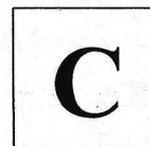M.Sc.(DS).II/04.24.001   Reg.No.

C

M.Sc. COMPUTER SCIENCE WITH SPECIALIZATION IN DATA SCIENCE
SECOND SEMESTER EXAMINATION APRIL2024

22-359-0201 NETWORKS AND DATA COMMUNICATIONS

(Regular)

Time: 3 Hours   Maximum Marks :50
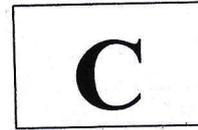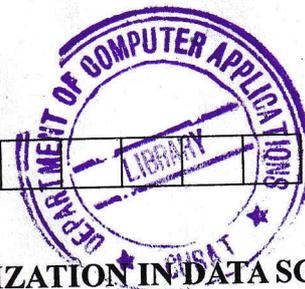
Each question carries 10 Marks

No		QUESTIONS	MARKS	CO	BL	PL
1.	a	With neat supporting diagrams, explain the MAC layer protocols.	10	CO1	L2	1.7.1
		OR				
2	a	There is the possibility of error occurrence in transmission of data in a network. Discuss the different types of such errors and how they affect data transmission. Also, suggest possible methods to overcome these. Explain any one of such methods.	10	CO1	L2	1.7.1
3	a	Discuss in detail the services provided by the Link Layer.	10	CO2	L2	1.7.1
		OR				
4	a	Draw the diagram and explain the 802.11 Architecture.	10	CO2	L2	1.7.1
5	a	Explain distance vector routing algorithm with an example.	7	CO3	L2	1.7.1
	b	Subnet the Class C IP Address 205.11.2.0 so that you have 30 subnets. i) What is the subnet mask for the maximum number of hosts? ii) How many hosts can each subnet have? iii) What is the IP address of host 3 on subnet 2 ?	3	CO3	L3	1.7.1
		OR				
6	a	What is CIDR notation? Provide an example and explain its significance in IP addressing.	3	CO3	L2	1.7.1
	b	Compare the IPv6 header format with that of IPv4. Highlight the main differences and improvements in IPv6.	7	CO3	L2	1.7.1

7	a	What are the two broad categories of congestion control mechanisms. Briefly explain all the techniques.	10	CO4	L1	1.7.1
		OR				
8	a	Explain in detail three way handshaking for connection establishment in TCP.	6	CO4	L1	1.7.1
	b	Explain Go-Back-N protocol with an example.	4	CO4	L2	1.7.1
9	a	Ameya is sending an e-mail to her friend Smitha. Ameya and Smitha are aware only of the interface that they are using. Explain to them technically, the entire process of how a mail sent by Ameya reaches Smitha.	10	CO5	L2	1.7.1
		OR				
10	a	How do processes communicate with each other? How can we trigger, manipulate and end communication between processes?	10	CO5	L2	1.7.1

****

M.Sc.(DS).II/04.24.004   Reg.No.

C

M.SC. COMPUTER SCIENCE WITH SPECIALIZATION IN DATA SCIENCE SECOND SEMESTER EXAMINATION, APRIL 2024

22-359-0212 DIGITAL IMAGE PROCESSING
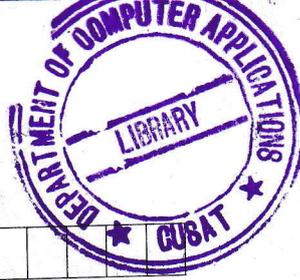### (Regular)

Time: 3 Hours   Maximum Marks :50

Each question carries 10 Marks

No		QUESTIONS	MARKS	CO	BL	PL
1	a.	Explain the spatial convolution with example.	6	CO1	L2	1.4.1
	b.	Explain contrast stretching with example.	4			
		OR				
2	a.	Explain image sampling and quantization with suitable examples.	6	CO1	L2	1.4.1
	b.	Design a first order derivative filter for image sharpening in spatial domain.	4			
3	a.	What is Discrete Fourier Transform (DFT)? Explain the basic steps for filtering in the frequency domain with DFT.	6	CO2	L2	1.4.1
	b.	Explain image smoothing using Gaussian lowpass filter in frequency domain.	4			
		OR				
4	a.	Explain different steps in the homomorphic filtering.	6	CO2	L2	1.4.1
	b.	Describe the band filters with examples.	4			
5	a.	Explain Adaptive Median Filter. Give the advantages of Adaptive Median Filter over traditional Median Filter.	6	CO3	L2	1.4.1
	b.	Explain the following: I. Midpoint Filter II. Periodic Noise	4			
		OR				
6	a.	Describe image degradation/restoration model with a neat diagram.	5	CO3	L2	1.4.1
	b.	What is salt and pepper noise? Explain how the salt and pepper noise can be reduced?	5			
7	a.	Explain in detail the Canny Edge Detector.	7	CO4	L3	1.4.1
	b.	Give the significance of Image Gradient in detection of edges.	3		L2	

		OR				
8	a.	Explain segmentation using K-Means clustering.	6	CO4	L3	1.4.1
	b.	Explain the following: I. Edge Models II. Thresholding	4		L2	
9		Explain in detail the Huffman Coding with example.	10	CO5	L3	1.4.1
		OR				
10		Explain in detail the Golomb Coding with example.	10	CO5	L3	1.4.1

****

B

M.Sc.(DS).II/04.24.002   Reg.No.

M.Sc. COMPUTER SCIENCE WITH SPECILIZATION IN DATA SCIENCE DEGREE
SECOND SEMESTER EXAMINATION, APRIL, 2024
22-359-0203-R FOR DATA ANALYTICS (REGULAR)

Time: 3 Hrs.   **Maximum Marks: 50**

Answer Any Five Questions
Each question carries 10 Marks

QNo	QUESTIONS	MARKS	CO	BL	PI
1.	a) With an example explain which looping structure can be used to skip the current iteration without executing the further statements in an R program?	3	CO1	L2	2.5.3
	b) What is a list in R language? c) Write an R program to create a list containing strings, numbers, vectors and logical values and do the following manipulations over the list. i. Access the first element in the list ii. Give names to the elements in the list iii. Add element at the end of the list iv. Remove the second element v. Print the fourth element vi. Update the third element vii. Write the updated list to a file.	7	CO1	L3	1.7.1
	OR				
2.	a) With examples explain the different ways in which matrices can be created in R.	2	CO1	L2	1.6.1
	b) With suitable examples explain the different data types and data structures used in R language.	8	CO1	L2	1.6.1
3.	a) Describe the steps in creating a histogram using R code, including the essential parameters needed for its construction.	3	CO4	L2	1.6.1
	b) i. What is the difference between noise and outliers in a dataset, give examples. ii. Find the five number summary of the dataset given below and also draw the boxplot in R.(Mark outliers if any) (6,65,68,72,78,87,95,137)	7	CO1	L3	1.7.1

4.	a) How scatter plots are created in R? How we can change the shape and size of the points in the scatter plot?	3	CO4	L2	1.6.1
	b) The sales of an item in a stall for the previous 9 months are given below. Forecast the sales of the 10th month using exponential smoothing method. Take $\alpha = 0.2$	7	CO1	L3	1.7.1

MONTH	SALES
1	34
2	38
3	46
4	41
5	43
6	48
7	51
8	50
9	56

5.	a) Which are the packages and functions required in R to perform decision tree classification?	3	CO2	L2	1.6.1
	b) Find which attribute would be chosen as the first splitting attribute if a decision tree is constructed for the below given dataset. Assume Profit as the class attribute. Use entropy and information gain as the metrics to find the best split.	7	CO2	L3	1.7.1

Age	Competition	Type	Profit
Old	yes	software	Down
Old	No	software	Down
Old	No	hardware	Down
mid	yes	software	Down
mid	yes	hardware	Down
mid	No	hardware	Up
mid	No	software	Up
new	yes	software	Up
new	No	hardware	Up
new	No	software	Up

6	a) Explain how an artificial neural network classifier can be implemented in R language	5	CO2	L2	1.6.1
	b) Given a training dataset as shown below, using Naïve Bayesian algorithm find whether a data object with the attribute values as given below have the class label 'played football' as Yes or No?	5	CO5	L3	1.7.1

outlook =sunny, Temperature =cool, wind=strong, Humidity=High

	Outlook	Temperature	Humidity	Wind	Played football(yes/no)
1	Outlook	Temperature	Humidity	Wind	Played football(yes/no)
2	Sunny	Hot	High	Weak	No
3	Sunny	Hot	High	Strong	No
4	Overcast	Hot	High	Weak	Yes
5	Rain	Mild	High	Weak	Yes
6	Rain	Cool	Normal	Weak	Yes
7	Rain	Cool	Normal	Strong	No
8	Overcast	Cool	Normal	Strong	Yes
9	Sunny	Mild	High	Weak	No
10	Sunny	Cool	Normal	Weak	Yes
11	Rain	Mild	Normal	Weak	Yes
12	Sunny	Mild	Normal	Strong	Yes
13	Overcast	Mild	High	Strong	Yes
14	Overcast	Hot	Normal	Weak	Yes
15	Rain	Mild	High	Strong	No

7.	a) Explain how Apriori rules can be generated from any transaction dataset in R language. How the rules can be arranged in descending order of their confidence value and the program should display only the first 3 rules.	4	CO3	L3	1.7.1
	b) Using Apriori algorithm find the frequent itemset from the below given transaction dataset. Assume minimum support count as 2.	6	CO3	L3	1.7.1

Tid	Items Bought
T1	Bread, Tea, Bun
T2	Eggs,Tea, Cake
T3	Bread,Eggs,Tea, Cake
T4	Eggs, Cake
T5	Cream

OR

8.	a) Write an R program to implement the preprocessing operations in text mining?	6	CO2	L2	1.6.1
	b) Assuming there are three different Word or HTML documents. Doc 1: Kareeshma studies about Cyber Security in CUSAT Doc 2: Renu works at Anna University. Doc 3: Data Engineers work on large datasets. Find the TF-IDF value of "Data Engineers" from the above Three documents.	4	CO2	L3	1.7.1

9.	a) How the accurate value for K can be selected in K means algorithm? How this can be implemented in R language?	4	CO2	L3	2.5.3
	b) Suppose we have 4 types of medicines and each has two attributes (pH and weight index). Our goal is to group these objects into K=2 group of medicine.(Assume A and B are randomly selected as initial centroids)	6	CO2	L3	1.7.1

Medicine	pH	Weight Index
A	1	1
B	2	1
C	4	3
D	5	4

OR

10.	a) Which function and packages are required for implementing hierarchal clustering in R.	4	CO2	L2	1.6.1
	b) Using single linkage/ MIN agglomerative clustering algorithm find the clusters of the data objects in the below given dataset.	6	CO2	L3	1.7.1

SAMPLE NO:	X	Y
P1	0.40	0.53
P2	0.22	0.38
P3	0.35	0.32
P4	0.26	0.19
P5	0.08	0.41
P6	0.45	0.30

****

M.Sc.(DS).II/04.24.003   Reg.No. [| | | | | | | |]   **A**
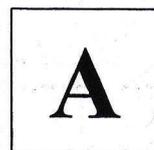
M.Sc. COMPUTER SCIENCE WITH SPECIALIZATION IN DATA SCIENCE
SECOND SEMESTER EXAMINATION, APRIL 2024

22-359-0204 MACHINE LEARNING

Time: 3 Hrs.   Maximum Marks: 50

Answer any five questions
Each question carries 10 Marks

QNo		QUESTIONS	MARKS	CO	BL	PI
1.		Explain Principal Component Analysis, Find the first principal component of the given data.	10	CO1	L3	1.6.1
		<table><tr><td>Y1</td><td>1.5</td><td>2.5</td><td>0.5</td><td>2.2</td><td>1.9</td><td>3.1</td><td>2.3</td><td>2.0</td><td>1.0</td></tr><tr><td>Y2</td><td>1.6</td><td>2.4</td><td>0.7</td><td>2.9</td><td>2.2</td><td>3.0</td><td>2.7</td><td>1.6</td><td>1.1</td></tr></table>				
		OR				
2.	a.	Suppose that the data for analysis includes the attribute height. The height values for the data are (in increasing order) 12,15,15,16,19,20,21,21,22,23,23,23,23,30,33,33,35,35,35,36,44, 45,46,53,80. Use smoothing by bin mean, smoothing by median, and smoothing by mode to smooth the data,using a bin depth of 3.	6	CO1	L3	1.6.1
	b.	Describe various data transformation strategies used in machine learning. How does normalization help in preparing data for modeling?	4			
3.		Consider the following set of transactions and answer the given questions assuming a minsup threshold of 2.	10	CO2	L3	1.6.1

TID	Items Bought
1	Milk, Tea, Cake
2	Eggs, Tea, Cold Drink
3	Milk, Eggs, Tea, Cold Drink
4	Eggs, Cold Drink
5	Juice

a. Is {Milk, Cold Drink} a frequent itemset?

b. Is Tea ⇒Eggs a valid association rule with 50% confidence?

c. Is {Milk,Eggs}⇒ Cold Drink a valid association rule?

d. Is {Cold Drink, Eggs}⇒Tea a valid association rule with 80% confidence?

OR

4.		Explain the FP-Growth algorithm used in frequent pattern mining. Describe the key steps involved in the algorithm's execution, including frequent itemset generation and association rule generation. Discuss the advantages and limitations of the FP-Growth algorithm compared to Apriori algorithms.	10	CO2	L2	2.5.1
5.	a.	Given a dataset containing information about house prices and various features such as square footage, number of bedrooms, and location. Design a simple linear regression model to predict house prices based on one of the given features. Explain the steps involved in training the model and evaluating its performance.	6	CO3	L3	1.6.1
	b.	Distinguish between overfitting and underfitting. How can it affect model generalization?	4	CO3	L3	1.6.1

OR

6.	a.	Consider the training data in the following table where Play is the class attribute. In the table, the Humidity attribute has values "L" (for low) or "H" (for high), Sunny has values "Y" (for yes) or "N" (for no), Wind has values "S" (for strong) or "W" (for weak), and Play has values "Yes" or "No".	5	CO3	L3	1.6.1

Humidity	Sunny	Wind	Play
L	N	S	No
H	N	W	Yes
H	Y	S	Yes
H	N	W	Yes
L	Y	S	No

What will be the class label for the following day (Humidity=L,Sunny=N,Wind=W), according to naive Bayesian classification? Explain the process of generating that class label.

	b.	Explain the mathematical idea behind the SVM classifier.	5	CO3	L3	1.6.1

7.	a.	Write down the algorithm for the K-means Clustering technique.	5	CO3	L2	1.6.1
	b.	What are the distance metrics used for quantitative and qualitative attributes? Explain each metric with the help of examples.	5			

OR

8.	Consider the following dataset with five data points in a two-dimensional space: **Data points: A(2,10), B(2,5), C(8,4), D(5,8), E(7,5)** Perform agglomerative clustering using single linkage (nearest neighbor) method on this dataset. Show the step-by-step process of clustering, including the distance matrix and dendrogram at each step. Use Euclidean distance as the distance metric.	10	CO3	L2	1.6.1

9.	a.	Explain the concept of Perceptron with a neat diagram	5	CO5	L2	2.5.1
	b.	Write the algorithm for Back propagation.	5			

OR

10	a.	What is the role of the Activation functions in Neural Networks? List down the names of some popular Activation Functions used in Neural Networks.	5	CO5	L2	2.5.1
	b.	With the help of a real-life example, explain how MLP can be used for classification. Write the corresponding Python code and necessary output generated by the model.	5			

****

M.Sc.(AI).II/04.24.005   Reg.No.

B

M.Sc. COMPUTER SCIENCE WITH SPECIALIZATION IN DATA SCIENCE DEGREE
SECOND SEMESTER EXAMINATION, APRIL 2024
22-359-0202 DATABASE MANAGEMENT SYSTEMS
(Regular)

Write any FIVE questions.
(Each Question Carries 10 Mark)

Time-3 Hours   **Maximum Marks :50**

No		QUESTIONS	MARKS	CO	BL	PL
1	a	What are the responsibilities of DBA?	4	CO1	L2	1.7.1
	b	How the following are represented in ER diagram. Give examples for each. a) Composite attributes. b) Stored attributes. c) Derived attributes. d) Multivalued attributes.	6	CO1	L2	1.7.1
OR						
2	a	What is the difference between a file system and a database system?	4	CO1	L2	1.7.1

	b	Assume the following application that models soccer teams, the games they play, and the players in each team. In the design, the main points to be considered are:
		
		 • There is a set of teams, each team has an ID (unique identifier), name, main stadium, and to which city this team belongs.
		 • Each team has many players, and each player belongs to one team. Each player has a number (unique identifier), name, DoB, start year, and shirt number that he uses.
		 • Teams play matches, in each match there is a host team and a guest team. The match takes place in the stadium of the host team.
		 • For each match you need to keep track of the following:
		   o The date on which the game is played.
		   o The final result of the match
		   o The players participated in the match. For each player, how many goals he scored, whether he took yellow card, and whether he took red card.
		   o During the match, one player may substitute another player. In this case capturing this substitution and the time at which it took place, are important.
		 • Each match has exactly three referees. For each referee we have an ID (unique identifier), name, DoB, years of experience. One referee is the main referee and the other two are assistant referee.
		
		 Design an ER diagram to capture the above requirements. State any assumptions you have, that affects your design. Make sure cardinalities and primary keys are clear.
			6	CO1	L3	1.7.1

3	**a**	Consider the following relations
		
		Product *(P_code, Description, Stocking_date, QtyOnHand, MinQty, Prices, Discount, V_code)*
		Vendor *(V_code, Name, City, Phone).*
		Here a vendor can supply more than one product but a product is supplied by only one vendor.
		
		Write SQL queries to
		   I. Display details of those vendors who supply products 'P1' and 'P2'.
		   II. List the names of all the vendors who supply more than one product.
		   III. List the details of the products whose prices exceed the average product price.
		   IV. Get details of those products that are supplied by the vendor who supplied product 'P5'.
		   V. Increase the discount by 5% on products whose vendors live in the city 'Chennai'.
			5	CO2	L3	1.7.1

	b	A table R (A, B, C, D, E) has the following functional dependencies: $\{ A \rightarrow C, B \rightarrow D, AB \rightarrow E \}$. Determine the normal form of the given table.	5	CO2	L3	1.7.1
		OR				
4	**a**	Consider a relation R(A, B, C, D, E, F) with A as the only key. Assume that the dependencies E→F and C→DE hold on R a. Is R in 2NF? If not decompose to 2NF. b. Is R in 3NF? If not decompose to 3NF.	5	CO2	L3	1.7.1

	b	Consider the Company database:
		Employee (emp_no, emp_name, age, e_location, mobile_no, dept_no, salary).
		Department (dept_no, dept_name, d_location, budget).
		Write SQL queries to
		   I. Display the details of all employees whose names start with letter C.
		   II. Find all employees who live in the same location as the department they work in.
		   III. Display names of employees whose salary is less than the average salary of all employees.
		   IV. Display names of all departments which have more than 2 employees.
		   V. Delete records of all employees who work in the 'Finance' department.
			5	CO2	L3	1.7.1

5	**a**	What are the problems that may occur with concurrent transactions in DBMS? Explain with examples.	4	CO3	L2	1.7.1

	b	How to check the serializability of a schedule in DBMS? Check whether the given schedule S_A is serializable or not? S_A: R1(A); R3(D); W1(B); R2(B); W3(B); R4(B); W2(C); R5(C); W4(E); R5(E); R5(A)	6	CO3	L3	1.7.1
		OR				
6	a	Discuss about the desirable properties of DBMS transactions.	4	CO3	L2	1.7.1
	b	Discuss about the two main schemes that prevent deadlock based on timestamp.	4	CO3	L2	1.7.1
	c	Explain the *unlock(X)* operation in shared lock.	2	CO3	L2	1.7.1
7		What is file organization? Explain about the following types of file organizations in detail. 1. Heap File Organization 2. Hash File Organization	10	CO4	L2	1.7.1
		OR				
8		Differentiate between different types of file organizations in DBMS.	10	CO4	L2	1.7.1
9	a	List the situation in which the NoSQL databases are preferred?	3	CO5	L1	1.7.1
	b	Consider MONGODB, with examples illustrate the given methods: 1. find() 2. findOne() 3. insertOne() 4. insertMany() 5. updateOne() 6. updateMany() 7. drop()	7	CO5	L1	1.7.1
		OR				
10	a	With examples list the main categories of NoSQL databases.	2	CO5	L1	1.7.1
	b	The collection "RESTAURANT " contains set of documents with the following Fields: {Name, Restaurant_Id, Location, ZIP_code, Cuisine, Rating} 1. Write a MongoDB query to display all the documents in the collection restaurants. 2. Write a MongoDB query to display the fields Restaurant_Id, name, Cuisine, but exclude the field _id for all the documents in the collection restaurant. 3. Write a MongoDB query to find the restaurants who achieved a rating more than 90. 4. Write MongoDB query to update all the documents in the collections where the Cuisine is "Bakery" and update it to "Multi Cusine"	8	CO5	L1	1.7.1

****