

task: open CR every 7 hours

PRN No.

300



Dr. Vishwanath Karad
MIT WORLD PEACE
UNIVERSITY | PUNE
TECHNOLOGY, RESEARCH, SOCIAL INNOVATION & PARTNERSHIPS

School of Computer Science and Engineering

Department of Computer Engineering and Technology

TY BTech (AIDS) (Academic Year 2025-26)

Mid Term Exam - Semester V

Course Name:- Machine Learning

Course Code:- AID3PM03A

Maximum Marks: 30

Time: 1 Hrs

Date: 6/10/2025

Instructions:-

1. Attempt any 3 questions from Q. 1 to Q. 4 AND Attempt any 3 questions from Q. 5 to Q. 8
2. Figure to the right indicates full marks.
3. Use of cell phone is prohibited in the examination hall.
4. Neat diagrams must be drawn wherever necessary.
5. Assume suitable data, if necessary and clearly state.
6. Use of scientific calculator is allowed

Attempt any 3 questions from Q. 1 to Q. 4			
Q.1	CO1/applying	State the significance of data normalization. Given a categorical feature 'Color = {Red, Blue, Green}', demonstrate how you would apply one-hot encoding to represent this feature.	[5 Marks]
Q.2	CO1/analyzing	State the various types of Machine Learning paradigms. Suppose, you have a large set of unlabelled customer data. Which Machine Learning paradigm would you use to discover patterns, and why?	[5 Marks]
Q.3	CO1/remembering	Why is Data preparation important? Explain any two methods to handle missing data for a given dataset.	[5 Marks]
Q.4	CO1/understanding	Critically evaluate the role of Exploratory Data Analysis (EDA) in Machine Learning. Describe any two EDA plots with its importance.	[5 Marks]
Attempt any 3 questions from Q. 5 to Q. 8			
Q.5	CO2/understanding	With the help of a neat diagram, explain the working principle of Bootstrap Aggregation (Bagging) algorithm in Machine Learning.	[5 Marks]

Q.6	CO2/ remembering	Explain Support Vector Classifier (SVC). State the significance of support vectors with the help of a diagram.	[5 Marks]									
Q.7	CO2/ analyzing	Illustrate the concept of bias and variance in the context of a dataset, and support your explanation with a clear diagram.	[5 Marks]									
Q.8	CO2/ evaluate	A quality control detection system of an industry, classifies the object assessment quality test as either "PASS" (Positive) or "FAIL" (Negative). Calculate: Accuracy, Recall (Sensitivity), Precision, Specificity and FalsePositiveRate (FPR). The confusion matrix below summarizes the model's performance on 1000 objects. Confusion Matrix: <table><tr><td></td><td>Predicted Positive</td><td>Predicted Negative</td></tr><tr><td>Actual Positive</td><td>400</td><td>150</td></tr><tr><td>Actual Negative</td><td>100</td><td>350</td></tr></table>		Predicted Positive	Predicted Negative	Actual Positive	400	150	Actual Negative	100	350	[5 Marks]
	Predicted Positive	Predicted Negative										
Actual Positive	400	150										
Actual Negative	100	350										

task: open CR every 7 hours
if CR == open

09/12/2025



Dr. Maheshwari Mahesh
**MIT WORLD PEACE
UNIVERSITY** | PUNE
TECHNOLOGY, BUSINESS, SOCIAL, ENVIRONMENTAL, LEADERSHIP



PRN: 1

600000

Term End Examinations December 2025

AID3PM03A/CET3006B - Machine Learning

Question Paper ID: 066100

Faculty/School	School of Computer Science and Engineering	Term	Semester V
Program	TY B.Tech AIDS CSE	Duration	1 hour 30 minutes
Specialization	-	Max. Marks	40

Section - 1 (8 X 5 Marks)

Answer any 8 questions

1	Compare between supervised learning and unsupervised learning.	5 marks	CO1	Understanding
2	With the help of an example describe the working of Decision Tree algorithm.	5 marks	CO2	Understanding
3	Explain the difference between centroid-based clustering and density-based clustering	5 marks	CO3	Understanding
4	A project uses DBSCAN with parameters $\epsilon = 0.5$ and $\text{minPts} = 5$. After clustering, you find many points labelled as "noise". Describe how you would apply a different parameter setting or algorithm to reduce noise points and obtain meaningful clusters.	5 marks	CO3	Applying
5	Describe the difference between L1 (LASSO) and L2 (Ridge) regularization.	5 marks	CO3	Understanding

task: open CR every 7 mins

(285)

6	You are a data scientist at a retail company. You have collected transaction data for 12 customers, each represented by two features: Feature X = number of items bought last month Feature Y = total spend last month (in thousands of currency units) The data (customers labelled A-L) are: A: (2, 1.5) B: (3, 2.0) C: (2, 1.0) D: (8, 5.5) E: (7, 6.0) F: (9, 5.0) G: (14, 9.0) H: (13, 8.5) I: (12, 9.5) J: (20, 12.0) K: (18, 11.5) L: (19, 12.5) Your task is to design a complete k-medoids clustering solution (choose $k = 3$ clusters) using Manhattan distance (L_1) as the dissimilarity measure.	5 marks	CO3, CO4	Evaluating
7	Elaborate the concept of hidden states and observation states in a Hidden Markov Model (HMM). Provide an example.	5 marks	CO3	Understanding
8	Brief two applications where Transfer Learning is commonly used	5 marks	CO4	Understanding
9	A logistics company deploys an autonomous robot in a large warehouse. The robot's goal is to pick and place items from storage to packaging stations. The environment is dynamic: new items arrive continuously, storage locations change, and multiple robots share the space causing possible collisions or traffic. The company wants to minimize order-fulfilment time, energy consumption, and avoid idle time. Identify and clearly describe the <i>states</i>, <i>actions</i>, <i>reward structure</i>, <i>policy</i>, and <i>environment transitions</i> in this scenario.	5 marks	CO4	Applying
10	Define Genetic Algorithm. Describe the role of fitness function in Genetic Algorithm	5 marks	CO1, CO4	Understanding

END OF QUESTION PAPER