CSE PhD/MSR Interview (Winter Entry 2023)

CSE IIT Delhi

Nov 2023

1 Programming

1.1 General Instructions

There are two programming questions in this test, each worth 15 marks. Your code will be evaluated using several test cases. One sample input and output has been provided for each question. Kindly note that input and output for the programming questions are taken from stdin and stdout respectively.

1.2 Efficient Parcel Delivery

You run a parcel delivery business and have n trucks to do so. Given a client's order to deliver m identical parcels, you find out that your i'th truck can hold at most c_i parcels. You wish to determine the fewest number of trucks that you must send to accommodate a total of m parcels.

Input format:

- The first line of input contains the value of m – the number of parcels to be delivered, and the value of n – the number of trucks, separated by a space.
- The second line of input contains n positive integers – the capacities c_i of the n trucks, separated by spaces.

Output format:
Print a single number, which is the minimum number of trucks needed to fit all the m parcels.

Sample input-output
Input:
8 3
4 4 4
Output:
2
Explanation:
The correct answer is 2 as we will need 2 trucks each having capacity 4 to fit 8 parcels.

Input:
19 6
4 5 4 5 2 1

Output:
5
Explanation:
The correct answer is 5 as we will need 5 trucks with capacities 4, 5, 4, 5 and 1 to fit 19(4+5+4+5+1) parcels.

1.3 Mad Over Uttapams

After a long long time, you happen to visit your favorite South Indian food joint, Mad Over Uttapams (MOU), with an infinite craving for uttapams. It's a self-service restaurant where you find n stacks of uttapams, each containing m uttapams, ready to be picked up. But in MOU, the crazy thing is that the cook puts a different amount of batter in every uttapam and so the weights of different uttapams are different. The restaurant requires you to pick the top few (possibly zero) uttapams from each stack, and pay a fixed cost for each uttapam picked, no matter what its weight. You find that you have enough money to pay for k uttapams, so subject to this constraint, you wish to maximize the total weight of the k uttapams you pick up. Remember that if you pick an uttapam, you must also pick all uttapams above it (and pay for them). The chutney is for free!

Input format:

- The first line of input contains the value of n – the number of stacks, the value of m – the number of uttapams in each stack, and the value of k – the number of uttapams to be bought, separated by a space.
- Each of the next n lines contains m space-separated non-negative integers. The i'th line lists the weights of the uttapams in the i'th stack from top to bottom.

Output format:
Print a single number, which is the maximum possible total weight of (at most) k uttapams, subject to the constraint that if an uttapam is picked, all uttapams above it are also picked.

Sample input-output
Sample input
```
3 4 6
3 1 1 2
1 4 4 2
3 1 1 3
```
Output
```
17
```
Explanation: We are given $n = 3$, $m = 4$, $k = 6$ followed by n lines depicting the n stacks of m uttapams. Here, we would pick the first uttapam in the first stack i.e of weight 3, followed by picking all uttapams of the second stack. Lastly, we pick the first uttapam of the last stack i.e 3 after which we have completed picking $k = 6$ uttapams. The total weight being $3 + 1 + 4 + 4 + 2 + 3 = 17$.
Hint:
We are looking for a *fast* (polynomial-time) implementation, so brute-force is out of question. Use dynamic programming. Compute the entries of a two dimensional array `T` so that `T[i][j]` is the maximum possible total mass of (at most) `j` uttappams, if you were to pick them only from the first `i` stacks. Do this computation as follows. Since you are picking `j` uttapams from the first `i` stacks, you may pick the top `x` uttapams from the `i`'th stack, where `x` is in $\{0, \ldots, \min(k, m)\}$. If

you decide to pick the top `x` uttapams from the `i`'th stack, you must pick `j-x` uttapams from the first `i-1` stacks optimally.

2 Probability and Maths

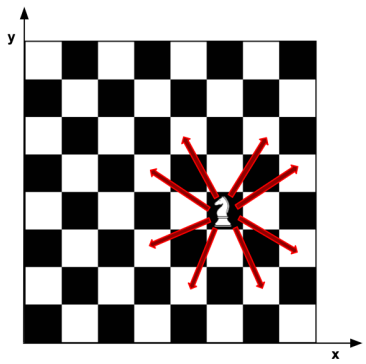
2.1 General Instructions

There are 15 questions in this section. Each question in this section has four alternatives, out of which exactly one is correct. For every correct attempt, you will get **+1 mark** and **-0.25 mark** for incorrect attempt.

Q1. S is a set with 8 elements. The number of pairs of sets (A, B) such that $\emptyset \subseteq A \subseteq B \subseteq S$ is equal to

a. 4^4
b. 6^4
c. 8^4
d. 9^4 (**Correct**)

Q2. A knight moves on an infinite chessboard in the following manner. From the cell (x, y), the knight can move to any of the cells $(x+2, y+1)$, $(x+1, y+2)$, $(x-1, y+2)$, $(x-2, y+1)$, $(x-2, y-1)$, $(x-1, y-2)$, $(x+1, y-2)$, $(x+2, y-1)$. Let $H(n)$ denote the minimum number of moves needed to move the knight from $(0, 0)$ to $(n, 0)$.



Let $D(n)$ denote the minimum number of moves needed to move the knight from $(0, 0)$ to (n, n). Select the correct statement from the following.

a. $H(n)$ and $D(n)$ are both $\Theta(n)$. (**Correct**)
b. $H(n)$ is $\Theta(n)$ but and $D(n)$ is not $\Theta(n)$.
c. $H(n)$ is not $\Theta(n)$ but $D(n)$ is $\Theta(n)$.
d. $H(n)$ and $D(n)$ are both not $\Theta(n)$.

Q3. Seven cards have one integer written on each of their sides and are placed on a table in front of you. You can see the numbers $0, 1, 3, 5, 6, 7, 11$ written on the cards, and you have no idea of the numbers written on the sides facing the table. Your friend says, "For each of these seven cards, if the number on any one side is divisible by 2, then the number on the other side is not divisible by 3". What is the minimum number of cards you need to turn to confirm that your friend's claim is true?

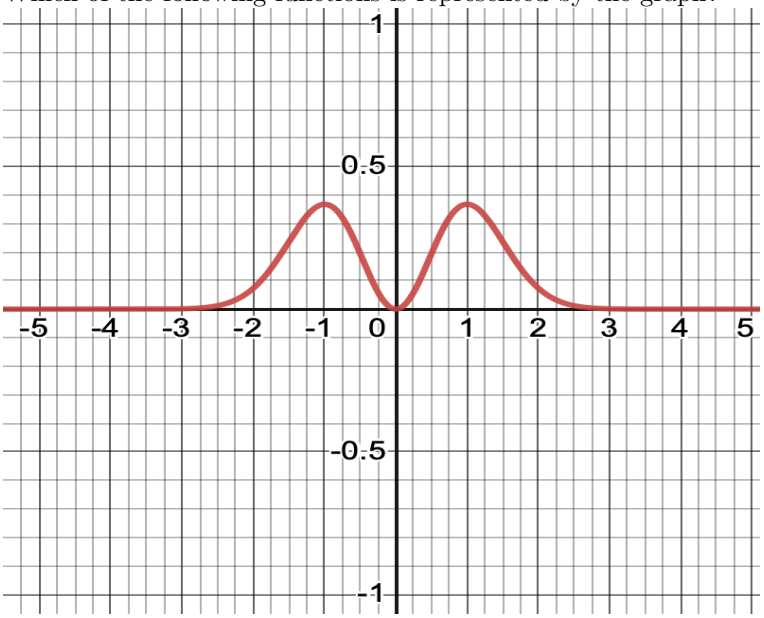
a. 2

b. 3 (**Correct**)
c. 6
d. 7

Q4. Suppose p, q and r are distinct primes and n is a non-negative multiple of pqr. Which of the following expressions gives the number of integers in the set $\{0, \ldots, n-1\}$ that are divisible by p or q?

a. $n/p + n/q + n/r$
b. $n/p + n/q + n/r - n/(pq) - n/(qr) - n/(rp)$
c. $n/p + n/q + n/r - n/(pq) - n/(qr) - n/(rp) + n/(rpq)$
d. $n/p + n/q - n/(pq)$ **(Correct)**

Q5. Which of the following functions is represented by the graph?



a. xe^{-x}
b. xe^{-x^2}
c. x^2e^{-x}
d. $x^2e^{-x^2}$ (**Correct**)

Q6. Which of the following statements about the roots of the polynomial $ax^2 + bx + c$ can be possibly true if a, b, c are all real numbers?

a. The polynomial has a unique root, and it is real. (**Correct**)
b. The polynomial has a unique root, and it is not real.
c. The polynomial has two roots, and one of them is real while the other isn't real.
d. The polynomial has three roots, and one of them is real while the other two aren't real.

Q7. A real-valued random variable X has expectation $\mathbb{E}[X] = \mu$. Which of the following statements is not necessarily true?

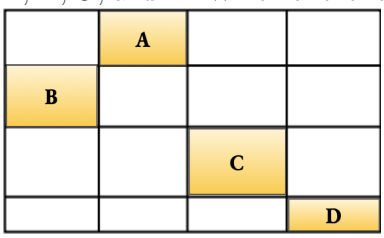
a. $\mathbb{E}[X - \mu] \geq 0$.
b. $\mathbb{E}[(X - \mu)^2] \geq 0$.
c. $\mathbb{E}[(X - \mu)^3] \geq 0$. (**Correct**)

d. $\mathbb{E}[(X-\mu)^4] \geq 0$.

Q8. You were a judge of a season of the reality show "Mastercook India" which started out with n contestants. In each episode of the show, all the surviving contestants cooked a dish for you. You tasted all dishes and eliminated the one contestant with the worst dish. After $n-1$ of the n contestants were eliminated, the unique surviving contestant was declared as the winner. You are unable to remember n but you do remember that the number of dishes you tasted is between 220 and 235. What is n?

a. 15
b. 21 (**Correct**)
c. 22
d. 23

Q9. In the figure below, a rectangle R is divided into 16 sub-rectangles, with four of them named A, B, C, and D. Which of the following statements is necessarily true?



a. The perimeter of R cannot be determined from the perimeters of A, B, C, D.
b. The perimeter of R is necessarily less than the sum of the perimeters of A, B, C, D.
c. The perimeter of R is necessarily equal to the sum of the perimeters of A, B, C, D. (**Correct**)
d. The perimeter of R is necessarily greater than the sum of the perimeters of A, B, C, D.

Q10. A (simple and undirected) graph G is called a *pseudoforest* if every connected component of G contains at most as many edges as the number of vertices in that component. Let m denote the maximum number of edges in a pseudoforest on a set of 2023 vertices. Which of the following is true?

a. $m \leq 2021$
b. $m = 2022$
c. $m = 2023$ (**Correct**)
d. $m \geq 2024$

Q11. Suppose A is an $m \times n$ matrix and B is an $n \times m$ matrix. Which of the following is necessarily true?

a. If $m \leq n$ then AB is invertible.
b. If AB is invertible then $m \leq n$. (**Correct**)
c. If $m \geq n$ then AB is invertible.
d. If AB is invertible then $m \geq n$.

Q12. For $n \in \mathbb{N}$, let $\ell(n)$ denote the number of ones in the binary representation of n. Which of the following is false?

a. For all $n \in \mathbb{N}$, $\ell(n) \leq 1 + \log_2 n$.
b. For all $n \in \mathbb{N}$, if n is even then $\ell(n+1) = \ell(n) + 1$.
c. For all $n \in \mathbb{N}$, if n is odd then $\ell((n-1)/2) = \ell(n) - 1$.

d. For all $n \in \mathbb{N}$, $\ell(2n) = 2\ell(n)$. (**Correct**)

Q13. We, the CSE department at IITD, are thinking of switching to the following algorithm for hiring research students: interview the n applicants in a uniformly random order, and hire every applicant who is better than all the previously interviewed applicants. Assume that no two applicants are ever found to be equally good, and the order of interviews has no bearing on whether a candidate is found to be better than another. Let $H(n)$ denote the expected number of candidates out of n that get hired out of this process. Please help us by specifying the asymptotic behavior of $H(n)$. Choose the correct alternative among the following.

a. $H(n)$ is $\Theta(1)$.
b. $H(n)$ is $\Theta(\log n)$. (**Correct**)
c. $H(n)$ is $\Theta(\sqrt{n})$.
d. $H(n)$ is $\Theta(n)$.

Q14. You have invited three friends, say A, B, and C, for a party. You know that they will arrive at your place in a uniformly random order. Consider the following three events.

- E_1: A arrives before B.
- E_2: A arrives before C.
- E_3: B arrives before C.

What is the probability of the event $E_1 \cap E_2 \cap E_3$?

a. 1/8
b. 1/6 (**Correct**)
c. 1/4
d. 1/3

Q15. I toss two standard 6-sided dice and take the sum of their outcomes. What is the expected value of the result?

a. 6
b. 7 (**Correct**)
c. 12
d. 14

3 Logical Reasoning and Comprehension

3.1 General Instructions

There are 5 paragraphs in this section. Each paragraph has multiple true-false based questions. For every correct answer, you will get **+1 mark** and **-1 mark** for an incorrect attempt. If you do not make an attempt, you will get 0.

Clarification. While the questions below are focused on a technical topic, we expect that a person with a good BTech-level foundation in Computer Science should be able to read and understand it *to the extent required to answer the questions asked* even if they are not familiar with the topic being discussed.

3.2 Data Management

Clustered and non-clustered indexes are two types of indexes used in relational databases for efficient data retrieval. A clustered index determines the physical order in which data is stored on disk, based on the values of one or more columns in the index. In other words, the rows of data in the table are physically sorted based on the values in the clustered index. A table can have only one clustered index, and it is automatically created if a primary key is defined on a column. On the other hand, a non-clustered index is a separate structure that stores the index data and a pointer to the actual data row in the table. Non-clustered indexes do not affect the physical order of the data on disk. A table can have multiple non-clustered indexes.

Consider a large table where records are frequently updated and queried for specific values of a a user selected column. Which of the following statements is/are correct?

(a) Inserting a new record in the table with a clustered index can involve large-scale data movement. **Correct**

(b) Every table must have a clustered index.

(c) Both clustered and non-clustered indexes will be equally effective for the above scenario.

3.3 Cryptography

Digital signatures are one of the most commonly used cryptographic primitives. A digital signature can be thought of as the electronic equivalent of a handwritten signature on a paper document. Just as a handwritten signature on a document verifies the identity of the signer and confirms their intent to sign the document, a digital signature serves the same purpose in the digital realm.

Formally, a digital signature scheme consists of three algorithms - Setup, Sign and Verify. Setup samples a secret signing key and a public verification key. The signing key is used to sign documents using the Sign algorithm. The corresponding verification key is used to verify the signature corresponding to this document, using Verify. A digital signature scheme is said to be secure if no polynomial time adversary can produce a fresh signature on a new document, even after seeing signatures on many documents. A real-world analogy for this security definition - someone may have seen your handwritten signature on multiple documents, but should not be able to sign a new document.

Here is one application of digital signatures - suppose you download an update for a software. How can you be sure that this update is indeed a genuine software update, and not a malicious virus? To ensure this, the software company has a signing key, and the corresponding verification key is stored on your machine when you first download the software. For each software update, the software company signs the update using its signing key, and this update is verified by your computer using the stored verification key.

Let (Setup, Sign, Verify) be a secure digital signature scheme. The Sign algorithm can sign documents of size at most 1GB, and the size of the resulting signature is 1MB. Which of the following statements is/are correct?

1. Given a signature on a document, it should not be possible to produce a signature on a another related (but different) document. (**Correct**)
2. Given just the verification key, it may be possible to compute the corresponding signature key efficiently.
3. Sign and Verify algorithms are independent of each other.

3.4 Deep Learning

Transformers provide general-purpose architectures for Natural Language Understanding and Natural Language Generation. A Transformer layer includes a multi-head self-attention sub-layer, where each token attends to all the tokens. Let the layer input $H = [h_1, h_2, ..., h_n]^T \in \mathbb{R}^{n \times d}$ corresponding to S, where d is the hidden dimension, $S = \{v_1, v_2, ..., v_n\}$ is a sequence of tokens, n is the length of the sequence, and $h_i \in \mathbb{R}^{d \times 1}$ is the hidden representation at position i. For a single-head self-attention sub-layer, the input H is projected by three matrices $W^Q \in \mathbb{R}^{d \times d_K}$, $W^K \in \mathbb{R}^{d \times d_K}$, and $W^V \in \mathbb{R}^{d \times d_V}$ to the corresponding representations Q, K, and V:

$$Q = HW^Q, \quad V = HW^V, \quad K = HW^K \tag{1}$$

The output of this single-head self-attention sub-layer is calculated as:

$$Attn(H) = softmax(A)V \tag{2}$$

$$A = \frac{QK^T}{\sqrt{d_K}} \tag{3}$$

Which of the following statements can be inferred from the above passage?

1. The dimension of the matrix A is $d_K \times d_K$ ($A \in \mathbb{R}^{d_K \times d_K}$).
2. The dimension of the matrix A is $n \times n$ ($A \in \mathbb{R}^{n \times n}$). (**Correct**)
3. The dimension of the matrix Q is $d_K \times n$ ($Q \in \mathbb{R}^{d_K \times n}$)
4. The dimension of the matrix K is $n \times d_K$ ($K \in \mathbb{R}^{n \times d_K}$) (**Correct**)

3.5 Blockchain

As the underlying support of Bitcoin, Blockchain is a distributed ledger technology that uses cryptographic techniques to secure and maintain a decentralized database. Blockchain is designed to provide independent internal verification, communication, transmission, and storage while maintaining a reliable and transparent environment. This technique has the potential to meet various data requirements as it allows any peer to add new data and maintain synchronized information according to specific rules. The Blockchain is structured as a series of blocks that store transactional information. Each block is comprised of two parts: the header and the body. The header includes hash values of both the previous block and its own, enabling the blocks to link and form a continuous chain. In addition to hash values, the header stores essential information about the block, such as timestamps. The body of the Blockchain holds detailed information about transactions, where the first block in the Blockchain is typically known as the "genesis" block. The features of Blockchain have led to rapid development in existing industries, which can be described as follows:

- Decentralization is the most significant feature of Blockchain. With the consensus algorithm, Blockchain can verify and execute information transactions without requiring a trusted third party.
- Immutability is an essential trait of Blockchain, as all peers approve the information newly added through a decentralized consensus. Hence, it is difficult and expensive to change the record of the Blockchain, which requires the consent of the majority.

- Auditability is also an important feature of Blockchain. Each transaction in the Blockchain is accompanied by a unique hash and timestamp, and a copy of the Blockchain is held by all peers, allowing every peer to audit any specific transaction.
- Blockchain is autonomous. With smart contracts, Blockchain can realize trust in physical machines, not bothered by anyone's interference.

Which of the following statements is a valid conclusion based on the features of blockchain described in the paragraph?

1. The consensus algorithm used in blockchain is a trusted third party that verifies and executes information transactions.
2. Smart contracts in blockchain allow for external interference, making it less autonomous than traditional methods of transaction.
3. Blockchain's reliance on hash values and timestamps for storing transactional information makes it vulnerable to cyber attacks.
4. Blockchain has the potential to revolutionize the way transactions are conducted across various industries due to its decentralized, immutable, auditable, and autonomous features. (**Correct**)

3.6 Networks

To improve reliable transport in computer networks, Forward Error Correction (FEC) methods transmit redundant information so that the original information can be recovered even if some parts of the transmitted information are lost or corrupted. 2/3 inter-packet FEC means that 3 packets are transmitted for every 2 packets, and if any two of the three packets are received then the original two packets can be reconstructed. Similarly, 2/3 intra-packet FEC means that within the same packet, for any 2 blocks of data (of say 512 bytes each), 3 blocks are written in the packet, and the original two blocks can be recovered if any two of the three blocks are received correctly. Which of the following statements is/are correct?

1. Intra-packet FEC is suited for networks having links on which bit corruptions are likely, like wireless links. (**correct**)
2. Inter-packet FEC can help improve reliable transport in congested networks but can also worsen congestion. (**correct**)
3. Inter-packet FEC can help improve reliable transport in congested networks that do multi-path routing. (**correct**)