Computer & Data Science

Sample test

Task 1

Compute the number of integer pairs (a, b) satisfying the constraints $19a + 14b = 1$ and $\{|a|, |b|\} < 100$.

Solution

The linear representation of $(19; 14)$ will help to find one of the solutions:

$19 = 14 \cdot 1 + 5$; $14 = 5 \cdot 2 + 4$; $5 = 4 \cdot 1 + 1$, therefore $1 = 5 \cdot 1 - 4 \cdot 1 = 5 \cdot 1 - (14 \cdot 1 - 5 \cdot 2) \cdot 1 = 5 \cdot 3 - 14 \cdot 1 = (19 \cdot 1 - 14 \cdot 1) \cdot 3 - 14 \cdot 1 = 19 \cdot 3 - 14 \cdot 4$.

This way, we have found the solution: $(3; -4)$. The others can be derived from this one by adding (resp. subtracting) 14 to a and subtracting (resp. adding) 19 from b:

$a = 3 + 14n,$

$b = -4 - 19n.$

A quick check shows that the appropriate values of n are $-5, -4, -3, -2, -1, 0, 1, 2, 3, 4, 5$.

Answer: 11.

Task 2

Suppose we remove 45 edges from a complete graph K_{11} on 11 vertices. What maximum number of connected components the resulting graph can have?

Solution

A complete graph on 11 vertices has $\frac{11 \cdot 10}{2} = 55$ edges. Upon the removal of 45 edges, there will be ten left. The number of connected components on 11 vertices with ten edges is the largest when all the edges belong to a clique on five vertices and leave the other six vertices isolated. The total number of components is thus $1 + 6 = 7$.

Answer: 7.

Task 3

In this problem, we use $\oplus$ to denote the exclusive OR (XOR) junction and a bar above variables to represent the Boolean negation. Consider the expression $x_1 \oplus x_2 \oplus x_3 \oplus \dots \oplus x_{20}$. Suppose that each of the symbols x_i is replaced with $true, false, a, \underline{a}, b, \underline{b}$. What is the probability that the value of the resulting expression will be constant $true$?

Solution

The probabilistic procedure here is equivalent to the following:

- Stage 1. In the original expression, every x_i is uniformly and independently replaced by $\{false, a, b\}$, resulting in an expression whose value is one of the following four: $\{false, a, b, a \oplus b\}$. The basic properties of the XOR junction suggest that $false \oplus p = p$ and $p \oplus p = false$.
- Stage 2. In the expression obtained at Stage 1, each of the 20 operands is negated with probability $\frac{1}{2}$. Thus, the value of the whole expression either stays intact or is negated, and the probability of each of these outcomes is $\frac{1}{2}$ (it is sufficient to change the negation state of x_{20} to alter the negation state of the whole expression).

The above implies that the probability that we need to determine is half the probability of getting a constant $false$ expression at Stage 1. This probability equals that of an even number of symbols in $x_1, \dots, x_{20}$ being changed to a, and another even number of the symbols, to b. The formula of total probability is used next, depending on what number of x_is was changed to $false$. Let k be the number of the x_is. Only even values of k would fit, as an odd k will make it impossible to divide the remaining x_is into two groups with an even number of members.

- For each fixed even $k < 20$, the conditional probability that the resulting expression is constant $false$ after changing all the remaining x_is (those that have not been replaced by $false$) to a or b is $\frac{1}{2}$. (The valid and invalid expressions are in 1-to-1 correspondence, which is produced by switching any x_i from a to b or vice versa.)
- For $k = 20$, the conditional probability is 1.

For every integer $k \in [0,20]$, the probability that, among all $x_1, \dots, x_{20}$, exactly k symbols are changed to $false$ is $\binom{20}{k} \cdot \left(\frac{1}{3}\right)^k \cdot \left(\frac{2}{3}\right)^{20-k}$.

By applying the total probability formula, we obtain:

$$\frac{1}{2}\cdot\left(\sum_{0\le k<20 k even}\binom{20}{k}\cdot\left(\frac{1}{3}\right)^{k}\cdot\left(\frac{2}{3}\right)^{20-k}\cdot\frac{1}{2}+\left(\frac{1}{3}\right)^{20}\cdot 1\right.$$

$$\left.=\frac{1}{4}\cdot\left(\left(\frac{1}{3}\right)^{20}+\sum_{0\le k\le 20 k even}\binom{20}{k}\cdot\left(\frac{1}{3}\right)^{k}\cdot\left(\frac{2}{3}\right)^{20-k}\cdot\right)\right)$$

The sum Σ can be simplified using the binomial summation formula:

$$\sum_{0\le k\le 20 k even}\binom{20}{k}\cdot\left(\frac{1}{3}\right)^{k}\cdot\left(\frac{2}{3}\right)^{20-k}$$

$$=\sum_{k=0}^{20}\binom{20}{k}\cdot\left(\frac{1}{3}\right)^{k}\cdot\left(\frac{2}{3}\right)^{20-k}$$

$$+\sum_{k=0}^{20}\binom{20}{k}\cdot\left(\frac{-1}{3}\right)^{k}\cdot\left(\frac{2}{3}\right)^{20-k}=\frac{1}{2}\cdot\left(\left(\frac{2}{3}+\frac{1}{3}\right)^{20}+\left(\frac{2}{3}-\frac{1}{3}\right)^{20}\right)=\frac{1}{2}\cdot\left(1+\left(\frac{1}{3}\right)^{20}\right).$$

Thus, the answer can be simplified so that it does not contain the Σ operator:

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

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

Task 4

Which disadvantages does a fully associative cache have?

a) **A high miss cost**
b) It is not compatible with all RAM types
c) A high percentage of misses
d) **The implementation complexity**

Task 5

The input to a program is strings of the form:

Car_brand, car_model, production_year, mileage.
All the strings end with the word "End".

Write a program that prints out the answer to the following question:
Which car brands have a mileage of over 100,000?

Output format
Lines (car_brand) are displayed in alphabetical order. If there are no such cars, do not display anything.

Brand_1
Brand_2
Brand_3

Solution

```
data = []
res = set()
s = input()
while s != 'End':
    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))))
```

Program input

```
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
```

Program output

Volkswagen
Volvo

Task 6

Write a template class MyVector, which is a dynamically expanding array, in C ++. The class destructor must free the memory allocated by the array. The following methods must be available to work with an array,:

A. Add to the end of the array
B. Add to the beginning of the array
C. Get the number of array elements
D. Print container content on the screen

Write a program that will use all these methods.

Solution

```
#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;
}
```

Task 7

What is a video page?

1) A web page that a video is playing on

2) The memory required to store a screen image

3) The memory in which a video file is stored

Task 8

Host IP address 192.168.15.158. Subnet mask 255.255.255.224. Determine the network address.

A. 192.168.15.144
B. 192.168.15.128
C. 192.168.15.152

Solution.

The specified IP address 192.168.15.158 must be masked with 255.255.255.224.

To do this, you need to write down the address and the mask in binary and perform the AND operation:

11000000.10101000.00001111.10011110

11111111.11111111.11111111.11100000

11000000.10101000.00001111.10000000 = 192.168.15.128

Answer: 192.168.15.128

Task 9

A database stores information about the employees of an organization. Each employee has a personal record containing gender (male or female), age, division (marketing department, logistics department or technical support department). The number of records received in response to certain queries to this database is as follows:

1. Employees aged> 40 in the logistics department: 3 entries.

2. Males over 30 in the logistics department: 10 entries.

3. Females aged> 30 in the department of logistics and marketing: 5 entries.

4. Females over 30 in the marketing department: 2 entries.

How many employees between the ages of 30 and 40 work in the logistics department?

Solution. From 3 and 4, we can deduce that there are three females aged> 30 in the logistics department, which also has ten males aged> 30. In total, the logistics department employs 13 people aged> 30. Of these, three people are aged> 40. Therefore, there are ten employees in the logistics department between the ages of 30 and 40.

Answer: 10.

Task 10

Firewall rules may have inconsistencies known as anomalies. A rule is shadowed by one or a set of preceding rules if each packet matched by this rule is also matched by preceding rules but takes a different action. Two rules are correlated if they have an intersection of rule spaces, i.e., one rule is a superset to the other in some fields. A rule is redundant if the action taken by the firewall on a packet does not change upon removing this rule from the ruleset.

Look at the firewall filtering rules. Determine if there are any anomalies in them.

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

- Rule 4 is shadowed by rule 2.
- Rules 1 and 3 are correlated.
- **Rule 4 is redundant (with respect to 2).**
- There are no anomalies in the rules.

Answer: rule 4 is redundant (with respect to 2).

Solution: Firewall rule anomalies relate to the intersection of fields of two rules or a set thereof and/or required action mismatch. Let us consider rules 2 and 4. The Source IP field of rule 4 is a subset of that of rule 2. The other fields of the two rules fully coincide. As the action in rules 2 and 4 is the same, rule 4 is not shadowed by rule 2. However, if we remove rule 4, the action of the firewall on the packet will not change. Therefore, Rule 4 is redundant.

Task 11

A 4-digit pin code (XXXX) is used to enter an office room. The pin code is a prime number, and the first digit and last one are the same (e.g. 1XX1).

The delay between login attempts is one (1) second. What is the maximum time (in seconds) within which a valid pin code will be obtained if entering the pin code does not take any time?

Answer: 117 s

Solution: 118 prime numbers meet these conditions. Therefore, it will take 118-1 =117 seconds maximum to enter them and obtain a valid pin code.

Task 12

You have received a datagram from some server:

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

The structure of the datagram is as follows:

2 bytes	2 bytes	2 bytes	...	2 bytes
Sender port	Receiver port	Datagram length	Data (payload)	Checksum

The length of a datagram is the sum of all its fields, including the length field. All network traffic is encrypted using a simple XOR cipher with a cyclically repeating 2-

byte key. The datagram header, including the destination and sender ports, is also encrypted.

Determine what data the server has sent if the port is 2021.

Supplementary materials: see hex codes of ASCII symbols at https://www.cisco.com/c/en/us/td/docs/ios/12_4/cfg_fund/command/reference/cfnapph.html

Answer: OPENDOORS_2021

Solution:

Knowing the datagram structure (1) and that it was sent by a server (2), we can deduce that **e0 ea** is the encrypted representation of the sender port, i.e. 2021. A simple XOR cipher is a bitwise XOR operation performed on every character using a given key. Therefore, an XOR operation on the first bytes of the datagram and the two bytes corresponding to the port number will help to obtain the key. To this end, we convert the number 2021 into hex to get 07e5 and calculate.

e0ea XOR **07e5 = e70f** (or 1110011100001111). **e70f** is the secret key.

Next, we sequentially use the key to decrypt the datagram data. First, we determine the size of the data by decrypting the size of the datagram **e7 18** XOR **e7 0f** = 17 (or 23 in decimal.) Thus, the length of the payload is 23 - 8 = 15, and it starts at the 7th byte **a8 5f a2 41 a3 40 a8 5d b4 50 d5 3f d5 3e c6.** By performing an XOR operation on this sequence with the **e7 0f** key, we obtain ASCII character codes in hex. Each ASCII character is encoded in one byte. The decryption result is ***OPENDOORS_2021.***

Task 13

A dataset contains information about characteristics (age, gender, profession, astigmatism, etc.) associated with using different types of lenses (soft, standard and hard):

What machine learning model (without data transformation) is the most appropriate for recommending lens type to new customers:

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

Russian Scholarship Project
open doors
Association "Global Universities"

- association rules

Task 14

A dataset contains readings from a car accelerator (acceleration_x, acceleration_y, acceleration_z,) and a gyroscope (gyro_x, gyro_y, gyro_z). The readings are taken every second. The state of the car is recorded for each time count: parking, starting the movement, stopping, left turn, right turn, braking, acceleration.

What type of neural network is the best for classifying vehicle states by sensor readings.

- convolutional neural network
- **recurrent neural network**
- self-organizing map
- back neural network
- direct neural network

Task 15

Look at the dataset:

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

Use a Naïve Bayes algorithm to calculate the normalized probability of the game being played. The desired conditions are as follows: Weather = Sun and

Temperature = Hot. Give your answer as a percent expressed as an integer from 0 to 100.

Solution:
P(Weather = Sun | Play = No) = 3/4;
P(Weather = Sun | Play = Yes) = 1/4;
P(Temperature = Hot | Play = No) = 2/3;
P(Temperature = Hot | Play = Yes) = 1/3;

P(Play = No) = 4/10;
P(Play = Yes) = 6/10.

P(Play = No | Weather = Sun,Temperature = Hot) =

P(Weather = Sun | Play = No) ×

P(Temperature = Hot | Play = No) ×

P(Play = No) =

3/4×2/3×4/10 =0,2

P(Play = Yes | Weather = Sun,Temperature = Hot) =

P(Weather = Sun | Play = Yes) ×

P(Temperature = Hot | Play = Yes) ×

P(Play = Yes) =

1/4×1/3×6/10 = 0,05

Normalization:

P(Play = Yes | Weather = Sun,Temperature = Hot) = 0,05/ (0,05+0,2) = 0,2

Answer: 20