence 1 length) x (sequence 2 length) by time taken for execution, we get a linear graph indicating each step is proportional to the step in time. Therefore, running time is $O(mn)$.

Starting Sequence Alignment for:

Seq1: TACGCA
Seq2: ATGCC

Scoring Matrix:

```
[[ 1. -0.8 -0.2 -2.3 -0.6]
 [-0.8 1. -1.1 -0.7 -1.5]
 [-0.2 -1.1 1. -0.5 -0.9]
 [-2.3 -0.7 -0.5 1. -1. ]
 [-0.6 -1.5 -0.9 -1.  0. ]]
```

Alignment found with score: 0.0

```
[[ 0. -0.6 -0.9 -1.5 -1. -1. ]
 [-0.9 -0.2  0.4 -1.1 -1.9 -1.5]
 [-0.6  0.1 -0.2 -0.4 -1.4 -2.1]
 [-1. -0.9 -0.4 -0.9  0.6 -0.4]
 [-1.5 -1.8 -1.9  0.6 -0.4 -0.1]
 [-1. -1.6 -2.3 -0.4  1.6  0.6]
 [-0.6  0. -0.9 -1.  1.  0. ]]
```

