# 1. Numpy

1.1 Dla dowolnego wektora liczb wyznacz (najlepiej wylosować 30 liczb z rozkładu jednostajnego z przedziału [-4, 4]):

A. Wszystkie wartości ze zbioru [-2,-1] i [1,2],
B. Wypisz na ekran liczbę wszystkich wartości nieujemnych,
C. Wyznacz średnią arytmetyczna wartości bezwzględnych elementów,
D. Największą i najmniejszą wartość,
E. Wyznacz wartość najbliższą i najdalsza od 0 (zachowując jej znak),
F. Wyznacz wartość najbliższą i najdalszą od 2 (zachowując jej znak),
G. Wypisz na ekran część ułamkową wszystkich elementów,
H. jego postać znormalizowaną - minimum przechodzi na -1, maks na 1, a pozostałe elementy mają zostać liniowo przeskalowane według poniższego wzoru, gdzie min, max to ekstrema wektora, a new max/min to docelowy zakres - w naszym przypadku -1 oraz 1

$$x' = \frac{x - \mathsf{min}}{\mathsf{max} - \mathsf{min}} \cdot (new_{\mathsf{max}} - new_{\mathsf{min}}) + new_{\mathsf{min}}$$

I. średnią wartość kwadratów liczb większych od 3 lub mniejszych od -2
J. wektor napisów składający się z dwóch wartości: "nieparzysta" oraz "parzysta" w zależności od tego czy dany element jest parzysty lub nie
K. Utwórz wektor napisów y o długości takiej samej, jaką ma x, dla ktorego yi przyjmuje wartosc 'nieujemna', jesli xi jest nieujemne oraz 'ujemna' w przeciwnym przypadku,
L. jego średnią (nie używając funkcji mean()),
M. jego wariancję (nie używając funkcji var oraz sd),
N. jego minimum i maksimum,ale nie używając funkcji min i max (rozwiązanie nie musi być optymalne)
O. Utwórz wektor liczbowy y o długosci takiej samej, jaką ma x, dla którego yi przyjmuje wartość k+0.5 wtedy i tylko wtedy, gdy xi należy do [k, k+1), gdzie k to liczba całkowita (prosty histogram).

1.2 Dla dwóch wektorów równej długości x i y oblicz ich korelację ze wzoru:

A. $$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

B. $$r = r_{xy} = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sqrt{\left(\sum x_i^2 - n\bar{x}^2\right)}\sqrt{\left(\sum y_i^2 - n\bar{y}^2\right)}}.$$

C. $$r = r_{xy} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n\sum x_i^2 - (\sum x_i)^2}\sqrt{n\sum y_i^2 - (\sum y_i)^2}}.$$

Sprawdź wyniki dla następujących par wektorów:
x = [1,2,3,4,5,6,7,8,9,10] i y = [10,9,8,7,6,5,4,3,2,1]
x = losowa próbka ze standardowego rozkładu normalnego, y = 5*x+2
x = losowa próbka ze standardowego rozkładu normalnego, y = losowa próbka ze standardowego rozkładu normalnego

1.3 Napisz funkcję, która standaryzuje wartości w zadanym wektorze numerycznym, tj. przeskalowuje elementy w taki sposób, że ich średnia jest równa 0, a odchylenie standardowe jest równe 1.

1.4 Oblicz iloczyn skalarny dwóch wektorów (suma iloczynów ich współrzędnych). $\mathbf{x}^T\mathbf{y} = \sum_{i=1}^{n} x_i y_i$

1.5 Oblicz root-mean-squared-error pomiędzy dwoma wektorami,

$$\mathrm{RMSE}(\mathbf{x},\mathbf{y}) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - x_i)^2}$$

1.6 Oblicz mean-absolute-error pomiędzy dwoma wektorami,

$$\mathrm{MAE}(\mathbf{x},\mathbf{y}) = \frac{1}{n}\sum_{i=1}^{n}|y_i - x_i|$$

1.7 Oblicz median-absolute-error pomiędzy dwoma wektorami,

$$\mathrm{MedAE}(\mathbf{x}, \mathbf{y}) = \mathrm{Median}\left(|y_1 - x_1|, \ldots, |y_n - x_n|\right)$$

1.8 Oblicz znormalizowany dystans Hamminga pomiędzy dwoma wektorami liczb całkowitych, x i y (o równej długości n). Najpierw policz, na ilu współrzędnych wektory x i y się różnią. Następnie podziel tę liczbę przez długość n.

1.9 Oblicz wartość cross-entropy loss pomiędzy wektorem y (liczby 0 albo 1) oraz $\hat{y}$ (wartości rzeczywiste pomiędzy 0 a 1), tj.

$$-\frac{1}{n}\left(\sum_{i=1}^{n} y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)\right)$$

1.10 Zaimplementuj funkcje lead(x, n) oraz lag(x, n). Pierwsza pozbywa się pierwszych n wartości, a następnie dodaje na koniec n wartości brakujących. Przykładowo lead([1,2,3,4,5],2) == [3,4,5,NaN, NaN]. Druga działa analogicznie: lag([1,2,3,4,5],2) == [NaN, NaN, 1,2,3].

1.11 Zaimplementuj cumall() i cumany(), czyli skumulowane wersje all() i any(). Tak więc, przykładowo, cumall([True, True, True, False, True, False]) == [True, True, True, False, False, False], podczas gdy cumany([False, False, True, False, True]) == [False, False, True, True, True].

1.12 Napisz funkcję losuj(), która przyjmuje (a) liczbę całkowitą n, (b) wektor numeryczny x o długości k o unikatowych wartościach, (c) wektor prawdopodobieństw p o długości k. Bazując tylko na rozkładzie jednostajnym, wygeneruj n losowych wartości z x zgodnie z prawdopodobieństwem p.
Algorytm: wygeneruj losową wartość u z [0,1]. Znajdź takie m, że wartość u znajduje się pomiędzy sumą m-1 pierwszych prawdopodobieństw z wektora p, a sumą m pierwszych prawdopodobieństw z wektora p. Zwróć m-tą wartość z x.

1.13 Napisz funkcję, która jako argumenty przyjmuje: (a) rosnąco posortowany wektor x o długości n, (b) dowolny wektor y o długości n, (c) wektor wartości z o długości k i elementach z przedziału tego samego, co wartości w x. Niech f będzie kawałkami liniową interpolacją punktów $(x_1, y_1), ..., (x_n, y_n)$. Zwróć wektor w o długości k, taki że $w_i = f(z_i)$.

1.14 Mając daną macierz n x k z elementami rzeczywistymi, zastosuj funkcję softmax do każdego wiersza, to znaczy $x_{i,j} \mapsto \frac{\exp(x_{i,j})}{\sum_{l=1}^{k} \exp(x_{i,l})}$. Następnie dokonaj kodowania one-hot dla każdego wiersza, czyli znajdź kolumnę z wartością najbliżej 1 (największą wartością). Zwróć wektor o długości n.

1.15 Dla danego wektora liczbowego x o nieparzystej długości n i wartości k <= (n−1)/2 wyznacz tzw. średnią k-winsorowską, tj. średnią arytmetyczną z podwektora x, w którym k najmniejszych i k największych elementów zostaje zastąpionych przez, odpowiednio, (k + 1)-szą wartość najmniejszą i największą.

1.16 Napisz funkcję factorial_stirling(), która zwraca przybliżoną wartość silni według wzoru Stirlinga: $n! \approx (\frac{n}{e})^n \sqrt{2\pi n}$

1.17 Korzystając ze wzoru Leibniza $\sum_{n=0}^{\infty} \frac{(-1)^n}{2n+1} = \pi/4$ oblicz przybliżoną wartość liczby pi dla 1 000, 10 000 i 100 000 początkowych wyrazów i porównaj uzyskane liczby z wartością R-owej stałej `pi`.

1.18 Podana wyżej metoda nie jest jedyną na przybliżanie liczby pi. Skorzystamy teraz z metody Monte Carlo, której algorytm wygląda następująco:

1. Wylosuj n punktów z dwuwymiarowej przestrzeni [-1,1] x [-1,1]
2. Sprawdź ile punktów jest oddalonych od punktu (0,0) o mniej niż 1
3. Podziel tę liczbę przez n i przemnóż przez 4

Do losowania punktów użyj funkcji np.random.uniform (sprawdź w dokumentacji jak)

1.19 Dla danego wektora liczbowego zawierającego braki danych (`NA`) należy uzupełnić je średnią wartością pozostałych, poprawnych wartości.

1.20 Wypisz w postaci wektora dziesięć pierwszych liczb rozwinięcia dziesiętnego liczby pi (korzystając ze stałej R-owej `pi`)
Wynik powinien być następujący:
[3 1 4 1 5 9 2 6 5 3 5]

1.21 Napisz funkcję, która zwraca modę, tzn. Najczęściej pojawiającą się wartość w wektorze. Jeśli moda nie jest unikalna, zwróć dowolną.

1.22 Napisz samodzielnie funkcję regresja(), która policzy współczynniki α, β w modelu prostej regresji liniowej, tj.y = α + βx. Funkcja na wejściu przyjmuje dwa wektory liczbowe x, y.

$$\beta = \frac{\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{\sum_{i=1}^{n}(x_i-\overline{x})^2}$$

$$\alpha = \overline{y} - \beta\overline{x}$$

1.23 Dla dowolnej macierzy odwróć kolejność jej wierszy
1.24 Dla dowolnej macierzy odwróć kolejność jej kolumn
1.25 Niech macierz X przechowuje n punktów (n wierszy) o d współrzędnych (d kolumn). Znaleźć boks okalający te punkty. Innymi słowy, zwrócić macierz B o 2 wierszach i d kolumnach taką że pierwszy wiersz to minimalne wartości w każdym wymiarze, a drugi wiersz to maksymalne wartości w każdym wymiarze.

1.26 Niech X przechowuje n punktów (n wierszy) o d współrzędnych (d kolumn). Niech Y przechowuje m punktów (m wierszy) o d współrzędnych (d kolumn). Zwróć wektor R, który ma długość m, taki że jego i-ta współrzędna oznacza indeks punktu w X, który ma najmniejszy dystans do i-tego punktu w Y.

1.27 Mamy dany wektor T o długości n liczb ze zbioru {0,...,k-1}. Napisz funkcję, która dokona kodowania one-hot-encode dla każej wartości z T. Innymi słowy zwróć macierz zer i jedynek R o n wierszach i k kolumnach, taką że jedynka znajduje się w i-tym wierszu na j-tej kolumnie wtedy i tylko wtedy, gdy i-ty element wektora T równa się j.

1.28 Prawą macierzą stochastyczną nazywamy macierz kwadratową, której elementami są nieujemne liczby rzeczywiste i w której każdy wiersz sumuje się do jedynki. Przykładem prawej macierzy stochastycznej jest macierz:

R =
0.5 & 0.3 & 0.2 \\
0.2 & 0.8 & 0 \\
0.3 & 0.3 & 0.4

Napisz funkcję doStochastycznej(), która przyjmuje kwadratową macierz (niekoniecznie prawą stochastyczną) i wykonuje następujące czynności:

1. Sprawdza, czy wszystkie elementy są nieujemne
2. Sprawdza, czy w każdym wierszu jest co najmniej jeden element większy od zera,
3. Tworzy nową macierz na bazie wejściowej macierzy, w której wiersze są ,,unormowane'', czyli sumują się do jedynki. Dla przykładu, dla następującego wiersza: 5, 3, 2, w macierzy zwracanej ten wierszy powinien wyglądać tak: 0.5, 0.3, 0.2.

1.29

Dana jest macierz $P \geqslant 0$ rozmiaru $n \times m$ taka, że $\sum_{i=1}^{n} \sum_{j=1}^{m} p_{i,j} = 1$ oraz posortowane rosnąco wektory liczbowe $\mathtt{x}$ ($n$-elementowy) i $\mathtt{y}$ ($m$-elementowy). Trójka $(\mathtt{x}, \mathtt{y}, P)$ opisuje rozkład prawdopodobieństwa pewnej dwuwymiarowej zmiennej losowej dyskretnej $(X, Y)$, tak jak w poniższym podzadaniu.

*W pewnej szkole rozkład prawdopodobieństwa uzyskania ocen z Filozofii Bytu i Wychowania Fizycznego przez tego samego studenta przedstawia się następująco.*

| | | *WF* | | | |
|---|---|---|---|---|---|
| | | *2* | *3* | *4* | *5* |
| *FB* | *2* | *0* | *0,01* | *0,1* | *0,2* |
| | *3* | *0,01* | *0,05* | *0,03* | *0,1* |
| | *4* | *0,1* | *0,03* | *0,05* | *0,01* |
| | *5* | *0,2* | *0,1* | *0,01* | *0* |

a) Zmienne losowe $X$ i $Y$ są niezależne wtedy i tylko wtedy, gdy dla każdego $i, j$ zachodzi $p_{i,j} = (\sum_{k=1}^{n} p_{k,j})(\sum_{l=1}^{m} p_{i,l})$. Napisz funkcję `niezaleznosc()`, która sprawdza, czy zachodzi ta własność dla danych $(\mathtt{x}, \mathtt{y}, P)$ (zwracamy wartość logiczną).
b) Ponadto napisz funkcję `podststat()`, która dla $(\mathtt{x}, \mathtt{y}, P)$ zwróci wektor liczbowy (z ustawionym atrybutem `names` – dowolne, lecz czytelne dla użytkownika etykiety) zawierający wartości podstawowych charakterystyk $(X, Y)$:

- Wartości oczekiwane: $\mathbb{E}\, X = \sum_{i=1}^{n} x_i \sum_{j=1}^{m} p_{i,j}$, $\mathbb{E}\, Y = \sum_{j=1}^{m} y_j \sum_{i=1}^{n} p_{i,j}$,
- Wariancje: $\operatorname{Var} X = \mathbb{E}\, X^2 - (\mathbb{E}\, X)^2$, gdzie $\mathbb{E}\, X^2 = \sum_{i=1}^{n} x_i^2 \sum_{j=1}^{m} p_{i,j}$ oraz $\operatorname{Var} Y = \mathbb{E}\, Y^2 - (\mathbb{E}\, Y)^2$, gdzie $\mathbb{E}\, X^2 = \sum_{j=1}^{m} y_j^2 \sum_{i=1}^{n} p_{i,j}$,
- Kowariancję: $\operatorname{Cov}(X, Y) = \mathbb{E}\,(XY) - \mathbb{E}\, X\, \mathbb{E}\, Y$ dla $\mathbb{E}\,(XY) = \sum_{i=1}^{n} \sum_{j=1}^{m} x_i y_j p_{i,j}$,
- Współczynnik korelacji: $\varrho(X, Y) = \operatorname{Cov}(X, Y) / \sqrt{\operatorname{Var} X \operatorname{Var} Y}$.