A. k-ближайших соседей

2 секунды, 256 мегабайт

Требуется ответить на несколько запросов вычисления среднего среди k-ближайших объектов к запросу. Все объекты одномерные, если не считать целевой признак.

Входные данные

Первая строка содержит одно целое положительное число n ($1 \le n \le 10^5$) — число объектов.

Следующие n строк содержат описание объектов. Каждая из этих строк содержит два разделённых пробелом целых числа: x_i ($|x_i| \le 10^9$) и y_i ($1 \le y_i \le 10^9$) — значения обычного и целевого признака i-го объекта. Гарантируется, что все x_i различны.

Далее следует строка с одним целым положительным числом m ($1 \le m \le 2 \cdot 10^4$) — число запросов.

Следующие m строк содержат описание запросов. Каждая из этих строк содержит два разделённых пробелом целых числа: x_q ($|x_q| \le 10^9$) и k_q ($1 \le k_q \le n$) — положение запроса и интересующее число ближайших объектов к нему.

Выходные данные

Для каждого запроса выведите одно число — среднее значение целевого признака k-ближайших объектов. Если нельзя однозначно выбрать k-ближайших объектов, то выведите -1.

входные данные

```
5
1 4
5 3
3 4
7 2
9 8
4
2 1
6 2
5 3
8 4
```

выходные данные

```
-1.0
2.5
3.0
4.25
```

B. 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 категориальный признак с двумя возможными событиями встретилось / не встретилось.

Входные данные

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

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

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

Следующая строка содержит целое положительное число N ($1 \le N \le 200$) — 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 ($1 \le C_i \le K$) — класса, к которому относится i-е сообщение. Далее следует целое положительное число L_i ($1 \le L_i \le 10^4$) — 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 ($1 \le M \le 200$) — 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 ($1 \le L_j \le 10^4$) — 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:

$p(w_x \mid c_y)$	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:

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

C. Марковская цепь

1 секунда, 256 мегабайт

Даны несколько строк. Известно, что почти все они были получены сэмплированием из одной марковской цепи, но одна строка получена из простого случайного распределения, в котором каждая буква выбирается независимо от остальных.

Найдите эту строку.

Входные данные

Первая строка содержит натуральное число N $(3 \le N \le 10)$ — число строк.

Далее следует N строк, которые состоят только из маленьких латинских букв и пробелов. Сумма длин всех строк не превышает 10^4.

Выходные данные

Выведите одно натуральное число — номер строки, которая была получена из простого случайного распределения. Строки нумеруются с единицы.

входные данные

```
3
sus asus
sus xsus
uss axss
```

выходные данные

```
3
```

D. Двумерная свёртка

1 секунда, 256 мегабайт

Дан вход и результат двумерной свёртки. Требуется найти ядро.

Решать можно только через свёртки.

Входные данные

Первая строка содержит разделённых пробелом два натуральных числа n и m $(1 < M < N \le 10)$ — размеры входа и результата свёртки соответственно.

Далее следует две квадратных матрицы — вход свёртки размера $n \times n$ и результат свёртки размера $m \times m$. Все числа целые и находятся в диапазоне $[0; 255]$.

Выходные данные

Выведите ядро свёртки. Если решений несколько, разрешается вывести любое. Допускаются вещественные числа.

Ответ засчитывается, если абсолютная погрешность полученного и ожидаемого результата свёртки не превышает 10^{-6}.

входные данные

```
3 2
1 2 3
4 5 6
7 8 9
23 33
53 63
```

выходные данные

```
4.0 3.0
2.0 1.0
```

E. Условная энтропия

1 секунда, 256 мегабайт

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 ($1 \le K_x, K_y \le 10^5$) — максимальное число различных значений признаков X и Y.

Следующая строка содержит целое положительное число N ($1 \le N \le 10^5$) — 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 ($1 \le x \le K_x$, $1 \le y \le K_y$) — 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
```

F. 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 ($1 \le K \le 20$). Далее идёт K строк — описание матрицы неточностей. Каждая строка c содержит K целых чисел — c-я строка матрицы неточностей. $\forall c,t : 0 \le CM[c,t] \le 100$ и $\exists c,t : CM[c,t] \ge 1$.

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.

G. Коэффициент ранговой корреляции Спирмена

1 секунда, 256 мегабайт

Посчитайте ранговую корреляцию Спирмена двух численных признаков.

Входные данные

Первая строка содержит целое положительное число N ($1 \le N \le 10^5$) — 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 ($-10^9 \le x_1, x_2 \le 10^9$) — 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

H. 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 ($1 \le K_1, K_2 \le 10^5$) — максимальное число различных значений первого и второго признака.

Следующая строка содержит целое положительное число N ($1 \le N \le 10^5$) — 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 ($1 \le x_1 \le K_1$, $1 \le x_2 \le K_2$) — 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 выглядит как $\begin{array}{cccc} & \mathbf{1} & \mathbf{2} & \mathbf{3} \\ \mathbf{1} & 1 & 1 & 1 \\ \mathbf{2} & 1 & 1 & 0 \end{array}$, а ожидаемое число наблюдений $\begin{array}{cccc} & \mathbf{1} & \mathbf{2} & \mathbf{3} \\ \mathbf{1} & 1.2 & 1.2 & 0.6 \\ \mathbf{2} & 0.8 & 0.8 & 0.4 \end{array}$.

I. 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 = $\sum_{i,j: y_i \ne y_j} |x_i - x_j|$

Входные данные

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

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

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

J. 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 ($1 \le K \le 10^5$) — максимальное число различных значений признака X.

Следующая строка содержит целое положительное число N ($1 \le N \le 10^5$) — 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 ($1 \le x \le K$, $|y| \le 10^9$) — 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
```

K. Категориальная корреляция

1 секунда, 256 мегабайт

Вычислите коэффициент корреляции Пирсона между категориальным и числовым признаком. Так как первый признак категориальный сперва требуется применить one-hot преобразование к нему, а затем вычислить среднее взвешенное значение корреляций между новыми признаками и b.

Входные данные

Первая строка содержит два натуральных числа N и K, разделённых пробелами: N ($1 \le N \le 10^5$) — число объектов, K ($1 \le K \le 10^5$) — число значений категории первого признака. Вторая строка содержит N натуральных чисел, разделённых пробелами: i-е из них a_i ($1 \le a_i \le K$) — значение первого признака i-го объекта. Третья строка содержит N целых чисел, разделённых пробелами: i-е из них b_i ($|b_i| \le 10^9$) — значение второго признака i-го объекта.

Выходные данные

Выведите одно вещественное число с плавающей точкой — коэффициент корреляции Пирсона между a и b. Абсолютная или относительная погрешность ответа не должна превышать 10^{-9}

входные данные

```
6 3
1 2 2 3 3 3
1 2 3 4 5 6
```

выходные данные

```
0.19203297584037293
```

В примере значение корреляции между первым новым признаком $(1,0,0,0,0,0)$ и b равно -0.654653671, а его вес равен единице, так как соответствующие значение встретилось только один раз. Значение корреляции между вторым новым признаком $(0,1,1,0,0,0)$ и b равно -0.414039336, а его вес равен двум. Значение корреляции между третьим новым признаком $(0,0,0,1,1,1)$ и b равно 0.878310066, а его вес равен трём.

L. 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 ($1 \le M \le 10$) — 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 ($1 \le D \le 2$) — число слоёв (преобразований) в вашей сети.

На следующей строке выведите D целых положительных чисел n_i ($1 \le n_i \le 512$ 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
```

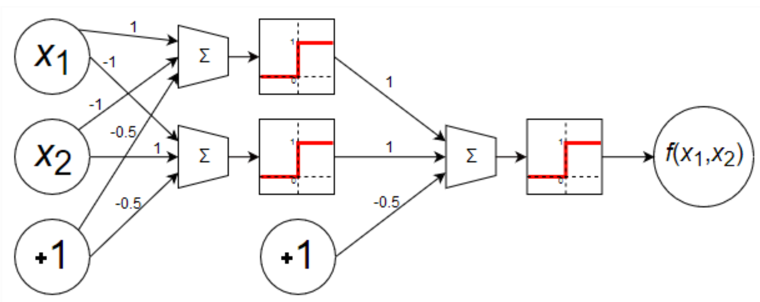
входные данные

```
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
```

Во втором примере в результате получается следующая сеть:



M. 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 $(2 \le M + K \le N \le 50)$ — 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 $(1 \le r, c \le 25)$ — входной параметр функции, матрица состоящая из r строк и c столбцов.
- **tnh** x $(1 \le x < i)$ — матрица из значений гиперболического тангенса вычисленного от соответствующих компонент матрицы, полученной из x-й вершины графа вычислений.
- **rlu** α^{-1} x $(1 \le \alpha^{-1} \le 100,\ 1 \le x < i)$ — матрица из значений функции параметрического линейного выпрямителя с параметром α, вычисленной от соответствующих компонент матрицы полученной из x-й вершины графа вычислений. α^{-1} — целое число. Производная в нуле равна единице.
- **mul** a b $(1 \le a, b < i)$ — произведение матриц, полученных из a-й b-й вершины графа вычислений соответственно.
- **sum** len u_1 u_2 ... u_{len} $(1 \le len \le 10,\ \forall_{1 \le j \le len} : 1 \le u_j < i)$ — сумма матриц, полученных из вершин $u_1, u_2, \ldots, u_{len}$ графа вычислений.
- **had** len u_1 u_2 ... u_{len} $(1 \le len \le 10,\ \forall_{1 \le j \le len} : 1 \le u_j < i)$ — произведение Адамара (покомпонентное) матриц, полученных из вершин $u_1, u_2, \ldots, u_{len}$ графа вычислений.

Гарантируется, что первые M вершин и только они имеют тип **var**. Последние K вершин считаются выходными. Никакие вершины не зависят от последних K вершин. 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.

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

$$\mathrm{ReLU}_{\alpha=0.1}\left(\begin{pmatrix} -2 & 3 & 5 \end{pmatrix} \times \begin{pmatrix} 4 & 2 \\ -2 & 0 \\ 2 & 1 \end{pmatrix} + \begin{pmatrix} 4 & -2 \end{pmatrix}\right),\text{ а } \begin{pmatrix} -1 & 1 \end{pmatrix}$$

производная по её выходу.

N. 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 $\forall x, y, c : |cnt(x, c) - cnt(y, c)| \le 1$ и $\forall x, y : |\sum_c cnt(x, c) - \sum_c cnt(y, c)| \le 1$.

Входные данные

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

Вторая строка: N целых чисел C_i $(1 \le C_i \le M)$ — 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.

Codeforces (c) Copyright 2010-2024 Михаил Мирзаянов
Соревнования по программированию 2.0