

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

TEST -2 EXAMINATION- 2024

M.Tech-I Semester (CSE(IS))

COURSE CODE (CREDITS): 13M1WCI331

MAX. MARKS: 25

COURSE NAME: Machine Learning

COURSE INSTRUCTORS: HRI

MAX. TIME: 1 Hour 30 Minutes

Note: (a) All questions are compulsory.

(b) The candidate is allowed to make Suitable numeric assumptions wherever required for solving problems

Q.No	Question	CO	Marks																				
Q1	<p>Compute R^2 for the following observations? Write all the steps involved.</p> <table border="1"> <thead> <tr> <th>Actual observed (y)</th> <th>Predicted (y')</th> <th>Irreducible error (E)</th> </tr> </thead> <tbody> <tr> <td>2</td> <td>3</td> <td>0.2</td> </tr> <tr> <td>4</td> <td>3</td> <td>0.4</td> </tr> <tr> <td>5</td> <td>6</td> <td>0.3</td> </tr> <tr> <td>7</td> <td>8</td> <td>0.5</td> </tr> <tr> <td>9</td> <td>9</td> <td>0.3</td> </tr> </tbody> </table>	Actual observed (y)	Predicted (y')	Irreducible error (E)	2	3	0.2	4	3	0.4	5	6	0.3	7	8	0.5	9	9	0.3	CO1	[3]		
Actual observed (y)	Predicted (y')	Irreducible error (E)																					
2	3	0.2																					
4	3	0.4																					
5	6	0.3																					
7	8	0.5																					
9	9	0.3																					
Q2	Why is R^2 a better metric as compared to RSE?	CO1	[2]																				
Q3	<p>Compute TP, TN, FP, and FN values for the class Setosa.</p> <table border="1"> <thead> <tr> <th></th> <th>Setosa</th> <th>Versicolor</th> <th>Verginica</th> </tr> </thead> <tbody> <tr> <th>Setosa</th> <td>16</td> <td>0</td> <td>0</td> </tr> <tr> <th>Versicolor</th> <td>0</td> <td>17</td> <td>1</td> </tr> <tr> <th>Verginica</th> <td>0</td> <td>0</td> <td>11</td> </tr> </tbody> </table>		Setosa	Versicolor	Verginica	Setosa	16	0	0	Versicolor	0	17	1	Verginica	0	0	11	CO2	[3]				
	Setosa	Versicolor	Verginica																				
Setosa	16	0	0																				
Versicolor	0	17	1																				
Verginica	0	0	11																				
Q4	Derive the gradient descent parameters update for regularized linear regression model.	CO2	[3]																				
Q5	Describe three ways of addressing over-fitting.	CO2	[2]																				
Q6	For a certain plat come from three different cities A, B and C where 30% are from A, 25% are from B and 45% are from C. Out of those 2% from A, 1% from B and 3% from C are found COVID positive. Find the probability that a worker chosen at random would be tested COVID positive and also find the probability that if a worker is tested COVID positive, what is the probability that he is coming from A or B or C?	CO3	[5]																				
Q7	<p>Draw a Decision Tree for the following data using (a) Information gain and (b) GINI Index for a Training set: 3 features and 2 classes (Y/N).</p> <table border="1"> <thead> <tr> <th>X</th> <th>Y</th> <th>Z</th> <th>C</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>1</td> <td>1</td> <td>Y</td> </tr> <tr> <td>1</td> <td>1</td> <td>0</td> <td>Y</td> </tr> <tr> <td>0</td> <td>0</td> <td>1</td> <td>N</td> </tr> <tr> <td>1</td> <td>0</td> <td>0</td> <td>N</td> </tr> </tbody> </table>	X	Y	Z	C	1	1	1	Y	1	1	0	Y	0	0	1	N	1	0	0	N	CO4	[5]
X	Y	Z	C																				
1	1	1	Y																				
1	1	0	Y																				
0	0	1	N																				
1	0	0	N																				
Q8	Assume that a dataset has two classes in the output (Yes/No). Describe the minimum and maximum values of probability(Yes), probability(No), entropy, and GINI Index. Draw a graph between the entropy and GINI Index vs probability (Yes/No)	CO4	[2]																				