

Черновик БВ: Большие данные, искусственный интеллект, финансовые технологии и машинное обучение

1 апр 2024 г., 00:00 — 1 май 2024 г., 23:59

№ 1 — 3

0.25 баллов

Квадратный трёхчлен $x^2 - 12x + 7$ имеет корни x_1 и x_2 . Чему равно $\frac{1}{x_1} + \frac{1}{x_2}$?

Ответ введите в виде дроби.

Число или дробь

0.25 баллов

Чему равно $x_1^2 + x_2^2$?

Число или дробь

0.5 баллов

Найдите $x_1^3 + x_2^3$.

Число или дробь

№ 4 — 5

0.25 баллов

Есть набор из 4 карточек с буквами «Д», «И», «М» и «А». Какова вероятность, что при случайном последовательном извлечении карточек из набора получится выложить слово «ДИМА»?

Ответ укажите в виде дроби.

Число или дробь

0.25 баллов

Допустим, что в наборе карточек представлены буквы «К», «А», «Р», «Т», «И», «Н» и «А». Какова вероятность в аналогичном эксперименте получить слово «КАРТИНА»?

Ответ введите в виде дроби.

Число или дробь

№ 6

0.5 баллов

Дима два раза подбросил шестигранный игральный кубик. Найдите вероятность того, что сумма величин, выпавших в результате подбрасываний, является простым числом.

Ответ укажите в виде дроби.

Число или дробь

№ 7 — 9

0.25 баллов

В классе учится 17 школьников: 8 мальчиков и 9 девочек. Сколькими способами из этого класса можно выбрать пару, состоящую из одного мальчика и одной девочки?

Число

0.25 баллов

В классе учится 17 школьников: 8 мальчиков и 9 девочек. Сколькими способами из этого класса можно выбрать трёх школьников для дежурства по школе?

Число

0.5 баллов

В классе учится 17 школьников: 8 мальчиков и 9 девочек. Сколькими способами из этого класса можно выбрать команду по волейболу из 5 человек, состоящую из трёх мальчиков и двух девочек?

Число

№ 10

0.5 баллов

Петя, Вася и Дима хотят распределить между собой 10 одинаковых яблок. Сколькими способами они могут это сделать?

В результате делёжки должны быть распределены *все* яблоки. Также возможна ситуация, при которой в результате делёжки кому-то не достанется ни одного яблока.

Число

№ 11

0.5 баллов

Сколькими способами можно расставить двух ладей и ферзя на шахматную доску таким образом, чтобы каждая ладья била ферзя?

Число

№ 12 — 13

1 балл

Дан Python-код функции, которая должна находить максимальное значение в двумерном массиве. Однако в коде содержатся некоторые ошибки.

На вход функции передаётся список списков целых чисел. Гарантируется, что входной список не пустой, а также что все внутренние списки не пустые и имеют одинаковую длину.

```
1:  def find_max(arr):
2:      found_max = float('inf')
3:
4:      for i in range(len(arr)):
5:          for j in range(len(arr[-1])):
6:              if arr[j][i] > found_max:
7:                  found_max = arr[i][j]
8:                  break
9:
10:     return found_max
```

Найдите все ошибки в реализации функции. Всего их 3.

В качестве ответа укажите номера строчек функции, в которых они допущены. Номера можно указывать в произвольном порядке. Нумерация строчек начинается с 1.

Число

Число

Число

1 балл

Дан Python-код функции, которая должна вычислять количество непустых строчек в файле. Однако в коде содержатся некоторые ошибки.

На вход функции подаётся путь до файла, который нужно обработать. Гарантируется, что данный файл существует и путь до него указан корректно.

Строчка называется непустой, если содержит символы, отличные от пробельных. Для проверки того, что все символы в строке являются пробельными, в Python удобно использовать метод `isspace`, который определён для объектов типа `str`.

```
1: def not_null_lines(file_path):
2:     cnt = 0
3:
4:     with open(file_path, 'w') as f:
5:         for line in f:
6:             if line.isspace():
7:                 pass
8:
9:             cnt += 1
10:
11:     return cnt
```

Найдите все ошибки в реализации функции. Всего их 2.

В качестве ответа укажите номера строчек функции, в которых они допущены. Номера можно указывать в произвольном порядке. Нумерация строчек начинается с 1.

Число

Число

№ 14 — 16

0.5 баллов

Матрица связности графа — матрица размера $N \times N$ (где N — число вершин в графе), в каждой ячейке которой стоит либо 1, либо 0. Значение 1 на пересечении строки i и столбца j соответствует тому, что между соответствующими вершинами графа (нумерация вершин начинается с 0) существует ребро.

Дан неориентированный граф, который задаётся матрицей связности, записанной в файл (📎 `graph.txt`). Строка в файле соответствует строке матрицы связности. Значения в строке файла разделяются с помощью запятой («,»).

Научитесь считывать в память компьютера матрицу связности.

В качестве ответа на задание укажите число связей у вершины графа, которая имеет наибольшее число связей.

Число

0.5 баллов

Дано несколько пар вершин графа. Для каждой из пар укажите, существует ли между ними путь в графе. **Важно**, что нумерация вершин в графе начинается с 0.

Номер первой вершины	Номер второй вершины	Есть путь	
81	52	☐ Да	☐ Нет
86	89	☐ Да	☐ Нет
21	12	☐ Да	☐ Нет
16	43	☐ Да	☐ Нет
1	6	☐ Да	☐ Нет

1 балл

Компонентой связности графа называется такое максимальное по включению подмножество его вершин, что между каждой парой из них существует путь. Компонент связности в графе может быть несколько, и они могут иметь разный размер (в частности, состоять только из одной вершины).

Найдите число компонент связности в представленном в задании графе.

Число

№ 17 — 18

0.5 баллов

В школе у Алисы совсем недавно началась химия, и она только делает первые шаги в изучении этого непростого предмета. При этом Алиса неплохо умеет программировать и хочет упростить себе изучение химии с помощью этих навыков.

Алиса не раз слышала от старших товарищей о таблице Менделеева, в которой собраны все существующие химические элементы. Алиса решила, что первым делом хочет изучить эту таблицу. Но для начала она хочет оценить, какой объём новых знаний придётся запомнить.

Для этого Алиса составила архив из файлов (📎 [elements.zip](#)). В каждом файле с названием химического элемента находится его короткое описание.

Посчитайте суммарное число *букв*, которые встречаются во всех файлах. В качестве ответа укажите полученное число.

Буквой в рамках задания будет считаться латинский символ из промежутка «a ... z» любого регистра. Для определения того, является ли какой-то символ такой буквой, в Python удобно воспользоваться методом `isalpha`, который есть у объектов типа `char`.

Число

1 балл

Определите, какие слова наиболее часто встречаются в представленных файлах.

При подсчёте не учитывайте регистр слов: например, если в данных встретятся слова «metal» и «Metal», то это будут разные вхождения *одного и того же* слова «metal».

Также при расчёте статистики важно убрать из слов лишние символы: «,», «(» и «)». Так, слово «metal,» должно быть приведено к виду «metal».

В качестве ответа укажите 3 наиболее часто встречаемых слова. Слова приведите к нижнему регистру, убрав из них обозначенные выше лишние символы. Слова в ответе расположите в порядке *убывания* частоты их встречаемости.

Ответ

Ответ

Ответ

№ 19 — 21

0.5 баллов

Представьте, что вы работаете аналитиком поискового сервиса. У вас на руках есть некоторая информация про действия, совершаемые его пользователями.

Эти данные содержатся в таблице (📎 [actions.csv](#), формат файла — [CSV](#)), в которой есть следующие колонки:

- **ip** — IP-адрес пользователя. Гарантируется, что каждый пользователь использует уникальный IP-адрес. То есть с помощью этого адреса можно уникальным образом идентифицировать конкретного пользователя.
- **date** — дата, когда пользователь совершал какие-либо действия. Формат представления даты — **%d.%m.%Y**. Например, запись «29.12.2023» соответствует 29 декабря 2023 года.
- **action** — строка со списком действий, которые совершал пользователь в указанную дату. Возможные действия — «search» (поиск) и «click» (переход по ссылке). Действия в списке записываются через запятую.

Например, строка «search, click, search» означает, что в указанную дату пользователь два раза произвёл поиск и один раз перешёл по ссылке в поисковой выдаче.

Определите, в какую дату было больше всего пользователей, совершавших какие-либо действия.

В качестве ответа укажите эту дату в том же формате, в котором она представлена в данных (т.е. **%d.%m.%Y**).

Ответ

0.5 баллов

Определите, сколько было *уникальных* пользователей, которые совершали какие-либо действия в марте 2023 года. То есть если один и тот же пользователь совершал действия в разные даты в течение указанного периода, то учесть его в статистике всё равно необходимо только один раз.

Число

1 балл

Определите, сколько в марте 2023 года было совершено действий типа «click».

Число

№ 22 — 24

0.5 баллов

Задача классификации — задача предсказания для некоторого объекта класса, к которому он принадлежит. В рамках задания будем рассматривать два возможных класса — «-1» (отрицательный класс) и «1» (положительный класс).

Логистическая регрессия — модель машинного обучения, которая часто используется для решения задачи классификации. Для данного объекта с характеристиками x_1 и x_2 модель предсказывает значение целевой характеристики по следующей формуле:

$$\hat{y} = \text{sign}(\sigma(w_0 + w_1 \cdot x_1 + w_2 \cdot x_2) - p),$$

где w_0 , w_1 и w_2 — параметры модели, $\sigma(x)$ — функция «сигмоида», p — некоторое пороговое значение, а $\text{sign}(x)$ — функция «знак». Подробнее про функции и пороговое значение можно прочитать ниже. Предсказанное моделью значение обозначают $\hat{y}$.

Сигмоида, знак и пороговое значение

Допустим, что в заданной модели логистической регрессии параметры имеют следующий вид:

$$w_0 = 1, \quad w_1 = 7, \quad w_2 = 7, \quad p = \frac{1}{2}.$$

Также дан объект со следующими характеристиками:

$$x_1 = -0.2, \quad x_2 = 0.3.$$

Вычислите для объекта значение $\sigma(w_0 + w_1 \cdot x_1 + w_2 \cdot x_2)$.

Ответ округлите до 2-х знаков после запятой.

Число или дробь

Вычислите для объекта значение итогового предсказания $\hat{y}$.

☐ 1

☐ -1

0.5 баллов

Для оценки качества работы модели для решения задачи классификации часто используется метрика *accuracy* («доля верных ответов»). На заданном множестве примеров метрика рассчитывается как отношение примеров, для которых модель дала правильный ответ, к общему числу примеров:

$$\text{accuracy} = \frac{\text{число правильно классифицированных примеров}}{\text{общее число примеров}}.$$

Дан набор примеров — таблица ( [validation_set.csv](#)) со следующими колонками:

- *x1* — значение характеристики *x*₁ для данного объекта,
- *x2* — значение характеристики *x*₂ для данного объекта,
- *y* — класс, которому на самом деле принадлежит объект.

Рассчитайте по представленным данным значение метрики *accuracy* для заданной модели логистической регрессии. Ответ округлите до 2-х знаков после запятой.

Число или дробь

1.5 баллов

При решении конкретной задачи для получения более оптимальных результатов значение параметра *p* в модели логистической регрессии можно варьировать.

Подберите значение *p* таким образом, чтобы значение метрики *accuracy* на представленных данных было максимальным.

В качестве ответа укажите полученное *значение параметра*, округлённое до 2-х знаков после запятой. Если таких значений несколько, укажите любое из них.

Число или дробь