НАЦИОНАЛНА ОЛИМПИАДА ПО ИЗКУСТВЕН ИНТЕЛЕКТ · 2026/2027

Общински кръг — Примерна тема 1

Време за работа: 3 часа · 3 задачи по 100 точки · Език: Python

Указания

- Решението чете от стандартния вход (`input()` или `sys.stdin`) и пише на стандартния изход (`print`).
- Ограничение по време: 1 секунда на тест (2 секунди за задача 1.3). Памет: 256 MB.
- Реалните числа в изхода се приемат при абсолютна грешка до 0.01.
- Точките се дават по подзадачи: за точките на една подзадача трябва да минат всички нейни тестове.
- Не се разрешава използването на генеративни асистенти и достъп до интернет (т. 1.9 от регламента).

Задача 1.1. Сензори

Метеорологична станция записва n показания на термометър, по едно на ред. Сензорът понякога греши и вместо число записва текст или празен ред. Преди да се обучи какъвто и да е модел, данните трябва да се почистят и опишат.

Едно показание е **валидно**, ако `float(ред)` се изпълнява без грешка и резултатът е крайно число (не е `nan`, `inf` или `-inf`). За валидните показания изчислете минимум, максимум, средно аритметично, медиана и стандартно отклонение.

- Медиана: средният елемент на сортираните стойности. При четен брой е средното аритметично на двата средни.
- Стандартно отклонение (за цялата съвкупност от v валидни стойности):

$$\sigma = \sqrt{\frac{1}{v}\sum_{i=1}^{v}\left(x_i - \bar{x}\right)^2}$$

Вход. На първия ред е n. Следват n реда, всеки с едно показание (произволен низ, възможно и празен).

Изход. На първия ред: броят валидни и броят невалидни показания. Ако няма валидни, на втория ред изведете `NO DATA`. Иначе на втория ред — минимума и максимума, а на третия — средното, медианата и стандартното отклонение, всички с точно 2 знака след десетичната точка.

Ограничения. $1 \le n \le 10^5$. Всеки ред е до 20 символа. Валидните стойности са между −1000 и 1000.

Подзадача	Точки	Допълнително условие
1	30	Всички показания са цели числа, n е нечетно
2	30	Невалидните редове съдържат само букви или са празни
3	40	Без допълнителни условия (включително `NaN`, `inf`, `12,5`)

Пример 1 — вход

```
8
21.5
19
err

22.75
NaN
12,5
-3
```

Изход

```
4 4
-3.00 22.75
15.06 20.25 10.52
```

Валидни са 21.5, 19, 22.75 и −3. Невалидни са `err`, празният ред, `NaN` (Python го преобразува, но не е крайно число) и `12,5` (запетая вместо точка).

Пример 2 — вход

```
3
inf
--
null
```

Изход

```
0 3
NO DATA
```

Задача 1.2. Речник

Езиковите модели не работят с думи, а с числа. Затова текстът се разделя на думи, строи се речник и всяка дума се заменя с номера си в речника.

1. **Дума** е максимална поредица от последователни букви (символи, за които `str.isalpha()` връща `True`; включва латиница и кирилица). Думите се преобразуват в малки букви. Цифри, интервали и пунктуация са разделители.
2. Думите от обучаващия текст се подреждат по брой срещания намаляващо, а при равенство — лексикографски (както сравнява низове Python).
3. В речника остават първите K думи (или всички, ако са по-малко). Те получават номера 1, 2, … в този ред.
4. Всяка заявка (изречение) се кодира: всяка дума се заменя с номера си, а дума извън речника — с 0.

Вход. На първия ред е K. Следват редовете на обучаващия текст до ред, съдържащ само `###`. След него е броят заявки q и q реда със заявки.

Изход. Размерът V на речника. След това V реда `номер дума брой`. Накрая за всяка заявка един ред с кодовете, разделени с интервал (празен ред, ако заявката няма думи).

Ограничения. $1 \le K \le 10^4$, $0 \le q \le 1000$. Общата дължина на входа е до 10^6 символа.

Подзадача	Точки	Допълнително условие
1	20	Текстът и заявките съдържат само малки латински букви и единични интервали
2	30	K е не по-малко от броя различни думи
3	50	Без допълнителни условия

Пример — вход

```
3
The cat sat on the mat.
The dog sat, too!
###
2
The cat and the dog
A mat? On the mat!
```

Изход

```
3
1 the 3
2 sat 2
3 cat 1
1 3 0 1 0
0 0 0 1 0
```

`cat`, `dog`, `mat`, `on` и `too` се срещат по веднъж; лексикографски първа е `cat`, така че тя заема третото място.

Задача 1.3. Най-светлият прозорец

Черно-бяло изображение е таблица n × m от яркости между 0 и 255. Конволюционните невронни мрежи обработват такива изображения през квадратни **прозорци**. Прозорец с размер k × k е квадрат от съседни пиксели, който се побира изцяло в изображението. Той се определя от горния си ляв ъгъл (ред r, стълб c, номерирани от 1). **Средната яркост** на прозореца е сумата на пикселите му, разделена на k·k.

За дадените k и T намерете:

- прозореца с най-голяма сума. При равенство се избира този с най-малък r, а при равни r — с най-малък c;
- броя прозорци със средна яркост поне T.

Вход. На първия ред са n, m, k и T. Следват n реда с по m цели числа.

Изход. На първия ред: r, c и сумата на най-светлия прозорец. На втория: броят прозорци със средна яркост поне T.

Ограничения. 1 ≤ n, m ≤ 700, 1 ≤ k ≤ min(n, m), 0 ≤ T ≤ 255.

Подзадача	Точки	Допълнително условие
1	20	n, m ≤ 30
2	30	n, m ≤ 150
3	50	Без допълнителни условия

Пример — вход

```
4 5 2 150
10 200 220 30 0
40 180 250 60 90
0 20 100 255 255
5 15 90 250 240
```

Изход

```
3 4 1000
5
```

Най-светъл е прозорецът в долния десен ъгъл: 255 + 255 + 250 + 240 = 1000. Средна яркост поне 150 означава сума поне 600. Такива са пет от дванадесетте прозореца — с горен ляв ъгъл (1, 2), (2, 3), (2, 4), (3, 3) и (3, 4).