

Отборочный этап  
олимпиады студентов "Я - Профессионал"  
по направлению "Искусственный интеллект"  
2020/2021 учебный год  
(объединенные КИМ для уровней  
"бакалавриат" и  
"магистратура/специалитет")

1. Задания 1-20 оцениваются исходя из 3 баллов. По одному баллу за каждый из 8 тестов (претесты не включены).
2. Задания 21-23 оцениваются исходя из 8 баллов. По одному баллу за каждый из 8 тестов (претесты не включены).
3. Задания 24-25 оцениваются исходя из 8 баллов. Каждая посылка участника имеет какое-то количество очков. Очки линейно нормируются в баллы так, чтобы 8 баллов получила посылка с максимальным количеством очков.
4. Апелляция должна содержать номер/номера посылок, с оценкой, которых вы не согласны.
5. Контесты проходили по адресам:
  - (a) <https://contest.yandex.ru/contest/17112/problems/>
  - (b) <https://contest.yandex.ru/contest/17109/problems/>
6. В задачах 1-20 приведены ответы и решения.
7. В задачах 21-23 приведены идеи решения и код программы, которая получает 8 баллов из 8.
8. В задачах 24-25 приведены идеи (без реализации), которых позволяет получить хотя бы 5 из 8 баллов.

# 1 Задачи на ответ. По 3 балла.

## 1. Комбинаторика

В регламенте олимпиады есть такие условия:

- 1) любой победитель олимпиады должен набрать больше баллов, чем любой призер олимпиады;
- 2) любой призер олимпиады должен набрать больше баллов, чем любой участник, не являющийся ни призером, ни победителем;
- 3) в олимпиаде должен быть хотя бы один призер и хотя бы один победитель.

Сколько существует способов назначить победителей и призеров олимпиады, не нарушая эти условия.

## 2. Теория вероятностей

Вася написал случайную последовательность из 0 и 1 длины 9. Найдите вероятность, что у него получился палиндром.

*Формат ответа - несократимая дробь  $p/q$ .*

## 3. Условное математическое ожидание

Международная олимпиада проводится в два тура.

Среди участников треть русских студентов, треть американских студентов, треть китайских студентов.

В каждом туре студенты из Америки получают результат от 51 до 80 баллов равномерно, из Китая - от 71 до 80 равномерно, из России - от 1 до 100 равномерно.

Результаты в разных турах независимы друг от друга.

Участник  $X$  набрал в первом туре 68 баллов, найдите ожидаемое количество баллов у него во втором туре.

## 4. Линейная алгебра

Решите систему уравнений

$$Ax = y$$

$$By = x$$

$$|x| = 1$$

Где матрицы  $A = \begin{pmatrix} 2 & 1 \\ 3 & 5 \end{pmatrix}$ ,  $B = \begin{pmatrix} 2 & 1 \\ 3 & 5 \end{pmatrix}$

*В ответе укажите векторы  $x, y$  в формате  $x_1 x_2 y_1 y_2$  - 4 числа, разделенных пробелами, где каждое число записано как несократимая дробь  $p/q$ .*

## 5. Конечные автоматы

Найдите самое короткое слово, которое принимается ровно 2 конечными автоматами из 3.

## 6. Инопланетный NLP

В инопланетном алфавите всего три буквы -  $a, b, c$ , а словом считается любая последовательность букв.

Ученые, изучающие инопланетный язык, считают, что:

- 1) после буквы  $a$  равномерно идет любая из трех букв; буква  $a$  не может быть последней;
- 2) после буквы  $b$  в половине случаев идет буква  $c$ , а в половине случаев это последняя буква;
- 3) после буквы  $c$  равномерно идет буква  $a, b$ , или эта буква последняя.

Найдите, в какой доле слов, начинающихся на  $a$ , есть хотя бы 10 букв и 10ая буква - это  $a$ .

## 7. Precision Recall

Вася и Петя решают задачу на определение, является ли письмо спамом.

Вася и Петя разделили датасет на две части и работают каждый со своей частью. Так получилось, что в каждой части ровно 10% писем являются спамом.

Каждый из них написал свой алгоритм бинарной классификации и запустил его на своей части. Известно, что оба алгоритма верно поместили как спам 5% писем (от размера соответствующей части).

Потом Вася и Петя объединили свои результаты. Можно ли найти Precision и Recall у получившейся классификации?

### 8. Подсчет действий

Какое количество операций сложения будет произведено при выполнении следующего кода:

```
def g(n):
    if n < 1:
        return 1
    return f(n-1)+f(n-1)

def f(n):
    if n < 2:
        return 1
    return g(n-2)+g(n-2)+g(n-2)
```

$f(30)$

### 9. Сверточная сеть

Дана матрица размера  $10^9 \times 10^9$ , значения в которой заданы формулой  $a_{ij} = |i - j|$ .

Пока размер матрицы больше 2, к ней применяется следующая операция свертки: каждое число заменяется на наиболее часто встречающееся число в подматрице  $3 \times 3$  с центром в этом числе (если таких чисел несколько, то берется наименьшее из них). Числа, стоящие в первом и последнем столбце/строке, удаляются - для них нет такой подматрицы  $3 \times 3$ .

В результате получается матрица, в которой на 2 строки и 2 столбца меньше, и дальше операция применяется уже к ней.

Какая сумма чисел будет в итоговой матрице  $2 \times 2$ ?

### 10. Оптимизация

Найдите значение

$$T = \min_{x \geq 1} \max_{3x \leq y \leq 8x} f(x, y), \text{ где}$$

$$f(x, y) = \ln x + \ln y + \ln(y - 3x) + \ln(8x - y).$$

В ответе запишите  $x_0 y_0$ , в котором это значение достигается, в формате несократимых дробей через пробел.

### 11. Статистика

Вася играет в игровой автомат.

Игра происходит следующим образом: игрок выбирает число  $X$  от 0 до 1, автомат генерирует случайное число  $Y$  от 0 до 1. Далее рассчитывается выигрыш, который зависит от того, насколько близко числа  $X$  и  $Y$ .

Вася узнал, что автомат генерирует число независимо от предыдущих, и берет его из следующего распределения: от 0 до  $1/2$  с плотностью  $3/2$  и от  $1/2$  до 1 с плотностью  $1/2$ .

Также известно, что выигрыш в автомате рассчитывается либо по формуле  $1 - |X - Y|$ , либо по формуле  $1 - (X - Y)^2$ , причем первая формула применяется в  $1/3$  случаев, а вторая - в остальных. Выбор формулы не зависит от чисел  $X, Y$ .

Какое число стоит назвать Васе?

### 12. Решающее дерево

На плоскости отмечено несколько точек (см. приложенную картинку), при этом каждая точка принадлежит одному из двух классов.

Вася придумал решающее дерево, которое верно классифицирует эти точки по двум классам. При этом все вопросы в этом дереве имеют вид  $ax + by < c$  для некоторых действительных  $a, b, c$ ,  $(x, y)$  координаты классифицируемой точки).

Какое минимальное количество вершин с вопросами может быть в этом дереве?

### 13. Калибровка робота

Дан клетчатый лабиринт.

Робот может выполнять команды  $U, D, R, L$  - сдвинуться на одну клетку вверх, вниз, вправо, влево соответственно. Если эту команду нельзя выполнить (на пути стена или препятствие), то робот остается на месте.

Напишите программу, калибрующую робота в данном лабиринте: после исполнения программы робот должен обязательно оказаться в одной и той же клетке вне зависимости от того, с какой клетки он стартовал.

Длина программы не должна превышать 100.

### 14. Где какая картинка

Вася решает задачу разделения двух кластеров точек на плоскости с помощью SVM.

Он рассматривает точки на плоскости как пары чисел  $(x, y)$ , затем, добавляя дополнительные признаки, выходит в пространство большей размерности и разделяет кластеры гиперплоскостью с помощью SVM, а затем рисует полученную разделяющую поверхность на плоскости.

Он делал это 4 разными способами, сопоставьте способы с полученными картинками:

- 1)  $(x, y) \rightarrow (x, y, 1)$
- 2)  $(x, y) \rightarrow (x, y, x^2)$
- 3)  $(x, y) \rightarrow (x, y, x^2, xy, y^2)$
- 4)  $(x, y) \rightarrow (x, y, |x|, |y|)$

### 15. Асимптотика

Отметьте все верные утверждения и только их:

- 1)  $n = O(n^2)$
- 2)  $2^n + n^5 = O(3^n)$
- 3)  $\log n = O(\log^2 n)$
- 4)  $\log n^2 3^n = O(n^2)$

### 16. Казино

Два математика играют в такую азартную игру:

- 1) Каждый ставит по 100 долларов.
- 2) Оба кидают по одному игральному кубику. Оба видят результат бросков.
- 3) Далее первый может повысить ставку на любое число долларов  $x$ .
- 4) Второй может либо сдаться, либо принять ставку и продолжить игру.
- 5) Оба кидают по одному игральному кубику. Тот, у кого сумма за два броска получилась больше, выигрывает и получает ставки от обоих игроков. Если суммы равны, каждый забирает свои деньги.

Оба математика играют рационально. Пусть после первых бросков у первого выпало 6, у второго 5. Найдите такое минимальное целое  $x$ , чтобы при повышении ставки на  $x$  первым игроком второй игрок сдался.

### 17. Регрессия

Известно, что почти по любым 3 точкам на плоскости можно однозначно восстановить квадратный трехчлен. То есть для любых трех точек  $(x_1, y_1), (x_2, y_2), (x_3, y_3)$ , таких что  $x_i \neq x_j$ , можно найти единственный  $f(x) = ax^2 + bx + c$  такой, что  $f(x_i) = y_i$ .

Вася узнал этот факт и теперь хочет решить задачу регрессии для набора из 10000 точек  $\{(i, i^3) \mid 1 \leq i \leq 10000\}$  своим алгоритмом: он записал квадратный трехчлен для каждой тройки точек из набора, после этого сложил их все и поделил на их количество.

Найдите трехчлен, который получил Вася.

Формат ответа для трехчлена  $ax^2 + bx + c$ :  $a \ b \ c$  - три несократимые дроби, записанные через пробел.

## 18. kNN

Вася классифицирует точки на плоскости на два класса с помощью алгоритма kNN.

Размеченные данные представлены на картинке и продублированы списком.

Первый класс:  $(0, 0)$ ,  $(0, \frac{1}{2})$ ,  $(\frac{1}{2}, 0)$ ,  $(1, 0)$

Второй класс:  $(\frac{1}{2}, \frac{1}{2})$ ,  $(\frac{1}{2}, 1)$ ,  $(0, 1)$ ,  $(\frac{1}{2}, 1)$ ,  $(1, 1)$ .

Параметр  $k = 3$ , то есть точка  $X$  относится к классу, который чаще встречается среди ближайших к  $X$  3 точек из размеченных данных.

Найдите площадь множества точек из квадрата  $[0, 1]^2$ , которые будут классифицироваться как класс 1.

*Формат ответа: несократимая дробь  $p/q$ .*

## 19. Теория игр

Три эксперта по машинному обучению работают вместе над решением задачи.

Каждый из них сделал по одному прототипу, идеи которых между собой скомбинировать не получилось, и нужно выбрать один из трех прототипов, который они будут развивать вместе.

Прототип эксперта  $A$  показывает качество 0.9, прототип эксперта  $B$  показывает качество 0.85, а прототип эксперта  $C$  - 0.8.

Выбирать итоговый прототип они будут следующим образом: эксперты по очереди презентуют свои решения, и после каждой презентации устраивается голосование (каждый голосует За/Против), и если хотя бы 2 эксперта проголосовали За, то прототип побеждает, и дальнейшие презентации не проводятся. Эксперты выступают в порядке  $A, B, C$ , и если ни в одном голосовании никто не победит, то избирается прототип, который презентуется последним - прототип эксперта  $C$ .

Каждый эксперт в первую очередь хочет, чтобы победил его прототип, но если этого не случится, то он будет сильнее рад, если победит прототип с большим качеством. Чей прототип будет в итоге выбран?

## 20. Восстановление графа

Вася взял клетчатую бумагу и пометил буквами некоторые узлы сетки - точки с целыми координатами, и потом выписал себе список пар точек, расстояние между которыми меньше 2.

Прошло много времени, и Вася потерял эту картинку, но сохранил список пар меток. При этом он помнит, что меткой  $A$  была помечена точка  $(0, 0)$ , меткой  $B$  точка  $(5, 5)$ , и все помеченные точки  $(x, y)$  были таковы, что  $x \geq y$ .

Помогите Васе понять, какая точка была отмечена меткой  $C$ , если список следующий:  $(A, X)$ ,  $(T, X)$ ,  $(R, T)$ ,  $(R, U)$ ,  $(U, B)$ ,  $(T, S)$ ,  $(U, Z)$ ,  $(S, C)$ ,  $(C, G)$ ,  $(G, Z)$

*Формат ответа для точки  $(x, y)$ : два целых числа, разделенных пробелом  $x y$ .*

## 21. (8)

Ограничение времени	1 секунда
Ограничение памяти	64Mb
Ввод	стандартный ввод или input.txt
Вывод	стандартный вывод или output.txt

На острове живут  $n$  человек. Каждый из них либо рыцарь, либо лжец. Все рыцари всегда говорят правду, все лжецы всегда лгут.

Каждый человек сделал по одному утверждению.  $i$ -тый островитянин ( $1 \leq i \leq n$ ) сказал: "Среди всех людей на острове не более  $k_i$  рыцарей".

По заданным утверждениям островитян найдите любое возможное количество рыцарей на острове или определите, что такая ситуация невозможна.

### Формат ввода

Первая строка входных данных содержит одно целое число  $n$  ( $1 \leq n \leq 1000$ ) — количество жителей острова.


Во второй строке задаются  $n$  целых чисел  $k_i$  ( $0 \leq k_i \leq n$ ), разделенных пробелом.

### Формат вывода


Если описанная ситуация невозможна, то выведите  $-1$ .

Иначе выведите одно целое число — количество рыцарей, живущих на острове. Если есть несколько возможных ответов, разрешается вывести любой из них.

## Пример 1

**Ввод** 

2  
2 2


**Вывод** 

2

## Пример 2

**Ввод** 

4  
0 1 2 3

**Вывод** 

2

## 22. (8)

Ограничение времени	1 секунда
Ограничение памяти	64Mb
Ввод	стандартный ввод или input.txt
Вывод	стандартный вывод или output.txt

Васе нужно сгенерировать прямоугольники со сторонами, параллельными осям координат. Он решил попробовать сделать это двумя способами.

1. Выбрать координаты  $x$  для левой и правой границы прямоугольника следующим образом: выбрать два случайных числа равномерно из отрезка  $[0, 1]$ , меньшее взять за координату  $x$  левой стороны прямоугольника, а большее за координату  $x$  правой стороны.

Аналогично выбрать координаты  $y$  для нижней и верхней сторон прямоугольника.

Построить прямоугольник по  $x$ -координатам его левой и правой сторон и  $y$ -координатам его нижней и верхней сторон.

2. Случайно выбрать центр будущего прямоугольника  $(x_0, y_0)$  из равномерного распределения на квадрате  $[0, 1]^2$ .

Рассчитать максимальную высоту  $h_{max}$  и ширину  $w_{max}$  прямоугольника, при которых он не выйдет из квадрата  $[0, 1]^2$  при том, что пересечение его диагоналей будет в точке  $(x_0, y_0)$ .

Выбрать реальную высоту равномерно из отрезка  $[0, h_{max}]$ , а реальную ширину равномерно из отрезка  $[0, w_{max}]$ .

Построить прямоугольник с заданными центром, высотой и шириной.

Вася сгенерировал одним из этих способов 1000 прямоугольников, записал их на листок и пошел отдохнуть. После отдыха он понял, что забыл, каким способом он генерировал прямоугольники.

Помогите ему понять, каким способом были сгенерированы прямоугольники.



## Формат ввода

В каждом тесте ровно 1000 строк.

На каждой из 1000 строк заданы 4 действительных числа из отрезка  $[0; 1]$ , записанные через пробел не более чем с 20 знаками после запятой:  $a, b, c, d$ . Они задают очередной прямоугольник  $\{(x, y) \mid a \leq x \leq b, c \leq y \leq d\}$ .

## Формат вывода

Выведите ровно одно число:

1 — если прямоугольники были сгенерированы первым способом;

2 — если прямоугольники были сгенерированы вторым способом.

## Примечания

Строки во входных данных выглядят так:

```
0.33388755706372086 0.3745512117320299 0.023735508485224266 0.1860715165653241  
0.7584918624144189 0.8981769737352442 0.6299305145914222 0.9853909408922319  
0.07179403452745871 0.16134669863190043 0.04432664205527381 0.8574621224696689
```

Тестовые примеры в условии не приведены, поскольку по маленьким выборкам невозможно надежно определить, каким способом были сгенерированы прямоугольники.

## 23. (8)

Ограничение времени	1 секунда
Ограничение памяти	64Mb
Ввод	стандартный ввод или input.txt
Вывод	стандартный вывод или output.txt

Вася узнал про решающие деревья и решил применить их в своей задаче классификации. Вася классифицирует точки на плоскости по двум классам и хочет построить решающее дерево. Первым шагом он решил подобрать такие целые числа  $a$  и  $b$ , что правило вида  $x \leq a$  AND  $y \leq b$  будет хорошим правилом для определения того, принадлежит ли точка первому классу.

Вася придумал два критерия, по которым он будет определять качество такого правила. Во-первых, правилу должны удовлетворять не больше 5 точек из второго класса. Во-вторых, из всех правил, удовлетворяющих первому условию, Вася выберет такое, что максимально возможное число точек первого класса подойдет под правило. Если таких правил несколько, то Вася выберет среди них правило с минимальным  $a$ . Если и таких правил несколько, то Вася выберет среди них правило с минимальным  $b$ .

Помогите Васе найти оптимальные целые числа  $a, b$  по известному набору точек.

## Формат ввода

В первой строке задано единственное целое число  $N$  — количество точек в датасете ( $1 \leq N \leq 10^5$ ).

Далее идет  $N$  строк, на каждой из которых содержится описание одной точки — три целых числа  $x, y, label$  ( $1 \leq x, y \leq 10^5$ ,  $label \in \{1, 2\}$ ), разделенные пробелом: координаты очередной точки и метка класса, которому она принадлежит.


## Формат вывода

На единственной строке выведите два целых числа  $a, b$ , разделенные пробелом.


Обратите внимание, что если оптимальных по количеству точек первого класса выборов несколько, то среди них нужно вывести лексикографически минимальную пару  $(a, b)$ .

Гарантируется, что при оптимальном выборе  $(a, b)$  хотя бы одна точка первого типа попадет под это правило.

## Пример

Ввод 

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

Вывод 

```
1 2
```

## 24. (8)

Вам даны синтетические данные вида  $(\mathbf{x}, f(\mathbf{x}))$ , где  $\mathbf{x}$  — вектор из 5 действительных чисел, а  $f(\mathbf{x})$  — действительное число.

Вам нужно решить задачу регрессии, обучившись на тренировочном датасете, и предсказать значение  $f(\mathbf{x})$  для всех  $\mathbf{x}$  из тестового датасета.

Тренировочный датасет состоит из 80000 строк  $id, \mathbf{x}_{id}, f(\mathbf{x}_{id})$  и заголовка. Ссылка на данные в формате csv: [https://yadi.sk/d/\\_zDavEQwLTq-\\_tQ](https://yadi.sk/d/_zDavEQwLTq-_tQ)

Тестовый датасет состоит из 20000 строк  $id, \mathbf{x}_{id}$  и заголовка. Ссылка на данные в формате csv: [https://yadi.sk/d/FVrLHez5\\_8UHxw](https://yadi.sk/d/FVrLHez5_8UHxw)

Отправленный файл должен состоять из 20000 строк вида  $id, f_{predict}(\mathbf{x}_{id})$  и заголовка  $id, y$ . Ссылка на пример: [https://yadi.sk/d/6-\\_Wz9YjJr561Gw](https://yadi.sk/d/6-_Wz9YjJr561Gw)

Метрика качества **MAE**: ошибка вычисляется по формуле  $S = \sum |f_{predicted}(\mathbf{x}_{id}) - f_{real}(\mathbf{x}_{id})|$ .

Баллы за посылку проставляются по такому правилу:

$S \leq 20000$  - 8 баллов

$S \leq 30000$  - 7 баллов

$S \leq 50000$  - 6 баллов

$S \leq 75000$  - 5 баллов

$S \leq 100000$  - 4 балла

$S \leq 200000$  - 3 балла

$S \leq 500000$  - 2 балла

$S \leq 1000000$  - 1 балл

# 25. (8)

**Задача предоставлена партнером олимпиады — лабораторией инноватики МФТИ.**

Дан датасет с картинками.

Картинки представляют из себя однотонный фон и некоторое количество разноцветных кругов (круги могут быть одного цвета).

**Размер** каждой картинки — **620 × 480** пикселей.

Ваша цель — научиться считать количество кругов на картинке.

Полный датасет разбит на 4 части. При генерации картинок были выполнены два условия:

1. Центр каждого круга находится в квадрате  $[0; 1]^2$ .
2. Радиус каждого круга лежит в промежутке  $[0.015; 0.165]$ .

Круг может обрезаться границей картинки и даже в этом случае нужно уметь его находить.

Далее описаны особенности генерации картинок в каждой из частей.

## *simple*

На каждой картинке случайное число кругов от 10 до 50.

Любые два круга не пересекаются.

## *normal*

На каждой картинке случайное число кругов от 20 до 60.

Круги могут пересекаться, но не очень сильно. Формально:

- а) Расстояние между центрами не меньше чем  $0.55 * \text{сумма радиусов}$ ;
- б) Нет двух кругов, таких что центр одного лежит внутри другого.

## *hard*

На каждой картинке случайное число кругов от 20 до 60.

Любые два круга либо не пересекаются, либо же один лежит строго внутри другого.

## *random*

На каждой картинке случайное число кругов от 20 до 70.

Центры и радиусы кругов выбираются случайно равномерно на описанных выше множествах.

## **Метрика**

Каждая из **2000** картинок вносит в общие очки за посылку свой вклад — действительное число от **0** до **1**.

Вклад за картинку с *target* кругами при предсказании *predicted* равен:

$$\max\left(1 - \frac{|target - predicted|}{10}, 0\right).$$

Баллы за посылку с *submission\_score* очками начисляются по следующей формуле:

$$\left[ 8 \cdot \frac{\textit{submission\_score}}{\textit{max\_score}} \right].$$

Где *max\_score* — максимальные очки по всем посылкам всех участников. В течение конкурса эта величина равна **2000.0**, а по окончании конкурса вычислится ее реальное значения и баллы всех посылок будут пересчитаны.

## Формат ввода

Тренировочный датасет состоит из **4** частей по **1000** картинок:

[https://yadi.sk/d/-\\_vxLaI5wGpk6Fg](https://yadi.sk/d/-_vxLaI5wGpk6Fg)

В каждой из 4 папок (*simple, normal, hard, random*) по **1000** пронумерованных картинок и *.csv* файл с разметкой.

Тестовый датасет состоит из **4** частей по **500** картинок:

<https://yadi.sk/d/EjCK1UdhVzTzMQ>

В каждой из 4 папок (*simple, normal, hard, random*) по **1000** пронумерованных картинок.

**Внимание!** Большой размер данных — **500** мб. Советуем скачивать данные порциями, и по возможности заранее.

## Формат вывода

В ответ введите 2004 строки:

4 блока по 501 строке в порядке *simple*, *normal*, *hard*, *random*. Первая строка в блоке — название блока, следующие 500 строк номер картинки ( $0 \leq i \leq 499$ ) и предсказанное количество кругов на  $i$ -той картинке. Все предсказания должны быть целыми числами.

*simple*

0      *answer*<sub>0</sub>

...      ...

499    *answer*<sub>499</sub>

*normal*

0      *answer*<sub>0</sub>

...      ...

499    *answer*<sub>499</sub>

*hard*

0      *answer*<sub>0</sub>

...      ...

499    *answer*<sub>499</sub>

*random*

0      *answer*<sub>0</sub>

...      ...

499    *answer*<sub>499</sub>