

### Randomized Motif Search

**Input:** Integers  $k$  and  $t$ , followed by a collection of strings  $Dna$ .

**Output:** A collection  $BestMotifs$  resulting from running  $RandomizedMotifSearch(Dna, k, t)$  1,000 times. Remember to use pseudocounts!

#### Pseudocode

RandomizedMotifSearch( $Dna, k, t$ ):

```
Motifs ← empty list
for each sequence seq in Dna
    add randomly selected k-mer from seq to Motifs
BestMotifs ← Motifs
while forever
    Profile ← Profile(Motifs)
    Motifs ← Motifs(Profile, Dna)
    if Score(Motifs) < Score(BestMotifs)
        BestMotifs ← Motifs
    else
        return BestMotifs
```

**SAMPLE DATASET:**

Input:

8 5

CGCCCCTCTCGGGGGTGTTCAGTAAACGGCCA  
GGGCGAGGTATGTGTAAGTGCCAAGGTGCCAG  
TAGTACCGAGACCGAAAGAAGTATACAGGCGT  
TAGATCAAGTTTCAGGTGCACGTCGGTGAACC  
AATCCACCAGCTCCACGTGCAATGTTGGCCTA

Output:

TCTCGGGG  
CCAAGGTG  
TACAGGCG  
TTCAGGTG  
TCCACGTG

The sample dataset is not actually run on your code.

## TEST DATASET 1:

Input:

6 8

```
AATTGGCACATCATTATCGATAACGATTCGCCGCATTGCC
GGTTAACATCGAATAACTGACACCTGCTCTGGCACCGCTC
AATTGGCGGGCGGTATAGCCAGATAGTGCCAATAATTCCT
GGTTAATGGTGAAGTGTGGGTTATGGGGAAAGGCAGACTG
AATTGGACGGCAACTACGGTTACAACGCAGCAAGAATATT
GGTTAACTGTTGTTGCTAACACCGTTAAGCGACGGCAACT
AATTGGCCAACGTAGGCGCGGCTTGGCATCTCGGTGTGTG
GGTTAAAAGGCGCATCTTACTCTTTTCGCTTTCAAAAAAA
```

Output:

```
CGATAA
GGTTAA
GGTATA
GGTTAA
GGTTAC
GGTTAA
GGCCAA
GGTTAA
```

This dataset checks if your code has an off-by-one error at the beginning of each sequence of Dna. Notice that the some of the motifs of the solution occur at the beginning of their respective sequences in Dna, so if your code did not check the first k-mer in each sequence of Dna, it would not find these sequences.

## TEST DATASET 2:

Input:

6 8

```
GCACATCATTA AACGATTCGCCGCATTGCCTCGATTAACC
TCATAACTGACACCTGCTCTGGCACCGCTCATCCAAGGCC
AAGCGGGTATAGCCAGATAGTGCCAATAATTCCTTAACC
AGTCGGTGGTGAAGTGTGGGTTATGGGGAAAGGCAAGGCC
AACCGGACGGCAACTACGGTTACAACGCAGCAAGTTAACC
AGGCGTCTGTTGTTGCTAACACCGTTAAGCGACGAAGGCC
AAGCTTCCAACATCGTCTTGGCATCTCGGTGTGTTAACC
AATTGAACATCTTACTCTTTTCGCTTTCAAAAAAAAGGCC
```

Output:

```
TTAACC
ATAACT
TTAACC
TGAAGT
TTAACC
TTAAGC
TTAACC
TGAACA
```

This dataset checks if your code has an off-by-one error at the end of each sequence of Dna. Notice that the some of the motifs of the solution occur at the end of their respective sequences in Dna, so if your code did not check the last k-mer in each sequence of Dna, it would not find these sequences.