Data Science and Artificial Intelligence (DA)

Note: The DA test paper is newly introduced in GATE 2024. This is a sample question paper for the candidates to practice the subject. We will not be able to answer any queries or provide answer keys to these questions.

Q.1 – Q.25 Carry ONE mark each.	
Q.1	Let b be the branching factor of a search tree. If the optimal goal is reached after d actions from the initial state, in the worst case, how many times will the initial state be expanded for iterative deepening depth first search (IDDFS) and iterative Deepening A* search (IDA*)?
(A)	IDDFS - d, IDA* -d.
(B)	IDDFS - d, IDA* -b^d
(C)	IDDFS - b^d, IDA* -d.
(D)	IDDFS - b^d, IDA* -b^d.
Q.2	Given 3 literals A, B, and C, how many models are there for the sentence $A \vee \neg B \vee C$?

Q.3	Which of the following first order logic sentence matches closest with the sentence "All students are not equal"?
(A)	$\forall x\ \exists y[student(x) \land student(y)] \Rightarrow \neg Equal(x, y)$
(B)	$\forall x\ \forall y[student(x) \land student(y)] \Rightarrow \neg Equal(x, y)$
(C)	$\forall x\ \exists y[student(x) \land student(y) \land \neg Equal(x, y)]$
(D)	$\forall x\ \forall y[student(x) \land student(y) \land \neg Equal(x, y)]$
Q.4	The mean of the observations of the first 50 observations of a process is 12. If the 51st observation is 18, then, the mean of the first 51 observations of the process is
(A)	12
(B)	12.12
(C)	12.36
(D)	18

Q.5	$\lim_{x \to 2} \frac{\sqrt{x} - \sqrt{2}}{x - 2}$
(A)	0
(B)	$\sqrt{2}$
(C)	$\frac{1}{2\sqrt{2}}$
(D)	$\frac{1}{\sqrt{2}}$
Q.6	Which among the following may help to reduce overfitting demonstrated by a model i) Change the loss function. ii) Reduce model complexity. iii) Increase the training data. iv) Increase the number of optimization routine steps.
(A)	ii and i
(B)	ii and iii
(C)	i, ii, and iii
(D)	i, ii, iii, and iv

Q.7	A fair coin is flipped twice and it is known that at least one tail is observed. The probability of getting two tails is
(A)	$\frac{1}{2}$
(B)	$\frac{1}{3}$
(C)	$\frac{2}{3}$
(D)	$\frac{1}{4}$
Q.8	Given *n* indistinguishable particles and *m* (> *n*) distinguishable boxes, we place at random each particle in one of the boxes. The probability that in n preselected boxes, one and only one particle will be found is
(A)	$$\frac{n!}{m^n}$$
(B)	$$\frac{(m-1)!\,n!}{(m+n-1)!}$$
(C)	$$\frac{1}{m^n}$$
(D)	$$\frac{1}{m}$$

Q.9	For two events A and B, $B \subset A$ Which of the following statement is correct?
(A)	$P(B \mid A) \geq P(B)$
(B)	$P(B \mid A) \leq P(B)$
(C)	$P(A \mid B) < 1$
(D)	$P(A \mid B) = 0$
Q.10	X is a uniform distribution random variable with support in [-2, 2] U [99.5, 100.5]. The mean of X is _______
Q. 11	You are reviewing four papers submitted to a conference on machine learning for medical expert systems. All the four papers validate their superiority on a standard benchmarking cancer dataset, which has only 5% of positive cancer cases. Which of the experimental setting is acceptable to you? paper i) We evaluated the performance of our model through a 5-fold cross validation process and report an accuracy of 93%. paper ii) The area under the ROC curve on a single left out test set of our model is around 0.8, which is the highest among all the different approaches. paper iii) We computed the average area under the ROC curve through 5-fold cross validation and found it to be around 0.75 – the highest among all the approaches.

	paper iv) The accuracy on a single left out test set of our model is 95%, which is the highest among all the different approaches.
(A)	paper i
(B)	paper i and paper iv
(C)	paper ii and paper iv
(D)	paper iii
Q.12	Increasing the regularizing coefficient value for a ridge regressor will i) Increase or maintain model bias. ii) Decrease model bias. iii) Increase or maintain model variance. iv) Decrease model variance.
(A)	i and iii
(B)	i and iv
(C)	ii and iii
(D)	ii and iv

Q.13	A decision tree classifier learned from a fixed training set achieves 100% accuracy. Which of the following models trained using the same training set will also achieve 100% accuracy? i) Logistic regressor. ii) A polynomial of degree one kernel SVM. iii) A linear discriminant function. iv) Naïve Bayes classifier.
(A)	i
(B)	i and ii
(C)	all of the above
(D)	none of the above
Q.14	Consider two relations R(x, y) and S(x,z). Relation R has 100 records, and relation S has 200 records. What will be the number of attributes and records of the following query? SELECT * from R CROSS JOIN S;
(A)	3 attributes, 20000 records
(B)	4 attributes, 20000 records

(C)	3 attributes, 200 records
(D)	4 attributes, 200 records
Q.15	Consider two relations R(x, y) and S(y), and perform the following operation R(x,y) DIVIDE S(Y) If X is the relation returns by the above operation, which of the following option(s) is/are always TRUE?
(A)	$\|X\| \leq \|R\|$
(B)	$\|X\| \leq \|S\|$
(C)	$\|X\| \leq \|R\|$ AND $\|X\| \leq \|S\|$
(D)	All of the Above
Q.16	Which of the following statements is/are TRUE?
(A)	Every relation with two attributes is also in BCNF.
(B)	Every relation in BCNF is also in 3NF.

(C)	No relation can be in both BCNF and 3NF.
(D)	None of Above
Q.17	For matrix $H = \begin{bmatrix} 9 & -2 \\ -2 & 6 \end{bmatrix}$, one of the eigenvalues is 5. Then, the other eigenvalue is
(A)	12
(B)	10
(C)	8
(D)	6
Q. 18	Two non-zero vectors **x** and **y** are perpendicular if
(A)	$\mathbf{x}^T\mathbf{y}=0$
(B)	$\mathbf{x}^T\mathbf{y}>0$
(C)	$\mathbf{x}^T\mathbf{y}<0$
(D)	None of the above.

GATE 2024
IISc Bengaluru

Q. 19	The function $f(x) = 1+x+x^2$ has a
(A)	Minima at $x=-0.5$
(B)	Maxima at $x=-0.5$
(C)	Saddle point at at $x=-0.5$
(D)	None of the above.
Q. 20	The Pearson's correlation coefficient between x and y rounded to the first decimal point for the given data in below table is

x	**y**
-6	6.4
2	4.7
0.2	8
7	2
-4	3.4

(A)	-0.5
(B)	0.5
(C)	0.3
(D)	-0.3

Q.21	The worst-case running times of Insertion sort, Merge sort and Quick sort respectively are
(A)	$\Theta(n\log n)$, $\Theta(n^2)$, $\Theta(n^2)$
(B)	$\Theta(n^2)$, $\Theta(n\log n)$, $\Theta(n\log n)$
(C)	$\Theta(n^2)$, $\Theta(n\log n)$, $\Theta(n^2)$
(D)	$\Theta(n^2)$, $\Theta(n^2)$, $\Theta(n\log n)$

Q.22	**Consider the following program.**
	`int func(int n)`
	`{`
	`    if(n <= 1)`
	`        return n;`
	`    else`
	`        return 3*func(n-3) - 3*func(n-2);`
	`    end`
	`}`
	The running time of the above function is

(A)	$\Theta(n)$
(B)	$\Theta(n^2)$
(C)	$\Theta(3^n)$
(D)	$\Theta(2^n)$

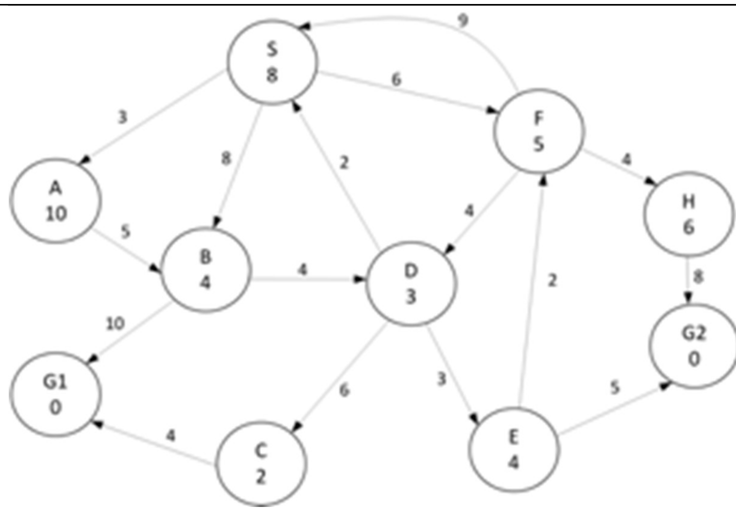
Q.23	Which of the following correctly describes the recurrence relation for the standard binary search algorithm on a sorted array of n numbers where c is a constant.
(A)	T(n) = 2*T(n/2) + c
(B)	T(n) = T(n/2)
(C)	T(n) = T(n-1) + c
(D)	T(n) = T(n/2) + c

Q.24	**Consider the following C program**
	`int func(int A[], int n, int m)`
	`{`
	`    int s = A[0];`
	`    for(int i = 1; i <= n - 1; i++)`
	`        total = m*s + A[i];`
	`    return m;`
	`}`
	Let Z be an array of 10 elements with Z[i] = 2 for all i such that 0<=i<=9; The value returned by func(Z,10,2) is _______

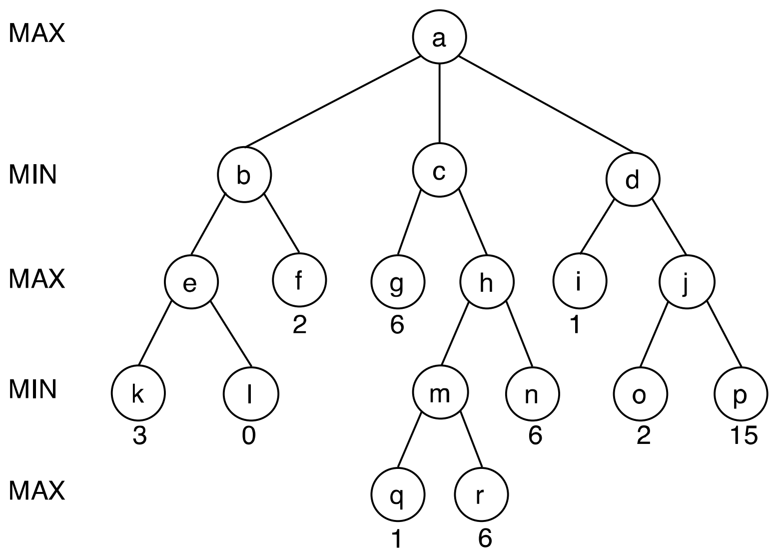
Q.25	Two eigenvalues of 3 x 3 matrix **X** are (1 + *i*) and 2. The determinant of the matrix X is _______

Q.26 – Q.55 Carry TWO marks each.	

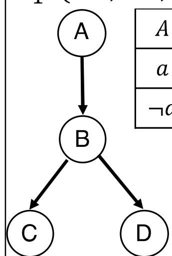
Q.26	Given the following relation instances X Y Z 1 4 2 1 5 3 1 4 3 1 5 2 3 2 1 Which of the following conditions is/are TRUE?
(A)	XY -> Z and Z -> Y
(B)	YZ -> X and X ->> Y
(C)	Y -> X and Y ->> X
(D)	XZ -> Y and Y -> X
Q.27	Consider the search space depicted in the Figure below. S is the initial state. G1 and G2 are two states that satisfy the goal test. The cost of traversing from one state to another is depicted by the numerical value close to the edge connecting the two states. The estimated cost to the goal is reported inside the states. Use alphabetical order of nodes to break ties. Which goal state is reached if you perform A* (graph) search? What is the largest value that the heuristic function can take for node A while still being admissible?

	
(A)	G1 and 16
(B)	G1 and 15
(C)	G2 and 16
(D)	G2 and 15
Q.28	Given a discrete K-class dataset containing N points, where sample points are described using D features with each feature capable of taking V values, how many parameters need to be estimated for Naïve Bayes Classifier?
(A)	$V^D K$
(B)	K^{V^D}

(C)	VDK
(D)	$K(V + D)$
Q.29	$K(x, x') = f(x)g(x') + f(x')g(x)$, where f and g are real-valued functions ($\mathcal{R}^D \rightarrow \mathcal{R}$) is not a valid kernel function. What additional terms would you include to the summation in K to make it a valid kernel?
(A)	$f(x) + g(x)$
(B)	$f(x)g(x) + f(x')g(x')$
(C)	$f(x)f(x') + g(x)g(x')$
(D)	$f(x') + g(x')$
Q.30	For perfectly spherical 2D data centered at the origin, which of the following the pairs of vectors are possible pairs of principal components? i) (1, 0) and (0, 1) ii) (0, -1) and (-1, 0) iii) (1, 1) and (1, -1) (-1, 1) and (-1, -1)
(A)	i
(B)	i and iii

(C)	i, ii, and iii
(D)	i, ii, iii and iv
Q.31	Consider the game tree shown below. The value below each node is the output of the utility function. The subtrees rooted at which of these nodes will be pruned because of alpha-beta pruning? 
(A)	m and j
(B)	r and j
(C)	h and p
(D)	no nodes are pruned

Q.32	Consider the Bayes Net containing four Boolean random variables $(A, B, C\ , D)$, with the following convention: $A = True \Rightarrow A = a, and\ A = False \Rightarrow A = \neg a$; and similarly for the other variables. The conditional probability tables for the nodes in the network are also indicated in the figure. The following samples were generated through likelihood weighting: $s_1: (\neg a, \neg b, \neg c, \neg d); s_2: (\neg a, b, \neg c, \neg d);\ s_3: (\neg a, \neg\ b, \neg c, d); s_4: (\neg a, b, \neg c, d)$



A	$P(A)$
a	1/5
$\neg a$	4/5

A	B	$P(B\|A)$
a	b	1/5
$\neg a$	b	1/2
a	$\neg b$	4/5
$\neg a$	$\neg b$	1/2

B	C	$P(C\|B)$
b	c	1/4
$\neg b$	c	3/4
b	$\neg c$	2/5
$\neg b$	$\neg c$	3/5

B	D	$P(D\|B)$
b	d	1/2
$\neg b$	d	4/5
b	$\neg d$	1/2
$\neg b$	$\neg d$	1/5

Estimate the likelihood weight of each sample and thereby estimate $P(b|\neg a, \neg c)$

(A)	$s_1: 0.48, s_2: 0.32, s_3: 0.48, s_4: 0.32, P(b|\neg a, \neg c) = 0.4$
(B)	$s_1: 0.48, s_2: 0.32, s_3: 0.48, s_4: 0.32, P(b|\neg a, \neg c) = 0.64$
(C)	$s_1: 0.32, s_2: 0.48, s_3: 0.48, s_4: 0.32, P(b|\neg a, \neg c) = 0.64$
(D)	$s_1: 0.48, s_2: 0.32, s_3: 0.32, s_4: 0.32, P(b|\neg a, \neg c) = 0.4$
Q.33	X is a uniformly distributed random variable from 0 to 1 $f(x) = \begin{cases} 1, & 0 \le x \le 1 \\ 0, & \text{otherwise} \end{cases}$

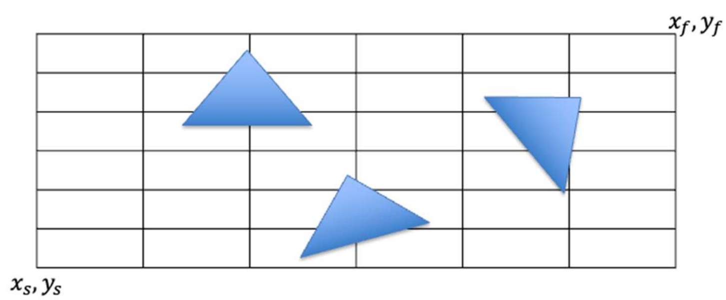
	The variance of X is
(A)	$\frac{1}{2}$
(B)	$\frac{1}{3}$
(C)	$\frac{1}{4}$
(D)	$\frac{1}{12}$
Q.34	The function $f(x) = 1 + 2x + 3x^2 + \cdots + 2026x^{2025}$. Which of the following statement is true?
(A)	*f(x)* has global minimum
(B)	*f(x)* has global maximum
(C)	*f(x)* does not have global minimum
(D)	None of the above

Q.35	Given a smooth sufficiently differentiable function, the following statements are given (P) A concave function can have a global minimum (Q) All convex functions have a global minimum
(A)	P and Q are true
(B)	P is true and Q is false
(C)	P is false and Q is true
(D)	P and Q are false
Q.36	Consider the following joint distribution of random variables X and Y: $$f(x,y) = \begin{cases} \dfrac{x(1+3y^2)}{4}, & 0 \le x \le 2, 0 \le y \le 1 \\ 0, & \text{otherwise} \end{cases}$$ Which one or more of the following statements is/are correct?
(A)	X and Y are mutually uncorrelated.
(B)	X and Y are mutually independent.
(C)	The mean of X is 1.
(D)	The mean of Y is 0.5

Q.37	For matrix $H = \begin{bmatrix} 3 & -1 \\ -1 & 3 \end{bmatrix}$, one of the eigenvectors is $\begin{bmatrix} -1 \\ -1 \end{bmatrix}$. Then, the other eigenvector is
(A)	$\begin{bmatrix} 1 \\ -1 \end{bmatrix}$.
(B)	$\begin{bmatrix} 1 \\ 1 \end{bmatrix}$.
(C)	$\begin{bmatrix} 1 \\ 0 \end{bmatrix}$.
(D)	$\begin{bmatrix} 0 \\ 1 \end{bmatrix}$.
Q.38	Given a matrix $\mathbf{A}_{m \times n}$. The following statements are made regarding the matrix **A**. P. The column space is orthogonal to the row space Q. The column space is orthogonal to the left null space R. The row space is orthogonal to the null space T. The null space is orthogonal to the left null space. Which of the statement(s) is/are true?
(A)	P and Q
(B)	P and R
(C)	Q and R

(D)	P and T
Q.39	Consider a matrix $\begin{bmatrix} 0 & 1 & 0 \\ a & 2 & d \\ b & 3 & c \end{bmatrix}$. The matrix cannot have rank
(A)	0
(B)	1
(C)	2
(D)	3
Q.40	A file with 100,000 records is indexed with B+ tree. If the size of a memory block is 2K bytes, the size of a key is 4 bytes, the size of a pointer is 4 bytes, what is the minimum possible height of the B+ tree index. Height is always greater than equal to 1. Hints: No records are store in the nodes, only keys are stored. The sizes of the pointers are same, irrespective of they point to a node of a record.
Q.41	Consider a schema R(A, B, C, D, E, F) and functional dependencies A -> B, C -> D, and E->F. What is the number of superkeys?

Q.42	Given the dataset: (1, 1), (3, 3), (4, 4), (5, 5), (6, 6), (9, 9), (0, 3), (3, 0) and assuming the initial centroids for ($K = 3$ – means clustering) to be $C_1 = (3, 3), \mathrm{C}_2 = (5, 5)$ and $\mathrm{C}_3 = (6, 6)$. One iteration of the Expectation Maximization Algorithm for K-means clustering, will update C_3 to (___, ___)
Q.43	Consider a Multi-Layer Perceptron (MLP) model with one hidden layer and one output layer. The hidden layer has 10 neurons, and the output layer has 3 neurons. The input to the MLP is a 5-dimensional vector. Each neuron is connected to every neuron in the previous layer, and a bias term is included for each neuron. The activation function used is the sigmoid function. Calculate the total number of trainable parameters in this MLP model.
Q.44	A company manufactures a product at the rate of P units per day. The cost per unit in Rs is $C = 50 + 0.1P + 9000/P$. The selling price per unit is Rs. 300. The production level minimizing the cost per unit and the total profit, respectively, are
(A)	300, 1250
(B)	150, 2500
(C)	300, 2500
(D)	150, 1250

Q.45	A class contains 60% students who are incapable of changing their opinions about anything, and 40% of students are changing their minds at random, with probability 0.3, between subsequent votes on the same issue. Then, the probability of a student randomly chosen voted twice in the same way is _______.
Q.46	Consider the grid world shown in the figure below. An agent is planning to move from the starting location (x_s, y_s) to the final location (x_f, y_f). The obstacles along the path are triangular in form. Consider the following heuristic functions to conduct A* search. - h_c assumes the obstacles are the smallest circles circumscribing the triangles. - h_r assumes the obstacles are smallest rectangles circumscribing the triangles. - $h_c{}'$ assumes the obstacles are largest circles inscribed in the triangles. - $h_r{}'$ assumes the obstacles are largest rectangles inscribed the triangles. Which of the following statement(s) is(are) true? 
(A)	h_c is an admissible heuristic
(B)	h_r is an admissible heuristic
(C)	$h_c{}'$ is an admissible heuristic
(D)	$h_r{}'$ is an admissible heuristic

Q.47	Let {O1, O2, O3, O4} represent the outcome of a random experiment, with P({O1})=P({O2})=P({O3})=P({O4}). Consider the following events: P={O1,O2}, Q={O2,O3}, R={O3,O4},S={O1,O2,O3}. Then, which of the following statements is true?
(A)	P and Q are independent
(B)	P and Q are not independent
(C)	R and S are independent
(D)	Q and S are not independent
Q.48	Consider the matrix **X** whose eigenvalues are 1, -1 and 3. Then Trace of $\mathbf{X}^3 - 3\mathbf{X}^2$ is ________

Q.49	**What is the output of the following program?**
	`int i = 1, j = 1;`
	`for(; i<= 10; i++)`
	`{`
	`if(i%3 != 0) { j += 2; continue; }`
	`if(j%3 == 0) break;`
	`}`
	`printf("%d",i+j);`

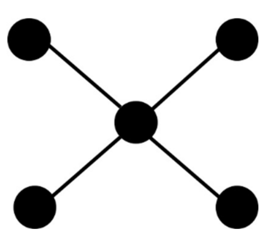
(A)	3
(B)	5
(C)	12
(D)	15

Q.50	Assume that S is a stack and Q1 and Q2 are two Queues which support the Enqueue and Dequeue operations. Consider the following pseudo code for implementing the Pop and Push operation on S.
	
	Push(S,x)
	    A(Q2,x)
	    while(Q1 not empty)
	        B(Q2,**C**(Q1));
	    Swap(Q1,Q2)
	
	Pop(S)
	    return(**D**(Q1))
	
	Which of the following options for the functions A, B, C, and D would correspond to correctly implementing the Push and Pop operations on the stack S?

(A)	A,B - Enqueue C,D - Dequeue
(B)	A,C - Enqueue B, D - Dequeue
(C)	A,C - Dequeue B,D- Enqueue
(D)	A,D - Enqueue B,C - Dequeue

Q.51	Consider the following program.
	
	int fun(float a[], float b[],int d)
	{
	  float n1 = 0; float n2 = 0; int flag = 1;
	
	  for (int i = 0;i< d; i++)
	    n1 = n1 + (a[i]*a[i]); n2 = n2 + (b[i]*b[i]);
	
	  for(int i = 0;i< d; i++)
	    a[i] = a[i]/sqrt(n1); b[i] = b[i]/sqrt(n2);

	<pre>for (int i = 0;i< d; i++)
	{
	   if(a[i] != b[i])
	       flag = 0; break;
	}
	return flag;
	}</pre>
	For which of the following inputs does the above algorithm produce 1 as an output?
	
	(P) a = {1,2,3,4}; b = {3,4,5,6}, d = 4
	(Q) a = {1,2,3,4}; b = {2,4,6,8}, d = 4
	(R) a = {1,2,3,4} b = {10,20,30,40}, d = 4
	(S) a = {1,2,3,4}, b = {1.1,2.1,3.1,4.1}, d = 4

(A)	P, Q, R, S
(B)	Q, R, S
(C)	Q, R
(D)	R, S
Q.52	Consider the following undirected graph on 5 nodes 

	Assume you are performing breadth first search on this graph using a queue data structure. How many unique breadth first orderings are possible on this graph?
(A)	9
(B)	24
(C)	48
(D)	120

Q.53	Let S^2 be the variance of a random sample of size $n > 1$ from a normal population with an unknown mean μ and an unknown finite variance $\sigma^2 < \infty$. Consider the following statements: (I) S^2 is an unbiased estimator of σ^2 , and S is an unbiased estimator of σ. (II) (n-1/n) S^2 is a maximum likelihood estimator of σ^2 , and $\sqrt{\frac{n-1}{n}}S$ is a maximum likelihood estimator of σ. Which of the above statements is/are true?
(A)	(I) only
(B)	(II) only
(C)	Both (I) and (II)
(D)	Neither (I) nor (II)

Q.54 The value of the real variable $x \geq 0$, which maximizes the function $f(x) = x^e e^{-x}$ is_____ (up to two decimal places)

Q.55 Consider the following relational schema:

```
employee(empId,empName,empDept)

customer(custId,custName,salesRepId,rating)
```

`salesRepId` is a foreign key referring to `empId` of the employee relation. Assume that each employee makes a sale to at least one customer. What does the following query return?

```
SELECT empName
FROM employee E
WHERE NOT EXISTS (SELECT custId
                  FROM customer C
                  WHERE C.salesRepId = E.empId
                  AND C.rating <> 'GOOD');
```

(A)	Names of all the employees with at least one of their customers having a ‘GOOD’ rating.
(B)	Names of all the employees with at most one of their customers having a ‘GOOD’ rating.
(C)	Names of all the employees with none of their customers having a ‘GOOD’ rating.
(D)	Names of all the employees with all their customers having a ‘GOOD’ rating.

End of Sample Question Paper