

Задание 1. В лаборатории «AI-Кузбасс» разрабатывают систему автоматической классификации изображений угля. Уже есть две обученные модели — Модель А (h_1) и Модель Б (h_2). Они делают предсказания для 10 изображений угля: каждая модель выдает класс угля (Д (длиннопламенный) или К (кокующийся)), к которому, по её мнению, принадлежит изображение. Окончательное решение по каждому снимку принимается большинством голосов систем автоматической классификации.

Известны истинные метки для этих изображений:

i	1	2	3	4	5	6	7	8	9	10
y	Д	К	Д	К	Д	К	Д	К	Д	К
h_1	Д	Д	Д	К	К	Д	Д	К	К	К
h_2	Д	Д	К	К	К	Д	Д	Д	К	К

Инженеры-программисты «AI-Кузбасс» разработали третью систему классификации Модель В (h_3), но как оказалось она конфликтует с Модель Б (h_2) и дает противоположный результат относительно этой модели. Посчитайте долю правильных ответов относительно известных меток и решений систем автоматической классификации угля.

Задача 2. Банковская система “Витязь” с внедренным ИИ способна выявлять мошеннические транзакции. Вероятность того, что случайная транзакция является мошеннической, составляет 0,1%. Алгоритм обнаруживает мошенничество с вероятностью 90% в случае мошенничества и ошибочно помечает не мошеннические транзакции как мошеннические с вероятностью 0,5%.

Какова вероятность, что транзакция действительно мошенническая, если алгоритм пометил ее как мошенническую?

Если процент будет дробным, округлите до целого числа.

Задача 3. В электронной почте есть 12 сообщений. Каждое сообщение оценивается по степени вероятности того, что это спам.

- **Легкие спамы:** вероятность < 0.4 — всего 4 сообщения.
- **Средние спамы:** вероятность от 0.4 до 0.7 — 3 сообщения.
- **Тяжелые спамы:** вероятность > 0.7 — 5 сообщений.

Создана модель, которая по порогу T (от 0 до 1) классифицирует сообщение как спам, если вероятность $\geq T$. Выбрать целое значение порога T (от 0 до 1, в десятичных долях), которая будет минимизировать сумму ошибок первого и второго рода.

Задача 4. В системе ИИ для обнаружения аномалий трафика используются три независимых скрипта (α , β , γ). Каждое событие срабатывает с вероятностью $1/2$ при обнаружении необычного поведения в сети. Система активируется, если хотя бы два скрипта сработали. Разработчик упростил логику: ИИ принимает решение, что аномалия есть, если сработает датчик β .

Какова вероятность, что ИИ правильно определит наличие аномалии, если по сути система должна активироваться при случае отработки от 2ух скриптов?

Задача 5. ИИ "Палиндром" анализирует тексты на предмет скрытых симметричных структур. В рамках исследования требуется определить для определенного символа в тексте максимальную длину зеркальной последовательности.

Формат входных данных.

Первая строка: анализируемый текст s (строчные латинские буквы a-z).

Вторая строка: один символ c — символ для анализа.

Третья строка: целое число k — номер вхождения символа c в тексте (нумерация с 1 и номер вхождения k гарантированно не превышает количество вхождений символа c в тексте.).

Формат выходных данных.

Первая строка: одно целое число — максимальную длину зеркальной последовательности с указанным вхождением символа c в центре.

Вторая строка: сама зеркальная последовательность. Если симметричной последовательности не существует, вывести 1 и символ c .

Ограничения.

Длина текста: $1 \leq |s| \leq 10^6$.

Пример.

Ввод

abacabadabacaba

c

2

Вывод

7

abacaba

Задача 6. ИИ обнаружил, что последние измерения в окне важнее для прогноза заморозков. Он применяет взвешенное среднее, где веса заданы. Вывести для каждого окна(b_i) как несократимую дробь или целое.

$$b_i = (w_1 * a_i + w_2 * a_{i+1} + \dots + w_k * a_{i+k-1}) / (w_1 + \dots + w_k)$$

Формат входных данных:

Первая строка: n, k ($1 \leq k \leq n \leq 100, k \leq 10$) — всего измерений температуры и размер окна.

Вторая строка: $a_1 \dots a_n$ (целые, $|a_i| \leq 1000$) — значения температур

Третья строка: $w_1 \dots w_k$ (целые положительные веса, сумма весов не равна 0) — веса для позиций в окне.

Формат выходных данных:

Значение окон в 1 строку. Если значением является дробь, вывести в формате {числитель}/{знаменатель}.

Пример:

Ввод	Вывод
5 3	2 3 4
1 2 3 4 5	
1 2 1	

Задача 7. Банк проанализировал $N=8000$ заявок на кредит. Из них $P=400$ в итоге оказались дефолтными (не вернули кредит), остальные 7600 — исправные заёмщики.

Скоринг-модель помечает заёмщика как «рисковый» или «надёжный».

TP — правильно предсказанные дефолтные;

FP — надёжные, ошибочно названные рисковыми;

FN — дефолтные, ошибочно названные надёжными;

TN — правильно предсказанные надёжные.

Recall: $R = TP/P$

Accuracy: $A = (TP+TN)/N$.

Определить **минимальное** и **максимальное** значение точности A .

Задача 8. Дан табличный файл в формате CSV. В таблице даны три столбца с натуральными числами: a, b, c (по одной строке на объект). Задан прототип $p=(3,2,1,d)$, для которого сумма равна $= 3+2+1+k=6+d$. Для каждой строки вычислите сумму $S=a+b+c+d$, где $d = (a+b+c)/2$. Строка считается похожей на прототип, если $|S-8| \leq 20$. Требуется подсчитать количество похожих на прототип строк в таблице и вывести одно это число.