

Matematyczne aspekty analizy danych (studia stacjonarne)

Dr Anna Muranova

Semestr zimowy 2025/2026, UWM w Olsztynie

Ćwiczenie 11

Zadanie 11.1. (a) Napisz program, który w pętli generuje 100 liczb całkowitych od -100 do 100 i zapisuje każdą z tych liczbę do odpowiedniego pliku: parzyste do even.txt, a nieparzyste do odd.txt.

(b) Napisz program, który wczyta pliki utworzone w poprzednim zadaniu i obliczy sumę oraz średnią arytmetyczną liczb znajdujących się w każdym z nich.

Zadanie 11.2. (a) Napisz program, który wypisuje w plik w plik 5000 liter, symulujących DNA, tzn. każda litera jest A, C, G lub T z równym prawdopodobieństwem albo losowym znakiem z prawdopodobieństwem 1/100 (błąd).

(b) Napisz program, który wczyta plik utworzony w poprzednim zadaniu i:

- usunie błędy
- obliczy ilość wystąpień każdej litery, wynik zapisz w słowniku
- obliczy ilość wystąpień TA i AT,
- przekształci DNA na białko według tabeli:

```
table = {  
    "ATA": "I", "ATC": "I", "ATT": "I", "ATG": "M",  
    "ACA": "T", "ACC": "T", "ACG": "T", "ACT": "T",  
    "AAC": "N", "AAT": "N", "AAA": "K", "AAG": "K",  
    "AGC": "S", "AGT": "S", "AGA": "R", "AGG": "R",  
    "CTA": "L", "CTC": "L", "CTG": "L", "CTT": "L",  
    "CCA": "P", "CCC": "P", "CCG": "P", "CCT": "P",  
    "CAC": "H", "CAT": "H", "CAA": "Q", "CAG": "Q",  
    "CGA": "R", "CGC": "R", "CGG": "R", "CGT": "R",  
    "GTA": "V", "GTC": "V", "GTG": "V", "GTT": "V",  
    "GCA": "A", "GCC": "A", "GCG": "A", "GCT": "A",  
    "GAC": "D", "GAT": "D", "GAA": "E", "GAG": "E",  
    "GGA": "G", "GGC": "G", "GGG": "G", "GGT": "G",  
    "TCA": "S", "TCC": "S", "TCG": "S", "TCT": "S",  
    "TTC": "F", "TTT": "F", "TTA": "L", "TTG": "L",  
    "TAC": "Y", "TAT": "Y", "TAA": "_", "TAG": "_",  
    "TGC": "C", "TGT": "C", "TGA": "_", "TGG": "W",  
}
```

i zapisuje wynik do nowego pliku.

- przekształca DNA na białko zaczynając od podanej przez użytkownika trojki.

Zadanie 11.3. Wczytać poszczególne wiersze z pliku

`http://wmii.uwm.edu.pl/~muranova/MAAD2025-26/dane10.txt`

do x , y , z jako `np.array`, i obliczyć po tym danym regresję liniową od dwóch zmiennych $z = f(x, y)$ oraz R^2 . *Narysować obrazek 3D.

Zadanie 11.4. Pobierz plik z cenami jaj w Polsce w wybranych sieciach. Źródła danych:

`http://www.dlahandlu.pl/koszyk/towar/10-jaj-najtansze,38.html`

`http://wmii.uwm.edu.pl/~muranova/MAAD2025-26/jajka2025.csv`.

Zapisz plik w folderze projektu. Załaduj go do środowiska. Dodaj encoding! Następnie przetwórz dane abyś można było wykonać operacje:

- obliczyć średnią cenę wszystkich jaj.
- wyznaczyć w którym mieście i w jakiej sieci są najtańsze a w jakich najdroższe jajka. Wynik zapisz w postaci dwuwymiarowej tablicy przechowującej pary (Miasto, nazwa sieci).

Wszystkie operacje wykonaj używając funkcji wbudowanych w interpretator lub biblioteki `numpy`.