

Machine Learning 2022-ITMO

А. Перекрёстная проверка

1 секунда, 256 мегабайт

Разбейте множество из  $N$  объектов, каждый из которых принадлежит к одному из  $M$  классов, на  $K$  частей. Каждый объект должен попасть ровно в одну часть так, чтобы размеры частей, а также распределение классов по этим частям было сбалансировано. Формально, пусть  $cnt(x, c)$  — число объектов с классом  $c$  попавших в часть  $x$ , тогда должно выполняться  $\forall x, y, c : |cnt(x, c) - cnt(y, c)| \leq 1$  и  $\forall x, y : |\sum_c cnt(x, c) - \sum_c cnt(y, c)| \leq 1$ .

Входные данные

Первая строка: три целых числа  $N, M, K$  ( $1 \leq N \leq 10^5, 1 \leq M, K \leq N$ ) — число объектов, классов и частей.

Вторая строка:  $N$  целых чисел  $C_i$  ( $1 \leq C_i \leq M$ ) — класс  $i$ -го объекта.

Выходные данные

Выведите  $K$  строк. Каждая строка  $x$  начинается с целого числа  $S$  — размера части  $x$ . Далее идут  $S$  целых чисел — номера объектов попавших в часть  $x$ . Объекты нумеруются с единицы.

входные данные
10 4 3 1 2 3 4 1 2 3 1 2 1
выходные данные
4 1 4 9 10 3 2 3 5 3 6 7 8

В первой части содержится четыре объекта, два из них первого класса, один второго и один четвёртого. Во второй и третьей части по три объекта первых трёх классов.

В. F-мера

1 секунда, 256 мегабайт

В результате эксперимента по классификации на  $K$  классов была получена матрица неточностей (Confusion matrix)  $CM$ , где  $CM[c, t]$  — число объектов класса  $c$ , которые были классифицированы как  $t$ . Посчитайте по данной матрице неточностей средневзвешенную по классам микро, макро и обычную F-меру.

Входные данные

Первая строка содержит целое число  $K$  — число классов ( $1 \leq K \leq 20$ ). Далее идёт  $K$  строк — описание матрицы неточностей. Каждая строка  $c$  содержит  $K$  целых чисел —  $c$ -я строка матрицы неточностей.  $\forall c, t : 0 \leq CM[c, t] \leq 100$  и  $\exists c, t : CM[c, t] \geq 1$ .

Выходные данные

Выведите три вещественных числа с плавающей точкой — взвешенно усреднённую по классам микро, макро и обычную F-меру. Абсолютная погрешность ответа не должна превышать  $10^{-6}$ .

входные данные
2 0 1 1 3
выходные данные
0.705882353 0.600000000 0.600000000

входные данные
3 3 1 1 3 1 1 1 3 1
выходные данные
0.333333333 0.326860841 0.316666667

В первом примере классы распределены как 1:4. Точность (precision), полнота (recall) и F-мера первого класса равны 0, а второго 0.75. При этом средняя точность, полнота и F-мера равны 0.6.

С. Непараметрическая регрессия

2 секунды, 256 мегабайт

Реализуйте алгоритм непараметрической регрессии, который бы поддерживал различные функции расстояний, ядер и окон. Описание ядер можно найти здесь: <https://en.wikipedia.org/w/index.php?oldid=911077090>. Обратите внимание, что определение Прямоугольного ядра в данной задаче отличается.

Входные данные

Первая строка содержит два целых числа  $N$  и  $M$  — число объектов и признаков ( $1 \leq N \leq 100, 1 \leq M \leq 10$ ).

Далее идёт  $N$  строк — описание набора данных. Каждая строка  $i$  содержит  $M + 1$  целое число  $d_{i,j}$  ( $-100 \leq d_{i,j} \leq 100$ ) — описание  $i$ -го объекта. Первые  $M$  из этих чисел признаки  $i$ -го объекта, а последнее — его целевое значение.

Следующая строка описывает объект запроса  $q$ . Она состоит из  $M$  целых чисел  $d_{q,j}$  ( $-100 \leq d_{q,j} \leq 100$ ) — признаки объекта  $q$ .

Далее идут три строки состоящих из строчных латинских букв.

Первая из них — название используемой функции расстояния: *manhattan, euclidean, chebyshev*.

Вторая — название функции ядра: *uniform, triangular, epanechnikov, quartic, triweight, tricube, gaussian, cosine, logistic, sigmoid*.

Третья — название типа используемого окна: *fixed* — окно фиксированной ширины, *variable* — окно переменной ширины.

Последняя строка содержит параметр окна: целое число  $h$  ( $1 \leq h \leq 100$ ) — радиус окна фиксированной ширины, либо целое число  $K$  ( $1 \leq K < N$ ) — число соседей учитываемое для окна переменной ширины.

Выходные данные

Выведите одно вещественное число с плавающей точкой — результат запроса. Допустимая абсолютная и относительная погрешность  $10^{-6}$ .

входные данные
3 2 0 2 1 1 1 0 2 0 1 0 0 euclidean uniform fixed 2
выходные данные
0.000000000

входные данные
3 2 0 2 1 1 1 0 2 0 1 0 0 euclidean gaussian variable 2
выходные данные
0.6090086848

В случае неопределённости, когда в окно не попало ни одного объекта, требуется вывести значение по умолчанию для задачи регрессии — среднее значение целевой переменной по всем объектам из обучающей выборки.

D. Линейная регрессия

2.0 с🕒, 256 мегабайт

Найдите градиент функции ошибки  $SMAPE(Y_i, \hat{Y}_i) = \frac{|Y_i - \hat{Y}_i|}{|Y_i| + |\hat{Y}_i|}$  для каждого объекта из набора данных. Гарантируется, что  $|Y_i| + |\hat{Y}_i| > 0$ .

Входные данные

Первая строка содержит два целых числа  $N$  ( $1 \leq N \leq 10^4$ ) — число объектов в наборе данных, и  $M$  ( $1 \leq M \leq \min(N, 1000)$ ) — число признаков у объектов исключая зависимую переменную.

Следующие  $N$  строк содержат описание объектов.  $i$ -я из этих строк содержит описание  $i$ -го объекта,  $M + 1$  целых чисел. Первые  $M$  из этих чисел:  $X_{i,j}$  ( $|X_{i,j}| \leq 10^9$ ) — признаки  $i$ -го объекта, а последнее  $Y_i$  ( $|Y_i| \leq 10^9$ ) — значение его зависимой переменной.

Последняя строка содержит  $M + 1$  целых числа  $A_j$  ( $|A_j| \leq 10^9$ ) — коэффициенты прямой из уравнения  $Y = A_0 \cdot X_0 + A_1 \cdot X_1 + \dots + A_{M-1} \cdot X_{M-1} + A_M$

Выходные данные

Выведите  $N$  строк из  $M + 1$  вещественных чисел.  $i$ -я из этих строк должна содержать градиент функции ошибки для  $i$ -го объекта.

входные данные
2 1 2015 2045 2016 2076 31 -60420
выходные данные
0.0 0.0 0.0 0.0

входные данные
4 1 1 0 1 2 2 2 2 4 1 1
выходные данные
0.0 0.0 0.0 0.0 0.32 0.16 -0.32653061224489793 -0.16326530612244897

E. Наивный байесовский классификатор

1 секунда🕒, 256 мегабайт

Реализуйте наивный байесовский классификатор.

Априорные вероятности классов оцениваются обыкновенным частотным методом.

Для оценки вероятности встречи слов в каждом классе используется модель Бернулли с аддитивным сглаживанием (сглаживание Лапласа)  $p(x) = \frac{count(x) + \alpha}{\sum_{y \in Q} count(y) + \alpha \cdot |Q|}$ , где  $x$  — рассматриваемое событие, а  $Q$  — множество всех событий.

Каждое слово это отдельный признак с двумя возможными событиями встретилось / не встретилось.

Входные данные

В первой строке содержится целое положительное число  $K$  ( $1 \leq K \leq 10$ ) — число классов.

Во второй строке содержится  $K$  целых положительных чисел  $\lambda_C$  ( $1 \leq \lambda_C \leq 10$ ) — штрафы за ошибки классификации сообщений соответствующих классов.

В третьей строке содержится целое положительное число  $\alpha$  ( $1 \leq \alpha \leq 10$ ) — интенсивность аддитивного сглаживания.

Следующая строка содержит целое положительное число  $N$  ( $1 \leq N \leq 200$ ) — число сообщений в обучающей выборке.

Следующие  $N$  строк содержат описания соответствующих сообщений из обучающей выборки. Каждое сообщение в ней начинается с целого положительного числа  $C_i$  ( $1 \leq C_i \leq K$ ) — класса к которому относится  $i$ -е сообщение. Далее следует целое положительное число  $L_i$  ( $1 \leq L_i \leq 10^4$ ) — число слов в  $i$ -м сообщении. Затем следует содержание сообщения —  $L_i$  слов состоящих из маленьких латинских букв.

Далее в отдельной строке содержится целое положительное число  $M$  ( $1 \leq M \leq 200$ ) — число сообщений в проверочной выборке.

Следующие  $M$  строк содержат описания соответствующих сообщений из проверочной выборки. Каждое сообщение в ней начинается с целого положительного числа  $L_j$  ( $1 \leq L_j \leq 10^4$ ) — число слов в  $j$ -м сообщении. Затем следует содержание сообщения —  $L_j$  слов состоящих из маленьких латинских букв.

Гарантируется, что сумма длин всех сообщений в обучающей и проверочной выборках меньше чем  $2 \cdot 10^6$ .

Выходные данные

Выведите  $M$  строк — результаты мягкой классификации оптимального наивного байесовского классификатора соответствующих сообщений из проверочной выборки. Допустимая абсолютная и относительная погрешность  $10^{-4}$ .

Каждый  $j$ -й результат мягкой классификации должен содержать  $K$  чисел  $p_C$  — вероятности того, что  $j$ -е сообщение относится к классу  $C$ .

входные данные
3 1 1 1 1 4 1 2 ant emu 2 3 dog fish dog 3 3 bird emu ant 1 3 ant dog bird 5 2 emu emu 5 emu dog fish dog fish 5 fish emu ant cat cat 2 emu cat 1 cat
выходные данные
0.4869739479 0.1710086840 0.3420173681 0.1741935484 0.7340501792 0.0917562724 0.4869739479 0.1710086840 0.3420173681 0.4869739479 0.1710086840 0.3420173681 0.4869739479 0.3420173681 0.1710086840

В примере условные вероятности выглядят следующим образом:

$p(w_x c_y)$	ant	bird	dog	emu	fish
$c_1$	3/4	1/2	1/2	1/2	1/4
$c_2$	1/3	1/3	2/3	1/3	2/3
$c_3$	2/3	2/3	1/3	2/3	1/3

Слово cat не рассматривается, так как оно ни разу не встретилось в обучающей выборке.

Для первого запроса  $X$ :

$$p(c_1) \cdot p(X|c_1) = \frac{2}{4} \cdot \left(1 - \frac{3}{4}\right) \cdot \left(1 - \frac{1}{2}\right) \cdot \left(1 - \frac{1}{2}\right) \cdot \left(\frac{1}{2}\right) \cdot \left(1 - \frac{1}{4}\right) = \frac{3/256}{3/256 + 1/243 + 2/243}$$

F. Дерево принятия решений

1.5 секунд, 256 мегабайт

Постройте дерево принятия решений.

Входные данные

Первая строка содержит три целых положительных числа  $M$  ( $1 \leq M \leq 100$ ) — число признаков у объектов (исключая класс),  $K$  ( $1 \leq K \leq 20$ ) — число классов и  $H$  ( $1 \leq H \leq 10$ ) — максимальная глубина (в рёбрах) дерева принятия решений.

Вторая строка содержит целое положительное число  $N$  ( $1 \leq N \leq 4000$ ) — число объектов в обучающей выборке.

Следующие  $N$  строк содержат описания объектов в обучающей выборке. В  $i$ -й из этих  $N$  строк перечислено  $M + 1$  целое число: первые  $M$  чисел  $A_{i,j}$  ( $|A_{i,j}| \leq 10^9$ ) — признаки  $i$ -го объекта, последнее число  $C_i$  ( $1 \leq C_i \leq K$ ) — его класс.

Выходные данные

Выведите построенное дерево принятия решений.

В первой строке выведите целое положительное число  $S$  ( $1 \leq S \leq 2^{11}$ ) — число вершин в дереве.

В следующих  $S$  строках выведите описание вершин дерева. В  $v$ -й из этих строк выведите описание  $v$ -й вершины:

- Если  $v$ -я вершина узел, выведите через пробел: заглавную латинскую букву 'Q', целое положительное число  $f_v$  ( $1 \leq f_v \leq M$ ) — индекс признака по которому происходит проверка в данном узле, вещественное число с плавающей точкой  $b_v$  — константа с которой происходит сравнения для проверки, два целых положительных числа  $l_v$  и  $r_v$  ( $v < l_v, r_v \leq S$ ) — индекс вершины дерева в которую следует перейти, если выполняется условие  $A[f_v] < b_v$ , и индекс вершины дерева в которую следует перейти, если условие не выполняется.
- Если  $v$ -я вершина лист, выведите через пробел: заглавную латинскую букву 'C' и целое положительное число  $D_v$  ( $1 \leq D_v \leq K$ ) — класс объекта попавшего в данный лист.

Вершины нумеруются с единицы. Корнем дерева считается первая вершина.

Система оценки

Решение будет проверено на секретном наборе данных. На основании предсказанных и реальных классов вычисляется усреднённая по классам микро  $F_1$ -мера.

Пусть  $Score = 100 \cdot \frac{F-B}{J-B}$ , где  $F$  —  $F_1$ -мера вашего решения,  $J$  —  $F_1$ -мера решения эталона с запасом  $\approx 1\%$ ,  $B$  —  $F_1$ -мера наивного решения с запасом  $\approx 2\%$ .

Тогда  $Verdict = \begin{cases} Ok & Score \geq 100 \\ PartiallyCorrect & 0 \leq Score \leq 100 \\ WrongAnswer & Score < 0 \end{cases}$

входные данные
2 4 2 8 1 2 1 2 1 1 3 1 2 4 2 2 3 4 3 4 3 3 1 3 4 2 4 4
выходные данные
7 Q 1 2.5 2 5 Q 2 2.5 3 4 C 1 C 4 Q 2 2.5 6 7 C 2 C 3

G. Логическое выражение

1 секунда, 256 мегабайт

Постройте искусственную нейронную сеть, вычисляющую логическую функцию  $f$ , заданную таблицей истинности.

Входные данные

Первая строка содержит целое число  $M$  ( $1 \leq M \leq 10$ ) — число аргументов  $f$ . Следующие  $2^M$  строк содержат значения  $f$  в таблице истинности (0 — ложь, 1 — истина). Строки в таблице истинности последовательно отсортированы по аргументам функции от первого к последнему. Например:

$M = 1$	$M = 2$	$M = 3$
$f(0)$	$f(0, 0)$	$f(0, 0, 0)$
$f(1)$	$f(1, 0)$	$f(1, 0, 0)$
	$f(0, 1)$	$f(0, 1, 0)$
	$f(1, 1)$	$f(1, 1, 0)$
		$f(0, 0, 1)$
		$f(1, 0, 1)$
		$f(0, 1, 1)$
		$f(1, 1, 1)$

Выходные данные

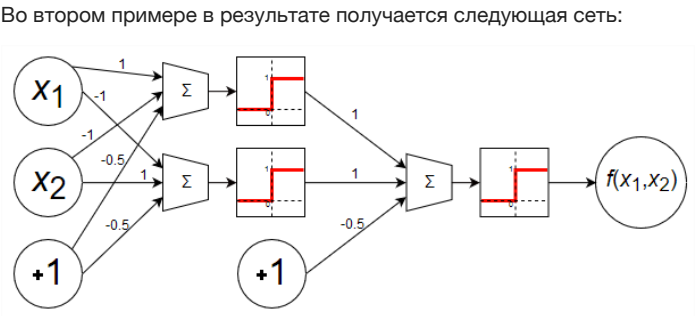
В первой строке выведите целое положительное число  $D$  ( $1 \leq D \leq 2$ ) — число слоёв (преобразований) в вашей сети.

На следующей строке выведите  $D$  целых положительных чисел  $n_i$  ( $1 \leq n_i \leq 512$  и  $n_D = 1$ ) — число искусственных нейронов на  $i$ -м слое. Предполагается, что  $n_0 = M$ .

Далее выведите описание  $D$  слоёв.  $i$ -й слой описывается  $n_i$  строками, описанием соответствующих искусственных нейронов на  $i$ -м слое. Каждый искусственный нейрон описывается строкой состоящей из  $n_{i-1}$  вещественных чисел с плавающей точкой  $w_j$  и одного вещественного числа  $b$  — описание линейной зависимости текущего нейрона от выходов предыдущего  $i$ -го слоя. Линейная зависимость задается по формуле:  $Y = \sum w_j \cdot x_j + b$ . Предполагается, что после каждого вычисления линейной зависимости к её результату применяется функция ступенчатой активации  $a(Y) = \begin{cases} 1 & Y > 0 \\ 0 & Y < 0 \end{cases}$ . Обратите внимание, что в нуле данная функция не определена, и если в ходе вычисления вашей сети будет вызвана активация от нуля, вы получите ошибку.

входные данные
2 0 1 0 1

выходные данные
2 2 1 1.0 -1.0 -0.5 1.0 1.0 -1.5 1 1 -0.5
входные данные
2 0 1 1 0
выходные данные
2 2 1 1.0 -1.0 -0.5 -1.0 1.0 -0.5 1 1 -0.5



Н. Матричная функция

1 секунда🕒, 256 мегабайт

Вычислите матричную функцию и её производную по заданному графу вычислений.

**Входные данные**  
В первой строке содержится три целых положительных числа  $N$ ,  $M$ ,  $K$  ( $1 \leq M, K \leq N \leq 50$ ) — число вершин в графе вычислений, число входных параметров (вершин) и число выходных параметров (вершин). Далее следует  $N$  строк — описание вершин графа вычислений.  $i$ -я из этих строк содержит описание  $i$ -й вершины:

- var**  $r$   $c$  ( $1 \leq r, c \leq 25$ ) — входной параметр функции, матрица состоящая из  $r$  строк и  $c$  столбцов.
- tnh**  $x$  ( $1 \leq x < i$ ) — матрица из значений гиперболического тангенса вычисленного от соответствующих компонент матрицы полученной из  $x$ -й вершины графа вычислений.
- rlu**  $\alpha^{-1} x$  ( $1 \leq \alpha^{-1} \leq 100, 1 \leq x < i$ ) — матрица из значений функции параметрического линейного выпрямителя с параметром  $\alpha$  вычисленной от соответствующих компонент матрицы полученной из  $x$ -й вершины графа вычислений.  $\alpha^{-1}$  — целое число. Производная в нуле равна единице.
- mul**  $a$   $b$  ( $1 \leq a, b < i$ ) — произведение матриц полученных из  $a$ -й  $b$ -й вершины графа вычислений соответственно.
- sum**  $len$   $u_1$   $u_2$  ...  $u_{len}$  ( $1 \leq len \leq 10, \forall 1 \leq j \leq len : 1 \leq u_j < i$ ) — сумма матриц полученных из вершин  $u_1, u_2, \dots, u_{len}$  графа вычислений.
- had**  $len$   $u_1$   $u_2$  ...  $u_{len}$  ( $1 \leq len \leq 10, \forall 1 \leq j \leq len : 1 \leq u_j < i$ ) — произведение Адамара (покомпонентное) матриц полученных из вершин  $u_1, u_2, \dots, u_{len}$  графа вычислений.

Гарантируется, что первые  $M$  вершин и только они имеют тип **var**. Последние  $K$  вершин считаются выходными. Гарантируется, что размеры матриц аргументов для каждой вершины согласованны.

Далее следует описание  $M$  матриц — входных параметров соответствующих вершин графа вычислений в порядке возрастания их индексов.

Затем следует описание  $K$  матриц — производных функции по соответствующим выходным вершинам в порядке возрастания их индексов. Обратите внимание, что производные вычислены только из некоторых скрытых вершин. Если какая-то выходная вершина зависит от другой выходной вершины, то соответствующую производную нужно досчитать.

Каждая строка, каждой матрицы расположена на отдельной строке. Матрицы состоят из целых чисел по модулю не превышающих 10.

**Выходные данные**  
Выведите  $K$  матриц — значение параметров соответствующих выходных вершин графа вычисления в порядке возрастания их индексов. Затем выведите  $M$  матриц производных функции по соответствующим входным вершинам в порядке возрастания их индексов. Допустимая абсолютная и относительная погрешность  $10^{-4}$ .

входные данные
6 3 1 var 1 3 var 3 2 var 1 2 mul 1 2 sum 2 4 3 rlu 10 5 -2 3 5 4 2 -2 0 2 1 4 -2 -1 1
выходные данные
0.0 -0.1 -3.8 2.0 -1.9 2.0 -0.2 -3.0 0.3 -5.0 0.5 -1.0 0.1

В примере вычисляется функция  $ReLU_{\alpha=0.1} \left( \begin{pmatrix} -2 & 3 & 5 \end{pmatrix} \times \begin{pmatrix} 4 & 2 \\ -2 & 0 \\ 2 & 1 \end{pmatrix} + \begin{pmatrix} 4 & -2 \end{pmatrix} \right), a \begin{pmatrix} -1 & 1 \end{pmatrix}$  производная по её выходу.

I. Свёрточная сеть

1 секунда🕒, 256 мегабайт

Посчитайте значение выхода свёрточной сети и пересчитайте её производную.

**Входные данные**  
В первой строке содержится описание входа свёрточной сети, трёхмерной матрицы. Высота этой матрицы совпадает с её шириной. Первое число  $N_0$  ( $1 \leq N_0 \leq 40$ ) — высота и ширина входной трёхмерной матрицы, второе число  $D_0$  ( $1 \leq D_0 \leq 10$ ) — её глубина. Следующие  $D_0 \times N_0 \times N_0$  чисел — описание трёхмерной матрицы, значения её ячеек выписанных в порядке: глубина, высота, ширина.

Следующая строка содержит одно число  $L$  ( $1 \leq L \leq 10$ ) — число слоёв (преобразований) в сети.

Следующие  $L$  строк содержат описания соответствующих преобразований:

- relu**  $\alpha^{-1}$  ( $1 \leq \alpha^{-1} \leq 100$ ) — функции параметрического линейного выпрямителя с параметром  $\alpha$ .
- pool**  $S$  ( $1 \leq S \leq 5$ ) — операция субдискретизации (подвыборки) по высоте и ширине размера  $S \times S$  с шагом  $S$ . В качестве свёртки используется операция максимума. Производная для максимума вычисляется как:  $\frac{\partial \max}{\partial x_i}(x) = 1$  если  $x_i = \max(x)$ , иначе 0.

- **bias**  $B_1, B_2, \dots, B_D$  ( $|B_i| \leq 10$ ) — операция сдвига, прибавляющая к каждой ячейке матрицы на глубине  $i$  значение  $B_i$ ,  $D$  — глубина матрицы до и после преобразования.
- **cnvm**  $H\ K\ S\ P\ A_{1,1,1,1}, A_{1,1,1,2}, \dots, A_{H,D,K,K}$  ( $1 \leq H \leq 10$ ,  $1 \leq K \leq 5$ ,  $1 \leq S \leq K$ ,  $0 \leq P < K$ ,  $|A_i| \leq 10$ ) — свёртка с ядром  $A$  размера  $H \times D \times K \times K$  с шагом  $S$  с зеркальным заполнением рамки размера  $P$ , где  $D$  — глубина матрицы до преобразования.  $H$  — глубина матрицы после преобразования. Значения ячеек  $A$  выписаны в порядке: глубина полученной матрицы, глубина исходной матрицы, высота ядра, ширина ядра.
- **cnve**  $H\ K\ S\ P\ A_{1,1,1,1}, A_{1,1,1,2}, \dots, A_{H,D,K,K}$  — свёртка с расширением границы. Аналогична предыдущей.
- **cnvc**  $H\ K\ S\ P\ A_{1,1,1,1}, A_{1,1,1,2}, \dots, A_{H,D,K,K}$  — свёртка с заполнением с циклическим сдвигом. Аналогична предыдущей.

Гарантируется, что размеры всех многомерных матриц согласованы с соответствующими гипер-параметрами преобразований.

В последней строке записана производная по выходу сети.

Все числа во входных данных целые.

**Выходные данные**  
Выведите значение выходной трёхмерной матрицы.

Далее выведите производную по входу сети.

Затем для каждого слоя сдвига и свёртки в возрастающем порядке номера слоя выведите производную по его параметрам.

Выходные матрицы могут содержать числа с плавающей точкой. Допустимая абсолютная и относительная погрешность  $10^{-4}$ .

входные данные																													
4	1	4	3	2	1	3	2	1	0	2	1	0	1	1	0	1	2												
4																													
cnvm	1	3	3	1	0	-1	0	-1	0	-1	0	-1	0																
bias	4																												
relu	8																												
pool	2																												
1																													
выходные данные																													
0.0																													
0.0	0.0	-2.0	0.0	0.0	0.0	0.0	0.0	-2.0	-2.0	0.0	0.0	-2.0	0.0	-2.0	0.0	0.0													
-2.0	-2.0	0.0																											
2.0	3.0	2.0	3.0	4.0	3.0	2.0	3.0	2.0																					
3.0																													

Пример заполнения угла рамки для свёрточного слоя:

cnvm	18	17	16	15	16	17	18	19	cnve	0	0	0	0	1	2	3	4	cnvc	12	13	14	10	11	12	13	14
	13	12	11	10	11	12	13	14		0	0	0	0	1	2	3	4		17	18	19	15	16	17	18	19
	8	7	6	5	6	7	8	9		0	0	0	0	1	2	3	4		22	23	24	20	21	22	23	24
	3	2	1	0	1	2	3	4		0	0	0	0	1	2	3	4		2	3	4	0	1	2	3	4
	8	7	6	5	6	7	8	9		5	5	5	5	6	7	8	9		7	8	9	5	6	7	8	9
	13	12	11	10	11	12	13	14		10	10	10	10	11	12	13	14		12	13	14	10	11	12	13	14
	18	17	16	15	16	17	18	19		15	15	15	15	16	17	18	19		17	18	19	15	16	17	18	19
	23	22	21	20	21	22	23	24		20	20	20	20	21	22	23	24		22	23	24	20	21	22	23	24

J. LSTM сеть

1 секунда🕒, 256 мегабайт

Дана сеть LSTM для обработки последовательностей.

Каждый блок этой сети вычисляет результат по формулам:  
 $f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f)$ ,  $i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i)$ ,  
 $o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o)$ ,  
 $c_t = f_t \circ c_{t-1} + i_t \circ \tanh(W_c x_t + U_c h_{t-1} + b_c)$  и  $h_t = o_t \circ c_t$ , где  $x_t$  — вход  $t$ -го блока,  $h_t$  и  $c_t$  — векторы краткосрочной и долгосрочной памяти,  $o_t$  — выход  $t$ -го блока, а  $\circ$  — произведение Адамара.

**Входные данные**  
В первой строке находится число  $N$  ( $1 \leq N \leq 20$ ) — размер векторов LSTM.

Задачи - Codeforces

Далее перечислены соответствующие матрицы и вектора  $W_f, U_f, B_f, W_i, U_i, B_i, W_o, U_o, B_o, W_c, U_c, B_c$ .

Затем следует число  $M$  ( $1 \leq M \leq 20$ ) — число элементов последовательности обрабатываемой LSTM сетью.

Далее следуют два вектора  $h_0$  и  $c_0$ , а также  $M$  векторов  $x_t$ .

Затем следует вектора производных сети по выходным векторам  $h_M$  и  $c_M$ , а также  $M$  векторов производных по выходам  $o_t$  в обратном порядке  $o_M, o_{M-1}, \dots, o_1$ .

Все вектора записаны  $N$  числами разделёнными пробелами на отдельной строке, а матрицы  $N$  векторами размера  $N$ . Все элементы векторов и матриц целые числа по модулю не превосходящие 10.

**Выходные данные**  
Сперва выведите  $M$  векторов выходов сети  $o_t$ .

Далее выведите два последних вектора памяти  $h_M$  и  $c_M$ .

Затем выведите  $M$  векторов производных сети по входам  $x_t$  в обратном порядке.

Далее выведите два вектора производных сети по  $h_0$  и  $c_0$ .

После выведите производные по соответствующим матрицам и векторам параметров LSTM:  $W_f, U_f, B_f, W_i, U_i, B_i, W_o, U_o, B_o, W_c, U_c, B_c$ .

Выходные вектора и матрицы могут содержать числа с плавающей точкой. Допустимая абсолютная и относительная погрешность  $10^{-6}$ .

входные данные															
1	-3	2	1	1	-2	-2	-3	-1	-2	1	-2	-1	1	1	-3
2	1	-2	-1	1	-3	2	1	-1	1	-3	2	1	-1	1	1
выходные данные															
1.	233945759863131E-4	-2.875857041962763E-5	-0.23306186831759548	-0.37692699674663843	0.21113860108361812	-0.047420021082055105	0.27102651105684017	0.13551325552842008	0.13551325552842008	0.159905268234481	0.0799526341172405	0.0799526341172405	1.8924865599381104E-4	9.462432799690552E-5	9.462432799690552E-5
	-0.10011198258925587	-0.050055991294627934	-0.050055991294627934												

К. Коэффициент корреляции Пирсона

1 секунда🕒, 256 мегабайт

Посчитайте корреляцию Пирсона двух численных признаков.



Входные данные

Первая строка содержит целое положительное число  $N$  ( $1 \leq N \leq 10^5$ ) — число объектов.

Следующие  $N$  строк содержат описания соответствующих объектов. Каждая из этих  $N$  строк содержит описание одного объекта: два целых числа  $x_1$  и  $x_2$  ( $-10^9 \leq x_1, x_2 \leq 10^9$ ) — значения первого и второго признака описываемого объекта.

Выходные данные

Выведите одно вещественное число с плавающей точкой — корреляцию Пирсона двух признаков у заданных объектов. Допустимая абсолютная и относительная погрешность  $10^{-6}$ .

входные данные
5 1 4 2 5 3 1 4 2 5 3
выходные данные
-0.500000000

L. Коэффициент ранговой корреляции  
Спирмена

1 секунда, 256 мегабайт

Посчитайте ранговую корреляцию Спирмена двух численных признаков.

Входные данные

Первая строка содержит целое положительное число  $N$  ( $1 \leq N \leq 10^5$ ) — число объектов.

Следующие  $N$  строк содержат описания соответствующих объектов. Каждая из этих  $N$  строк содержит описание одного объекта: два целых числа  $x_1$  и  $x_2$  ( $-10^9 \leq x_1, x_2 \leq 10^9$ ) — значения первого и второго признака описываемого объекта. Гарантируется, что все значения каждого признака различны.

Выходные данные

Выведите одно вещественное число с плавающей точкой — коэффициент ранговой корреляции Спирмена двух признаков у заданных объектов. Допустимая абсолютная и относительная погрешность  $10^{-6}$ .

входные данные
5 1 16 2 25 3 1 4 4 5 9
выходные данные
-0.500000000

M. Расстояния

1 секунда, 256 мегабайт

Посчитайте зависимость категориального признака  $Y$  от числового  $X$  по внутриклассовому и межклассовому расстоянию:

- Внутриклассовое расстояние =  $\sum_{i,j: y_i=y_j} |x_i - x_j|$
- Межклассовое расстояние =  $\sum_{i,j: y_i \neq y_j} |x_i - x_j|$

Входные данные

Первая строка содержит одно целое положительное число  $K$  ( $1 \leq K \leq 10^5$ ) — максимальное число различных значений  $Y$  второго признака.

Следующая строка содержит одно целое положительное число  $N$  ( $1 \leq N \leq 10^5$ ) — число объектов.

Следующие  $N$  строк содержат описания соответствующих объектов. Каждая из этих  $N$  строк содержит описание одного объекта: два целых числа  $x$  и  $y$  ( $|x| \leq 10^7, 1 \leq y \leq K$ ) — значения первого и второго признака описываемого объекта.

Выходные данные

В первой строке выведите одно целое число — внутриклассовое расстояние.  
Во второй строке выведите одно целое число — межклассовое расстояние.

входные данные
2 4 1 1 2 2 3 2 4 1
выходные данные
8 12

N. Условная дисперсия

1 секунда, 256 мегабайт

Вычислите критерий связи двух признаков категориального  $X$  и числового  $Y$  на основе математического ожидания условной дисперсии  $D(Y|X)$ . Вероятности для  $X$  оцениваются обыкновенным частотным методом.

Входные данные

Первая строка содержит одно целое положительное число  $K$  ( $1 \leq K \leq 10^5$ ) — максимальное число различных значений признака  $X$ .

Следующая строка содержит целое положительное число  $N$  ( $1 \leq N \leq 10^5$ ) — число объектов.

Следующие  $N$  строк содержат описания соответствующих объектов. Каждая из этих  $N$  строк содержит описание одного объекта: два целых положительных числа  $x$  и  $y$  ( $1 \leq x \leq K, |y| \leq 10^9$ ) — значения признаков  $X$  и  $Y$ .

Выходные данные

Выведите одно вещественное число с плавающей точкой — математическое ожидание условной дисперсии. Допустимая абсолютная и относительная погрешность  $10^{-6}$ .

входные данные
2 4 1 1 2 2 2 3 1 4
выходные данные
1.25

O. Хи-квадрат

1 секунда, 256 мегабайт

Посчитайте зависимость двух категориальных признаков согласно критерию хи-квадрат (критерий согласия Пирсона).

Входные данные

Первая строка содержит два целых положительных числа  $K_1$  и  $K_2$  ( $1 \leq K_1, K_2 \leq 10^5$ ) — максимальное число различных значений первого и второго признака.

Следующая строка содержит целое положительное число  $N$  ( $1 \leq N \leq 10^5$ ) — число объектов.

Следующие  $N$  строк содержат описания соответствующих объектов. Каждая из этих  $N$  строк содержит описание одного объекта: два целых положительных числа  $x_1$  и  $x_2$  ( $1 \leq x_1 \leq K_1$ ,  $1 \leq x_2 \leq K_2$ ) — значения первого и второго признака описываемого объекта.

Выходные данные

Выведите одно вещественное число с плавающей точкой — критерий хи-квадрат зависимости двух признаков у заданных объектов. Допустимая абсолютная и относительная погрешность  $10^{-6}$ .

входные данные			
2	3		
5			
1	2		
2	1		
1	1		
2	2		
1	3		
выходные данные			
0.833333333			

В примере реальное число наблюдений наблюдений выглядит как

	1	2	3
1	1	1	1
2	1	1	0
	1	2	3
1	1.2	1.2	0.6
2	0.8	0.8	0.4

Вычислите критерий связи двух категориальных признаков  $X$  и  $Y$  на основе математического ожидания условной энтропии  $H(Y|X)$ . Вероятности оцениваются обыкновенным частотным методом. При расчётах используйте натуральный логарифм  $\ln(x)$ , либо логарифм идентичный натуральному  $\log_e(x)$ .

Входные данные

Первая строка содержит два целых положительных числа  $K_x$  и  $K_y$  ( $1 \leq K_x, K_y \leq 10^5$ ) — максимальное число различных значений признаков  $X$  и  $Y$ .

Следующая строка содержит целое положительное число  $N$  ( $1 \leq N \leq 10^5$ ) — число объектов.

Следующие  $N$  строк содержат описания соответствующих объектов. Каждая из этих  $N$  строк содержит описание одного объекта: два целых положительных числа  $x$  и  $y$  ( $1 \leq x \leq K_x$ ,  $1 \leq y \leq K_y$ ) — значения признаков  $X$  и  $Y$ .

Выходные данные

Выведите одно вещественное число с плавающей точкой — математическое ожидание условной энтропии. Допустимая абсолютная и относительная погрешность  $10^{-6}$ .

входные данные			
2	3		
5			
1	2		
2	1		
1	1		
2	2		
1	3		
выходные данные			
0.9364262454248438			

Р. Условная энтропия