



**Avinashilingam Institute for Home Science and Higher Education for Women**  
Deemed to be University Estd. u/s 3 of UGC Act 1956, Category A by MHRD [now MoE]  
Re-accredited with 'A++' Grade by NAAC CGPA 3.64/4 Category I by UGC  
Coimbatore - 641 043, Tamil Nadu, India

**Continuous Internal Assessment- II October 2025**  
**SEMESTER V**

**Class: III B.Sc**  
**Major: Computer Science**

**Time: 1 hour**  
**Max. Marks: 30**

**23BCSDE2 Fundamentals of Data Science**

**Course Outcomes:**

At the end of the course, students will:

- CO1: Ability to apply the basic Data Science Knowledge on various problems.
- CO2: Contribution of the knowledge gained on choosing and evaluation on different models.
- CO3: Understanding the application and use of regression Techniques.
- CO4: Attaining knowledge to identify the application areas of various unsupervised methods.
- CO5: analysing the documents for final deployment and effective presentation.

**Part A**

**Choose the correct answer**

**10 x 1 = 10**

1. Linear Regression is primarily used for CO3K1
  - a) Classification of categorical outcomes
  - b) Predicting continuous dependent variables
  - c) Clustering similar data points
  - d) Reducing dimensionality
2. In Logistic Regression, the output is transformed using which function? CO3K2
  - a) Linear function
  - b) Exponential function
  - c) Sigmoid (logistic) function
  - d) Step function
3. In a regression model summary, the R-squared ( $R^2$ ) value indicates CO3K2
  - a) The correlation between independent and dependent variables
  - b) The proportion of variance in the dependent variable explained by the model
  - c) The slope of the regression line
  - d) The standard error of estimate
4. The learner is trying to predict housing prices based on the size of each house. What type of regression is this? CO3K1
  - a) Multivariate Logistic Regression
  - b) Logistic Regression
  - c) Linear Regression
  - d) Multivariate Linear Regression
5. What is the output of Logistic Regression? CO3K1
  - a) Continuous values
  - b) Probability values between 0 and 1
  - c) Categorical labels directly
  - d) Clusters
6. The main objective of clustering algorithms in unsupervised learning is to CO4K2
  - a. Predict continuous values
  - b. Determine the optimal number of clusters
  - c. Assign input data to predefined categories or classes
  - d. Identify patterns in unlabeled data
7. Which algorithm is used to identify patterns or relationships in unlabeled data? CO4K1
  - a. Decision tree
  - b. Association rule mining
  - c. K-means clustering
  - d. Principal Component Analysis (PCA)
8. Which of the following is the most important step while delivering results in datascience? CO5K1
  - a) Using complex mathematical notations
  - b) Communicating insights clearly to stakeholders
  - c) Focusing only on the algorithm used
  - d) Showing all raw data without summarization

9. When presenting results to a non-technical audience, the best approach is CO5K2
- Use technical jargon to prove expertise
  - Focus on the business impact of the results
  - Provide only statistical significance values
  - Skip visualizations and share code outputs
10. Which visualization is best suited to show the distribution of a continuous variable? CO5K1
- Bar chart
  - Histogram
  - Pie chart
  - Scatter plot

**Part B**

**Answer All Questions**

**4 x 5 = 20**

**Each answer should not exceed 400 words or two pages**

- Illustrate on the logistic regression CO3K3

(Or)

  - Discuss on the process of model summary. CO3K2
- Discuss on the need of data preparation for cluster analysis. CO4K2

(Or)

  - Explain on the Advantages of hierarchical clustering CO4K2
- Explain on the clustering takeaways. CO4K2

(Or)

  - Highlight on the Association Algorithm with neat sketch and examples. CO5K2
- Explain the importance of producing presentations. CO5K2

(Or)

  - Define how comments and version control helps to run the document? CO5K1

\*\*\*\*\*

**Staff in-charge : Dr. N. Valliammal ( Campus I ) , Mrs.S.Preema (Campus II)**  
**No. of Copies: 55+58= 100 copies**