

Frequency Array

Input: A DNA string *Text* followed by an integer *k*

Output: *FrequencyArray(Text, k)*

Pseudocode

```
FrequencyArray(Text, k)
  array  $\leftarrow$  0 for all  $4^k$  kmers
  for i  $\leftarrow$  0 to |Text| - k
    num = PatternToNumber(Text(i, l))
    array[num]  $\leftarrow$  array[num] + 1
  return array
```

SAMPLE DATASET:

Input:

ACGCGGCTCTGAAA

2

Output:

2 1 0 0 0 0 2 2 1 2 1 0 0 1 1 0

The sample dataset is not actually run on your code.

TEST DATASET 1:

Input:

AAAAC

2

Output:

3 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0

This dataset checks if you have an off-by-one error at the end of Text (i.e. you are not counting the last kmer in Text). There are three instances of AA (~~AAAAC~~, AAAAC, and AAAAC), but there is one instance of AC at the end (~~AAAAC~~).

TEST DATASET 2:

Input:

TTAAA

2

Output:

2 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 1

This dataset checks if you have an off-by-one error at the beginning of Text (i.e. you are not counting the first kmer in Text). There are two instances of AA (TTAAA and TTAAA), but there is one instance of TTA (TTAAA) and one instance of TAA (TTAAA).

TEST DATASET 3:

Input:

AAA

2

Output:

2 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0

This dataset checks if your code actually increments each count, or if your code instead just sets the count equal to one each time. In other words, this dataset checks if your code is doing something like `array[kmer] = 1` instead of `array[kmer] += 1`.