

TE / AI & DS / SEM-VI / e-2019 / DEC. 2023

(3 Hours)

(Total Marks: 80)

- N.B.: 1. Question No. 1 is compulsory.
 2. Answer any three out of the remaining questions.
 3. Assume suitable data if necessary.
 4. Figures to the right indicate full marks.

Q1. Attempt the following (any 4):

(20)

- Why is data analytics lifecycle essential?
- The regression lines of a sample are $x + 6y = 6$ and $3x + 2y = 10$.
Find (i) sample means \bar{x} and \bar{y} .
(ii) coefficient of correlation between x and y .
- Differentiate between linear regression and logistic regression.
- What is Pandas? State and explain key features of Pandas.
- Explain term frequency (TF), document frequency (DF), and inverse document frequency (IDF).

Q2. Attempt the following:

- Explain the data analytics lifecycle. (10)
- Find two lines of regression from the following data: (10)

Age of husband (x)	25	22	28	26	35	20	22	40	20	18
Age of wife (y)	18	15	20	17	22	14	16	21	15	14

Estimate (i) the age of husband when the age of wife is 19 and (ii) the age of wife when the age of the husband is 30.

Attempt the following:

- Explain Box-Jenkins intervention analysis. (10)
- What is text mining? Enlist and explain the seven practice areas of text analytics. (10)

Q4. Attempt the following:

- Explain different types of data visualizations in R programming language. (10)
- Fit a regression equation to estimate β_0 , β_1 , and β_2 to the following data of a transport company on the weights of 6 shipments, the distances they were moved and the damage of the goods that was incurred. (10)

Weight X_1 (1000 kg)	4.0	3.0	1.6	1.2	3.4	4.8
Distance X_2 (100 km)	1.5	2.2	1.0	2.0	0.8	1.6
Damage Y (Rs.)	160	112	69	90	123	186

Estimate the damage when a shipment of 3700 kg is moved to a distance of 260 km.

39481

Page 1 of 2

Q.P. CODE

Q5. Attempt the following:

- a. From the following results, obtain two regression equations and estimate the yield when the rainfall is 29 cm and the rainfall when the yield is 600 kg. (10)

	Yield in Kg.	Rainfall in cm
Mean	508.4	26.7
SD	36.8	4.6
Coefficient of Correlation	0.52	

- b. What is stepwise regression? State and explain different types of stepwise regression. (10)

Q6. Write short notes on (any 2):

- Time series analysis
- Exploratory data analysis
- Regression plot
- Generalized linear model (GLM)

(20)

TE | sem-VI | AI&DS | C-2019 | Dec 2023

Duration: 3 Hrs.

Marks: 80

Note:

1. Question 1 is compulsory.
2. Attempt any 3 questions out of the remaining questions.

Q1. Attempt any Four.

- a. Explain the different modes of block ciphers. 05
- b. List with examples the different mechanisms to achieve security. 05
- c. Differentiate MD5 and SHA-1 algorithms. 05
- d. List and explain security requirements of database. 05
- e. Explain phishing and list different types of phishing techniques. 05

Q2.

- a. User A and B want to use RSA to communicate securely. A chooses public key as (7, 119) and B chooses public key as (13, 221). Calculate their private keys. A wishes to send message $m = 10$ to B. Produce the ciphertext. Formulate the key using which A encrypt the message "m" if A need to authenticate itself to B. 10
- b. Explain memory and address protection in detail. Write a note on file protection. 10

Q3.

- a. List the functions of the different protocols of SSL. Explain the handshake protocol. 10
- b. List different poly-alphabetic substitution ciphers. Encrypt "The key is hidden under the door" using playfair cipher with keyword "domestic". 10

Q4.

- a. Define digital signature. Explain any digital signature algorithm in detail. 10
- b. Give the format of X.509 digital certificate and explain the use of a digital signature in it. 10

Q5.

- a. Explain session hijacking and management. 10
- b. What is need of Diffie-Hellman algorithm. User A and B decide to use Diffie-Hellman algorithm to share a key. They choose $p = 23$ and $g = 5$ as the public parameters. Their secret keys are 6 and 15 respectively. Compute the secret key that they share. 10

Q6. Attempt any Four.

- a. Explain the different types of firewalls and mention the layer in which they operate. 05
- b. List and explain vulnerabilities in windows operating system. 05
- c. List and explain characteristics needed in secure hash function. 05
- d. Explain Triple DES in short. 05
- e. Explain with examples, keyed and keyless transposition ciphers. 05

G.P. code.

41717

TE / sem-VI / AI&DS / C-2019 / Dec-2023

Time: 3 Hours

Max. Marks: 80

Instructions:

- Figures to the right indicate max marks.
- Draw appropriate diagram whenever applicable.
- Assume suitable data wherever applicable. State your assumptions clearly.
- **Question number 1 is compulsory.**
- Attempt **any Three** questions from remaining questions

Q.1 Solve any Four

- A. Explain SVD and its applications? [05]
 B. Differentiate between supervised and unsupervised learning. [05]
 C. Explain Hebbian Learning rule [05]
 D. Explain Perceptron model with Bias. [05]
 E. Differentiate between Ridge and Lasso Regression [05]

Q.2 Solve the following

- A. Draw a block diagram of the Error Back Propagation Algorithm and explain with the flow chart the Error Back Propagation Concept. [10]
 B. Find a linear regression equation for the following two sets of data: [10]

X	Y
3	12
5	18
7	24
9	30

Q.3 Solve the following

- A. Diagonalize the matrix A [05]

$$\begin{bmatrix} 1 & 5 \\ 4 & 2 \end{bmatrix}$$

- B. Write short note on overfitting and underfitting of model [05]
 C. What are activation functions? Explain Binary, Bipolar, Continuous, and Ramp activation functions. [10]

Q.P. code

38169

Page 1 of 2

4. Solve the following

- A. Explain Least-Squares Regression for classification. [10]
- B. What is the curse of dimensionality? Explain PCA dimensionality reduction technique in detail. [10]

Q. 5 Solve the following

- A. How to calculate Performance Measures by Measuring Quality of model. [10]
- B. Explain the Perceptron Neural Network [10]

Q. 6.

- A. Discuss the various steps of developing a Machine Learning Application. [10]
- B. Write a short note on LMS-Widrow Hoff [05]
- C. Explain the Maximization algorithm for clustering. [05]

Time: 3 hours

Max. Marks: 80

Instructions:

- 1) Only **Four question** need to be solved.
- 2) All question carries equal marks.
- 3) Illustrate your answers with neat sketches wherever necessary.
- 4) Figures to the right indicate full marks.
- 5) Assume suitable additional data, if necessary and clearly state it.
- 6) All sub-questions of the same question should be grouped together.

- Q.1** Solve any **four**
- (a) What are various issues of distributed system? 05
 - (b) Justify how Ricart-Agrawala's algorithm optimized the Message overhead in achieving mutual exclusion. 05
 - (c) Explain the election algorithm? 05
 - (d) Explain Suzuki-Kasami algorithm? 05
 - (e) Difference between RMI and RPC? 05
- Q.2**
- (a) What is distributed computing? Explain various system models of distributed computing? 10
 - (b) Define Remote Procedure Call (RPC). Explain the working of RPC in Detail. 10
- Q.3**
- (a) What is a logical clock? Why are logical clocks needed in a distributed system? Explain Lamport algorithm. 10
 - (b) Describe code migration issues in detail? 10
- Q.4**
- (a) Explain Hadoop Distributed File System (HDFS). 10
 - (b) Differentiate between message-oriented communication and stream-oriented communication. 10
- Q.5**
- (a) Compare Load sharing to Task Assignment and Load balancing strategies for scheduling processes in a distributed system. 10
 - (b) Discuss various client centric consistency models. 10
- Q.6** Write Short note (Any 2)
- (a) Physical Clock Synchronization 10
 - (b) Load balancing techniques 10
 - (c) Andrew File System (AFS) 10
 - (d) Fault tolerance 10

G.P. Code
39044