

Демонстрационный вариант заданий второго этапа по Профилю «Компьютерные науки и науки о данных»

Задача 1

Сколько существует пар целых чисел a и b , таких что $19a + 14b = 1$, и $\{|a|, |b|\} < 100$.

Решение

Одно из решений находится через вычисление коэффициентов линейного представления НОД(19; 14) как вычисление коэффициентов Безу:

$$19 = 14 \cdot 1 + 5; 14 = 5 \cdot 2 + 4; 5 = 4 \cdot 1 + 1, \text{ следовательно } 1 = 5 \cdot 1 - 4 \cdot 1 = 5 \cdot 1 - (14 \cdot 1 - 5 \cdot 2) \cdot 1 = 5 \cdot 3 - 14 \cdot 1 = (19 \cdot 1 - 14 \cdot 1) \cdot 3 - 14 \cdot 1 = 19 \cdot 3 - 14 \cdot 4.$$

Таким образом $(3; -4)$ — одно из решений. Другие решения получаются добавлением (вычитанием) 14 к значению a с одновременным вычитанием (добавлением) 19 из b :

$$a = 3 + 14n,$$

$$b = -4 - 19n.$$

Решения будут удовлетворять условию при $n = -5, -4, -3, -2, -1, 0, 1, 2, 3, 4, 5$.

Ответ: 11.

Задача 2

Из полного графа K_{11} на 11 вершинах удалили 45 рёбер, чтобы получить как можно больше компонент связности. Сколько компонент связности получилось?

Решение

В полном графе на 11 вершинах $\frac{11 \cdot 10}{2} = 55$ рёбер. После удаления 45 рёбер их останется 10.

Если удалить рёбра так, чтобы остался полный граф K_5 на 5 вершинах, в котором 10 рёбер, то оставшиеся 6 вершин будут другими компонентами связности.

Ответ: 7.

Задача 3

В данной задаче символом \oplus обозначено исключающее ИЛИ (XOR), а чертой сверху — булева инверсия. Рассмотрим выражение $x_1 \oplus x_2 \oplus x_3 \oplus \dots \oplus x_{20}$, в котором каждый из символов x_i заменяется равновероятно и независимо на один из символов

$true, false, a, \underline{a}, b, \underline{b}$. Какова вероятность того, что полученное выражение окажется тождественно равно $true$ (независимо от значений a и b)?

Решение

Заметим, что модель равносильна такой двухэтапной процедуре:

- ЭТАП 1. В исходном выражении случайным образом каждый x_i можно заменить на $false, a$ или b , получив выражение, тождественно равное одному из следующих четырёх: $\{false, a, b, a \oplus b\}$. Это вытекает из основных свойств связки XOR: $false \oplus p = p$ и $p \oplus p = false$.
- ЭТАП 2, в выражении, полученном на предыдущем этапе, над каждым из двадцати "слагаемых" независимо с вероятностью $\frac{1}{2}$ поставить инверсию. В результате этого второго шага значение всего выражения либо останется неизменным, либо инвертируется, при этом вероятность получить неинвертированное и инвертированное выражение совпадают (достаточно изменить инвертированность x_{20} , чтобы изменить инвертированность всего результата).

Из сказанного следует, что ответ равен половине от вероятности получить $false$ на первом этапе выбора. Эта вероятность, в свою очередь, равна вероятности заменить среди x_1, \dots, x_{20} чётное количество символов на a и чётное на b . Удобно посчитать это по формуле полной вероятности, в зависимости от того, сколько x_i было заменено на $false$. Пусть ровно k таких x_i . Нас интересуют только чётные k , поскольку если k нечётно, то невозможно оставшиеся x_i разбить на две чётные группы.

- При каждом фиксированном чётном $k < 20$ условная вероятность того, что при замене x_i , не заменённых на $false$, одним из символов a, b , итог выражения получится $false$, равна $\frac{1}{2}$ (между допустимыми и недопустимыми заменами можно установить биекцию переменной одного из x_i с a на b и наоборот).
- При $k = 20$ указанная условная вероятность равна 1.

При каждом целом $k \in [0, 20]$ вероятность того, что среди всех x_1, \dots, x_{20} на первом этапе ровно k будут заменены на $false$, равна $\binom{20}{k} \cdot \left(\frac{1}{3}\right)^k \cdot \left(\frac{2}{3}\right)^{20-k}$.

Итоговый ответ, полученный по формуле полной вероятности, равен:

$$\begin{aligned} & \frac{1}{2} \cdot \left(\sum_{0 \leq k < 20 \text{ чётное}} \binom{20}{k} \cdot \left(\frac{1}{3}\right)^k \cdot \left(\frac{2}{3}\right)^{20-k} \cdot \frac{1}{2} + \binom{1}{3}^{20} \cdot 1 \right) \\ &= \frac{1}{4} \cdot \left(\left(\frac{1}{3}\right)^{20} + \sum_{0 \leq k \leq 20 \text{ чётное}} \binom{20}{k} \cdot \left(\frac{1}{3}\right)^k \cdot \left(\frac{2}{3}\right)^{20-k} \right). \end{aligned}$$

Заметим, что сумму можно свернуть, воспользовавшись формулой бинома:

$$\begin{aligned}
 & \sum_{0 \leq k \leq 20, k \text{ чётное}} \binom{20}{k} \cdot \left(\frac{1}{3}\right)^k \cdot \left(\frac{2}{3}\right)^{20-k} \\
 &= \sum_{k=0}^{20} \binom{20}{k} \cdot \left(\frac{1}{3}\right)^k \cdot \left(\frac{2}{3}\right)^{20-k} + \sum_{k=0}^{20} \binom{20}{k} \cdot \left(-\frac{1}{3}\right)^k \cdot \left(\frac{2}{3}\right)^{20-k} \\
 &= \frac{1}{2} \cdot \left(\left(\frac{2}{3} + \frac{1}{3}\right)^{20} + \left(\frac{2}{3} - \frac{1}{3}\right)^{20} \right) = \frac{1}{2} \cdot \left(1 + \left(\frac{1}{3}\right)^{20} \right).
 \end{aligned}$$

Ответ можно, тем самым, переписать в упрощённом виде:

$$\frac{1}{4} \cdot \left(\left(\frac{1}{3}\right)^{20} + \frac{1}{2} \cdot \left(1 + \left(\frac{1}{3}\right)^{20} \right) \right) = \frac{1}{8} \cdot \left(1 + \left(\frac{1}{3}\right)^{19} \right).$$

Ответ: $\frac{1}{8} \cdot \left(1 + \left(\frac{1}{3}\right)^{19} \right)$.

Задача 4

Отметьте недостатки полностью ассоциативного кэша

- a) **Высокие затраты при промахе**
- b) Подходит не для всех видов оперативной памяти
- c) Высокий процент промахов
- d) **Сложность реализации**

Задача 5

На вход программе подаются строки вида:

Марка_авто, модель_авто, год_выпуска, пробег

которые заканчиваются словом "Конец".

Напишите программу, которая выводит ответ на следующий вопрос:

Автомобили каких марок имеют пробег больше 100000?

Формат вывода

Строки (марки_авто) выводятся в отсортированном виде в алфавитном порядке. Если таких автомобилей нет, не выводить ничего

Марка_1

Марка_2

Марка_3

Решение

```
data = []
```

```
res = set()
```

```
s = input()
```

```
while s != 'Конец':
```

```
    data.append(s.split(','))
```

```
    s = input()
```

```
for i in data:
```

```
    if int(i[3]) > 100000:
```

```
        res.add(i[0])
```

```
print('\n'.join(sorted(list(res))))
```

Входные данные

Audi, A8, 2015, 89516

Opel, Astra, 2012, 100000

Volkswagen, Passat, 2010, 245789

Audi, A4, 2018, 98564

Audi, A4, 2019, 45782

Volvo, XC90, 2015, 198456

Audi, A6, 2020, 21041

Audi, A8L, 2017, 68745

Volvo, XC90, 2015, 269458

Volkswagen, Touareg, 2016, 97856

End

Выходные данные

Volkswagen

Volvo

Задача 6

Напишите на языке C++ шаблонный класс MyVector, представляющий собой динамически расширяющийся массив. Деструктор класса должен освобождать память,

занимаемую массивом. Для работы с массивом должны быть доступны следующие методы:

- a. Добавить в конец массива
- b. Добавить в начало массива
- c. Получить количество элементов массива
- d. Печать содержимого контейнера на экран

Приведите пример программы, использующей все эти методы.

Решение

```
#include <iostream>
```

```
#include <cstring>
```

```
template <class T>
```

```
class MyVector{
```

```
    const int delta = 3;
```

```
    T *arr;
```

```
    size_t capacity;
```

```
    size_t size;
```

```
    void expansion(){
```

```
        capacity+=delta;
```

```
        T *tmp = new T[capacity];
```

```
        std::memcpy(tmp, arr, sizeof(T) * size);
```

```
        delete[]arr;
```

```
    arr = tmp;  
}
```

public:

```
MyVector(){  
  
    size = 0;  
  
    capacity = delta;  
  
    arr = new T[capacity];  
}
```

```
~MyVector(){  
  
    delete[] arr;  
}
```

```
void pushBack(T elem){  
  
    if(size >= capacity)  
  
        expansion();  
  
    arr[size] = elem;  
  
    size++;  
}
```

```
void pushFront(T elem){  
  
    insert(elem, 0);  
  
}  
  
int count(){  
  
    return size;  
  
}  
  
void print(){  
  
    for(int i=0; i<size; i++){  
  
        std::cout<<arr[i]<<' '  
  
    }  
  
    std::cout<<std::endl;  
  
}  
  
};  
  
int main(){  
  
    MyVector<int> v;  
  
    std::cout<<v.count()<<std::endl;  
  
    v.print();  
  
}
```



```
v.pushBack(10);  
  
v.print();  
  
v.pushFront(5);  
  
v.print();  
  
v.pushBack(15);  
  
v.print();  
  
v.pushFront(2);  
  
v.print();  
  
std::cout<<v.count()<<std::endl;  
  
v.pushBack(20);  
  
v.print();  
  
v.pushFront(1);  
  
v.print();  
  
v.insert(77, 3);  
  
v.print();  
  
std::cout<<v.count()<<std::endl;  
  
}
```

Задача 7

Что такое видеостраница?

Веб-страница, на которой воспроизводится видео

Память, необходимая для хранения полного образа экрана

Память, в которой хранится видеофайл

Задача 8

IP-адрес хоста 192.168.15.158. Маска подсети 255.255.255.224. Определите адрес сети.

192.168.15.144

192.168.15.128

192.168.15.152

Решение.

На заданный IP-адрес 192.168.15.158 необходимо наложить маску 255.255.255.224.

Для этого необходимо записать адрес и маску в двоичном виде и совершить операцию И:

11000000.10101000.00001111.10011110

11111111.11111111.11111111.11100000

11000000.10101000.00001111.10000000 = 192.168.15.128

Ответ: 192.168.15.128

Задача 9

В базе данных хранится информация о сотрудниках организации. Каждому сотруднику соответствует запись, содержащая пол сотрудника (мужчина или женщина), его возраст,

подразделении (отдел маркетинга, отдел логистики, отдел технической поддержки).
Известно количество записей, полученных в ответ на ряд запросов к этой базе:

1. Возраст сотрудника > 40 в отделе логистики – 3 записи.
2. Мужчин в возрасте >30 в отделе логистики – 10 записей
3. Женщин в возрасте >30 в отделе логистики и маркетинга – 5 записей.
4. Женщин в возрасте >30 в отделе маркетинга – 2 записи.

Сколько сотрудников в возрасте от 30 до 40 работает в отделе логистики?

Решение. Из п. 3 и п. 4 можно сделать вывод, что женщин в возрасте >30 в отделе логистики 3. Мужчин в возрасте >30 в отделе логистики – 10. Итого в отделе логистики работает 13 человек в возрасте >30 . Из них 3 человека в возрасте >40 . Следовательно, сотрудников отдела логистики в возрасте от 30 до 40 – 10.

Ответ: 10.

Задача 10

При проектировании правил межсетевого экрана могут возникать различные ошибки, связанные с противоречиями в них. Такие ошибки называются аномалиями. Выделяют следующие аномалии. Аномалия затенения возникает, если правило дублировано одним или набором предыдущих правил, и при этом выполняет другое действие. Два правила считаются пересекающимися, если одно правило является надмножеством другого в некоторой части полей и наоборот. Правило объявляется избыточным, если действие, выполняемое межсетевым экраном над пакетом, не изменяется даже после его удаления из набора правил.

Вам даны правила фильтрации межсетевого экрана. Определите, если ли в них противоречия (аномалии).

Rule	Protocol	Source IP	Source port	Destination IP	Destination port	Action
1	*	10.19.55.124	*	10.12.32.21– 10.12.32.22	1–80	Allow

2	TCP	172.19.55.*	*	10.12.32.21	80	Allow
3	TCP	192.168.5.64	*	10.12.32.23	23	Allow
4	TCP	172.19.55.121	*	10.12.32.21	80	Allow

- затенение правил 2 и 4
- пересечение правил 1 и 3
- **избыточность правила 4 (относительно 2)**
- в правилах нет противоречий

Ответ: избыточность правила 4 (относительно 2)

Решение: аномалии правил межсетевых экранов связаны с пересечением полей правил и/или не совпадением требуемого действия. Рассмотрим правила 2 и 4, в поле Source IP правила 4 является подмножеством правила 2, в остальных полях они полностью совпадают. Поэтому можно сказать, что правила 2 и 4 не затеняют друг друга (т.е. нет противоречия в поле Action). Однако, если удалить правило 4, то действие межсетевого экрана над пакетом не изменится. Следовательно, правило 4 является избыточным.

Задача 11

Для входа в помещение используется 4-х значный пин-код - XXXX. Известно, что пин-код является простым числом, при этом первая цифра равна последней цифре. Например, 1XX1.

Задержка между попытками входа в систему равна 1 секунде. За какое максимальное количество времени (в секундах) можно гарантированно получить пин-код, если на ввод пин-кода время не тратится?

Ответ: 117 с

Пояснение: всего подходящих под условие простых чисел - 118, следовательно, на их ввод понадобится 117 секунд, т.к. время тратится только на задержку между попытками.

Задача 12

Было получена дейтаграмма от некоторого сервера:

e0 ea eb 3b e7 18 a8 5f a2 41 a3 40 a8 5d b4 50 d5 3f d5 3e c6 0a a6

Известна структура дейтаграммы, которая представлена ниже:

2 байта	2 байта	2 байта	...	2 байта
порт отправител я	порт получателя	Длина дейтаграммы	Данные	Контрольная сумма

Длина дейтаграммы представляет собой сумму всех ее полей, включая поле с длиной. Весь сетевой трафик шифруется методом «двоичного гаммирования», то есть путём применения операции XOR к байтам дейтаграммы (включая заголовок с портами получателя и отправителя) и байтам, полученными циклическим повторением последовательности байт секретного ключа. Известно, что для шифрования используется 2-байтовый ключ.

Определите, что послал сервер, если известно, что он находится на порту 2021.

Ответ: OPENDOORS_2021!

Вспомогательные материалы: hex коды ASCII-символов

https://www.cisco.com/c/en/us/td/docs/ios/12_4/cfg_fund/command/reference/cfnapp.html

Решение:

Зная структуру дейтаграммы (1) и то, что она отправлена сервером (2), мы можем сделать вывод, что **e0 ea** это закодированное представление порта отправителя, т.е. 2021. Следовательно, если выполнить операцию XOR первых двух байт сообщения и двух байт, соответствующих номеру порта, то можно получить ключ. Для этого представим число 2021 в hex - 07e5, вычислим

e0ea XOR 07e5 = e70f (или 1110011100001111), что и является секретным ключом.

Далее последовательно применим ключ для дешифрации данных дейтаграммы. Сначала определим размер данных: расшифруем размер дейтаграммы **e7 18 XOR e70f = 17** (или 23 в десятичной системе.) Значит длина полезной нагрузки равна $23 - 8 = 15$ и она расположена начиная с 7-го байта **a8 5f a2 41 a3 40 a8 5d b4 50 d5 3f d5 3e c6**. Применяя к этой последовательности операции гаммирования ключом **e70f** получаем коды символов ASCII в hex. Каждый символ ASCII кодируется одним байтом. Результат дешифрования **OPENDOORS_2021!**

Задача 13

Имеется набор данных с информацией о характеристиках человека (возраст, пол, профессия, астигматизм, и т.п.), при которых рекомендуется использовать разные типы линз (soft, standart, hard):

Какая модель машинного обучения должна быть построена (без преобразования данных) для формирования рекомендации типа линз новым клиентам:

- regression function
- cluster model
- **decision tree**
- association rules

Задача 14

Имеется набор данных показаний акселератора (acceleration_x, acceleration_y, acceleration_z), гироскопа (gyro_x, gyro_y, gyro_z) установленных на автомобиле. Показания сняты с интервалом в одну секунду. Для каждого отсчета времени указано состояние автомобиля: стоянка, начало движения, остановка, левый поворот, правый поворот, торможение, разгон.

Какой тип нейронной сети лучше использовать для классификации состояний автомобиля по показаниям датчиков.

- a. полносвязанные нейронные сети
- b. рекуррентные нейронные сети**
- c. самоорганизующиеся карты Кахонена
- d. обратные нейронные сети
- e. прямые нейронные сети

Задача 15

Задан набор данных:

Weather	Temperature	Play
Sun	Hot	No
Sun	Hot	No
Cloudiness	Hot	Yes
Rain	Standard	Yes
Rain	Cold	Yes
Rain	Cold	No
Cloudiness	Cold	Yes
Sun	Standard	No
Sun	Cold	Yes
Rain	Standard	Yes

Вычислить с помощью алгоритма Naïve Bayes нормализованную вероятность проведения игры для условий (в процентах целое число от 0 до 100): Weather = Sun и Temperature = Hot

Решение:

$$P(\text{Weather} = \text{Sun} \mid \text{Play} = \text{No}) = 3/4;$$

$$P(\text{Weather} = \text{Sun} \mid \text{Play} = \text{Yes}) = 1/4;$$

$$P(\text{Temperature} = \text{Hot} \mid \text{Play} = \text{No}) = 2/3;$$

$$P(\text{Temperature} = \text{Hot} \mid \text{Play} = \text{Yes}) = 1/3;$$

$$P(\text{Play} = \text{No}) = 4/10;$$

$$P(\text{Play} = \text{Yes}) = 6/10.$$

$$P(\text{Play} = \text{No} \mid \text{Weather} = \text{Sun}, \text{Temperature} = \text{Hot}) =$$

$$\begin{aligned} &P(\text{Weather} = \text{Sun} \mid \text{Play} = \text{No}) \cdot \\ &P(\text{Temperature} = \text{Hot} \mid \text{Play} = \text{No}) \cdot \\ &P(\text{Play} = \text{No}) = \\ &3/4 \cdot 2/3 \cdot 4/10 = 0,2 \end{aligned}$$

$$P(\text{Play} = \text{Yes} \mid \text{Weather} = \text{Sun}, \text{Temperature} = \text{Hot}) =$$

$$\begin{aligned} &P(\text{Weather} = \text{Sun} \mid \text{Play} = \text{Yes}) \cdot \\ &P(\text{Temperature} = \text{Hot} \mid \text{Play} = \text{Yes}) \cdot \\ &P(\text{Play} = \text{Yes}) = \\ &1/4 \cdot 1/3 \cdot 6/10 = 0,05 \end{aligned}$$

Нормализация:

$$P(\text{Play} = \text{Yes} \mid \text{Weather} = \text{Sun}, \text{Temperature} = \text{Hot}) = 0,05 / (0,05 + 0,2) = 0,2$$

Ответ: 20%