

Roll Number: \_\_\_\_\_

**Thapar Institute of Engineering & Technology, Patiala**  
Department of Computer Science and Engineering  
**Mid Semester Test (MST)**

B.E. (5 <sup>th</sup> Semester)	Course Code: UML501
	Course Name: Machine Learning
Date: October 03, 2023	Time: 03:00 PM – 05:00 PM
Duration: 2 Hours, M. Marks: 35	Name of Faculty: Dr. Jatin, Dr. Harpreet, Dr. Ashutosh, Dr. Jyoti, Dr. Anjula, Dr. Arun, Dr. Sumit

**Note: Attempt all questions in a proper sequence. Assume missing data, if any, suitably.**

Q1	<p>Categorize the following real-world tasks as classification, regression, clustering or reinforcement learning. Justify your answer in one line:</p> <p>a) Predicting the electricity consumption of a household based on the number of occupants, time of day, and weather conditions.</p> <p>b) Predicting the likelihood of a student passing an exam based on study hours, previous grades, and attendance.</p> <p>c) Training an autonomous vehicle to navigate through city traffic, obeying traffic rules and optimizing travel time.</p> <p>d) Predicting the duration of a flight based on its origin, destination, and other flight-related factors.</p> <p>e) Teaching a chatbot to recommend movies to users based on their viewing history and preferences.</p> <p>f) Organizing a collection of articles into topics based on their content to aid in content management.</p> <p>g) Training a robot to sort and categorize recyclable materials into different bins based on material type and shape.</p>	(7)														
Q2 (a)	<p>The values of a given set of data are grouped into intervals. Find first, second and third quartile of the below given dataset.</p> <table><tr><td><i>age</i></td><td><i>frequency</i></td></tr><tr><td>1–5</td><td>200</td></tr><tr><td>5–15</td><td>450</td></tr><tr><td>15–20</td><td>300</td></tr><tr><td>20–50</td><td>1500</td></tr><tr><td>50–80</td><td>700</td></tr><tr><td>80–110</td><td>44</td></tr></table>	<i>age</i>	<i>frequency</i>	1–5	200	5–15	450	15–20	300	20–50	1500	50–80	700	80–110	44	(3)
<i>age</i>	<i>frequency</i>															
1–5	200															
5–15	450															
15–20	300															
20–50	1500															
50–80	700															
80–110	44															
Q2 (b)	<p>Given the following data matrix A,</p> $A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & -1 & 0 \end{bmatrix}$ <p>a) Compute the covariance matrix of A</p> <p>b) Compute Eigen Values and Eigen Vectors of Covariance(A).</p>	(4)														
Q3(a)	<p>In the Multiple Linear Regression model, the error (e) can be represented as: <b>e=y-Xb</b>.</p> <p>Write down the cost (objective) function equation for the multiple linear regression model with ridge regularization using the Least square error method. Derive the equation to compute optimal values of the coefficient's matrix for which the total cost function is minimum.</p>	(4)														

Q3(b)	<p>Consider the hours spent in studying ML(x) affect the scores in the MST (y) and the regression line equation to define this relationship is given by:</p> $y_{pred} = 0.152 + 4.316x.$ <table><tr><th>x</th><th>y</th></tr><tr><td>2</td><td>8</td></tr><tr><td>3</td><td>12</td></tr><tr><td>5</td><td>20</td></tr><tr><td>7</td><td>30</td></tr></table> <p>Find MAE, MSE, R<sup>2</sup> score using the above dataset.</p>	x	y	2	8	3	12	5	20	7	30	(3)																							
x	y																																		
2	8																																		
3	12																																		
5	20																																		
7	30																																		
Q4	<p>Imagine you are working with a dataset of loan approvals. You are using a classifier with the regression equation <math>Y = (-1) + 0.2x</math>, where 'Y' represents regression outcome, and 'x' represents the applicant's credit score. No preprocessing is applied to the data.</p> <table><tr><th>Applicant Id</th><th>Credit Score</th><th>Loan Approval (1=Yes, 0=No)</th></tr><tr><td>1</td><td>630</td><td>0</td></tr><tr><td>2</td><td>680</td><td>1</td></tr><tr><td>3</td><td>720</td><td>1</td></tr><tr><td>4</td><td>610</td><td>0</td></tr><tr><td>5</td><td>690</td><td>1</td></tr><tr><td>6</td><td>650</td><td>0</td></tr><tr><td>7</td><td>700</td><td>1</td></tr><tr><td>8</td><td>660</td><td>1</td></tr><tr><td>9</td><td>680</td><td>1</td></tr><tr><td>10</td><td>640</td><td>0</td></tr></table> <p>a) Which type of regression model you can use here; Linear Regression or Logistic Regression?</p> <p>b) Calculate the probability of a loan being approved for an applicant with a credit score of 650.</p> <p>c) Determine the minimum credit score required for an applicant to have a loan approval probability of 80% or higher.</p>	Applicant Id	Credit Score	Loan Approval (1=Yes, 0=No)	1	630	0	2	680	1	3	720	1	4	610	0	5	690	1	6	650	0	7	700	1	8	660	1	9	680	1	10	640	0	(1+3+3)
Applicant Id	Credit Score	Loan Approval (1=Yes, 0=No)																																	
1	630	0																																	
2	680	1																																	
3	720	1																																	
4	610	0																																	
5	690	1																																	
6	650	0																																	
7	700	1																																	
8	660	1																																	
9	680	1																																	
10	640	0																																	
Q5(a)	<p>Show that the logistic sigmoid function satisfies the property <math>\sigma(-a) = 1 - \sigma(a)</math> and that its inverse is given by <math>\sigma^{-1}(y) = \ln \{y/(1 - y)\}</math>.</p>	(3)																																	
Q5(b)	<p>A machine learning model is trained to predict tumour in patients. The test dataset consists of 100 people. Out of them, 18 people are actually having the tumour. According to the ML model predictions there are 32 tumour cases; however, on comparison with actual observations, it has been identified that among these predicted cases only 10 people are actually having tumour. Consider the people with tumour as 'positive' class and the people without tumour as 'negative'. With the help of given information, design and calculate the following:</p> <p>a) Confusion matrix</p> <p>b) Precision</p> <p>c) Recall</p> <p>d) F-1 score</p>	(4)																																	