

# Metric based classification

ограничение по времени на тест: 20 секунд  
ограничение по памяти на тест: 512 мегабайт  
ввод: стандартный ввод  
вывод: стандартный вывод

## Задача

Требуется построить метрический классификатор на обучающем наборе данных с известными классами и классифицировать с его помощью тестовый набор данных с неизвестными классами. Ваш классификатор должен содержать несколько метрик, сглаживающих ядер и стратегий выбора ширины окна ядра (числа ближайших соседей). Требуется выбрать оптимальную комбинацию гипер-параметров для каждого отдельного набора данных.

## Входные данные

Первая строка содержит целое число  $M$  ( $5 \leq M \leq 200$ ) — число признаков у объектов исключая класс.

Вторая строка содержит целое число  $K$  ( $2 \leq K \leq 25$ ) — число классов.

Третья строка содержит целое число  $N$  ( $50 \leq N \leq 400$ ) — число объектов в обучающем множестве.

Следующие  $N$  строк содержат описание объектов.  $i$ -тая из этих строк содержит описание  $i$ -того объекта  $M+1$  целых чисел: первые  $M$  чисел  $A_{ij}$  ( $A_{ij} \leq 10^9$ ) — признаки объекта, последнее число  $C_i$  ( $1 \leq C_i \leq K$ ) — его класс.

Следующая строка содержит целое число  $Q$  ( $50 \leq Q \leq 400$ ) — число объектов в тестовом множестве.

Следующие  $Q$  строк содержат описание объектов.  $t$ -тая из этих строк содержит описание  $t$ -того объекта:  $M$  целых чисел  $A_{tj}$  ( $|A_{tj}| \leq 10^9$ ) — признаки объекта.

## Выходные данные

Выведите  $Q$  строк. Каждая  $t$ -тая строка из них должна содержать результат классификации  $t$ -того объекта из тестового множества: целое число  $S_t$  ( $1 \leq S_t \leq 20$ ) — число соседей классифицируемого объекта, затем следует  $S_t$  пар чисел  $i$  и  $w$  ( $1 \leq i \leq N$ ,  $0 \leq w \leq 10^6$ ), где  $i$  — целое число, индекс объекта из тренировочного множества, а  $w$  — вещественное число с плавающей точкой, вес с которым учитывается этот объект. Числа  $S_t$  не обязательно должны быть одинаковыми для всех объектов.

# Система оценки

Для каждого объекта  $t$  будет рассчитан его предсказанный класс  $p_t = \operatorname{argmax}(\sum w[C_i = c])$  исходя из соответствующего множества пар  $(i, w)$ , в случае неопределённости класс выбирается псевдослучайно. Далее на основании предсказанных и реальных классов вычисляется усреднённая по классам микро F мера. Тест считается пройденным, если эта F мера будет выше определённого порога, рассчитанного с 5% запасом с помощью базового метрического классификатора.

## Пример

Входные данные	Выходные данные
2	3 1 0.75 3 0.21 4 0.21
2	3 2 0.75 3 0.65 4 0.65
4	
2 2 1	
6 2 1	
4 1 2	
4 3 2	
2	
1 2	
5 2	

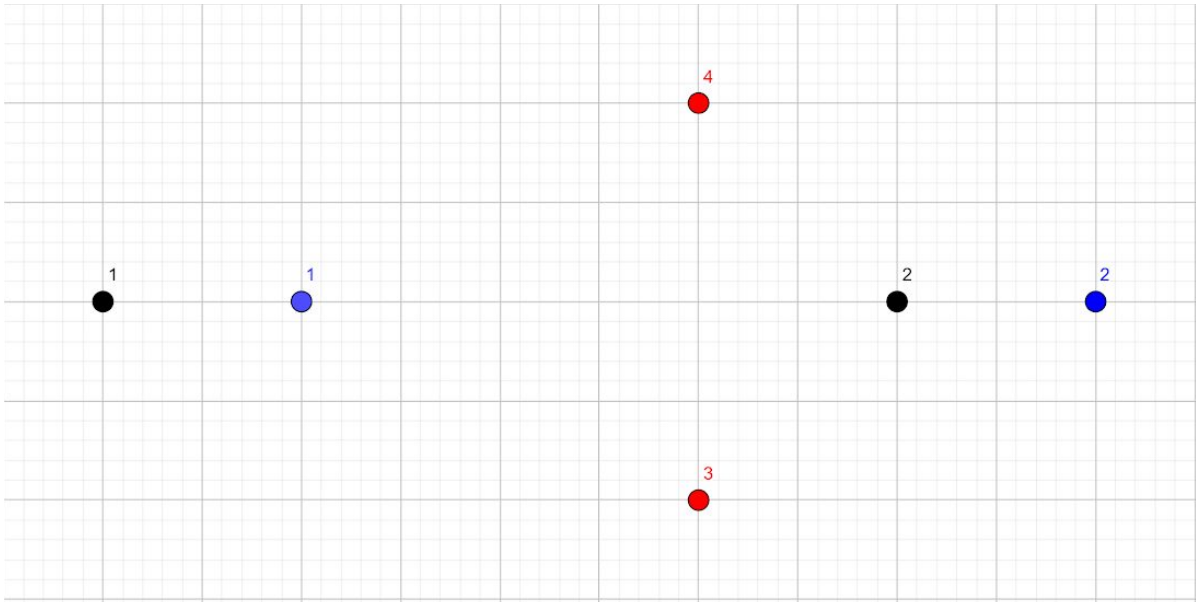


Иллюстрация примера

## Пояснение

Обратите внимание, что экземпляр задачи из данного примера не подходит под нижние ограничения на число объектов и признаков!

В данном примере в качестве метрики используется Евклидово расстояние, а в качестве взвешивающей функции Треугольное ядро. Используется три ближайших соседа с шириной окна 4.

Для первого запроса:

Номер объекта:	1	3	4
Расстояние:	1	$\sqrt{10}$	$\sqrt{10}$
Полученный вес:	0.75	0.21	0.21
Класс:	1	2	2

Суммарный вес первого класса **0.75**, второго 0.42.

Для второго запроса:

Номер объекта:	2	3	4
Расстояние:	1	$\sqrt{2}$	$\sqrt{2}$
Полученный вес:	0.75	0.65	0.65
Класс:	1	2	2

Суммарный вес первого класса 0.75, второго **1.3**.

# SVM

ограничение по времени на тест: 4 секунды  
ограничение по памяти на тест: 512 мегабайт  
ввод: стандартный ввод  
вывод: стандартный вывод

## Задача

Дан набор данных для бинарной классификации. Требуется построить разделяющие правило опирающиеся на объекты из заданного набора.

## Входные данные

Первая строка содержит целое число  $M$  ( $2 \leq M \leq 10$ ) — число признаков у объектов исключая класс.

Вторая строка содержит целое число  $N$  ( $M \leq N \leq 100$ ) — число объектов в обучающем множестве.

Следующие  $N$  строк содержат описание объектов.  $i$ -тая из этих строк содержит описание  $i$ -того объекта:  $M$  целых чисел  $D_{i,j}$  ( $|D_{i,j}| \leq 10^5$ ) — признаки объекта и его класс ('+' если объект положительный  $\text{class}(D_i) = +1$  и '-' если он отрицательный  $\text{class}(D_i) = -1$ ).

## Выходные данные

Выведите решающие правило формата:

$$\text{class}(Q) = \text{sign}((\sum \lambda_i \cdot \text{class}(D_i) \cdot f(D_i, Q)) - b)$$

В первой строке выведите симметричную функцию ядра  $f: \mathbb{R}^M \times \mathbb{R}^M \rightarrow \mathbb{R}$ , состоящую из не более чем 1000 символов. Данная функция должна удовлетворять грамматике:

- $E \rightarrow \text{pow}(E, E)$ , где  $\text{pow}(a, b) = a^b$
- $E \rightarrow \text{sub}(E, E)$ , где  $\text{sub}(a, b) = a - b$
- $E \rightarrow \text{sum}(E, E, \dots, E)$ , где  $\text{sum}(a, b, \dots, z) = a + b + \dots + z$
- $E \rightarrow \text{prod}(E, E, \dots, E)$ , где  $\text{prod}(a, b, \dots, z) = a \times b \times \dots \times z$
- $E \rightarrow A0 \mid A1 \mid \dots \mid A9$ , где  $Ai$  значение  $i$ -той координаты первого вектора ( $0 \leq i < M$ , координаты нумеруются с нуля).
- $E \rightarrow B0 \mid B1 \mid \dots \mid B9$ , где  $Bi$  значение  $i$ -той координаты второго вектора.
- $E \rightarrow$  число с плавающей точкой.

Запрещено использовать пробелы и иные символы.  $E, E, \dots, E$  — это непустое перечисление через запятую.

В следующих  $N$  строках выведите  $N$  вещественных чисел  $\lambda_i$  ( $0 \leq \lambda_i \leq 10^6$ ,  $\sum \lambda_i \cdot \text{class}(D_i) = 0$ ) — веса объектов в порядке их перечисления во входных данных.

В последней строке выведите одно вещественное число  $b$  — коэффициент сдвига.

## Система оценки

Решение будет проверено на секретном наборе данных. На основании предсказанных и реальных классов вычисляется усреднённая по классам микро F-мера. Тест считается пройденным, если эта F-мера будет выше определённого порога, рассчитанного с 5% запасом с помощью базового решения.

## Пример

Входные данные	Выходные данные
2	<code>pow(sum(0.0,prod(sub(A0,2.5),sub(B0,2.5),0.73)</code>
12	<code>,prod(sub(A1,2.5),sub(B1,2.5),0.3)),1.0)</code>
1 2 -	0.3823
1 3 -	0
1 4 -	0
2 1 +	0.3823
2 3 -	1.0
2 4 -	0
3 1 +	0
3 2 +	1.0
3 4 -	0.3823
4 1 +	0
4 2 +	0
4 3 +	0.3823
	0

# Linear regression

ограничение по времени на тест: 4 секунды  
ограничение по памяти на тест: 512 мегабайт

ввод: стандартный ввод  
вывод: стандартный вывод

## Задача

Найдите уравнения прямой аппроксимирующей положение объектов из заданного набора данных.

## Входные данные

Первая строка содержит целое число  $M$  ( $1 \leq M \leq 1000$ ) — число признаков у объектов исключая зависимую переменную.

Вторая строка содержит целое число  $N$  ( $M \leq N \leq 10000$ ) — число объектов в обучающем множестве.

Следующие  $N$  строк содержат описание объектов.  $i$ -тая из этих строк содержит описание  $i$ -того объекта,  $M$  целых чисел:  $X_{i,j}$  ( $|X_{i,j}| \leq 10^9$ ) — признаки объекта и  $Y_i$  ( $|Y_i| \leq 10^9$ ) — значение его зависимой переменной.

## Выходные данные

Выведите  $M + 1$  вещественных чисел с плавающей точкой  $A_j$  — коэффициенты прямой из уравнения  $y = a_0 \cdot x_0 + a_1 \cdot x_1 + \dots + a_{M-1} \cdot x_{M-1} + a_M$

## Система оценки

Решение будет проверено на секретном наборе данных. На основании предсказанных значений  $Y'_t$  и реальных  $Y_t$  вычисляется ошибка предсказания — нормированная сумма квадратов  $E = \sum (Y'_t - Y_t)^2 / D(Y)$ , где  $D(Y)$  — это дисперсия зависимой величины. Решение засчитывается если полученная ошибка  $E$  отличается от ошибки полученной базовым решением не более чем на 0.01

Пример

Входные данные	Выходные данные
1 4 1 0 1 2 2 2 2 4 3 4 3 6 4 6 4 8	2 -1