Profile-most Probable _k_-mer Problem

Input: A string _Text_, an integer k, and a $4 \times k$ matrix _Profile_

Output: A _Profile_-most probable k-mer in _Text_

SAMPLE DATASET:

<u>Input</u>:

ACCTGTTTATTGCCTAAGTTCCGAACAAACCCAATATAGCCCGAGGGCCT

5

0.2 0.2 0.3 0.2 0.3

0.4 0.3 0.1 0.5 0.1

0.3 0.3 0.5 0.2 0.4

0.1 0.2 0.1 0.1 0.2

<u>Output</u>:

CCGAG

The sample dataset is not actually run on your code.

TEST DATASET 1:

Input:
AGCAGCTTTGACTGCAACGGGCAATATGTCTCTGTGTGGATTAAAAAAAGAGTGTCT
GATCTGAACTGGTTACCTGCCGTGAGTAAAT
8
0.7 0.2 0.1 0.5 0.4 0.3 0.2 0.1
0.2 0.2 0.5 0.4 0.2 0.3 0.1 0.6
0.1 0.3 0.2 0.1 0.2 0.1 0.4 0.2
0.0 0.3 0.2 0.0 0.2 0.3 0.3 0.1
Output:
AGCAGCTT

This dataset checks for off-by-one errors at the beginning of Text. Notice that the optimal solution ("AGCAGCTT") occurs at the very beginning of Text, so if your code does not check this k-mer, then your code will output a different (incorrect) k-mer as the solution.

TEST DATASET 2:

Input:

TTACCATGGGACCGCTGACTGATTTCTGGCGTCAGCGTGATGCTGGTGTGGATGACA
TTCCGGTGCGCTTTGTAAGCAGAGTTTA

12

0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.1 0.2 0.3 0.4 0.5

0.3 0.2 0.1 0.1 0.2 0.1 0.1 0.4 0.3 0.2 0.2 0.1

0.2 0.1 0.4 0.3 0.1 0.1 0.1 0.3 0.1 0.1 0.2 0.1

0.3 0.4 0.1 0.1 0.1 0.1 0.0 0.2 0.4 0.4 0.2 0.3

Output:

AAGCAGAGTTTA

       This dataset checks for off-by-one errors at the end of Text. Notice that the optimal solution ("AAGCAGAGTTTA") occurs at the very end of Text, so if your code does not check this k-mer, then your code will output a different (incorrect) k-mer as the solution.