

Data Science and Artificial Intelligence - DA

General Aptitude

Q.1 - Q.5 Carry ONE mark Each

Q.1	Courage : Bravery :: Yearning : _____ Select the most appropriate option to complete the analogy.
(A)	Longing
(B)	Yelling
(C)	Yawning
(D)	Glaring

Q.2	We _____ tennis in the lawn when it suddenly started to rain. Select the most appropriate option to complete the above sentence..
(A)	have been playing
(B)	had been playing
(C)	would have been playing
(D)	could be playing

Q.3	A 4×4 digital image has pixel intensities (U) as shown in the figure. The number of pixels with $U \leq 4$ is:																
	<table><tr><td>0</td><td>1</td><td>0</td><td>2</td></tr><tr><td>4</td><td>7</td><td>3</td><td>3</td></tr><tr><td>5</td><td>5</td><td>4</td><td>4</td></tr><tr><td>6</td><td>7</td><td>3</td><td>2</td></tr></table>	0	1	0	2	4	7	3	3	5	5	4	4	6	7	3	2
0	1	0	2														
4	7	3	3														
5	5	4	4														
6	7	3	2														
(A)	3																
(B)	8																
(C)	11																
(D)	9																

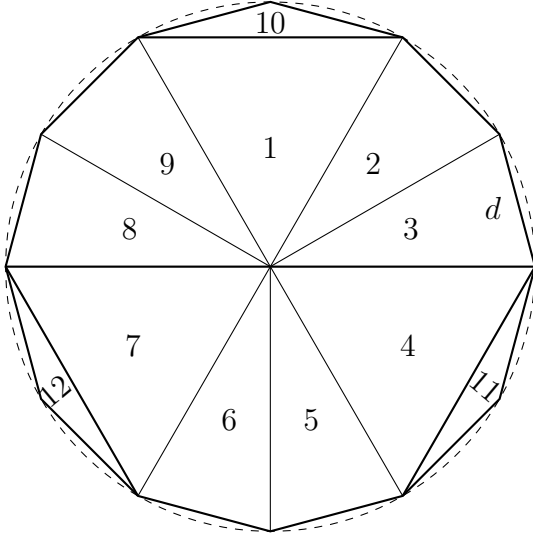
Q.4	In the given figure, the numbers associated with the rectangle, triangle, and ellipse are 1, 2, and 3, respectively. Which one among the given options is the most appropriate combination of P, Q, and R?
(A)	$P = 6; Q = 5; R = 3$
(B)	$P = 5; Q = 6; R = 3$
(C)	$P = 3; Q = 6; R = 6$
(D)	$P = 5; Q = 3; R = 6$

Q.5	A rectangle has a length L and a width W , where $L > W$. If the width, W , is increased by 10%, which one of the following statements is correct for all values of L and W ?
(A)	Perimeter increases by 10%.
(B)	Length of the diagonals increases by 10%.
(C)	Area increases by 10%.
(D)	The rectangle becomes a square.

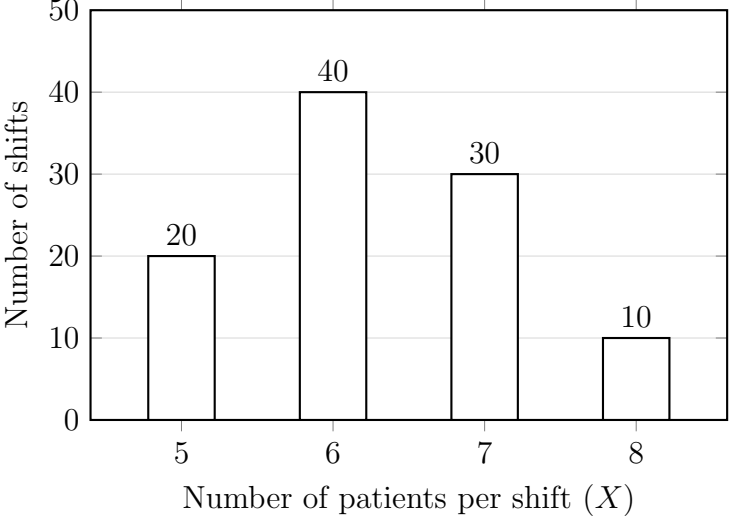
Q.6 - Q.10 Carry TWO marks Each

Q.6	Column-I has statements made by Shanthala; and, Column-II has responses given by Kanishk.			
	Column-I		Column-II	
	P.	This house is in a mess.	1.	Alright, I won't bring it up during our conversations.
	Q.	I am not happy with the marks given to me.	2.	Well, you can easily look it up.
	R.	Politics is a subject I avoid talking about.	3.	No problem, let me clear it up for you.
	S.	I don't know what this word means.	4.	Don't worry, I will take it up with your teacher.
Identify the option that has the correct match between Column-I and Column-II.				
(A)	P – 2; Q – 3; R – 1; S – 4			
(B)	P – 3; Q – 4; R – 1; S – 2			
(C)	P – 4; Q – 1; R – 2; S – 3			
(D)	P – 1; Q – 2; R – 4; S – 3			

Q.7	<p>Weight of a person can be expressed as a function of their age. The function usually varies from person to person. Suppose this function is identical for two brothers, and it monotonically increases till the age of 50 years and then it monotonically decreases. Let a_1 and a_2 (in years) denote the ages of the brothers and $a_1 < a_2$.</p> <p>Which one of the following statements is correct about their age on the day when they attain the same weight?</p>
(A)	$a_1 < a_2 < 50$
(B)	$a_1 < 50 < a_2$
(C)	$50 < a_1 < a_2$
(D)	Either $a_1 = 50$ or $a_2 = 50$

Q.8	<p>A regular dodecagon (12-sided regular polygon) is inscribed in a circle of radius r cm as shown in the figure. The side of the dodecagon is d cm. All the triangles (numbered 1 to 12) in the figure are used to form squares of side r cm and each numbered triangle is used only once to form a square.</p> <p>The number of squares that can be formed and the number of triangles required to form each square, respectively, are:</p> <p>Note: The figure shown is representative.</p>
	
(A)	3; 4
(B)	4; 3
(C)	3; 3
(D)	3; 2

Q.9	If a real variable x satisfies $3^{x^2} = 27 \times 9^x$, then the value of $\frac{2^{x^2}}{(2^x)^2}$ is:
(A)	2^{-1}
(B)	2^0
(C)	2^3
(D)	2^{15}

Q.10	<p>The number of patients per shift (X) consulting Dr. Gita in her past 100 shifts is shown in the figure. If the amount she earns is ₹1000($X - 0.2$), what is the average amount (in ₹) she has earned per shift in the past 100 shifts?</p> <p>Note: The figure shown is representative.</p>										
	 <table border="1" data-bbox="475 533 1206 1048"> <thead> <tr> <th>Number of patients per shift (X)</th> <th>Number of shifts</th> </tr> </thead> <tbody> <tr> <td>5</td> <td>20</td> </tr> <tr> <td>6</td> <td>40</td> </tr> <tr> <td>7</td> <td>30</td> </tr> <tr> <td>8</td> <td>10</td> </tr> </tbody> </table>	Number of patients per shift (X)	Number of shifts	5	20	6	40	7	30	8	10
Number of patients per shift (X)	Number of shifts										
5	20										
6	40										
7	30										
8	10										
(A)	6,100										
(B)	6,300										
(C)	6,000										
(D)	6,500										

Q. 11 - Q. 35 carry one mark each.

Q. 11 Suppose X and Y are random variables. The conditional expectation of X given Y is denoted by $E[X|Y]$. Then $E[E[X|Y]]$ equals

- (A) $E[X|Y]$
- (B) $\frac{E[X]}{E[Y]}$
- (C) $E[X]$
- (D) $E[Y]$

- Q. 12 The number of additions and multiplications involved in performing Gaussian elimination on any $n \times n$ upper triangular matrix is of the order
- (A) $O(n)$
 - (B) $O(n^2)$
 - (C) $O(n^3)$
 - (D) $O(n^4)$
- Q. 13 The sum of the elements in each row of $A \in \mathbb{R}^{n \times n}$ is 1. If $B = A^3 - 2A^2 + A$, which one of the following statements is correct (for $x \in \mathbb{R}^n$)?
- (A) The equation $Bx = 0$ has no solution
 - (B) The equation $Bx = 0$ has exactly two solutions
 - (C) The equation $Bx = 0$ has infinitely many solutions
 - (D) The equation $Bx = 0$ has a unique solution
- Q. 14 Let $f(x) = \frac{e^x - e^{-x}}{2}, x \in \mathbb{R}$. Let $f^{(k)}(a)$ denote the k^{th} derivative of f evaluated at a . What is the value of $f^{(10)}(0)$? (Note: $!$ denotes factorial)
- (A) 0
 - (B) 1
 - (C) $\frac{1}{10!}$
 - (D) $\frac{2}{10!}$

- Q. 15 Let p and q be any two propositions. Consider the following propositional statements.

$$S_1 : p \rightarrow q, \quad S_2 : \neg p \wedge q, \quad S_3 : \neg p \vee q, \quad S_4 : \neg p \vee \neg q,$$

where \wedge denotes conjunction (AND operation), \vee denotes disjunction (OR operation), and \neg denotes negation (NOT operation). Which one of the following options is correct? (Note: \equiv denotes logical equivalence)

- (A) $S_1 \equiv S_3$
 - (B) $S_2 \equiv S_3$
 - (C) $S_2 \equiv S_4$
 - (D) $S_1 \equiv S_4$
- Q. 16 If a relational decomposition is not dependency-preserving, which one of the following relational operators will be executed more frequently in order to maintain the dependencies?
- (A) Selection
 - (B) Projection
 - (C) Join
 - (D) Set union

Q.17 Consider the following three relations:

Car (model, year, serial, color)
Make (maker, model)
Own (owner, serial)

A tuple in **Car** represents a specific car of a given **model**, made in a given **year**, with a **serial** number and a **color**. A tuple in **Make** specifies that a **maker** company makes cars of a certain **model**. A tuple in **Own** specifies that an **owner** owns the car with a given **serial** number. Keys are underlined; (owner, serial) together form key for **Own**. (\bowtie denotes natural join)

$$\pi_{\text{owner}}(\text{Own} \bowtie (\sigma_{\text{color}=\text{"red"}}(\text{Car} \bowtie (\sigma_{\text{maker}=\text{"ABC"}}\text{Make}))))$$

Which one of the following options describes what the above expression computes?

- (A) All owners of a red car, a car made by ABC, or a red car made by ABC
- (B) All owners of more than one car, where at least one car is red and made by ABC
- (C) All owners of a red car made by ABC
- (D) All red cars made by ABC

Q.18 Consider a hash table of size 10 with indices $\{0, 1, \dots, 9\}$, with the hash function

$$h(x) = 3x \pmod{10},$$

where linear probing is used to handle collisions. The hash table is initially empty and then the following sequence of keys is inserted into the hash table: 1, 4, 5, 6, 14, 15. The indices where the keys 14 and 15 are stored are, respectively

- (A) 2 and 5
- (B) 2 and 6
- (C) 4 and 5
- (D) 4 and 6

- Q.19** Let X be a continuous random variable whose cumulative distribution function (CDF) $F_X(x)$, for some t , is given as follows:

$$F_X(x) = \begin{cases} 0 & x \leq t \\ \frac{x-t}{4-t} & t \leq x \leq 4 \\ 1 & x \geq 4 \end{cases}$$

If the median of X is 3, then what is the value of t ?

- (A) 2
- (B) 1
- (C) -1
- (D) 0

- Q.20** Let $X = aZ + b$, where Z is a standard normal random variable, and a, b are two unknown constants. It is given that

$$E[X] = 1, \quad E[(X - E[X])Z] = -2, \quad E[(X - E[X])^2] = 4,$$

where $E[X]$ denotes the expectation of random variable X . The values of a, b are:

- (A) $a = -2, b = 1$
- (B) $a = 2, b = -1$
- (C) $a = -2, b = -1$
- (D) $a = 1, b = 1$

Q.21 It is given that $P(X \geq 2) = 0.25$ for an exponentially distributed random variable X with $E[X] = \frac{1}{\lambda}$, where $E[X]$ denotes the expectation of X . What is the value of λ ? (\ln denotes natural logarithm)

- (A) $\ln 2$
- (B) $\ln 4$
- (C) $\ln 3$
- (D) $\ln 0.25$

Q.22 Consider designing a linear classifier

$$y = \text{sign}(f(x; w, b)), \quad f(x; w, b) = w^\top x + b$$

on a dataset $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$, $x_i \in \mathbb{R}^d$, $y_i \in \{+1, -1\}$, $i = 1, 2, \dots, N$. Recall that the sign function outputs $+1$ if the argument is positive, and -1 if the argument is non-positive. The parameters w and b are updated as per the following training algorithm:

$$w_{\text{new}} = w_{\text{old}} + y_n x_n, \quad b_{\text{new}} = b_{\text{old}} + y_n$$

whenever $\text{sign}(f(x_n; w_{\text{old}}, b_{\text{old}})) \neq y_n$. In other words, whenever the classifier wrongly predicts a sample (x_n, y_n) from the dataset, w_{old} gets updated to w_{new} , and likewise b_{old} gets updated to b_{new} . Consider the case $(x_n, +1)$, $f(x_n; w_{\text{old}}, b_{\text{old}}) < 0$. Then

- (A) $f(x_n; w_{\text{new}}, b_{\text{new}}) > f(x_n; w_{\text{old}}, b_{\text{old}})$
- (B) $f(x_n; w_{\text{new}}, b_{\text{new}}) < f(x_n; w_{\text{old}}, b_{\text{old}})$
- (C) $f(x_n; w_{\text{new}}, b_{\text{new}}) = f(x_n; w_{\text{old}}, b_{\text{old}})$
- (D) $y_n f(x_n; w_{\text{old}}, b_{\text{old}}) > 1$

Q.23 Consider the following Python declarations of two lists.

A = [1, 2, 3]

B = [4, 5, 6]

Which one of the following statements results in A = [1, 2, 3, 4, 5, 6]?

(A) A.extend(B)

(B) A.append(B)

(C) A.update(B)

(D) A.insert(B)

Q.24 Consider two functions $f : \mathbb{R} \rightarrow \mathbb{R}$ and $g : \mathbb{R} \rightarrow (1, \infty)$. Both functions are differentiable at a point c . Which of the following functions is/are ALWAYS differentiable at c ? The symbol \cdot denotes product and the symbol \circ denotes composition of functions.

(A) $f \pm g$

(B) $f \cdot g$

(C) $\frac{f}{g}$

(D) $f \circ g + g \circ f$

Q.25 Which of the following statements is/are correct?

(A) \mathbb{R}^n has a unique set of orthonormal basis vectors

(B) \mathbb{R}^n does not have a unique set of orthonormal basis vectors

(C) Linearly independent vectors in \mathbb{R}^n are orthonormal

(D) Orthonormal vectors in \mathbb{R}^n are linearly independent

Q.26 Which of the following statements is/are correct in a Bayesian network?

- (A) Variable elimination is an approximate inference algorithm
- (B) Gibbs sampling is an exact inference algorithm
- (C) Variable elimination is used to determine conditional probabilities
- (D) Rejection sampling is an approximate inference algorithm

Q.27 For which of the following inputs does binary search take time $O(\log n)$ in the worst case?

- (A) An array of n integers in any order
- (B) A linked list of n integers in any order
- (C) An array of n integers in increasing order
- (D) A linked list of n integers in increasing order

Q.28 Let $A = I_n + xx^\top$, where I_n is the $n \times n$ identity matrix and $x \in \mathbb{R}^n, x^\top x = 1$. Which of the following options is/are correct?

- (A) Rank of A is n
- (B) A is invertible
- (C) 0 is an eigenvalue of A
- (D) A^{-1} has a negative eigenvalue

Q.29 Suppose that insertion sort is applied to the array $[1, 3, 5, 7, 9, 11, x, 15, 13]$ and it takes exactly two swaps to sort the array. Select all possible values of x .

- (A) 10
- (B) 12
- (C) 14
- (D) 16

Q.30 Let C_1 and C_2 be two sets of objects. Let $D(x, y)$ be a measure of dissimilarity between two objects x and y . Consider the following definitions of dissimilarity between C_1 and C_2 .

$$\mathbf{DIS-1}(C_1, C_2) = \max_{x \in C_1, y \in C_2} D(x, y)$$

$$\mathbf{DIS-2}(C_1, C_2) = \min_{x \in C_1, y \in C_2} D(x, y)$$

Which of the following statements is/are correct?

- (A) Single Linkage Clustering uses **DIS-1**
- (B) Single Linkage Clustering uses **DIS-2**
- (C) Complete Linkage Clustering uses **DIS-2**
- (D) Complete Linkage Clustering uses **DIS-1**

Q.31 There are three boxes containing white balls and black balls.

- Box-1 contains 2 black and 1 white balls.
- Box-2 contains 1 black and 2 white balls.
- Box-3 contains 3 black and 3 white balls.

In a random experiment, one of these boxes is selected, where the probability of choosing Box-1 is $\frac{1}{2}$, Box-2 is $\frac{1}{6}$, and Box-3 is $\frac{1}{3}$. A ball is drawn at random from the selected box. Given that the ball drawn is white, the probability that it is drawn from Box-2 is _____
(Round off to two decimal places)

Q.32 $\lim_{t \rightarrow +\infty} \sqrt{t^2 + t} - t = \underline{\hspace{2cm}}$

(Round off to one decimal place)

Q.33 On a relation named **Loan** of a bank:

Loan		
loan_number	branch_name	amount
L11	Banjara Hills	90000
L14	Kondapur	50000
L15	SR Nagar	40000
L22	SR Nagar	25000
L23	Balanagar	80000
L25	Kondapur	70000
L19	SR Nagar	65000

the following SQL query is executed.

```
SELECT L1.loan_number
```

```
FROM Loan L1
```

```
WHERE L1.amount > (SELECT MAX (L2.amount)
```

```
FROM Loan L2
```

```
WHERE L2.branch_name = 'SR Nagar');
```

The number of rows returned by the query is _____ (Answer in integer)

Q.34 Given data $\{(-1, 1), (2, -5), (3, 5)\}$ of the form (x, y) , we fit a model $y = wx$ using linear least-squares regression. The optimal value of w is _____

(Round off to three decimal places)

- Q.35** The naive Bayes classifier is used to solve a two-class classification problem with class-labels y_1, y_2 . Suppose the prior probabilities are $P(y_1) = \frac{1}{3}$ and $P(y_2) = \frac{2}{3}$. Assuming a discrete feature space with

$$P(x|y_1) = \frac{3}{4} \quad \text{and} \quad P(x|y_2) = \frac{1}{4}$$

for a specific feature vector x . The probability of misclassifying x is

(Round off to two decimal places)

Q. 36 – Q. 65 carry two marks each.

- Q.36** Let $Y = Z^2, Z = \frac{X-\mu}{\sigma}$, where X is a normal random variable with mean μ and variance σ^2 . The variance of Y is

- (A) 1
- (B) 2
- (C) 3
- (D) 4

- Q.37** Let $A \in \mathbb{R}^{n \times n}$ be such that $A^3 = A$. Which one of the following statements is ALWAYS correct?

- (A) A is invertible
- (B) Determinant of A is 0
- (C) The sum of the diagonal elements of A is 1
- (D) A and A^2 have the same rank

Q.38 Let $\{x_1, x_2, \dots, x_n\}$ be a set of linearly independent vectors in \mathbb{R}^n . Let the (i, j) -th element of matrix $A \in \mathbb{R}^{n \times n}$ be given by $A_{ij} = x_i^\top x_j$, $1 \leq i, j \leq n$. Which one of the following statements is correct?

- (A) A is invertible
- (B) 0 is a singular value of A
- (C) Determinant of A is 0
- (D) $z^\top A z = 0$ for some non-zero $z \in \mathbb{R}^n$

Q.39 Consider the cumulative distribution function (CDF) of a random variable X :

$$F_X(x) = \begin{cases} 0 & x \leq -1 \\ \frac{1}{4}(x+1)^2 & -1 \leq x \leq 1 \\ 1 & x \geq 1 \end{cases}$$

The value of $P(X^2 \leq 0.25)$ is

- (A) 0.625
- (B) 0.25
- (C) 0.5
- (D) 0.5625

- Q.40** A random variable X is said to be distributed as *Bernoulli*(θ), denoted by $X \sim \text{Bernoulli}(\theta)$, if

$$P(X = 1) = \theta, \quad P(X = 0) = 1 - \theta$$

for $0 < \theta < 1$. Let $Y = \sum_{i=1}^{300} X_i$, where $X_i \sim \text{Bernoulli}(\theta)$, $i = 1, 2, \dots, 300$ be independent and identically distributed random variables with $\theta = 0.25$. The value of $P(60 \leq Y \leq 90)$, after approximation through Central Limit Theorem, is given by

(Recall that $\phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt$)

- (A) $\phi(2) - \phi(-2)$
- (B) $\phi(1) - \phi(-1)$
- (C) $\phi(3) - \phi(-3)$
- (D) $\phi(90) - \phi(60)$

- Q.41** For $x \in \mathbb{R}$, the floor function is denoted by $f(x) = \lfloor x \rfloor$ and defined as follows

$$\lfloor x \rfloor = k, \quad k \leq x < k + 1,$$

where k is an integer. Let $Y = \lfloor X \rfloor$, where X is an exponentially distributed random variable with mean $\frac{1}{\ln 10}$, where \ln denotes natural logarithm. For any positive integer ℓ , one can write the probability of the event $Y = \ell$ as follows

$$P(Y = \ell) = q^\ell(1 - q)$$

The value of q is

- (A) 0.1
- (B) 0.01
- (C) 0.5
- (D) 0.434

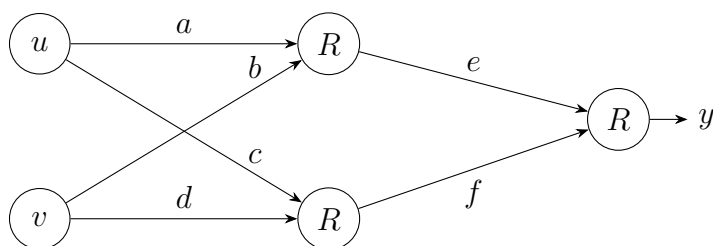
Q.42 Consider the neural network shown in the figure with

inputs: u, v

weights: a, b, c, d, e, f

output: y

R denotes the ReLU function, $R(x) = \max(0, x)$.



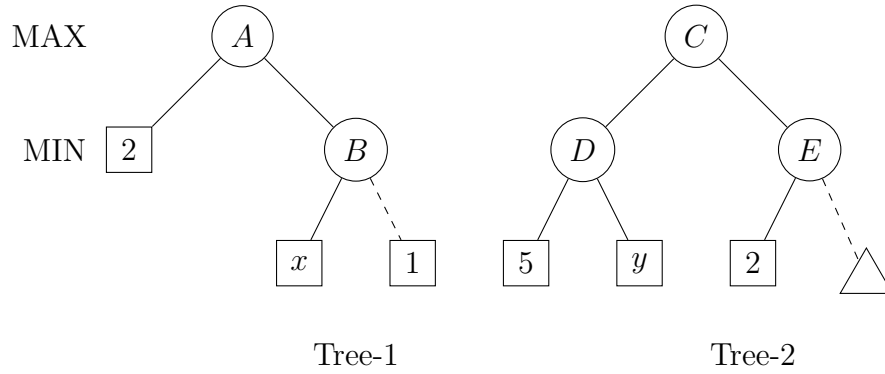
Given $u = 2, v = 3$,

$a = 1, b = 1, c = 1, d = -1, e = 4, f = -1$,

which one of the following is correct?

- (A) $\frac{\partial y}{\partial a} = 8, \frac{\partial y}{\partial f} = 0$
- (B) $\frac{\partial y}{\partial a} = 1, \frac{\partial y}{\partial f} = 0$
- (C) $\frac{\partial y}{\partial a} = 1, \frac{\partial y}{\partial f} = -1$
- (D) $\frac{\partial y}{\partial a} = 2, \frac{\partial y}{\partial f} = -1$

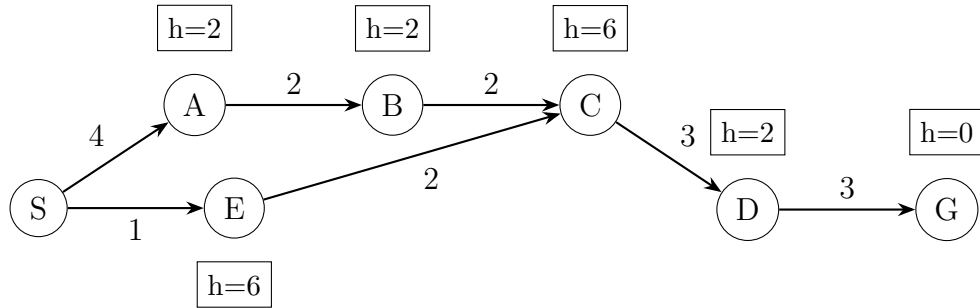
- Q.43** Consider game trees Tree-1 and Tree-2 as shown. The first level is a MAX agent and the second level is a MIN agent. The value in the square node is the output of the utility function.



For what ranges of x and y , the right child of node B and the right child of node E will be pruned by alpha-beta pruning algorithm?

- (A) $x \in [1, \infty)$ and $y \in (-\infty, 2]$
- (B) $x \in (-\infty, 2]$ and $y \in (-\infty, 5]$
- (C) $x \in (-\infty, 2]$ and $y \in [2, \infty)$
- (D) $x \in [1, \infty)$ and $y \in (-\infty, 5]$

- Q.44** The state graph shows the action cost along the edges and the heuristic function h associated with each state.



Suppose A^* algorithm is applied on this state graph using priority queue to store the frontier. In what sequence are the nodes expanded?

- (A) S, A, E, C, B, D, G
- (B) S, E, A, C, B, D, G
- (C) S, A, E, B, C, D, G
- (D) S, A, B, E, C, D, G
- Q.45** A random experiment consists of throwing 100 fair dice, each die having six faces numbered 1 to 6. An event A represents the set of all outcomes where at least one of the dice shows a 1. Then, $P(A) =$
- (A) 0
- (B) 1
- (C) $1 - \left(\frac{5}{6}\right)^{100}$
- (D) $\left(\frac{5}{6}\right)^{100}$

- Q.46** Consider a fact table in an OLAP application: **Facts**(D1, D2, val), where D1 and D2 are its dimension attributes and val is a dependent attribute. Suppose attribute D1 takes 3 values and D2 takes 2 values, and all combinations of these values are present in the table **Facts**. How many tuples are there in the result of the following query?

```
SELECT D1, D2, sum(val)

FROM Facts

GROUP BY CUBE (D1, D2);
```

- (A) 1
- (B) 6
- (C) 9
- (D) 12

Q.47 Consider the following Python code snippet.

```
A={"this", "that"}
B={"that", "other"}
C={"other", "this"}

while "other" in C:

    if "this" in A:

        A, B, C = A-B, B-C, C-A

    if "that" in B:

        A, B, C = C|A, A|B, B|C
```

When the above program is executed, at the end, which of the following sets contains "this"?

- (A) Only A
- (B) Only B
- (C) Only C
- (D) A, C

Q.48 Which of the following statements is/are correct about the rectified linear unit (ReLU) activation function defined as $\text{ReLU}(x) = \max(x, 0)$, where $x \in \mathbb{R}$?

- (A) ReLU is continuous everywhere
- (B) ReLU is differentiable everywhere
- (C) ReLU is not differentiable at $x = 0$
- (D) $\text{ReLU}(x) = \text{ReLU}(ax)$, for all $a \in \mathbb{R}$

Q.49 Consider the function $f(x) = \frac{x^3}{3} + \frac{7}{2}x^2 + 10x + \frac{133}{2}$, $x \in [-8, 0]$. Which of the following statements is/are correct?

- (A) The maximum value of f is attained at $x = -5$
- (B) The minimum value of f is attained at $x = -2$
- (C) The maximum value of f is $\frac{133}{2}$
- (D) The minimum value of the derivative of f is attained at $x = -\frac{7}{2}$

Q.50 Let x_1, x_2, x_3, x_4, x_5 be a system of orthonormal vectors in \mathbb{R}^{10} . Consider the matrix $A = x_1x_1^\top + \cdots + x_5x_5^\top$. Which of the following statements is/are correct?

- (A) Singular values of A are also its eigenvalues
- (B) Singular values of A are either 0 or 1
- (C) Determinant of A is 1
- (D) A is invertible

Q.51 Let $f : \mathbb{R} \rightarrow \mathbb{R}$ be a twice-differentiable function and suppose its second derivative satisfies $f''(x) > 0$ for all $x \in \mathbb{R}$. Which of the following statements is/are ALWAYS correct?

- (A) f has a local minima
- (B) There does not exist x and $y, x \neq y$, such that $f'(x) = f'(y) = 0$
- (C) f has at most one global minimum
- (D) f has at most one local minimum

Q.52 An $n \times n$ matrix A with real entries satisfies the property: $\|Ax\|^2 = \|x\|^2$, for all $x \in \mathbb{R}^n$, where $\|\cdot\|$ denotes the Euclidean norm. Which of the following statements is/are ALWAYS correct?

- (A) A must be orthogonal
- (B) $A = I$, where I denotes the identity matrix, is the only solution
- (C) The eigenvalues of A are either $+1$ or -1
- (D) A has full rank

Q.53 Consider designing a linear binary classifier $f(x) = \text{sign}(w^\top x + b)$, $x \in \mathbb{R}^2$ on the following training data:

$$\text{Class-1: } \left\{ \begin{pmatrix} 2 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ 2 \end{pmatrix} \right\}, \text{ Class-2: } \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix} \right\}$$

Hard-margin support vector machine (SVM) formulation is solved to obtain w and b . Which of the following options is/are correct?

- (A) $w = \begin{pmatrix} 4 \\ 4 \end{pmatrix}$ and $b = 1$
- (B) The number of support vectors is 3
- (C) The margin is $\sqrt{2}$
- (D) Training accuracy is 98%

- Q.54** Consider a coin-toss experiment where the probability of head showing up is p . In the i^{th} coin toss, let $X_i = 1$ if head appears, and $X_i = 0$ if tail appears. Consider

$$\hat{p} = \frac{1}{n} \sum_{i=1}^n X_i$$

where n is the total number of independent coin tosses. Which of the following statements is/are correct?

- (A) $E[\hat{p}] = p$
- (B) $E[\hat{p}] = \frac{p}{n}$
- (C) As n increases, variance of \hat{p} decreases
- (D) Variance of \hat{p} does not depend on n

- Q.55** Consider a two-class problem in \mathbb{R}^d with class labels *red* and *green*. Let μ_{red} and μ_{green} be the means of the two classes. Given test sample $x \in \mathbb{R}^d$, a classifier calculates the squared Euclidean distance (denoted by $\|\cdot\|^2$) between x and the means of the two classes and assigns the class label that the sample x is closest to. That is, the classifier computes

$$f(x) = \|\mu_{\text{red}} - x\|^2 - \|\mu_{\text{green}} - x\|^2$$

and assigns the label *red* to x if $f(x) < 0$, and *green* otherwise. Which of the following statements is/are correct?

- (A) The sample $x = 0$ is assigned the label *green* if $\|\mu_{\text{red}}\| < \|\mu_{\text{green}}\|$
- (B) f is a linear function of x
- (C) $f(x) = w^\top x + b$, where w and b are functions of μ_{red} and μ_{green}
- (D) f is a quadratic polynomial in x

Q.56 Consider the following two relations, named **Customer** and **Person**, in a database:

Person (

aadhaar CHAR(12) PRIMARY KEY,

name VARCHAR(32));

Customer (

name VARCHAR(32),

email VARCHAR(32) PRIMARY KEY,

phone CHAR(10),

aadhaar CHAR(12),

 FOREIGN KEY (**aadhaar**) REFERENCES **Person**(**aadhaar**));

Which of the following statements is/are correct?

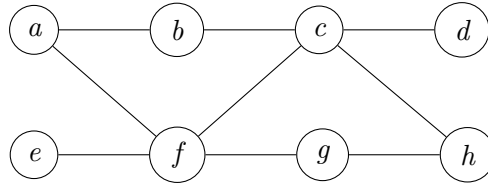
- (A) **aadhaar** is a candidate key in the **Customer** relation
- (B) **phone** can be NULL in the **Customer** relation
- (C) **aadhaar** is a candidate key in the **Person** relation
- (D) **aadhaar** can be NULL in the **Person** relation

- Q.57** Consider a database relation R with attributes $ABCDEFG$, and having the following functional dependencies:

$$A \rightarrow BCEF, \quad E \rightarrow DG, \quad BC \rightarrow A$$

Which of the following statements is/are correct?

- (A) A is the only candidate key of R
 - (B) A, BC are the candidate keys of R
 - (C) A, BC, E are the candidate keys of R
 - (D) Relation R is not in Boyce-Codd Normal Form (BCNF)
- Q.58** Let G be a simple, unweighted, and undirected graph. A subset of the vertices and edges of G are shown below.



It is given that $a - b - c - d$ is a shortest path between a and d ; $e - f - g - h$ is a shortest path between e and h ; $a - f - c - h$ is a shortest path between a and h . Which of the following is/are NOT the edges of G ?

- (A) (b, d) (B) (b, g) (C) (b, h) (D) (e, g)
- Q.59** Let $f : \mathbb{R} \rightarrow \mathbb{R}$ be such that $|f(x) - f(y)| \leq (x - y)^2$ for all $x, y \in \mathbb{R}$. Then $f(1) - f(0) = \underline{\hspace{2cm}}$ (Answer in integer)

- Q.60** Let $D = \{x^{(1)}, \dots, x^{(n)}\}$ be a dataset of n observations where each $x^{(i)} \in \mathbb{R}^{100}$. It is given that $\sum_{i=1}^n x^{(i)} = 0$. The covariance matrix computed from D has eigenvalues $\lambda_i = 100^{2-i}, 1 \leq i \leq 100$. Let $u \in \mathbb{R}^{100}$ be the direction of maximum variance with $u^\top u = 1$.

The value of

$$\frac{1}{n} \sum_{i=1}^n (u^\top x^{(i)})^2 = \underline{\hspace{2cm}}$$

- Q.61** A bag contains 5 white balls and 10 black balls. In a random experiment, n balls are drawn from the bag one at a time with replacement. Let S_n denote the total number of black balls drawn in the experiment.

The expectation of S_{100} denoted by $E[S_{100}] = \underline{\hspace{2cm}}$

(Round off to one decimal place)

Q.62 Consider the following tables, **Loan** and **Borrower**, of a bank.

Loan		
loan_num	branch_name	amount
L11	Banjara Hills	90000
L14	Kondapur	50000
L15	SR Nagar	40000
L22	SR Nagar	25000
L23	Balanagar	80000
L25	Kondapur	70000
L19	SR Nagar	65000

Borrower	
customer_name	loan_num
Anand	L11
Karteek	L11
Karteek	L14
Ankita	L15
Gopal	L19
Karteek	L22
Karteek	L23
Sunil	L23
Sunil	L25

Query: $\pi_{\text{branch_name, customer_name}}(\text{Loan} \bowtie \text{Borrower}) \div \pi_{\text{branch_name}}(\text{Loan})$

where \bowtie denotes natural join.

The number of tuples returned by the above relational algebra query is _____

(Answer in integer)

Q.63 Consider the following Python code snippet.

```
def f(a, b):
    if (a == 0):
        return b
    if (a % 2 == 1):
        return 2 * f((a - 1) / 2, b)
    return b + f(a - 1, b)

print(f(15, 10))
```

The value printed by the code snippet is ____.

(Answer in integer)

Q.64 Consider the following pseudocode.

```
Create empty stack S
Set x=0, flag=0, sum=0
Push x onto S
while (S is not empty){
    if (flag equals 0){
        Set x = x+1
        Push x onto S}
    if (x equals 8):
        Set flag=1
    if (flag equals 1){
        x = Pop(S)
        if (x is odd):
            Pop(S)
        Set sum = sum + x}
}
Output sum
```

The value of **sum** output by a program executing the above pseudocode is ____.

(Answer in integer)

Q.65 Consider a directed graph $G = (V, E)$, where $V = \{0, 1, 2, \dots, 100\}$ and $E = \{(i, j) : 0 < j - i \leq 2, \text{ for all } i, j \in V\}$. Suppose the adjacency list of each vertex is in *decreasing* order of vertex number, and depth-first search (DFS) is performed at vertex 0. The number of vertices that will be discovered after vertex 50 is ____.

(Answer in integer)