

Cross-Validation

ограничение по времени на тест: 2 секунды
ограничение по памяти на тест: 256 мегабайт
ввод: стандартный ввод
вывод: стандартный вывод

Задача

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 $\text{cnt}(x, c)$ — число объектов с классом c попавших в часть x , тогда должно выполняться ($\forall x, y, c : |\text{cnt}(x, c) - \text{cnt}(y, c)| \leq 1$) и ($\forall x, y : |\sum \text{cnt}(x, c) - \sum \text{cnt}(y, c)| \leq 1$).

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 -того 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.

Пример

Входные данные	Выходные данные
10 4 3 1 2 3 4 1 2 3 1 2 1	4 1 4 9 10 3 2 3 5 3 6 7 8

Пояснение

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. Обратите внимание, что форматирование в примере используется только для наглядности!

F₁ score

ограничение по времени на тест: 2 секунды
ограничение по памяти на тест: 256 мегабайт
ввод: стандартный ввод
вывод: стандартный вывод

Задача

В 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 усреднённую по классам макро и микро 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 -тая 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 два вещественных числа с плавающей точкой — усреднённую по классам макро и микро F-меру. Абсолютная или относительная погрешность ответа не должна превышать 10^{-6} .

Примеры

Входные данные	Выходные данные
2 0 1 1 3	0.6 0.6
3 3 1 1 3 1 1 1 3 1	0.3268608414239482 0.31666666666666665

Пояснение

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.

Metric based classification

ограничение по времени на тест: 20 секунды

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	

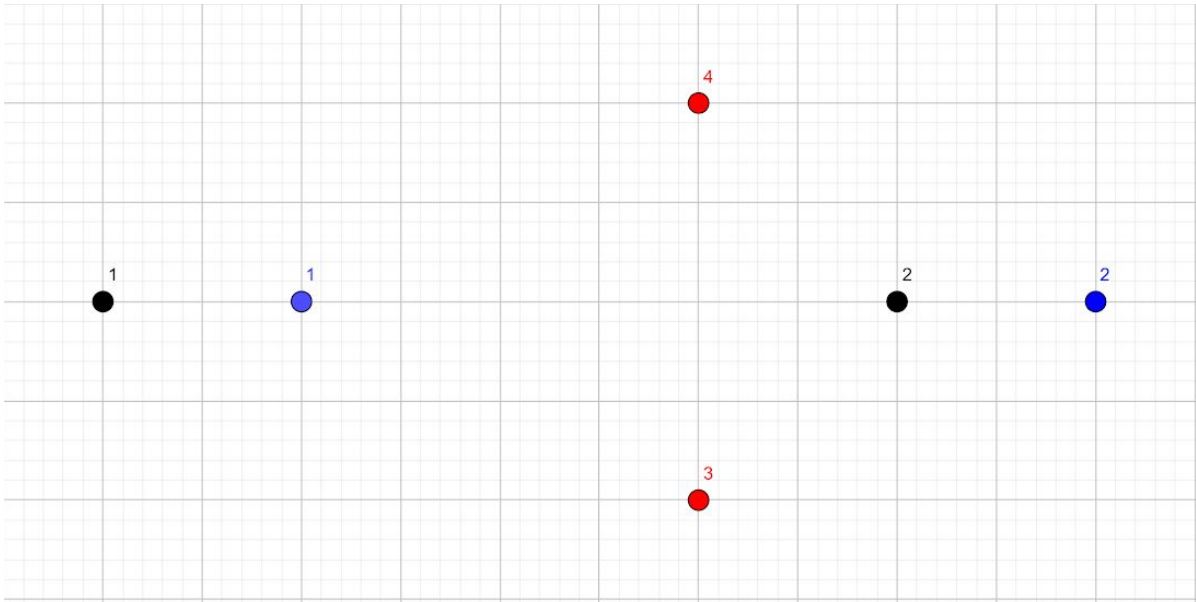


Иллюстрация примера

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**.