

Задача 1

Кофейня

В небольшой кофейне используются единые технологические карты приготовления напитков. Каждый напиток состоит из нескольких ингредиентов, а его себестоимость равна сумме стоимостей всех использованных ингредиентов.

После сбоя в системе учёта была утеряна информация о стоимости каждого ингредиента, однако технологические карты и себестоимость приготовленных напитков сохранились.

Известно, что в кофейне используются только следующие ингредиенты (для каждого указано количество в одной упаковке):

- молотый кофе — 1 кг;
- вода — 19 литров;
- молоко — 1,8 л;
- сливки — 480 мл;
- шоколадный сироп — 400 мл;
- карамельный сироп — 400 мл;
- ванильный сироп — 400 мл;
- сахар — 1 кг;
- лёд — 5 кг;
- какао-порошок — 480 г.

Для сыпучих и твёрдых ингредиентов количество указано в граммах, для жидких — в миллилитрах.

Также известно, что в среднем за день продаётся по 250 чашек каждого напитка. Ингредиенты закупаются в упаковках (пачках или бутылках) фиксированного объёма или веса. Даже если к концу дня упаковка используется не полностью, она считается израсходованной целиком.

Используя данные технологических карт, определите, сколько денег необходимо тратить в день на закупку всех ингредиентов.

Напиток	Состав	Себестоимость
Эспрессо	Кофе — 18 г, вода — 36 мл	36,72 Р
Доппио	Кофе — 36 г, вода — 72 мл	73,44 Р
Американо	Кофе — 18 г, вода — 180 мл	39,60 Р
Капучино	Кофе — 18 г, вода — 30 мл, молоко — 140 мл	50,60 Р
Латте	Кофе — 18 г, вода — 30 мл, молоко — 220 мл	58,60 Р
Флэт уайт	Кофе — 36 г, вода — 30 мл, молоко — 160 мл	88,60 Р
Ванильный раф	Кофе — 18 г, вода — 30 мл, сливки — 180 мл, ванильный сироп — 20 мл, сахар — 10 г	84,80 Р
Мокка	Кофе — 18 г, вода — 30 мл, молоко — 180 мл, шоколадный сироп — 25 мл, какао-порошок — 5 г	69,35 Р
Карамельный латте	Кофе — 18 г, вода — 30 мл, молоко — 210 мл, карамельный сироп — 25 мл	67,60 Р
Ванильный латте	Кофе — 18 г, вода — 30 мл, молоко — 210 мл, ванильный сироп — 25 мл	67,60 Р

Напиток	Состав	Себестоимость
Айс-латте	Кофе — 18 г, вода — 30 мл, молоко — 180 мл, лёд — 120 г	56,40 Р
Айс-карамельный макиато	Кофе — 18 г, вода — 30 мл, молоко — 170 мл, карамельный сироп — 20 мл, ванильный сироп — 15 мл, лёд — 120 г	69,40 Р
Горячий шоколад	Молоко — 220 мл, шоколадный сироп — 35 мл, сахар — 10 г, какао-порошок — 8 г	43,95 Р
Какао	Молоко — 220 мл, сахар — 12 г, какао-порошок — 18 г	35,32 Р
Карамельный фраппе	Кофе — 18 г, вода — 30 мл, молоко — 140 мл, карамельный сироп — 20 мл, сахар — 10 г, лёд — 180 г	61,90 Р

Задача 2

Сбой датчика

В ходе эксперимента было получено n точек на плоскости:

$$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n).$$

Предполагается, что результаты эксперимента подчиняются зависимости

$$y = kx,$$

где k — некоторое вещественное число.

Однако один из датчиков сработал неправильно, поэтому ровно одна точка является сильным выбросом. Если удалить эту точку, то для оставшихся $n-1$ измерений можно подобрать коэффициент k , при котором сумма квадратов ошибок минимальна:

$$E(k) = \sum (y_i - kx_i)^2.$$

Для каждого возможного удаления точки i обозначим через E_i минимальное значение суммы квадратов ошибок для оставшихся точек.

Гарантируется, что существует единственный индекс p , для которого значение E_p минимально.

Найдите ошибочное измерение.

Задача 3

Пороговый классификатор

В сервисе рекомендаций используется простой автоматический классификатор. Для каждого пользователя вычисляется числовой показатель активности. По этому показателю пользователь должен быть отнесён к одной из двух групп: 0 или 1. Классификатор работает очень просто: он выбирает некоторое пороговое значение. Все пользователи с показателем меньше или равным порогу относятся к группе 0, а все остальные — к группе 1.

Однако исторические данные были размечены не идеально: у части пользователей указана ошибочная группа.

Требуется выбрать такой порог, чтобы количество пользователей, для которых классификатор не совпадает с указанной группой, было минимальным.

Задача 4

Классификация объектов

Вам предоставлен набор данных, содержащий числовые характеристики объектов. Каждый объект относится к одному из пяти классов.

Для части объектов класс уже известен. Используя эти данные, необходимо построить модель классификации и определить классы объектов из тестовой выборки.

В рамках задачи рекомендуется использовать метод k **ближайших соседей** — **kNN**, однако разрешается применять любые алгоритмы классификации.

Чем точнее модель классифицирует объекты скрытой тестовой выборки, тем больше баллов вы получите.

Данные

К условию задачи приложены следующие файлы:

- `train.csv` — обучающая выборка;
- `public_test.csv` — открытая тестовая выборка с известными правильными ответами;
- `private_test.csv` — скрытая тестовая выборка, для которой необходимо сформировать ответ;
- `sample_submission.csv` — пример файла с ответом.

Файл `train.csv`

Файл `train.csv` содержит 10 000 объектов.

Каждая строка имеет следующий формат:

```
id,feature_1,feature_2,...,feature_12,class
```

где:

- `id` — уникальный номер объекта;
- `feature_1`–`feature_12` — числовые признаки объекта;
- `class` — правильный класс объекта.

Класс задаётся целым числом от 0 до 4.

Файл `public_test.csv`

Файл `public_test.csv` содержит 2 500 объектов и имеет такую же структуру, как `train.csv`.

Правильные классы объектов указаны в столбце `class`.

Эту выборку можно использовать для локальной оценки качества модели, выбора параметра k , способа нормализации и других настроек алгоритма.

Объекты из `public_test.csv` не используются при итоговой проверке решения.

Файл `private_test.csv`

Файл `private_test.csv` содержит 5 000 объектов.

Каждая строка имеет следующий формат:

```
id,feature_1,feature_2,...,feature_12
```

Столбец `class` в этом файле отсутствует.

Для каждого объекта необходимо предсказать его класс.

Формат ответа

Необходимо сформировать CSV-файл со следующей структурой:

```
id,class
1,3
2,0
3,4
...
```

В первой строке должны находиться названия столбцов:

```
id,class
```

После этого для каждого объекта из файла `private_test.csv` необходимо вывести:

- `id` — номер объекта;
- `class` — предсказанный класс от 0 до 4.

Ответ должен содержать предсказание для каждого объекта из `private_test.csv`.

Порядок строк должен соответствовать порядку объектов в тестовом файле.

Оценивание

Качество решения определяется долей правильно классифицированных объектов скрытой тестовой выборки:

$$\text{Accuracy} = \frac{\text{количество правильных предсказаний}}{\text{общее количество объектов}}.$$

Количество баллов вычисляется по формуле:

$$\text{Score} = 100 \cdot \text{Accuracy}.$$

Например, если правильно классифицировано 4 250 объектов из 5 000, решение получит

$$100 \cdot \frac{4250}{5000} = 85$$

баллов.

Максимальная оценка за задачу — 100 баллов.

Баллы начисляются частично: даже решение, которое классифицирует правильно только часть объектов, получит соответствующее количество баллов.

Требования к ответу

Ответ считается некорректным, если:

- отсутствует заголовок `id,class`;
- количество строк не совпадает с количеством объектов в `private_test.csv`;
- некоторые значения `id` отсутствуют или повторяются;
- в столбце `class` находятся значения, отличные от целых чисел от 0 до 4;
- файл имеет неправильный формат.

Пример

Пусть необходимо определить классы трёх объектов.

Пример файла ответа:

```
id,class
1,2
2,0
3,4
```

Этот пример показывает только формат файла и не соответствует приложенной тестовой выборке.

Примечание

Признаки имеют разные масштабы. Поэтому перед применением метода k ближайших соседей может быть полезно выполнить нормализацию или стандартизацию данных.

Также на качество классификации могут влиять:

- значение параметра k ;
- способ вычисления расстояния;
- использование весов ближайших соседей;
- выбор признаков;
- обработка выбросов и шумовых признаков.

Разрешается использовать библиотеки NumPy, pandas, scikit-learn и другие инструменты для анализа данных и построения модели.

Задача 5

Bopl Battle

Вы — преданный фанат игры *Bopl Battle*. Приехав на смену в Алгонову, вы обнаружили, что вместе с вами на смене находится N учеников, каждый из которых тоже хочет поиграть. Известно, что за смену все ученики суммарно хотят сыграть не менее X раундов.

Однако Андрей Александрович постоянно ругается на тех, кто играет. Поэтому каждые K_i дней он забирает i -го ученика решать задачи. В такие дни этот ученик не играет ни одного раунда.

Для каждого ученика известны два числа:

- A_i — количество раундов, которые i -й ученик может сыграть за один день;
- K_i — как часто Андрей Александрович отправляет его решать задачи. Таким образом, i -й ученик не играет в дни $K_i, 2K_i, 3K_i, \dots$

Все ученики начинают играть с первого дня одновременно.

В каждый день суммарное количество сыгранных раундов равно сумме A_i по всем ученикам, которые в этот день играют. Если в некоторый день Андрей Александрович забрал всех учеников, то в этот день суммарное количество сыгранных раундов не увеличивается.

Определите минимальное количество дней, необходимое для того, чтобы ученики суммарно сыграли не менее X раундов. Если сделать это невозможно, выведите -1 .

Входные данные

В первой строке заданы два целых числа N и X — количество учеников и требуемое суммарное количество сыгранных раундов:

$$1 \leq N \leq 2 \cdot 10^5, \quad 0 \leq X \leq 10^{12}.$$

В следующих N строках содержится по два целых числа A_i и K_i :

- A_i — количество раундов, которые i -й ученик может сыграть за один день;
- K_i — период, с которым Андрей Александрович забирает i -го ученика решать задачи.

Выходные данные

Выведите одно целое число — минимальное количество дней, необходимое для того, чтобы ученики суммарно сыграли не менее X раундов, либо -1 , если это невозможно.

Подзадачи

№	Баллы	Дополнительные ограничения	Зависимости
1	8	$N = 1, K_i = 2$	—
2	12	$K_i \leq 2$	1
3	25	$N \leq 50, X \leq 10^6$	—
4	55	Без дополнительных ограничений.	3