

Eksploracja danych

Piotr Lipiński

Lista zadań nr 2 – Minikurs Matlaba

Zadanie 0. (rozgrzewkowe, 1 punkt, zadanie należy przesłać emailiem)

Utwórz skrypt definiujący poniższe zmienne:

$a = [1, 2, 3, 4, \dots, 100]$ (wektor złożony z liczb całkowitych od 1 do 100)

$b = [1, 3, 5, 7, \dots, 99]$ (wektor złożony z liczb całkowitych nieparzystych od 1 do 99)

$c = [-1.00 * \pi, -0.99 * \pi, \dots, -0.01 * \pi, 0, 0.01 * \pi, \dots, 0.99 * \pi, 1.00 * \pi]$

$d = [-1.00 * \pi, -0.99 * \pi, \dots, -0.01 * \pi, 0.01 * \pi, \dots, 0.99 * \pi, 1.00 * \pi]$

$e = [e_1, e_2, \dots, e_{100}]$, gdzie $e_i = \sin(i)$, jeśli $\sin(i) > 0$, lub $e_i = 0$ w przeciwnym przypadku

A = macierz rozmiaru 10×10 zawierająca liczby całkowite od 1 do 100: w pierwszym wierszu od lewej 1, 2, ..., w drugim wierszu od lewej 11, 12, ..., itd. (wskazówka: użyć polecenia `reshape`)

B = macierz trójdzielna rozmiaru 100×100 mająca na głównej przekątnej liczby całkowite od 1 do 100, a poniżej i powyżej głównej przekątnej liczby od 99 do 1

C = macierz trójkątna górna wypełniona jedynkami (łącznie z główną przekątną)

D = macierz rozmiaru 2×100 , w której pierwszy wiersz zawiera elementy $d_{1i} = 1 + 2 + \dots + i$, a drugi wiersz zawiera elementy $d_{2i} = i!$

E = macierz rozmiaru 100×100 mająca 1 w pozycji (i, j) , jeśli i dzieli j , lub 0 w przeciwnym przypadku.

Zadanie 1. (1 punkt)

a) Niech \mathbf{x} , \mathbf{y} , \mathbf{w} będą wektorami kolumnowymi ustalonej długości d . Policz:

- długość wektora \mathbf{x} ,
- średnią ważoną wektora \mathbf{x} z wagami \mathbf{w} ,
- odległość euklidesową między wektorami \mathbf{x} i \mathbf{y} ,
- iloczyn skalarny wektorów \mathbf{x} i \mathbf{y} .

Obliczenia przeprowadź dla losowo wygenerowanych wektorów \mathbf{x} , \mathbf{y} , \mathbf{w} (dla $d = 100$).

b) Niech \mathbf{X} będzie macierzą ustalonego rozmiaru $d \times N$ zawierającą N wektorów kolumnowych długości d . Niech \mathbf{y} i \mathbf{w} będą wektorami kolumnowymi długości d . Policz

- długości kolejnych wektorów z macierzy \mathbf{X} (wyznacz wektor długości N zawierający te długości),
- średnią ważoną kolejnych wektorów z macierzy \mathbf{X} z wagami \mathbf{w} (wyznacz wektor długości N zawierający te średnie),
- odległości euklidesowe między kolejnymi wektorami z macierzy \mathbf{X} i wektorem \mathbf{y} (wyznacz wektor długości N zawierający te odległości),
- iloczyny skalarnie kolejnych wektorów z macierzy \mathbf{X} i wektora \mathbf{y} (wyznacz wektor długości N zawierający te iloczyny).

Obliczenia przeprowadź dla losowo wygenerowanej macierzy \mathbf{X} i losowo wygenerowanych wektorów \mathbf{y} i \mathbf{w} (dla $d = 100$ i $N = 1000$).

Zadanie 2. (1 punkt)

Napisz funkcję, która dla danych macierzy \mathbf{X} i \mathbf{Y} ustalonego rozmiaru $d \times N$ i $d \times M$ odpowiednio, zawierających N i M wektorów kolumnowych długości d , wyznacza macierz odległości euklidesowych między wektorami tych macierzy, tzn. macierz \mathbf{D} rozmiaru $N \times M$, gdzie $\mathbf{D}(i, j)$ to odległości między i -tym wektorem z macierzy \mathbf{X} i j -tym wektorem z macierzy \mathbf{Y} . Oblicz czas działania napisanej funkcji dla losowo wygenerowanych macierzy \mathbf{X} i \mathbf{Y} dla $d = 100$ i $N = M = 1\,000$ oraz dla $d = 100$, $N = 10\,000$ i $M = 1\,000$.

Zadanie 3. (1 punkt)

Napisz funkcję, która dla danych macierzy \mathbf{X} i \mathbf{Y} ustalonego rozmiaru $d \times N$ i $d \times M$ odpowiednio, zawierających N i M wektorów kolumnowych długości d , wyznacza dla każdego wektora z macierzy \mathbf{X} najbliższego sąsiada spośród wektorów z macierzy \mathbf{Y} , tzn. zwraca wektor \mathbf{h} rozmiaru $1 \times N$, gdzie $\mathbf{h}(i)$ to numer wektora z macierzy \mathbf{Y} będącego najbliższym sąsiadem i -tego wektora z macierzy \mathbf{X} . Napisz też wersję tej funkcji, która wyznacza k najbliższych sąsiadów, tzn. zwraca macierz \mathbf{H} rozmiaru $k \times N$, gdzie $\mathbf{H}(i, j)$ to numer wektora z macierzy \mathbf{Y} będącego i -tym najbliższym sąsiadem j -tego wektora z macierzy \mathbf{X} (czyli $\mathbf{h}(j) = \mathbf{H}(1, j)$).

Zadanie 4. (1 punkt)

Utwórz skrypt generujący losowe sudoku, tzn. macierz M rozmiaru 9×9 zawierającą liczby $1, 2, \dots, 9$, taką, że w każdym wierszu każda z liczb występuje dokładnie jeden raz, w każdej kolumnie każda z liczb występuje dokładnie jeden raz oraz w każdej klatce 3×3 , powstałej przez podział macierzy M dwoma liniami pionowymi i dwoma liniami poziomymi, każda z liczb występuje dokładnie jeden raz.

Zadanie 5. (bonusowe, 1 punkt)

Zapoznaj się z paradoksem Monty'ego Halla (który był podstawą teleturniejów telewizyjnych "Let's make a deal", w polskiej wersji "Idź na całość"). Napisz skrypt symulujący taki teleturniej. Przeprowadź minimum 10 000 prób i oszacuj prawdopodobieństwo wygranej dla strategii pozostawiania przy swoim wyborze oraz dla strategii zmiany wyboru.

UWAGA: Proszę nie korzystać z żadnych funkcji wbudowanych ani bibliotecznych liczących odległości, iloczyny skalarne, itp. Proszę sprawdzić działanie swoich funkcji na przykładowych danych (najlepiej dość dużych rozmiarów). Proszę spróbować ocenić efektywność swoich obliczeń.