

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 2024/25 учебного года состоял из трёх задач разных типов в рамках тематики Олимпиады и 5 тестовых заданий на анализ данных. 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 в соответствии с критериями оценки.

Задача 1: Моделирование виртуальной среды

Формулировка (Вариант 1)

Артур запрограммировал симуляцию естественного отбора. В этой виртуальной “эко-системе” различные кибер-организмы, обладающие набором меняющихся характеристик, приспособляются и выживают в окружающей их кибер-среде. Время в кибер-среде меняется дискретно, будем измерять его в условных днях.

Наблюдая за развитием одного из кибер-видов, он заметил, что организмы этого вида живут 3 дня. Всю популяцию этого вида Артур разбил на 3 части.

1. Группа 1 состоит из организмов возрастом 0 дней.
2. Группа 2 состоит из организмов возрастом 1 день.
3. Группа 3 состоит из организмов возрастом 2 дня.

При смене дня организмы каждой возрастной группы либо переходят в следующую группу, либо погибают. А именно, по прошествии 1 дня:

1. 80% организмов из Группы 1 перейдёт в Группу 2, остальные погибают;
2. 70% организмов из Группы 2 перейдёт в Группу 3, остальные погибают;
3. 100% организмов Группы 3 погибает.

Все возрастные группы каждый день производят потомство.

1. Коэффициент рождаемости Группы 1 равен 0.3;
2. Коэффициент рождаемости Группы 2 равен 0.9;
3. Коэффициент рождаемости Группы 3 равен 0.1.

В 1-ый день наблюдения было:

1. 1000 организмов Группы 1;
2. 2000 организмов Группы 2;
3. 500 организмов Группы 3.

Оцените количество организмов Группы 1 на 50-ый день наблюдения. Ответ округлите до целых.

Ответ:

12560

Формулировка (Вариант 2)

Артур запрограммировал симуляцию естественного отбора. В этой виртуальной “эко-системе” различные кибер-организмы, обладающие набором меняющихся характеристик, приспосабливаются и выживают в окружающей их кибер-среде. Время в кибер-среде меняется дискретно, будем измерять его в условных днях.

Наблюдая за развитием одного из кибер-видов, он заметил, что организмы этого вида живут 3 дня. Всю популяцию этого вида Артур разбил на 3 части.

1. Группа 1 состоит из организмов возрастом 0 дней.
2. Группа 2 состоит из организмов возрастом 1 день.

3. Группа 3 состоит из организмов возрастом 2 дня.

При смене дня организмы каждой возрастной группы либо переходят в следующую группу, либо погибают. А именно, по прошествии 1 дня:

1. 85% организмов из Группы 1 перейдёт в Группу 2, остальные погибают;
2. 80% организмов из Группы 2 перейдёт в Группу 3, остальные погибают;
3. 100% организмов Группы 3 погибает.

Все возрастные группы каждый день производят потомство.

1. Коэффициент рождаемости Группы 1 равен 0.25;
2. Коэффициент рождаемости Группы 2 равен 0.85;
3. Коэффициент рождаемости Группы 3 равен 0.15.

В 1-ый день наблюдения было:

1. 1000 организмов Группы 1;
2. 2000 организмов Группы 2;
3. 500 организмов Группы 3.

Оцените количество организмов Группы 1 на 50-ый день наблюдения. Ответ округлите до целых.

Ответ:

10552

Задача 2: Построение линейной модели

Формулировка (Вариант 1)

Борис открыл свой бизнес по 3D-печати маленьких фигурок и различных деталей. С 2020 года данные об объёме продаж Борис почти каждый день записывал в файл (некоторые дни оказались не заполненными). Данные хранятся в виде пар $(\text{day}_i, \text{sales}_i)$, $i = 1, \dots, M$, где day_i — это i -ый день с 1 января 2020 года, а sales_i — это объём продаж за i -ый день. Бизнес за 4 года активно развивался и Борис решил узнать средний прирост продаж за месяц.

Для этого он построил линейную модель данных продаж, которая имеет вид $f(t) = a * t + b$, где t - это номер дня с 1 января 2020 года. Параметры a, b Борис подобрал так, чтобы среднеквадратичная ошибка модели на данных была минимальной, то есть следующая функция имеет наименьшее значение

$$I(a, b) = \frac{1}{M} \sum_{t=1}^M (\text{sales}_t - (a t + b))^2$$

Найденный параметр a будет означать средний прирост продаж за день согласно построенной модели.

В качестве ответа запишите прирост продаж за месяц (30 дней) с точностью до 1 знака после запятой.

Данные находятся в файле “sales.csv” и представляют собой массив размера 1674×2 . В первом столбце содержатся даты, во втором - объёмы продаж.

Ответ:

5.7

Формулировка (Вариант 2)

Борис открыл свой бизнес по 3D-печати маленьких фигурок и различных деталей. С 2020 года данные об объёме продаж Борис почти каждый день записывал в файл (некоторые дни оказались не заполненными). Данные хранятся в виде пар $(\text{day}_i, \text{sales}_i)$, $i = 1, \dots, M$, где day_i — это i -ый день с 1 января 2020 года, а sales_i — это объём продаж за i -ый день. Бизнес за 4 года активно развивался и Борис решил узнать средний прирост продаж за месяц.

Для этого он построил линейную модель данных продаж, которая имеет вид $f(t) = a * t + b$, где t - это номер дня с 1 января 2020 года. Параметры a, b Борис подобрал так, чтобы среднеквадратичная ошибка модели на данных была минимальной, то есть следующая функция имеет наименьшее значение

$$I(a, b) = \frac{1}{M} \sum_{t=1}^M (\text{sales}_t - (a t + b))^2$$

Найденный параметр a будет означать средний прирост продаж за день согласно построенной модели.

В качестве ответа запишите прирост продаж за месяц (30 дней) с точностью до 1 знака после

запятой.

Данные находятся в файле “sales2.csv” и представляют собой массив размера 1674×2 . В первом столбце содержатся даты, во втором - объёмы продаж.

Ответ:

4.5

Задача 3: Анализ изображений

Формулировка (Вариант 1)

У Андрея был цифровой телескоп, с помощью которого он фотографировал Луну на протяжении своего обучения в школе. Наконец, он решил выяснить количество фотографий Луны в момент полнолуния.

Фотографии Луны хранятся в формате PNG, они имеют размер 128 на 128 пикселей и находятся в архиве “moon.zip”.

Придумайте и реализуйте алгоритм, позволяющий найти количество фотографий, изображающих полную Луну. Ответом должно быть целое число.

Ответ:

1146

Формулировка (Вариант 2)

У Андрея был цифровой телескоп, с помощью которого он фотографировал Луну на протяжении своего обучения в школе. Наконец, он решил выяснить количество фотографий Луны в момент полнолуния.

Фотографии Луны хранятся в формате PNG, они имеют размер 128 на 128 пикселей и находятся в архиве “moon2.zip”.

Придумайте и реализуйте алгоритм, позволяющий найти количество фотографий, изображающих полную Луну. Ответом должно быть целое число.

Ответ:

1210

Задания на анализ данных

Формулировки и решения

Файл “flats.csv” содержит таблицу данных с характеристиками и стоимостью квартир в различных районах города N.

1. Отметьте, в каких столбцах таблицы данных есть пропущенные значения.

Ответ: ‘этаж’, ‘метро_км’

2. Укажите число квартир, цена которых меньше средней цены квартиры.

Ответ: 19329

3. Отметьте три признака, которые имеют наиболее высокую корреляцию с признаком “доп_образование_район”

Ответ: число_школ_район, число_спорт_объектов_район, торговые_центры_район

4. Найти количество квартир, у которых площадь больше или равна 70 м², а расстояние до метро меньше 1 км.

Ответ: 1736

5. Укажите значение цены самой дешёвой квартиры, у которой площадь больше или равна 90 м², а расстояние до КАД меньше 1 км.

Ответ: 990000

6. Укажите значение цены самой дешёвой квартиры, расположенной в районе с наибольшим суммарным числом школ, спортивных объектов, объектов дополнительного образования и торговых центров.

Ответ: 19329

7. Укажите значение цены самой дешёвой квартиры, у которой суммарное расстояние до детского сада, школы, парка и КАДа меньше 2 км.

Ответ: 990000

8. Укажите значение средней цены квадратного метра среди квартир, у которых расстояние до центра меньше 3 км.

Ответ: 156180.47

9. Найти количество квартир, у которых расстояние до КАД в 2 раза больше, чем сумма расстояний до школы и до метро.

Ответ: 14024