

А. Прямая снова крутится

Рассмотрим линейный классификатор на плоскости $\mathbb{R}^2$, задаваемый прямой $x + Ay = 0$. Точка (x, y) относится к первому классу, если $x + Ay > 0$. В противном случае, если $x + Ay \leq 0$, точка относится ко второму классу.

Приведите пример таких **целых** чисел A , b и c , что для трёхчлена $P(x) = x^2 + bx + c$ точка $(2, P(2))$ относится к первому классу, а точки $(1, P(1))$ и $(3, P(3))$ — ко второму классу.

Решение.

Один из примеров: $A = -4$, $b = -4$, $c = 4$, то есть $P(x) = (x - 2)^2 = x^2 - 4x + 4$. Положим $L(x, y) = x - 4y$. Тогда

$$L(1, P(1)) = L(1, 1) = -3 \leq 0 \implies \text{точка } (1, P(1)) \text{ относится ко второму классу};$$

$$L(2, P(2)) = L(2, 0) = 2 > 0 \implies \text{точка } (2, P(2)) \text{ относится к первому классу};$$

$$L(3, P(3)) = L(3, 1) = -1 \leq 0 \implies \text{точка } (3, P(3)) \text{ относится ко второму классу}.$$

В. Генератор случайности

Генератор случайных чисел работает следующим образом. Если задать ему натуральное число $m \geq 6$, он выдаст последовательно 10 натуральных чисел. Каждое число выбирается им независимо и равновероятно из множества $\{1, 2, \dots, m\}$.

В силу системной ошибки, результат генератора выводится на экран следующим образом: числа 1, 2, 3, 4, 5 выводятся без изменений, а вместо любого другого числа выводится число 6.

При каком натуральном $m \geq 6$ вероятность получить на экране последовательность чисел

$$6, 2, 6, 6, 1, 6, 4, 6, 3, 5$$

максимальна?

Решение.

Ответ: $m = 10$.

Вероятность того, что на данном месте в последовательности будет число $k \in \{1, 2, 3, 4, 5\}$ составляет

$$\mathbb{P}(k) = \frac{1}{m}.$$

В связи с тем, что все остальные числа заменены на 6, та же вероятность для числа 6 составляет

$$\mathbb{P}(6) = \frac{m-5}{m}.$$

В изучаемой последовательности значение 6 встретилось 5 раз, и ещё 5 раз встретились числа $1, \dots, 5$. Поэтому вероятность получить эту последовательность равна

$$\left(\frac{1}{m}\right)^5 \left(\frac{m-5}{m}\right)^5 = \left(\frac{m-5}{m^2}\right)^5.$$

Максимальное значение этой вероятности достигается, когда наибольшее возможное значение принимает величина

$$\frac{m-5}{m^2} = \frac{1}{m} - \frac{5}{m^2}.$$

Положим $x = \frac{1}{m}$. Поскольку максимум квадратичной функции $f(x) = x - 5x^2$ достигается при $x = \frac{1}{10}$, искомое значение $m = 10$.

С. Градиентный спуск на листочке

Рассмотрим функцию $f(x, y) = x^{20} + y^{26}$. Ниже приведён алгоритм, который по начальной точке (a, b) строит последовательность точек (X_i, Y_i) , $i = 0, 1, 2, \dots$.

Ниже в коде `decrease_lr = False` для пункта (а), и `decrease_lr = True` для пункта (б).

```
X0, Y0 ← a, b
dx, dy ← 0, 1
lr ← 1
for i = 1, 2, 3, ... do
    dx, dy ← dy, -dx
    if decrease_lr and i · lr ≥ 2 then:
        lr ← lr / 2
    end if
    Xi, Yi ← Xi-1, Yi-1
    if f(Xi + dx · lr, Yi + dy · lr) < f(Xi, Yi) then
        Xi ← Xi + dx · lr
        Yi ← Yi + dy · lr
    end if
end for
```

Докажите, что, какая бы точка ни была начальной, справедливы следующие утверждения.

- (а) Найдётся такой номер N , что расстояние от точки (X_N, Y_N) до начала координат не превосходит 1.
- (б) Найдётся такой номер N_0 , что для любого номера $N \geq N_0$ расстояние от точки (X_N, Y_N) до начала координат не превосходит $\frac{1}{1000}$.

Решение.

Изучим очередной шаг алгоритма. Делается *попытка* изменить значение одной из координат на некоторую величину lr . Это изменение происходит в том и только в том случае, когда уменьшается модуль соответствующей координаты, поскольку

$$f(x + lr, y) < f(x, y) \iff |x + lr| < |x|,$$

$$f(x, y + lr) < f(x, y) \iff |y + lr| < |y|.$$

Пункт (а).

По координате x алгоритм делает попытки изменения $+1, -1, +1, -1, \dots$, по координате y , наоборот, $-1, +1, -1, +1, \dots$.

Рассмотрим две последовательные попытки изменения x -координаты.

- Если $|x| > \frac{1}{2}$, то ровно одна из двух попыток приводит к изменению координаты, при этом $|x|$ уменьшается ровно на 1.
- Если $|x| \leq \frac{1}{2}$, то ни одна из попыток к изменению координаты не приводит.

Таким образом, после конечного числа таких пар шагов алгоритма будет выполнено неравенство $|x| \leq \frac{1}{2}$, так как $|x|$ не может бесконечное количество раз уменьшаться на 1.

То же верно и для y -координаты, значит, для некоторого номера N будут выполнены неравенства

$$|X_N| \leq \frac{1}{2}, \quad |Y_N| \leq \frac{1}{2}.$$

Тогда и расстояние до начала координат в этого момент будет меньше, чем 1:

$$X_N^2 + Y_N^2 \leq \frac{1}{4} + \frac{1}{4} = \frac{1}{2} < 1.$$

В силу сказанного выше, это неравенство будет справедливо и для любого номера $M \geq N$, это соображение нам потребуется для решения другого пункта.

Пункт (b).

Разделим шаги алгоритма на блоки последовательных с одним и тем же значением величины lr . Рассмотрим m -й блок, состоящий из шагов с номерами $i = 2^m + 1, \dots, 2^{m+1}$, где $lr = \frac{1}{2^m}$. Далее изучаем блоки для $m \geq 3$.

Тогда аналогично пункту а) внутри каждого блока каждая вторая попытка изменения x -координаты уменьшает значение $|x|$, пока выполняется неравенство $|x| > \frac{lr}{2} = \frac{1}{2^{m+1}}$.

Значит, для m -го блока шагов справедливо хотя бы одно из следующих утверждений.

1. После выполнения всех шагов m -го блока $|x| \leq \frac{lr}{2}$.
2. В результате каждого четвёртого шага алгоритма (то есть каждой второй попытки изменения x -координаты) величина $|x|$ уменьшалась на lr . Таким образом, за этот блок шагов она суммарно уменьшилась на $2^{m-2} \cdot lr = \frac{1}{4}$.

Заметим, что случай 2 мог возникнуть лишь для конечного числа блоков, так как $|x|$ не может уменьшаться бесконечное количество раз на $\frac{1}{4}$.

Следовательно, существует такое число M_x , что для блоков с номерам, большими M_x , справедливо первое утверждение.

Итого при $N > N_x = \max(2^{M_x+1}, 10)$ будут выполнены неравенства

$$|X_N| \leq \frac{1}{2^{N_x+1}} \leq \frac{1}{2^{11}} < \frac{1}{2000}.$$

Аналогично и для координаты y найдётся такой номер N_y , что при $N > N_y$ выполнено неравенство $|Y_N| < \frac{1}{2000}$.

Тогда при $N > \max(N_x, N_y)$ имеют место неравенства $|X_N| < \frac{1}{2000}$ и $|Y_N| < \frac{1}{2000}$. Значит,

$$\sqrt{|X_N|^2 + |Y_N|^2} \leq \sqrt{|X_N|^2 + 2 \cdot |X_N||Y_N| + |Y|^2} = |X_N| + |Y_N| < \frac{1}{2000} + \frac{1}{2000} = \frac{1}{1000}.$$

Таким образом, номер $N_0 = \max(N_x, N_y)$ удовлетворяет условиям задачи.

Замечание. Из приведённого решения следует, что последовательность точек (X_i, Y_i) сходится к точке минимума функции $f(x, y)$ — началу координат.

Д. Лазер

Пол прямоугольной камеры с зеркальными стенами имеет форму клетчатого прямоугольника 1000×3000 . В **узлах сетки** расположены датчики температуры (всего $1001 \cdot 3001$ датчиков). Изначально все датчики показывают температуру 0°C .

В **центрах некоторых клеток** расположено по одному устройству, испускающему лазерные лучи. Каждое устройство испускает лазерный луч по диагонали клетки, в центре которой оно находится (то есть в одном из четырех возможных направлений, в сторону одного из узлов соответствующей клетки). При этом то же устройство поглощает лучи, приходящие из противоположного узла. На лучи трёх других направлений устройство не влияет.

Всего установлено 300 устройств, которые излучают по одному лучу. Каждый луч отражается зеркально от стенок (попадая в угол, луч отражается в противоположном направлении), пока не поглотится каким-то из устройств.

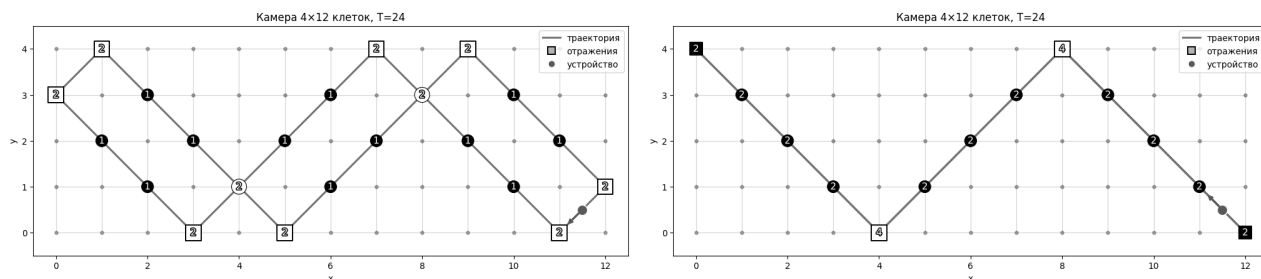
Лазерный луч, «посещая» очередной датчик температуры, увеличивает его показание на 1°C . Если же в точке, где располагается датчик, луч отражается, он увеличивает показание соответствующего датчика сразу на 2°C . Назовём *тепловой картой* матрицу A размера 1001×3001 , элементы которой равны итоговым показаниям соответствующих датчиков температуры.

Для проверки корректности показаний датчиков, полученные значения анализируют следующим образом. Выбирается квадратная матрица K размера $s \times s$, называемая *ядром* размера s , после чего вычисляется *свёртка* матрицы A с ядром K , то есть матрица $B = A \star K$ размера $(1002 - s) \times (3002 - s)$, где

$$B[i][j] = \sum_{u=0}^{s-1} \sum_{v=0}^{s-1} A[i+u][j+v] K[u][v], \quad 0 \leq i \leq 1001 - s, \quad 0 \leq j \leq 3001 - s.$$

Назовём ядро K *определяющим*, если для любой тепловой карты A все элементы матрицы B равны нулю, при этом не все элементы ядра K равны 0. При каком наименьшем значении s существует определяющее ядро размера s ?

Ниже на рисунке вы можете видеть как запущенный луч изменит температуру в датчиках до момента поглощения.



Решение.

Ответ: $s = 3$.

Рассмотрим ядро

$$K = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}.$$

Покажем, что оно подходит. Заметим, что траектории всех лучей разбиваются на циклы потому что каждый датчик принимает и испускает ровно один луч (отметим, что один цикл может быть объединением траекторий сразу нескольких лучей). Заметим, что достаточно проверить условие для одной циклической траектории (в силу линейности свёртки). Будем мыслить такую траекторию как набор диагональных отрезков между последовательными отражениями, где каждый отрезок увеличивает значения каждого датчика, который на нём расположен, на 1°C . Легко видеть, что тепловая карта траектории равна сумме тепловых карт этих отрезков.

Значит, достаточно проверить, что в результате свёртки с ядром K любой матрицы, в которой единицы расположены на одном диагональном ряду (остальные элементы — нули), получается нулевая матрица. Это следует из того, что сумма элементов матрицы K на любом диагональном ряду равна нулю.

Остаётся доказать, что меньшие значения s не подходят. Достаточно доказать это для $s = 2$. Предположим обратное, что есть определяющее ядро

$$K = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

Рассмотрим циклическую траекторию, аналогичную приведённой в условии на втором рисунке (проходящую через два противоположных угла). Легко видеть, что если расставить в центры клеток, которые она пересекает, 300 устройств, можно добиться, чтобы получилась **только** эта траектория (нужно направить все лучи из устройств «вдоль» неё).

Пусть она проходит через левый нижний угол $(0, 0)$. Тогда после свёртки с ядром K мы получим $B[1][0] = 2a$ и $B[0][1] = 2d$. Итого $a = d = 0$. Рассуждая аналогично для траектории, проходящей через два других угла, мы получим, что $b = c = 0$, противоречие.

Е. Новая выборка

Модель по вещественному числу x предсказывает вещественное число методом *линейной регрессии*, то есть по формуле $f(x) = ax + b$, где $a, b \in \mathbb{R}$ — параметры модели f .

Дана *обучающая выборка* $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, где $x_1, \dots, x_n, y_1, \dots, y_n \in \mathbb{R}$. Модель f обучается на этих данных: параметры a и b выбираются *методом наименьших квадратов*, то есть так, чтобы значение выражения

$$\sum_{i=1}^n (y_i - f(x_i))^2$$

было минимальным.

Определим *коэффициент детерминации* R_0^2 , посчитанный на этой выборке, следующим образом:

$$R_0^2 = 1 - \frac{\sum_{i=1}^n (y_i - f(x_i))^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad \text{где} \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i.$$

Построим новую выборку, добавив к исходной n объектов $(x_i, f(x_i)), i = 1, 2, \dots, n$. Пусть на получившейся выборке из $2n$ объектов обучается модель $g(x) = cx + d$ также методом наименьших квадратов. Обозначим через R_1^2 коэффициент детерминации для модели g , посчитанный на выборке из всех $2n$ объектов.

Предполагается, что параметры обеих моделей удалось подобрать методом наименьших квадратов единственным образом.

Выразите R_1^2 через R_0^2 .

Решение.

Ответ: $R_1^2 = \frac{2R_0^2}{1 + R_0^2}$.

Положим $\hat{y}_i = f(x_i)$, (предсказание); $SSE_0 = \sum_{i=1}^n (y_i - \hat{y}_i)^2$ (сумма квадратов ошибок), $SST_0 = \sum_{i=1}^n (y_i - \bar{y})^2$ (общая сумма квадратов). Тогда $R_0^2 = 1 - \frac{SSE_0}{SST_0}$.

Аналогично определим величины SSE_1 и SST_1 для второй модели.

Лемма 1. Модель, обученная на объединённой выборке, совпадает с исходной, и потому $SSE_1 = SSE_0$.

Доказательство. Для модели f на добавленных точках $(x_i, \hat{y}_i)$ ошибки $(\hat{y}_i - f(x_i))^2$ равны нулю, а для исходных точек сумма квадратов ошибок — наименьшая возможная. Значит, минимум на объединённой выборке достигается для тех же значений параметров (тем самым, модели совпадают). $\square$

Лемма 2. Имеют место равенства $\sum_{i=1}^n (y_i - \hat{y}_i) = 0$ и $\sum_{i=1}^n x_i (y_i - \hat{y}_i) = 0$.

Доказательство. Точка (a, b) выбирается методом наименьших квадратов, то есть совпадает с наименьшим значением функции

$$F(u, v) = \sum_{i=1}^n (y_i - ux_i - v)^2.$$

Значит, если зафиксировать $u = a$ у получившейся квадратичной функции от переменной v наименьшее значение достигается в точке b . Это означает, что v -координата вершины соответствующей параболы равна b , то есть

$$-b = \frac{1}{n} \left(\sum_{i=1}^n (ax_i - y_i) \right) \implies 0 = \sum_{i=1}^n (ax_i + b) - \sum_{i=1}^n y_i = \sum_{i=1}^n (\hat{y}_i - y_i).$$

Рассуждая аналогично для $v = b$, получаем второе заявленное соотношение:

$$-a \cdot \sum_{i=1}^n x_i^2 = \sum_{i=1}^n (b - y_i)x_i \implies 0 = \sum_{i=1}^n x_i(ax_i + b - y_i) = \sum_{i=1}^n x_i(\hat{y}_i - y_i).$$

□

Лемма 3. *Имеет место равенство $SST_1 = 2SST_0 - SSE_0$.*

Доказательство. Согласно лемме 2, справедливо равенство $\frac{1}{n} \sum_{i=1}^n \hat{y}_i = \bar{y}$. Значит,

$$SST_1 = \sum_{i=1}^n (y_i - \bar{y})^2 + \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 = SST_0 + \sum_{i=1}^n (\hat{y}_i - \bar{y})^2.$$

Кроме того, $y_i - \bar{y} = (y_i - \hat{y}_i) + (\hat{y}_i - \bar{y})$. Возведём все такие равенства в квадрат, после чего просуммируем. Мы получим, что

$$SST_0 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + 2 \sum_{i=1}^n (y_i - \hat{y}_i)(\hat{y}_i - \bar{y}).$$

Последний член равен нулю. Действительно, воспользуемся леммой 2 следующим образом:

$$\sum_{i=1}^n (y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) = \sum_{i=1}^n (y_i - \hat{y}_i)\hat{y}_i = a \sum_{i=1}^n x_i(y_i - \hat{y}_i) + b \sum_{i=1}^n (y_i - \hat{y}_i) = 0.$$

Тогда $SST_0 = SSE_0 + \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$, что и даёт требуемое равенство. □

Собирая вместе результаты лемм 1–3, имеем:

$$R_1^2 = 1 - \frac{SSE_1}{SST_1} = 1 - \frac{SSE_0}{2SST_0 - SSE_0} = \frac{2(SST_0 - SSE_0)}{2SST_0 - SSE_0}.$$

Разделив числитель и знаменатель на SST_0 , мы получаем заявленный ответ:

$$R_1^2 = \frac{2(1 - \frac{SSE_0}{SST_0})}{2 - \frac{SSE_0}{SST_0}} = \frac{2R_0^2}{1 + R_0^2}.$$

Комментарий. Лемму 2 можно доказать по-другому, воспользовавшись тем, что частные производных функции $F(u, v)$ в точке (a, b) равны нулю.

Г. Вымышленная ситуация

Артур и Таня готовят вычислительные мощности к практическому туру для 1000 участников финала всероссийской олимпиады школьников по ИИ. Изначально они хотели вернуть всю инфраструктуру в облаке и выдавать видеокарты (GPU) из общего набора, но, чтобы исключить сетевые задержки, было решено собрать каждому участнику персональный кластер из 8 устройств.

На складе партнёров олимпиады есть по 8000 устройств каждого из трех типов.

- Тип 1 (зеленые) — CUDA, быстрые, но горячие.
- Тип 2 (красные) — поддерживают открытые стандарты, подходят для обучения больших языковых моделей (LLM), но требуют сложной настройки драйверов.
- Тип 3 (желтые) — экспериментальные ускорители (TPU).

По условиям поставки устройства типа 1 можно получать только партиями по 7 штук; устройств каждого типа должно быть не меньше, чем 2026; всего со склада следует взять ровно 8000 устройств. Если в соответствии с этими условиями можно привезти со склада y_1 устройств первого типа, y_2 — второго и y_3 — третьего, будем называть тройку целых чисел (y_1, y_2, y_3) *допустимой*, то есть:

$$y_1 + y_2 + y_3 = 8000, \quad y_t \geq 2026 \ (t = 1, 2, 3), \quad y_1 \div 7.$$

Каждый участник имеет персональный шифр $i = 1, 2, \dots, 1000$. Устройства, поставленные со склада, распределяются между участниками. Пусть участник i получает персональный кластер $S_i = (x_{i,1}, x_{i,2}, x_{i,3})$, в котором $x_{i,1}, x_{i,2}, x_{i,3}$ — количество устройств первого, второго и третьего типа соответственно. Числа $x_{i,1}, x_{i,2}, x_{i,3}$ — целые неотрицательные, причём $x_{i,1} + x_{i,2} + x_{i,3} = 8$.

Оказалось, что кто-то из участников предпочитает «зеленых», кто-то фанат открытых стандартов и выбирает «красных», а кто-то любит необычных «жёлтых». Для каждого участника заданы *предпочтения* устройств $u_{i,1}, u_{i,2}, u_{i,3}$, указанные во входном файле `gpu.csv`.

Файл содержит заголовок и 1000 строк с 3 столбцами с данными: в i -й строке записаны три целых числа $u_{i,1}, u_{i,2}, u_{i,3}$ (именно в таком порядке), соответствующие предпочтениям участника i .

Полезность кластера $S = (x_1, x_2, x_3)$ для участника i задаётся формулой

$$V_i(S) = u_{i,1}x_1 + u_{i,2}x_2 + u_{i,3}x_3.$$

Распределение устройств называется *честным*, если неравенство $V_i(S_i) \geq V_i(S_j)$ справедливо для любых двух участников i, j .

- Приведите пример допустимой тройки значений (y_1, y_2, y_3) , при которой честного распределения не существует, и докажите это.
- Существует ли допустимая тройка (y_1, y_2, y_3) , при которой существует честное распределение?

Решение.

Будем называть величины $u_{i,j}$ *полезностями* устройств типа j для участника i .

Лемма 1. Если умножить значения полезностей произвольного участника i на одно и то же положительное число, любое честное распределение останется честным, а любое нечестное — нечестным.

Доказательство. Обе части неравенств, в которых фигурирует полезность V_i , мы умножим на одну и ту же положительную константу, тем самым выполнение системы новых и изначальных неравенств равносильно. $\square$

Наблюдение. По данным файла `gru.csv`, школьники делятся на две группы.

- группа A : 511 школьников, набор полезностей которых пропорционален

$$u^A = (2053, 2063, 2070);$$

- группа B : 489 школьников, набор полезностей которых пропорционален

$$u^B = (2081, 2027, 2003).$$

В силу леммы 1, можно считать, что всем участникам группы A соответствует набор полезностей u^A , а всем участникам группы B — набор полезностей u^B .

Лемма 2. В любом честном распределении все школьники группы B получают один и тот же набор устройств.

Доказательство. Возьмём двух участников группы B . Пусть первый получил b_i устройств типа i , второй — b'_i устройств типа i . Поскольку предпочтения у этих участников совпадают, полезности кластеров (b_1, b_2, b_3) и (b'_1, b'_2, b'_3) для них также совпадают и равны между собой. То есть

$$2081b_1 + 2027b_2 + 2003b_3 = 2087b'_1 + 2027b'_2 + 2003b'_3.$$

Поскольку в обоих кластерах по 8 устройств, вычитая из обеих частей $2003 \cdot 8$, получаем

$$(2081 - 2003)b_1 + (2027 - 2003)b_2 = (2081 - 2003)b'_1 + (2027 - 2003)b'_2 \Rightarrow \\ 78(b_1 - b'_1) + 24(b_2 - b'_2) = 0 \Rightarrow 13(b_1 - b'_1) + 4(b_2 - b'_2) = 0$$

Значит, число $b_2 - b'_2$ кратно 13. Однако, поскольку $0 \leq b_2, b'_2 \leq 8$, это возможно лишь в случае $b_2 = b'_2$, а тогда и $b_1 = b'_1$, $b_3 = b'_3$, что завершает доказательство леммы. $\square$

Лемма 3. В любом честном распределении все школьники группы A получают один и тот же набор устройств.

Доказательство. Рассуждение, в целом, аналогично предыдущей лемме. Вновь, предполагая противная, имеем два кластера (a_1, a_2, a_3) и (a'_1, a'_2, a'_3) одинаковой полезности для участника группы A . Теперь, поскольку $2053 - 2070 = -17$ и $2063 - 2070 = -7$, имеем: $-17(a_1 - a'_1) - 7(a_2 - a'_2) = 0$. Следовательно, $a_2 - a'_2$ делится на 17, поэтому $a_2 = a'_2$, а тогда и $a_1 = a'_1$, $a_3 = a'_3$, лемма доказана. $\square$

Итого все 511 школьников группы A получают один и тот же кластер $a = (a_1, a_2, a_3)$, а все 489 школьников группы B — один и тот же кластер $b = (b_1, b_2, b_3)$.

Посчитаем количество устройств каждого типа:

$$y_t = 511a_t + 489b_t \quad (t = 1, 2, 3).$$

Пункт (а).

Рассмотрим набор $(y_1, y_2, y_3) = (2100, 2100, 3800)$. Легко видеть, что он допустим: 2100 делится на 7, все y_i не меньше, чем 2026 и $2100 + 2100 + 3800 = 8000$. Докажем, что для этого набора честного распределения не существует.

Предположим противное. В силу сказанного выше, для некоторых кластеров $a = (a_1, a_2, a_3)$ и $b = (b_1, b_2, b_3)$ имеет место равенство $2100 = y_1 = 511a_1 + 489b_1$. Но числа вида $2100 - 511n$ при $n = 0, 1, 2, 3$ не делятся на 489, а при $n \geq 4$ — отрицательны, противоречие.

Пункт (b).

Ответ: такой набор существует.

Рассмотрим следующее распределение: каждому школьнику группы A соберём кластер $(0, 4, 4)$, а каждому школьнику группы B — кластер $(7, 1, 0)$.

Вычислим количество устройств каждого типа:

$$y_1 = 511 \cdot 0 + 489 \cdot 7 = 3423, \quad y_2 = 511 \cdot 4 + 489 \cdot 1 = 2044 + 489 = 2533, \quad y_3 = 511 \cdot 4 + 489 \cdot 0 = 2044.$$

Отметим, что y_1 делится на 7, и количества устройств каждого типа не меньше, чем 2026. Тем самым, такое распределение соответствует допустимой тройке

$$(y_1, y_2, y_3) = (3423, 2533, 2044).$$

Проверим, что получилось честное распределение. Для любого школьника i из группы A :

$$V_i(a) = 2053 \cdot 0 + 2063 \cdot 4 + 2070 \cdot 4 = 8252 + 8280 = 16532,$$

$$V_i(b) = 2053 \cdot 7 + 2063 \cdot 1 + 2070 \cdot 0 = 14371 + 2063 = 16434.$$

Значит, $V_i(a) > V_i(b)$. Прделаем то же вычисление для школьника участника j из группы B :

$$V_j(a) = 2081 \cdot 0 + 2027 \cdot 4 + 2003 \cdot 4 = 8108 + 8012 = 16120,$$

$$V_j(b) = 2081 \cdot 7 + 2027 \cdot 1 + 2003 \cdot 0 = 14567 + 2027 = 16594.$$

Следовательно, $V_j(b) > V_j(a)$. Итого каждый участник группы B предпочитает свой кластер кластеру группы A и наоборот. Поскольку внутри каждой из групп участники получают одинаковые кластеры, распределение честное. Таким образом, подходит набор

$$(y_1, y_2, y_3) = (3423, 2533, 2044).$$

Замечание. Можно доказать, что указанный набор — единственный, подходящий под условия пункта (b). Тем самым, любой другой допустимый набор подойдёт в качестве примера для пункта (a).

А. Прямая снова крутится

А. Верный ответ без проверки — 45 баллов.

В. Генератор случайности

А. Найдено $\mathbb{P}(6) = (m - 5)/m - 5$ баллов

В. Найдена вероятность последовательности из условия при фиксированном m — 15 баллов.

С. Заявлено, что указанная в предыдущем критерии вероятность возрастает при целых m от 6 до 10 и убывает далее — 15 баллов

Д. Вероятность посчитана с ошибкой в константу раз (и другие не влияющие на ход решения неточности) — снимается не менее 5 баллов.

М. Доказано, что число 10 является экстремумом. Используется, но не доказано, что это именно максимум — снимается не менее 5 баллов.

- Все приведенные продвижения и штрафы суммируются.

С. Градиентный спуск на листочке

Общая часть.

С1 (5 баллов) Сформулировано чередование направлений спуска с периодом 4.

С2 (5 баллов) Установлено убывание $|X|$ и $|Y|$.

Пункт (а).

А. Верное доказательство пункта (а) — 20 баллов.

МА. Неверно указана область остановки или используется без доказательства существование такой области — снимается 10 баллов.

Пункт (b).

В. Верное доказательство пункта (b) — 20 баллов.

В1. Введена структура блоков, с указанием связи их размера и общим значением lr — 5 баллов.

В2. Доказано, что вклад каждого блока в изменение параметров $|X|$ и $|Y|$ составляет не меньше некоторой константы, если все шаги этого блока приводили к изменению соответствующих координат.

МВ. Ошибки в изложении оставшейся части доказательства — снимается не менее 5 баллов.

- Продвижения и штрафы по трем приведенным частям задачи суммируются.

Д. Лазер

Часть А (пример). Часть В (оценка).

А1. Приведена правильная матрица K — 10 баллов.

А2. Показано, что матрица обнуляет диагональные ряд — 5 баллов.

А3. Завершение обоснования, почему матрица K является определяющим ядром — 10 баллов.

В1. Получено линейное уравнение на элементы ядра — 5 баллов.

М. Ошибки в построении «подходящего луча» (он рассматривается только локально, ставится меньшее число устройств, чем требует условие и т.д.) — снимается не менее 5 баллов.

- Все приведенные продвижения и штрафы суммируются.
-

Е. Новая выборка

Общеизвестные свойства линейной регрессии принимаются без доказательства **при наличии верной формулировки**.

А. Верный ответ — 10 баллов.

В. Доказано, что $f = g$ — 10 баллов.

- Если только сформулировано, что $f = g$, выставляется 5 баллов.
- Если доказательство опирается на свойства линейной регрессии, должно быть явно сформулировано свойство добавляемых данных, которое позволяет сделать вывод о равенстве коэффициентов моделей.
- Если это используется и даже не формулируется, баллы по критерию B не начисляются.
- За вывод о равенстве числителей в выражениях коэффициента детерминации дополнительные баллы не начисляются.

С1. Установлены зависимости между $y_i, x_i, f(x_i)$ (лемма 2) — 10 баллов.

С2. Получены выражения для a и b (формула коэффициентов линейной регрессии) с доказательством равенства средних — 10 баллов.

- Должны быть установлены **обе** зависимости (в том виде, как это указано С1 или С2) и явно сформулировано равенство средних значений старой и новой выборки. За получения любого одного из соотношений начисляется 5 баллов за часть C .

М. Лемма 3 (или аналогичное равенство) используется в решении без доказательства — снимается 15 баллов.

- За алгебраические преобразования, не приводящие к зависимости R_1^2 **только** от R_0^2 баллы не начисляются.

Ф. Вымышленная ситуация

А. Верный ответ в пункте (а) — 0 баллов.

В. Лемма 1 (о домножении полезностей на константу) вместе с наблюдением о разделении участников на две группы — 10 баллов.

В1. Указано разделение агентов делится на 2 группы по предпочтениям — 5 баллов.

В2. Лемма 1 только сформулирована — 5 баллов.

С. Формулировка и доказательство леммы 2 и 3 (о равенстве кластеров внутри каждой из групп участников) — 15 баллов.

С1. Формулировка леммы 2 и 3 — 5 баллов.

Д. Вывод пункта (а) из лемм 2 и 3 — 5 баллов.

Е. Ответ в пункте (б) — 10 баллов.

Ф. Проверка ответа в пункте (б) — 10 баллов.
