

Time: 03 Hours**Marks: 80****Note:** 1. Question 1 is compulsory

2. Answer any three out of remaining questions.
3. Assume suitable data wherever required and justify the same.

- Q1 a) Describe how logistic regression can be used as a classifier. [5]
 b) Explain cross-validation for accuracy estimation. [5]
 c) What is data journalism? [5]
 d) What is data leakage with respect to big data? [5]
- Q2 a) What is type I and type II errors in hypothesis testing? Is one always more serious than the other? Why? [10]
 b) Describe the working of the Map-Reduce with an example. [10]
- Q3 a) Explain Gaussian (normal) distribution with respect to pdf and cdf and its use in statistics. [10]
 b) Explain time series mining with an appropriate example. [10]
- Q4 a) You have collected a data of about ten thousand rows of tweet text. With help of text mining how you will create a tweet classification model that categorizes each of the tweets in three different classes. What could be the challenges while performing text mining to this context? [10]
 b) Explain the process of content based RS with suitable example. [10]
- Q5 a) Explain singular value decomposition (SVD) with an example. [10]
 b) What infrastructure is most appropriate for Hadoop? Draw and describe Hadoop Ecosystem Architecture. [10]
- Q6 a) Given $S = \begin{bmatrix} 1 & -1 \\ 0 & 1 \\ -1 & 0 \end{bmatrix}$ [10]
 Find principal components.
 b) Draw and describe the information visualization process. [10]

* * *

(Three Hours)

Total Marks: 80

Instructions:

- Attempt any four questions out of six questions
- Assume suitable data wherever necessary
- Figures to the right indicate full marks.

- Q.1 Answer any Four. 20
- a. Role of SPSS in data analysis
 - b. Foot notes and Bibliography
 - c. Importance of t-tests
 - d. Descriptive statistics
 - e. Testing of hypothesis
 - f. Non-parametric tests
- Q.2 a. What is the research methodology? Explain the steps in scientific research process. Briefly explain about literature review. 10
- b. State the sources of research problem. How a problem is identified? Enumerate the criteria for the selection of a problem. 10
- Q.3 a. Explain the concept of attitude scale. Explain the Likert's scale to measure data attitude. 10
- b. Explain Quantitative vs. Qualitative type of research. Explain Post Facto research and Motivation in research. 10
- Q.4 a. Explain critically interpretation and Organization of the data. 10
- b. Hypothesis is a statement which involves a relationship of variable. Enumerate the types of variables included in stating a hypothesis. 10
- Q.5 a. What are the characteristics of research? What are the factors affecting research design? 10
- b. "A systematic bias results from errors in the sampling procedures". What do you mean by such a systematic bias? Describe the important causes responsible for such a bias. 10
- Q.6 a. What are the differences between observation and interviewing as methods of data collection? Explain with two specific examples of situations where either observation or interviewing would be more appropriate. 10
- b. You have been asked to research setting up of a roadside hotel. Design a questionnaire to find out the prospects of proceeding with the venture. 10

(Time: 3 Hours)

[Total Marks 80]

N. B:

1. Question No. 1 is Compulsory.
2. Solve any THREE from Question No. 2 to 6.
3. Draw neat well labeled diagram wherever necessary.

Q. 1 a) Discuss in detail SISD, SIMD, MISD, MIMD, SIMT, SPMD, MPMD. (10)

b) Write a MPI program to find sum of N numbers. (10)

Q. 2 a) Derive the expression for speedup and efficiency by Amdahl's law and comment on the same. Assume suitable data if required. (10)

b) Discuss different levels of parallel processing. (10)

Q. 3 a) Explain about process synchronization mechanism with Semaphore. (10)

b) Define MPI. Explain in detail principles of Message Passing Programming. (10)

Q. 4 a) Explain in brief Quantum Computers. (10)

b) Define CUDA? Explain in CUDA processor architecture. (10)

Q. 5 a) Explain speedup, efficiency and scalability with suitable example. (10)

b) Explain in detail Architecture of NVIDIA GPU. (10)

Q. 6 Write short note on the following:

1. Nanotechnology
 2. Grain packing and scheduling in parallel processing
 3. Data Race
 4. OpenMP
- (20)