Lab Session 2

14.08.2024

BT 3051 - DSA Biology Lab

1. **Read about these functions**

   - `range()`
   - `enumerate()`
   - `zip()`
   - `any()`, `all()`
   - `chr()`, `ord()`
   - `map()`
   - `filter()`
   - `str.join()`

2. **Problem 1:**

   Given two positive integers a and b, each less than 1000, find the integer corresponding to the square of the hypotenuse of the right triangle whose legs have lengths a and b.

   Input: 3 4
   Output: 25

3. **Problem 2:**

   Given two positive integers a and b ($a < b < 10000$), find the sum of all odd integers from a through b, inclusively.

   Input: 100 200
   Output: 7500

4. **Problem 3:**

   Write a python code to perform the following (without using numpy and built-in functions):

   (a) Dot product of two vectors
   (b) Outer product of two vectors

(c) Multiplication of two matrices

(d) Hadamard product of two matrices

In all these cases, the function should check the compatibility of vector or matrix shapes.

5. **Problem 4:**

Write a python code to find k closest neighbors of the given element in a sorted array.

Input: [2, 3, 4, 5, 6, 7], $k = 3$, $a = 5$
Output: [3, 4, 6]

Rule: The integer a may or may not be present in the input array. If a is less than or equal to the first element in the input array, return the first k elements. Similarly, if a is more than or equal to the last element in the input array, return the last k elements. The returned elements should be in the same order as present in the input array. Consider the smaller element in the case of odd equally distant neighbors.

6. **Problem 5:**

Write a Python program that takes a DNA sequence as input and performs the following tasks:

- Calculate the length of the sequence
- Count the occurrences of each nucleotide (A, T, C, G)
- Find the reverse complement of the sequence (replace A with T, T with A, C with G, and G with C)
- Transcript RNA string (replace T with U)

7. **Problem 6:**

Consider a group of 5 synthetic organisms , each with a genetic code composed exclusively of two nucleotides: A and B. Each organism's genetic code is represented as a string of fixed length, where each position in the string can be either 'A' or 'B'.

The task is to:

- Calculate the genetic similarity between every possible pair of organisms in the group. The similarity between two organisms is defined as the percentage of positions where the nucleotides are identical.
- Identify the pair of organisms that are the most closely related, i.e., the pair with the highest percentage of matching positions in their genetic codes.

Input:

A list of genetic sequences (strings) representing the genetic codes of the organisms. All sequences are of the same length.

Output:

A matrix or list of similarity percentages between each pair of organisms. The pair of organisms with the highest similarity percentage.