Macierze

Wszystkie zadania rozwiązujemy bez używania pętli.

Zadanie 5.1 (MG)**.**
Niech `t` będzie wektorem o n elementach będących liczbami całkowitymi ze zbioru $\{1, \ldots, k\}$. Napisz funkcję, która dokona kodowania elementów t_i (*one-hot-encode*). Funkcja powinna zwracać macierz zero-jedynkową R wymiaru $n \times k$ taką, że $r_{i,j} = 1$ wtedy i tylko wtedy gdy $t_i = j$. Taka reprezentacja jest przydatna np. w problemie klasyfikacji wieloetykietowej przy użyciu k klasyfikatorów.

Let `t` *be vector of n integers in $\{1, \ldots, k\}$. Write a function to one-hot-encode each t_i. Return a 0-1 matrix R of size $n \times k$ such that $r_{i,j} = 1$ if and only if $t_i = j$. By the way, such a representation is useful when solving, e.g., a multiclass classification problem by means of k binary classifiers.*

Zadanie 5.2 (MG)**.**
Dokonaj przekształcenia *softmax* każdego wiersza macierzy $\mathbf{X} \in \mathbb{R}^{n\times k}$, tzn. przekształcenia postaci:

$$x_{i,j} \mapsto \frac{\exp(x_{i,j})}{\sum_{l=1}^{k} \exp(x_{i,l})}.$$

Następnie dokonaj odkodowania każdego wiersza (*one-hot decode*), tj. dla każdego wiersza należy znaleźć numer kolumny o wartości najbardziej zbliżonej do 1. Zwróć wektor n-elementowy.

Given an $n \times k$ matrix with elements in $\mathbb{R}$, apply the softmax function to each row, i.e., $x_{i,j} \mapsto \frac{\exp(x_{i,j})}{\sum_{l=1}^{k} \exp(x_{i,l})}$. Then one-hot decode the values in each row, i.e., find the column number with the value most close to 1. Return a vector of size n.

Zadanie 5.3 (MG)**.**
Niech dana będzie macierz $\mathbf{X} \in \mathbb{R}^{n\times d}$. Wyznacz przedział wielowymiarowy ograniczający wartości n punktów reprezentowanych jako $\mathbf{X}$. Dokładniej, wyznacz i zwróć macierz $\mathbf{B} \in \mathbb{R}^{2\times d}$ taką, że $b_{1,j} = \min_i x_{i,j}$ oraz $b_{2,j} = \max_i x_{i,j}$.

For a given $\mathbf{X} \in \mathbb{R}^{n\times d}$, determine the bounding hyperrectangle of the n points. Return a matrix $\mathbf{B} \in \mathbb{R}^{2\times d}$ with $b_{1,j} = \min_i x_{i,j}$ and $b_{2,j} = \max_i x_{i,j}$.

Zadanie 5.4 (MG)**.**
Niech macierz $\mathbf{X}$ wymiaru $n \times d$ reprezentuje n punktów w $\mathbb{R}^d$. Napisz funkcję, która wyznaczy i zwróci odległości między punktami z $\mathbf{X}$ oraz (danym jako drugi argument funkcji) punktem $\mathbf{y} \in \mathbb{R}^d$. Funkcja powinna zwrócić wektor $\mathbf{d} \in \mathbb{R}^n$ taki, że $d_i = \|\mathbf{x}_{i,\cdot} - \mathbf{y}\|_2$.

Assume that an $n \times d$ matrix $\mathbf{X}$ represents n points in $\mathbb{R}^d$. Write a function that determines the pairwise distances between all the points in $\mathbf{X}$ and a given $\mathbf{y} \in \mathbb{R}^d$. Return a vector $\mathbf{d} \in \mathbb{R}^n$ with $d_i = \|\mathbf{x}_{i,\cdot} - \mathbf{y}\|_2$.

Zadanie 5.5 (*MG; Dyskretna dwuwymiarowa zmienna losowa*)**.**
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 `x` (n-elementowy) i `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: $\varrho(X, Y) = \operatorname{Cov}(X, Y) / \sqrt{\operatorname{Var} X \operatorname{Var} Y}$.

Zadanie 5.6 (*MG; Średnia ruchoma*)**.**
Napisz funkcję `movingavg()`, która dla danego szeregu czasowego `x` o n elementach oraz nieparzystej liczby naturalnej k wyznaczy k-średnią ruchomą, $k < n$, tj. szereg czasowy $(w_1, \ldots, w_{n-k+1})$, dla którego $w_i = \sum_{j=1}^{k} x_{i+j-1}/k$. Jednostki czasu dla wynikowego szeregu dobierz wedle uznania.

Wbudowany szereg czasowy `UKgas` zawiera dane na temat kwartalnej konsumpcji gazu w Zjednoczonym Królestwie. Użyj go do przetestowania zaimplementowanego algorytmu. Szereg możesz narysować, wywołując `plot(UKgas)`.