

Максимальное количество баллов за олимпиаду — 600

Задание 1. Картинки в папке

В папке на компьютере размещено 3000 изображений трёх типов: с кошкой, с собакой, с кошкой и собакой одновременно. Изображений всех типов одинаковое количество. Модель ИИ ответила на два вопроса к каждой картинке: есть ли на ней кошка и есть ли на ней собака. На первый вопрос было получено 2790 ответов «да», а на второй — 2861. Удалили все изображения, для которых хотя бы один ответ ИИ был неверным. Какое наименьшее число изображений могло остаться в папке?

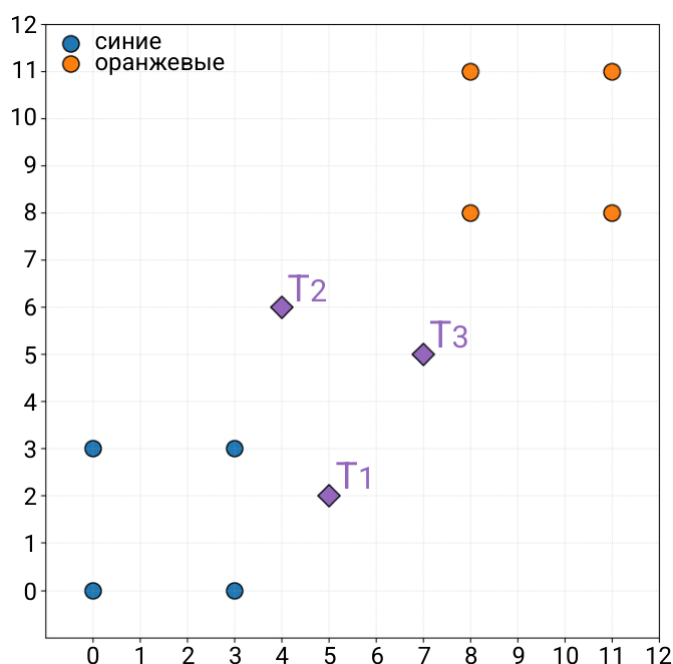
Задание 2. Сколько расстановок дают положительный определитель

Матрицей $n \times m$ будем называть таблицу из чисел, состоящую из n строк и m столбцов. Определитель матрицы 2×2 — это число, вычисляемое по формуле. Для $\begin{pmatrix} p & q \\ r & s \end{pmatrix}$ его значение равно $ps - qr$.

Сколько существует способов расставить числа 1, 2, 3, 6 по клеткам матрицы 2×2 (каждое число используется ровно один раз) так, чтобы $ps - qr > 0$?

Задание 3. Свой среди чужих

На плоскости даны восемь обучающих точек — четыре синие и четыре оранжевых, а также три тестовые точки T_1 , T_2 , T_3 , обозначенные фиолетовыми ромбиками.



Цвет тестовой точки будем предсказывать по цветам её соседей.

Иногда разные координаты имеют очень разный масштаб: одна может меняться от 0 до 10000, а другая — от 0 до 0.00001. Тогда при вычислении расстояний первая координата начинает полностью «перебивать» вторую. Чтобы обе координаты влияли честно, первую иногда растягивают или сжимают.

В этой задаче первая координата умножается на положительное число $\lambda > 0$:

$(x, y) \mapsto (\lambda x, y)$ Расстояние между точками считаем евклидовым в новых координатах $(\lambda x, y)$. Цвет тестовой точки предсказываем так:

- находим 3 ближайших к ней обучающих точки;
- если среди них больше синих, считаем тестовую точку синей, если больше оранжевых — оранжевой;
- если расстояния совпадают, то более близкой считается точка с меньшим λx , а при равных λx — с меньшим y .

Зафиксируем $\lambda = 1$ и посмотрим, какого цвета при этом каждая тестовая точка T_i .

Назовём значение $\lambda > 0$ *интересным* для точки T_i , если при таком λ предсказанный цвет T_i отличается от её цвета при $\lambda = 1$.

Рассмотрим для каждой точки все её интересные значения λ .

Оказывается, что для каждой T_i возможны только два случая:

- интересных значений нет вообще — тогда цвет точки не меняется ни при каком $\lambda > 0$;
- интересные значения существуют и устроены так: есть число $L_i > 0$, что при всех $0 < \lambda < L_i$ цвет точки T_i другой, чем при $\lambda = 1$, а при всех $\lambda > L_i$ цвет уже такой же, как при $\lambda = 1$. При $\lambda = L_i$ цвет может совпадать или не совпадать с цветом при $\lambda = 1$ — это надо аккуратно проверить.

Число L_i будем называть *границей* интересных значений для точки T_i . Для каждой тестовой точки T_i :

- выведите -1 , если интересных значений λ нет;
- найдите её границу L_i и выведите число $(L_i)^2$ в ином случае.

Задание 4. Мониторинг энергопотребления магазина

В течение нескольких дней подряд в магазине измеряли энергопотребление каждые 15 минут. Таким образом, за одни сутки получается 96 измерений. Результаты вы можете скачать в форматах [XLSX](#), [ODS](#) или [CSV](#).

В таблице содержится непрерывный ряд измерений без пропусков; в каждой строке указаны t (номер измерения, начиная с 0) и **power** (потребление в ваттах). Обозначим $x_t = \text{power}[t]$.

Чтобы проанализировать динамику потребления, рассмотрим несколько производных величин.

Во-первых, будем отслеживать изменение мощности от шага к шагу. Для каждого $t \geq 1$ определим разность: $d_t = x_t - x_{t-1}$. Во-вторых, нас интересуют сглаженные характеристики. Скользящие средние вычисляются по «хвостовым» окнам, то есть берут текущий момент и несколько предыдущих:

- среднее потребление за последние 8 шагов: $\text{ma_8}[t] = \frac{1}{8} \sum_{j=0}^7 x_{t-j} \quad (t \geq 7)$,
- среднее изменение потребления за последние 4 шага: $\text{ma_diff_4}[t] = \frac{1}{4} \sum_{j=0}^3 d_{t-j} \quad (t \geq 4)$.

Кроме того, по всему ряду значений x_t необходимо найти медиану **med**. Медиана — это число, которое стоит посередине после сортировки всех значений (или среднее двух центральных при чётном числе точек).

Поскольку измерения выполнялись много дней подряд, выражение $t \bmod 96$ указывает, на каком 15-минутном интервале суток находится момент t . Значения от 36 до 84 соответствуют промежутку с 09:00 до 21:00 включительно.

Сколько моментов времени t одновременно удовлетворяют трём условиям:

$$\text{ma_diff_4}[t] \geq 3.0, \text{ma_8}[t] \geq \text{med}, 36 \leq (t \bmod 96) \leq 84?$$

Задание 5. Оптимальная конфигурация

Ограничение по времени: 1 секунда

Ограничение по памяти: 256 мегабайт

Обучение больших нейронных сетей происходит на кластере — вычислительной инфраструктуре, которая позволяет использовать одновременно большое количество видеокарт. Видеокарты на кластере объединяются в ноды (блоки по 8 карт).

Для обучения больших языковых моделей методом обучения с подкреплением видеокарты делятся на два типа: одни генерируют данные, другие обновляют веса модели на основе генерированных данных. Вторая задача гораздо сложнее, поэтому при наилучшем распределении ресурсов на неё должно выделяться в k раз больше видеокарт. Обучение с оптимальным распределением ресурсов мы будем называть оптимальным.

Вы работаете в очень большой компании, поэтому ваши ресурсы не ограничены. Вы хотите выделить минимальное количество нод так, чтобы их видеокарты можно было разбить на два вышеописанных типа так, чтобы обучение при этом было оптимальным.

Формат входных данных

В первой и единственной строке входных данных вводится одно целое число k ($1 \leq k \leq 1000000$).

Формат выходных данных

Выведите одно число: минимальное число нод, которые нужно выделить так, чтобы обучение было оптимальным.

Примеры

стандартный ввод	стандартный вывод
5	3

Задание 6. Распознавание объектов

Ограничение по времени: 1 секунда

Ограничение по памяти: 256 мегабайт

Одна из задач машинного обучения — распознавание объектов. Она заключается в том, чтобы выделить на изображении зоны, в которых находятся объекты, и сказать, что это за объекты.

Гоша запустил умную нейронную сеть для распознавания объектов. Она вернула таблицу $n \times m$, где в каждой клетке указан номер класса, к которому принадлежит находящийся в клетке объект по мнению модели, или 0, если модель считает, что в этой клетке нет никаких интересующих её объектов.

К сожалению, этого недостаточно: Гоша хочет выделить в таблице прямоугольники-рамки со сторонами, параллельными её краям, так, чтобы внутри этих рамок были распознанные объекты. Объектами Гоша считает все клетки, которые модель распознала как клетки одного конкретного класса.

Помогите Гоше: выделите в таблице все распознанные умной нейронной сетью объекты. Так как объектов в таблице может быть очень много, посчитайте и выведите всего одно число: сумму площадей всех прямоугольников-рамок объектов.

Формат входных данных

В первой строке вводятся два целых числа n, m ($1 \leq n, m \leq 1000$).

В следующих n строках вводится по m чисел a_{ij} — номера классов объектов, которые модель нашла в конкретной клетке, или 0, если в этой клетке не был распознан ни один объект.

Формат выходных данных

Выведите одно целое число: сумму площадей прямоугольников-рамок для всех распознанных объектов.

Примеры

стандартный ввод	стандартный вывод
1 0 2 0 0 0 1 2 0 3 1 0 2 4 0 2 2 2 2 3	26

Замечание

Рассмотрим первый пример из условия.

