

PSG COLLEGE OF ARTS & SCIENCE
(AUTONOMOUS)

MSc (SS) DEGREE EXAMINATION MAY 2025
(Ninth Semester)

Branch – SOFTWARE SYSTEMS (Five Years Integrated)

DATA SCIENCE

Time: Three Hours

Maximum: 75 Marks

SECTION-A (10 Marks)

Answer ALL questions

ALL questions carry EQUAL marks (10 x 1 = 10)

1. What is data science primarily concerned with?
(i) Analyzing and interpreting data (ii) Collecting data only
(iii) Storing data in database (iv) All the mentioned above
2. Which of the following characteristic of big data is relatively more concerned data science?
(i) Volume (ii) Velocity (iii) Variety (iv) None of the mentioned
3. A _____ node acts as the Slave and is responsible for executing a Task assigned to it by the Job Tracker.
(i) MapReduce (ii) Mapper (iii) Task Tracker (iv) Job Tracker
4. Point out the correct statement.
(i) MapReduce tries to place the data and the compute as close as possible
(ii) Map Task in MapReduce is performed using the Mapper() function
(iii) Reduce Task in MapReduce is performed using the Map() function
(iv) All of the mentioned
5. Linear regression models are preferable for
(i) Interpretation (ii) Predictions (iii) Conclusion (iv) Both a and b
6. When the projectors are parallel to each other and also perpendicular to the plane, the projection is called _____.
(i) Perspective projection (ii) Oblique projection
(iii) Isometric projection (iv) Orthographic projection
7. Which one of these statements is not a Gauss-Markov assumption?
(i) That the error term has a conditional mean of zero
(ii) Absence of influential observations.
(iii) That the error term has constant variance
(iv) That the errors are uncorrelated
8. Which of the following things can be accomplished with linear model?
(i) Flexibly fit complicated functions
(ii) Uncover complex multivariate relationships
(iii) Build accurate prediction models
(iv) All of the mentioned
9. Which of the following is a good way of performing experiments in data science?
(i) Generalize to the problem (ii) Have Replication
(iii) Measure variability (iv) All of the mentioned
10. Which of the following is performed by Data Scientist?
(i) Challenge results (ii) Create reproducible code
(iii) Define the question (iv) All of the mentioned

Cont...

SECTION - B (25 Marks)

Answer **ALL** questions
ALL questions carry **EQUAL** Marks

(5 x 5 = 25)

- 11 a. Analyze the need to study Data Science.
OR
11 b. Justify the importance of studying Mathematics for understanding data Science.
- 12 a. Identify the advantages of parallel query processing.
OR
12 b. Appreciate the features of Hadoop.
- 13 a. Evaluate Linear Regression with an example.
OR
13 b. Elucidate EM algorithm.
- 14 a. Illustrate Guass Markov theorem with an example.
OR
14 b. What is the SVD technique?
- 15 a. Discuss Page Rank algorithm.
OR
15 b. Interpret Opinion Minig in Data science.

SECTION -C (40 Marks)

Answer **ALL** questions
ALL questions carry **EQUAL** Marks
Question No. 16 is Compulsory

(5 x 8 = 40)

- 16 Evaluate Data wrangling, and Data Cleaning.
- 17 a. Compare and contrast SQL and No- SQL.
OR
17 b. Enumerate Map reduce.
- 18 a. Assess Normal Equations.
OR
18 b. Explain Rank deficiency and state the methods to deal it.
- 19 a. Criticize Gram-Schmidt orthogonalization process in Mathematical Model.
OR
19 b. Determine the difference between univariate and multivariate visualization.
- 20 a. Evaluate Collaborative network and data Science.
OR
20 b. State the role of Semantic Web in data Science.

Z-Z-Z

END