

Laboratorium 1 — Wprowadzenie do środowiska obliczeniowego

Moduł NUMPY jest zestawem narzędzi umożliwiającym zaawansowane obliczenia matematyczne. Rozszerza on możliwości PYTHONa o nowe typy danych, operacje na nich oraz funkcje przyspieszające obliczenia. Jednym z podstawowych obiektów jest 'array', który stanowi kontener dla macierzy i tablic oraz ułatwia m.in. obliczenia na dużych zbiorach danych.

Zadanie 1 - Tablice

1. Utwórz tablicę jednowymiarową zawierającą kolejne cyfry od 1 do 7.
2. Utwórz tablicę dwuwymiarową o postaci:

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \end{bmatrix}$$

3. Utwórz macierz transponowaną na podstawie macierzy z pkt. 2.
4. Za pomocą polecenia `arange` utwórz wektor zawierający wartości od 1 do 20 z krokiem 0.5.
5. Za pomocą `linspace` utwórz wektor zawierający 100 równomiernie rozłożonych wartości z przedziału $[0, 5]$.

Zadanie 2 - Liczby losowe

1. Za pomocą funkcji `random` utwórz tablicę z 20 liczb losowych rozkładu normalnego, zaokrąglonych do dwóch miejsc po przecinku.
2. Wygeneruj losowo 100 liczb całkowitych w zakresie od 1 do 1000.
3. Za pomocą funkcji `'zeros'` i `'ones'` wygeneruj dwie macierze o rozmiarze 3×2 .
4. Utwórz macierz losową złożoną z liczb całkowitych o rozmiarze 5×5 i nadaj jej typ 32bit.
5. Wygeneruj tablicę złożoną z losowo wybranych liczb dziesiętnych od 0 - 10:
 - (a) Zamień wartości na `'integer'` i wstaw w nową tablicę.
 - (b) Znajdź funkcję `'numpy'`, która zaokrągli tablicę (a) do liczb całkowitych. Zamień je następnie na typ `'integer'`.
 - (c) Porównaj wyniki z a i b.

Zadanie 3 - Selekcja danych

1. Utwórz macierz: `b = np.array([[1,2,3,4,5], [6,7,8,9,10]], dtype = np.int32)`
2. Za pomocą pola klasy `'ndim'` sprawdź ile wymiarów ma tablica b.
3. Za pomocą pola klasy `'shape'`, sprawdź wymiary tablicy b.
4. Wybierz wartości 2 i 4 z tablicy b.
5. Wybierz pierwszy wiersz tablicy b.
6. Wybierz wszystkie wiersze z kolumny 1.
7. Wygeneruj macierz losową o rozmiarze 20×7 , złożoną z liczb całkowitych w przedziale 0-100. Wyświetl wszystkie wiersze dla czterech pierwszych kolumn.

Zadanie 4 - Operacje matematyczne i logiczne

1. Stwórz dwie macierze w przedziale 0-10 o rozmiarach 3x3 (a i b). Dodaj, pomnóż, podziel, spotęguj ich elementy przez siebie.
2. Sprawdź czy jakakolwiek element z macierzy a jest większy lub równy 4.
3. Sprawdź czy jakakolwiek element z macierzy a należy do przedziału [1, 4].
4. Znajdź funkcję w 'numpy' do obliczenia sumy głównej przekątnej macierzy b.

Zadanie 5 - Dane statystyczne

1. Oblicz sumę, wartość minimum, maksimum, odchylenie standardowe dla macierzy b.
2. Oblicz średnią dla wierszy w macierzy b.
3. Oblicz średnią dla kolumn macierzy b.

Zadanie 6 - Rzutowanie wymiarów za pomocą shape lub resize:

1. Utwórz wektor składający się z 50 liczb.
2. Za pomocą funkcji 'reshape' utwórz macierz 10x5. Spróbuj dokonać tego samego za pomocą funkcji resize.
3. Sprawdź do czego służy komenda ravel.
4. Stwórz dwie tablice o rozmiarach 5 i 4 i dodaj je do siebie. Sprawdź do czego służy funkcja 'NEWAXIS' i wykorzystaj ją.

Zadanie 7 - Sortowanie danych:

1. Sprawdź składnię funkcji sort i argsort.
2. Utwórz macierz `a = np.random.randn(5, 5)`, a następnie posortuj wiersze rosnąco. Posortuj kolumny malejąco.

Moduł Pandas jest zestawem narzędzi umożliwiającym sprawne manipulowanie zestawem danych. Rozszerza on możliwości PYTHONa o łatwy import i eksport danych m.in. do plików w formatach tekstowych (csv) czy konkretnych aplikacji (excel). Dane są przechowywane w tabelach tzw. DataFrame. Pandas dostarcza wiele narzędzi do selekcji, łączenia i sortowania danych.

Zadanie 1 - DataFrame

Zapoznaj się z obiektem 'pandas.DataFrame' i utwórz tabelę składającą się z liczb losowych wygenerowanych z rozkładu normalnego złożoną z trzech kolumn z nagłówkami (A,B,C) i pięciu wierszy z indeksem o nazwie "data" złożonym z dat w przedziale od 2020-03-01 do 2020-03-05 np.

data	A	B	C
2020-03-01	0,720184489	0,657752214	0,297794824
2020-03-02	0,775572034	0,532276741	0,051391112
2020-03-03	0,346071872	0,22057064	0,640917109
2020-03-04	0,475104139	0,228407696	0,507173083
2020-03-05	0,475104139	0,899298837	0,541077542

Zadanie 2 - Indeksacja

Wygeneruj tabelę złożoną z liczb losowych i indeksu "id" w formacie 'integer' złożoną z 20 wierszy i trzech kolumn ('A','B','C'). Następnie:

1. wybierz trzy pierwsze wiersze z tabeli,
2. wybierz trzy ostatnie wiersze z tabeli,
3. wyświetl nazwę indeksu tabeli,
4. wyświetl nazwy kolumn,
5. wyświetl tylko dane bez indeksów i nagłówek kolumn,
6. wybierz wartości kolumny 'A' a następnie 'A' i 'B',
7. zapoznaj się z funkcją 'iloc' i wyświetl:
 - (a) trzy pierwsze wiersze i kolumny 'A' i 'B'
 - (b) wiersz piąty
 - (c) wiersze 0, 5, 6, 7 i kolumny 1 i 2

Zadanie 3 - Operacje na danych

Utwórz tabelę `df = pd.DataFrame("x": [1, 2, 3, 4, 5], 'y': ['a', 'b', 'a', 'b', 'b'])`, a następnie:

1. Zgrupuj tabelę po zmiennej symbolicznej Y, a następnie wyznacz średnią wartość atrybutu numerycznego X w grupach wyznaczonych przez Y,
2. Wyznacz rozkład licznosci atrybutów (`value_counts`).