Time: 03 Hours **Marks: 80**

Note: 1. Question 1 is compulsory
2. Answer any three out of remaining five questions.
3. Assume any suitable data wherever required and justify the same.

Q1 a) Explain the Expectation Maximization Algorithm (EMA) [5]

b) Explain kernel functions and kernel trick [5]

c) What are the issues in decision tree learning? [5]

d) "Entropy is a thermodynamic function used to measure the disorder of a system in Chemistry." Clarify the concept of entropy in Machine Learning? [5]

Q2 a) Compare and contrast Linear and Logistic regressions with respect to their mechanisms of prediction. [10]

b) Consider 2-D dataset given in the table below. Construct a SVM classifier model. Given (2, 1), (2, -1) and (4, 0) as support vectors; estimate the parameters of the model and classify (4, 2). Why is SVM called as optimal binary hyper plane classifier? [10]

(X1, X2)	(1, -1)	(2, -1)	(5, -1)	(4, 0)	(6, 0)	(1, 1)	(2, 1)	(5, 1)
Class	C1	C1	C2	C2	C2	C1	C1	C2

Q3 a) You are given a data set on cancer detection. You have built a classification model and achieved an accuracy of 96%. Why shouldn't you be happy with your model performance? What can you do about it? [10]

b) What is a HMM? What are the issues in Hidden Markov Model (HMM)? [10]

Q4 a) You came to know that your model is suffering from low bias and high variance. Which algorithm should you use to tackle it? Why? [10]

b) Differentiate between simple linkage, average linkage and complete linkage algorithms. Use complete linkage algorithm to find the clusters from the following dataset. [10]

X	4	8	15	24	24
Y	4	4	8	4	12

Q5 a) Draw the block diagram of Error Back Propagation Algorithm and explain with flow chart the concept of Back Propagation. [10]

b) The following table consists of training data from an employee database. The data have been generalized. For example, "31 . . . 35" for age represents the age range of 31 to 35. For a given row entry, count represents the number of data tuples having the values for department, status, age, and salary given in that row. Let the status be the class-label attribute. [10]

(i) Design a multilayer feed-forward neural network for the given data. Label the nodes in the input and output layers.

(ii) Using the multilayer feed-forward neural network obtained in (i), show the weight values after one iteration of the back propagation algorithm, given the training instance "(sales, senior, 31 . . . 35, 46K . . . 50K)".

Assume initial weight values and biases. Assume learning rate to be 0.9. Use binary input and draw (*one input layer, one output layer and one hidden layer*) neural network. Solve the problem for one epoch.

department	*status*	*age*	*salary*	*count*
sales	senior	31 … 35	46K … 50K	30
sales	junior	26 … 30	26K … 30K	40
sales	junior	31 … 35	31K … 35K	40
systems	junior	21 … 25	46K … 50K	20
systems	senior	31 … 35	66K … 70K	5
systems	junior	26 … 30	46K … 50K	3
systems	senior	41 … 45	66K … 70K	3
marketing	senior	36 … 40	46K … 50K	10
marketing	junior	31 … 35	41K … 45K	4
secretary	senior	46 … 50	36K … 40K	4
secretary	junior	26 … 30	26K … 30K	6

Q6 Write short notes on any two of the following: [20]

a) Temporal Difference Learning in Reinforcement Learning

b) Over fitting' in Machine learning

c) Independent Component Analysis

***

Time: 03 Hours **Marks: 80**

Note: 1. Question 1 is compulsory
2. Answer any three out of remaining five questions.
3. Assume any suitable data wherever required and justify the same.

Q1 a) What is dimensional modeling? Design the data warehouse for wholesale furniture Company. The data warehouse has to allow to analyze the company's situation at least with respect to the Furniture, Customer and Time. Moreover, the company needs to analyze: The furniture with respect to its type, category and material. The customer with respect to their spatial location, by considering at least cities, regions and states. The company is interested in learning the quantity, income and discount of its sales. **[10]**

b) i. Explain the architecture of data mining. **[10]**
ii. Explain different steps involved in data processing.

Q2 a) Differentiate top-down and bottom-up approaches for building data warehouse. Discuss the merits and limitations of each approach. **[10]**

b) Explain frequent pattern growth mining with example. **[10]**

Q3 a) For the following dataset, apply decision tree classification algorithm and show the generated rules **[10]**

Id	Age	Income	Student	Credit-rating	buys computer
1	young	high	no	fair	no
2	young	high	no	good	no
3	middle	high	no	fair	yes
4	old	medium	no	fair	yes
5	old	low	yes	fair	yes
6	old	low	yes	good	no
7	middle	low	yes	good	yes
8	young	medium	no	fair	no
9	young	low	yes	fair	yes
10	old	medium	yes	fair	yes
11	young	medium	yes	good	yes
12	middle	medium	no	good	yes
13	middle	high	yes	fair	yes
14	old	medium	no	good	No

b) Explain steps of ETL process in detail **[10]**

Q4 a) What is Clustering Techniques? Discuss the Agglomerative algorithm using following data and plot a Dendrogram using single link and complete link approach. The following figure contains sample data items indicting the distance between the elements: **[10]**

Item	E	A	C	B	D
E	0	1	2	2	3
A	1	0	2	5	3
C	2	2	0	1	6
B	2	5	1	0	3
D	3	3	6	3	0

b) i. Explain different OLAP models. **[10]**

ii. Differentiate Online transaction processing (OLTP) and Online analytical processing (OLAP)

Q5 a) Consider a data warehouse for a hospital where there are three dimension a) Doctor b) Patient c) Time. Consider a measure charge fee that the doctor charges to a patient for a visit. Create a cube and illustrate the following OLAP operations: **[10]**

1) Rollup 2) Drill down 3) Slice 4) Dice 5) Pivot.

2)

b) Discuss Association Rule Mining (AR) and Apriori Algorithm. Apply AR Mining to find all frequent item sets and association rules for the following dataset: **[10]**

Q6 Write short notes on any *two* of the following: **[20]**

a) Linear Regression

b) Data Visualization

c) DBSCAN clustering

(3 Hours) **(Total Marks : 80)**

N.B.: (1) Question No.1 is **compulsory**.

(2) Attempt **any three** questions from the remaining **five** questions.

(3) Make suitable assumptions wherever **necessary** but **justify your assumptions**.

1. (a) Explain the category "cybercrimes against persons". **05**

(b) Define term Digital Forensic and Digital forensic investigation. **05**

(c) Define digital evidence and its types of digital evidences. **05**

(d) What is DOS attack? How to achieve recovery from DOS attack? **05**

2. (a)What steps or activities are done in an initial response phase? **10**

(b) What are the steps involved in computer evidence handling? Explain in detail. **10**

3. (a) What is Address Spoofing explain it types ? **10**

(b) What are possible investigation phase carried out in Data Collection and Analysis? **10**

4. (a) What are the requirements of forensic duplication tools? Elaborate different ways of creating a forensic duplicate of a hard-disk. **10**

(b) Difference Between Network based IDS and Host based IDS. **10**

5. (a) Explain how law enforcement is done in computer forensics. **10**

(b) What are the goals of network monitoring? What are the different types of network monitoring? Explain with examples. **10**

6. Write a short note on **20**

(1) Steps of Unix system investigation

(2) How to collect network based evidence Log files?

(3 Hours) **[Total Marks: 80]**

Note

Q.1 is compulsory

Solve any three from remaining.

All questionscarry 20 marks.

Q 1 **solves any four** **[20]**

A] Explain three level of processing in details.

B] Explain goal directed design in details.

C] How images and graphics are important in graphics.

D] Explain different categories of Users.

E] What do you mean by keyboard accelerator?

F] What are the factors to be considered to choose colors.

Q 2 a] what are the various factors considered for user interface design? **[10]**

Give example for the same.

B] What are the advantages and disadvantages of digital or graphical system? **[10]**

Q 3 A] Explain in details about Gesalts principles. **[10]**

B] Explain in details about response time with salient feature. **[10]**

Q 4 a] what are different presentation style of windows with advantages and disadvantages. **[10]**

B] What do you mean by personas? Mention steps in constructing personas. **[10]**

Q 5 a] what do you mean by device based and screen based control. **[10]**

B] Explain different behavioral pattern in details. **[10]**

Q 6 a] Explain various menus in HMI. **[10]**

B] Differentiate between web page navigation & printed page navigation. **[10]**

(3 hours)

[Total Marks: 80]

NB :

1) Question No.1 is **compulsory.**
2) Attempt any **three** questions out of the remaining questions.
3) Make suitable assumptions wherever necessary.

Q.1 Solve Any four **(5 Marks X 4)**

a) Compare parallel and distributed system models by giving example of each. **5**

b) State the goals of a distributed system. **5**

c) Compare and contrast between message oriented and stream oriented communication. **5**

d) Discuss Amdahl's law for measuring speed up performance of parallel system. **5**

e) Enlist and discuss desirable features of global scheduling algorithm **5**

Q.2 a) Illustrate 4 stage pipeline architecture. **10**

b) What is Remote Procedure Call. Discuss the working of RPC in detail. **10**

Q.3 a) Discuss the role consistency in distributed system. What is the need of client centric consistency models. Explain any two data centric consistency models. **10**

b) Illustrate the implementation details of pipelined floating point adder. **10**

Q.4 a) Discuss the need of process migration. Explain the role of resource to process and process to resource binding in process migration. **10**

b) Explain Raymond's Tree based algorithm of token based distributed mutual exclusion. **10**

Q.5 a) Describe code migration issues in detail. **10**

b) Explain the load balancing approach. Explain static and dynamic load balancing algorithm. **10**

Q.6 Attempt any two **(10X2)** **20**

a) Pipeline hazards and techniques to eliminate those hazards

b) Lamport algorithm

c) Election Algorithm

d) Andrew File System(AFS)

****