

# **Задания отборочного этапа олимпиады школьников СПбГУ по математическому моделированию и искусственному интеллекту 9-11 классы**

Вариант задания отборочного этапа Олимпиады школьников СПбГУ по математическому моделированию и искусственному интеллекту 2023/24 учебного года состоял из трёх задач разных типов в рамках тематики Олимпиады. Каждая задача оценивалась определенным количеством баллов в зависимости от уровня сложности.

Наибольшая итоговая сумма баллов, которой оценивались ответы всех задач, была равна 100. Подсчет итоговой оценки осуществлялся путем суммирования баллов, выставленных за ответы каждой из задач. Оценка ответов по каждой задаче формировалась в зависимости от близости введённого ответа к точному ответу.

## Задача 1: Параметрическое моделирование

### Формулировка (Вариант 1)

Уравнение ван дер Ваальса описывает то, как давление  $P$  (в паскалях) некоторого газа с количеством вещества  $\nu$  моль зависит от объёма  $V$  (в  $\text{м}^3$ ) в ходе изотермического процесса при температуре  $T$  (в кельвинах):

$$\left(P + \frac{a\nu^2}{V^2}\right)(V - \nu b) = \nu R T,$$

где  $R$  — универсальная газовая постоянная,  $a$ ,  $b$  постоянные ван дер Ваальса, которые различны для различных газов (см. таблицу с газами).

| Название газа    | $a$ , $\text{Па} \cdot \text{м}^6 \cdot \text{моль}^{-2}$ | $b$ , $\text{м}^3 \cdot \text{моль}^{-1}$ |
|------------------|---|---|
| Ацетилен         | 0.4516  | $52.2 \cdot 10^{-6}$                      |
| Криптон          | 0.2349  | $39.78 \cdot 10^{-6}$                     |
| Бромоводород     | 0.451   | $44.31 \cdot 10^{-6}$                     |
| Пропан           | 0.8779  | $84.45 \cdot 10^{-6}$                     |
| Фторид кремния   | 0.4251  | $55.71 \cdot 10^{-6}$                     |
| Моносилан        | 0.4377  | $57.86 \cdot 10^{-6}$                     |
| Ксенон           | 0.425   | $51.05 \cdot 10^{-6}$                     |
| Фторметан        | 0.4692  | $52.64 \cdot 10^{-6}$                     |
| Аммиак           | 0.4225  | $37.1 \cdot 10^{-6}$                      |
| Сероводород      | 0.449   | $42.87 \cdot 10^{-6}$                     |
| Диоксид углерода | 0.364   | $42.67 \cdot 10^{-6}$                     |
| Хлор             | 0.6579  | $56.22 \cdot 10^{-6}$                     |
| Циан             | 0.7769  | $69.01 \cdot 10^{-6}$                     |

Над некоторым газом с количеством вещества  $\nu = 1$  моль при температуре  $T = 350 \text{ K}$  был проведён эксперимент. В ходе эксперимента для различных объёмов газа  $V$  было измерено давление газа  $P$  в паскалях. Эти данные представлены в таблице данных (см. таблицу ниже). В первом столбце таблицы расположены объёмы от  $V_1 = 10^{-4} \text{ м}^3$  до  $V_{21} = 4 \cdot 10^{-4} \text{ м}^3$ , во втором столбце соответствующие измеренные значения давления от  $P_1$  до  $P_{21}$  в атмосферах ( $1 \text{ атм} = 101325 \text{ Па}$ ).

Требуется определить постоянные ван дер Ваальса  $a$ ,  $b$ , при которых обеспечивается наименьшая квадратичная ошибка между набором давлений ( $P_1, \dots, P_{21}$ ) в атмосферах, полученных в ходе эксперимента, и набором соответствующих теоретических значений давления ( $P(V_1, a, b), \dots, P(V_{21}, a, b)$ ) в атмосферах, найденных из уравнения ван дер Ваальса.

Квадратичная ошибка вычисляется так:

$$\text{Error}(a, b) = \sum_{k=1}^{21} (P_k - P(V_k, a, b))^2.$$

В качестве ответа выберите из выпадающего списка название газа, над которым проводился эксперимент, и укажите найденное значение наименьшей квадратичной ошибки на

данных с точностью до 2-х знаков после запятой.

Таблица данных:

| $V, \text{ м}^3$     | $P, \text{ атм}$ |
|----------------------|------------------|
| $1. \cdot 10^{-4}$   | 150.29           |
| $1.15 \cdot 10^{-4}$ | 118.64           |
| $1.3 \cdot 10^{-4}$  | 98.23            |
| $1.45 \cdot 10^{-4}$ | 88.78            |
| $1.6 \cdot 10^{-4}$  | 80.75            |
| $1.75 \cdot 10^{-4}$ | 89.45            |
| $1.9 \cdot 10^{-4}$  | 77.45            |
| $2.05 \cdot 10^{-4}$ | 80.86            |
| $2.2 \cdot 10^{-4}$  | 75.21            |
| $2.35 \cdot 10^{-4}$ | 71.82            |
| $2.5 \cdot 10^{-4}$  | 76.27            |
| $2.65 \cdot 10^{-4}$ | 68.59            |
| $2.8 \cdot 10^{-4}$  | 71.32            |
| $2.95 \cdot 10^{-4}$ | 72.39            |
| $3.1 \cdot 10^{-4}$  | 66.19            |
| $3.25 \cdot 10^{-4}$ | 57.32            |
| $3.4 \cdot 10^{-4}$  | 57.7             |
| $3.55 \cdot 10^{-4}$ | 62.04            |
| $3.7 \cdot 10^{-4}$  | 62.93            |
| $3.85 \cdot 10^{-4}$ | 59.12            |
| $4. \cdot 10^{-4}$   | 50.37            |

**Ответ:**

правильный газ — это фторметан.

Значение наименьшей квадратичной ошибки на данных: 332.056

### Формулировка (Вариант 2)

Уравнение ван дер Ваальса описывает то, как давление  $P$  (в паскалях) некоторого газа с количеством вещества  $\nu$  моль зависит от объёма  $V$  (в  $\text{м}^3$ ) в ходе изотермического процесса при температуре  $T$  (в кельвинах):

$$\left(P + \frac{a \nu^2}{V^2}\right)(V - \nu b) = \nu R T,$$

где  $R$  — универсальная газовая постоянная,  $a$ ,  $b$  постоянные ван дер Ваальса, которые различны для различных газов (см. таблицу с газами).

Out[ ]=

| Название газа    | $a, \text{ Па} \cdot \text{ м}^6 \cdot \text{ моль}^{-2}$ | $b, \text{ м}^3 \cdot \text{ моль}^{-1}$ |
|------------------|---|--|
| Ацетилен         | 0.4516  | $52.2 \cdot 10^{-6}$                     |
| Криптон          | 0.2349  | $39.78 \cdot 10^{-6}$                    |
| Бромоводород     | 0.451   | $44.31 \cdot 10^{-6}$                    |
| Пропан           | 0.8779  | $84.45 \cdot 10^{-6}$                    |
| Фторид кремния   | 0.4251  | $55.71 \cdot 10^{-6}$                    |
| Моносилан        | 0.4377  | $57.86 \cdot 10^{-6}$                    |
| Ксенон           | 0.425   | $51.05 \cdot 10^{-6}$                    |
| Фторметан        | 0.4692  | $52.64 \cdot 10^{-6}$                    |
| Аммиак           | 0.4225  | $37.1 \cdot 10^{-6}$                     |
| Сероводород      | 0.449   | $42.87 \cdot 10^{-6}$                    |
| Диоксид углерода | 0.364   | $42.67 \cdot 10^{-6}$                    |
| Хлор             | 0.6579  | $56.22 \cdot 10^{-6}$                    |
| Циан             | 0.7769  | $69.01 \cdot 10^{-6}$                    |

Над некоторым газом с количеством вещества  $\nu = 1$  моль при температуре  $T = 350 \text{ K}$  был проведён эксперимент. В ходе эксперимента для различных объёмов газа  $V$  было измерено давление газа  $P$  в паскалях. Эти данные представлены в таблице данных (см. таблицу ниже). В первом столбце таблицы расположены объёмы от  $V_1 = 10^{-4} \text{ м}^3$  до  $V_{21} = 4 \cdot 10^{-4} \text{ м}^3$ , во втором столбце соответствующие измеренные значения давления от  $P_1$  до  $P_{21}$  в атмосферах ( $1 \text{ атм} = 101325 \text{ Па}$ ).

Требуется определить постоянные ван дер Ваальса  $a, b$ , при которых обеспечивается наименьшая квадратичная ошибка между набором давлений ( $P_1, \dots, P_{21}$ ) в атмосферах, полученных в ходе эксперимента, и набором соответствующих теоретических значений давления ( $P(V_1, a, b), \dots, P(V_{21}, a, b)$ ) в атмосферах, найденных из уравнения ван дер Ваальса.

Квадратичная ошибка вычисляется так:

$$\text{Error}(a, b) = \sum_{k=1}^{21} (P_k - P(V_k, a, b))^2.$$

В качестве ответа выберите из выпадающего списка название газа, над которым проводился эксперимент, и укажите найденное значение наименьшей квадратичной ошибки на данных с точностью до 2-х знаков после запятой.

Таблица данных:

| V, м <sup>3</sup>       | P, атм |
|-------------------------|--------|
| 1. · 10 <sup>-4</sup>   | 173.89 |
| 1.15 · 10 <sup>-4</sup> | 140.71 |
| 1.3 · 10 <sup>-4</sup>  | 115.71 |
| 1.45 · 10 <sup>-4</sup> | 108.33 |
| 1.6 · 10 <sup>-4</sup>  | 103.18 |
| 1.75 · 10 <sup>-4</sup> | 98.37  |
| 1.9 · 10 <sup>-4</sup>  | 94.31  |
| 2.05 · 10 <sup>-4</sup> | 85.22  |
| 2.2 · 10 <sup>-4</sup>  | 80.66  |
| 2.35 · 10 <sup>-4</sup> | 89.15  |
| 2.5 · 10 <sup>-4</sup>  | 79.12  |
| 2.65 · 10 <sup>-4</sup> | 70.19  |
| 2.8 · 10 <sup>-4</sup>  | 70.65  |
| 2.95 · 10 <sup>-4</sup> | 59.12  |
| 3.1 · 10 <sup>-4</sup>  | 59.45  |
| 3.25 · 10 <sup>-4</sup> | 66.6   |
| 3.4 · 10 <sup>-4</sup>  | 68.27  |
| 3.55 · 10 <sup>-4</sup> | 61.68  |
| 3.7 · 10 <sup>-4</sup>  | 66.1   |
| 3.85 · 10 <sup>-4</sup> | 60.41  |
| 4. · 10 <sup>-4</sup>   | 47.02  |

Out[\*]=

**Ответ:**

правильный газ — это ксенон.

Значение наименьшей квадратичной ошибки на данных: 498.996

## Задача 2: Кластеризация данных

### Формулировка (Вариант 1)

Витя тщательно собирал записи об 11700 объектах, каждая запись об объекте представляла собой строку из 4 чисел, являющихся измерениями некоторых параметров объекта, и метку класса, к которому этот объект принадлежит. Каждый объект принадлежит одному из 4 классов.

Перед сохранением этой базы данных Витя случайно удалил столбец с метками классов. В итоге, сохранилась только матрица размера 11700 на 4 с числовыми значениями. Файл с данными находится по ссылке: "UnlabeledVar1.xlsx"

При составлении этой базы данных Витя заметил, что роль разных параметров в определении класса объекта не одинакова и некоторые параметры позволяют весьма точно определить класс объекта. Помогите Вите восстановить классификацию объектов по числовой матрице данных.

В качестве ответа введите количество объектов того класса, к которому принадлежит самая первая запись в числовой матрице:

(1.473619495, -1.467503773, 0.80425779, 0.170893091).

Ответом должно быть целое число.

**Ответ:**

2250

### Формулировка (Вариант 2)

Витя тщательно собирал записи об 11700 объектах, каждая запись об объекте представляла собой строку из 4 чисел, являющихся измерениями некоторых параметров объекта, и метку класса, к которому этот объект принадлежит. Каждый объект принадлежит одному из 4 классов.

Перед сохранением этой базы данных Витя случайно удалил столбец с метками классов. В итоге, сохранилась только матрица размера 11700 на 4 с числовыми значениями. Файл с данными находится по ссылке: "UnlabeledVar2.xlsx"

При составлении этой базы данных Витя заметил, что роль разных параметров в определении класса объекта не одинакова и некоторые параметры позволяют весьма точно определить класс объекта. Помогите Вите восстановить классификацию объектов по числовой матрице данных.

В качестве ответа введите количество объектов того класса, к которому принадлежит самая первая запись в числовой матрице:

(1.43909236, -0.219649159, 1.094699851, -0.62520536)

Ответом должно быть целое число.

**Ответ:**

3550

## Задача 3: Анализ текстовых данных

### Формулировка (Вариант 1)

При формировании текстовой базы данных для обучения модели, выделяющей отзывы из текстовых сообщений, дата-аналитик Василий решил протестировать навыки своих коллег в анализе текстовых данных. Для этого он сформировал набор текстовых данных, состоящий из двух типов текстов: новостные статьи и отзывы на отели. Файл с текстами находится по ссылке "TextsUnlabeledVar1.xlsx"

Задача состоит в том, чтобы выделить отзывы на отели из этого набора текстов. Предложите свой способ решения этой задачи.

В качестве ответа напишите оценку количества отзывов на отели в этой базе данных. Ответом должно быть целое число.

**Ответ:**

7750

### Формулировка (Вариант 2)

При формировании текстовой базы данных для обучения модели, выделяющей отзывы из текстовых сообщений, дата-аналитик Василий решил протестировать навыки своих коллег в анализе текстовых данных. Для этого он сформировал набор текстовых данных, состоящий из двух типов текстов: новостные статьи и отзывы на отели. Файл с текстами находится по ссылке "TextsUnlabeledVar2.xlsx"

Задача состоит в том, чтобы выделить отзывы на отели из этого набора текстов. Предложите свой способ решения этой задачи.

В качестве ответа напишите оценку количества отзывов на отели в этой базе данных. Ответом должно быть целое число.

**Ответ:**

6250