

B.E. / B.Tech. - DEGREE EXAMINATIONS, NOV / DEC 2025

Seventh Semester

Information Technology

20ITPC702 - DATA SCIENCE WITH MACHINE LEARNING

Regulations - 2020

Duration: 3 Hours

Max. Marks: 100

PART - A (MCQ) (10 × 1 = 10 Marks)

Answer ALL Questions

	<i>Marks</i>	<i>K- Level</i>	<i>CO</i>
1. What is the first step in the Data Science process? (a) Data Analysis (b) Data Cleaning (c) Data Collection (d) Data Visualization	1	K1	CO1
2. Which SQL command is used to retrieve data from a database? (a) INSERT (b) SELECT (c) UPDATE (d) DELETE	1	K1	CO1
3. In R, which function is used to combine elements into a vector? (a) combine() (b) c() (c) vector() (d) append()	1	K1	CO2
4. Which function is used to read a CSV file into R? (a) read.csv() (b) load.csv() (c) read. file() (d) import.csv()	1	K2	CO2
5. Which type of machine learning uses labeled data? (a) Supervised learning. (b) Unsupervised learning. (c) Reinforcement learning. (d) Semi-supervised learning.	1	K1	CO3
6. A decision tree is primarily used for (a) Clustering. (b) Regression and classification. (c) Reinforcement learning. (d) Dimensionality reduction.	1	K1	CO3
7. In agglomerative clustering, clusters are formed by (a) Dividing data points into smaller groups. (b) Merging smaller clusters into larger ones. (c) Assigning data points to random clusters. (d) Separating clusters based on their labels.	1	K1	CO4
8. Which of the following is a common application of unsupervised learning? (a) Email spam detection. (b) Customer segmentation. (c) Predicting stock prices. (d) Image classification.	1	K1	CO4
9. What does a multivariate decision tree do differently compared to a univariate decision tree? (a) It uses a single feature to make splits at each node. (b) It uses a combination of features for splits at each node. (c) It is used only for regression problems. (d) It only considers categorical features for splits.	1	K1	CO5
10. What is the primary goal of Reinforcement Learning? (a) To minimize the error in supervised learning. (b) To learn the optimal policy for maximizing cumulative reward. (c) To cluster data into distinct groups. (d) To generate synthetic data samples.	1	K1	CO6

PART - B (12 × 2 = 24 Marks)

Answer ALL Questions

11. State the role of Machine Learning in Data Science.	2	K1	CO1
12. Summarize the main steps involved in Data Exploration.	2	K2	CO1
13. Outline the use of length() function.	2	K2	CO2
14. Demonstrate how to create a data frame in R with Name, Age, and Gender.	2	K2	CO2

15. Distinguish between univariate and multivariate decision trees. 2 K2 CO3
16. Differentiate classification and regression. 2 K2 CO3
17. State the advantages and disadvantages of Hierarchical Clustering. 2 K1 CO4
18. Calculate the centroid between the three points. A(4, 9), B(6, 15), and C(2, 6). 2 K2 CO4
19. Clarify how rewards guide the learning process in reinforcement learning. 2 K2 CO5
20. What is overfitting and underfitting? 2 K1 CO5
21. Summarize the importance of subset selection in regression models. 2 K2 CO6
22. How Factor Analysis used to interpret underlying structures in psychological testing? 2 K1 CO6

PART - C (6 × 11 = 66 Marks)

Answer ALL Questions

23. a) Summarize the essential technologies that support Data Science process. 11 K2 CO1
- OR**
- b) Explain the various techniques for exploring data. 11 K2 CO1
24. a) Apply probability distribution functions to simulate and analyze random data. Interpret the outcomes using suitable plots. 11 K3 CO2
- OR**
- b) Construct a data frame to represent student details (Name, Roll No, Marks in 3 subjects, and Grade). Demonstrate how to add new columns, subset data, and compute average marks. 11 K3 CO2
25. a) Apply the concepts of Machine Learning types to design a basic learning system and explain its components. 11 K3 CO3
- OR**
- b) Classify the given sample using a decision tree algorithm. Apply the same to construct a decision tree for the data given below. 11 K3 CO3

Size	Color	Shape	Class
Small	Yellow	Round	A
Big	Yellow	Round	A
Big	Red	Round	A
Small	Red	Round	A
Small	Black	Round	B
Big	Black	Cube	B
Big	Yellow	Cube	B
Big	Black	Round	B
Small	Yellow	Cube	B

26. a) Perform K means clustering on these datapoints: 11 K3 CO4
 A1(2,10), A2(2,5), A3(8,4), B1(5,8), B2(7,5), B3(6,4), C1(1,2), C2(4,9) with
 Initial Centroids: A1(2,10), B1(5,8), C1(1,2).
 Perform one full iteration of K-means by computing
 1. Euclidean distance of each point.
 2. Cluster Assignment.
 3. New Centroid for each cluster

OR

- b) Given the following five data points, perform Hierarchical Clustering using the 11 K3 CO4
 Single Linkage method and construct a Dendrogram.

Point	P1	P2	P3	P4	P5
P1	0	3	8	4	7
P2	3	0	7	6	5
P3	8	7	0	9	4
P4	4	6	9	0	5
P5	7	5	4	5	0

27. a) Help the Robot to reach the target block using Reinforcement learning. 11 K3 CO5

s1	s2	s3	Diamond s4
s5	Wall s6	s7	Fire s8
□ s9	s10	□ s11	s12

□ → Robot

Fire (Obstacle) -1

Target (Diamond) +1

OR

- b) Justify why pruning is required in decision tree algorithms. Also, discuss the different 11 K3 CO5
 types of pruning techniques with suitable examples.
28. a) Illustrate the Bayesian graphical models in detail. 11 K2 CO6
- OR**
- b) Interpret how PCA transforms correlated variables into principal components to 11 K2 CO6
 reduce dimensions while keeping key information intact. Give an example.