

May-June 2024	
Program: B.Tech	Scheme : II
Regular: TY	Semester: VI
Course Code: AIC602 and Course Name: Machine Learning	
Date of Exam: 17/5/2024	Duration: 02.5 Hours
Max. Marks:60	

Instructions:				
(1) All questions are compulsory. (2) Draw neat diagrams wherever applicable. (3) Assume suitable data, if necessary				
		Max. Marks	CO	BT level
Q 1	Solve any six questions out of eight:	12		
i)	Define ROC. What is the significance of ROC-AUC.	2	CO1	Un
ii)	Define overfitting and underfitting.	2	CO1	Un
iii)	What are the benefits and limitations of KNN algorithms?	2	CO3	Un
iv)	What is logistic regression?	2	CO2	Un
v)	What is the main difference between <i>k-Means</i> and <i>k-Nearest neighbors</i> ?	2	CO3	Un
vi)	Define information Gain and Gini index.	2	CO2	Un
vii)	What is the learning rate?	2	CO4	Un
viii)	What is a major goal of the dimensionality reduction technique?	2	CO5	Un
Q.2	Solve any four questions out of six.	16		
i)	Explain the supervised learning method with an example.	4	CO1	Un
ii)	Discuss entropy in the ID3 algorithm with an example.	4	CO2	Un
iii)	What's the "kernel trick" and how is it useful?	4	CO3	Un
iv)	What Is Principal Component Analysis (PCA) & How Does It Work?	4	CO5	Un
v)	Explain the random search method for optimizing models.	4	CO4	Un
vi)	What are the different attribute selection measures in the decision tree?	4	CO2	Un
Q.3	Solve any two questions out of three.	16		
i)	The sales (in million dollars) of the company for each year are shown in the table below.	8	CO2	Ap

K. J. Somaiya Institute of Technology, Sion, Mumbai-22
(Autonomous College Affiliated to University of Mumbai)

	<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 15%;">x(year)</td> <td style="width: 15%;">2005</td> <td style="width: 15%;">2006</td> <td style="width: 15%;">2007</td> <td style="width: 15%;">2008</td> <td style="width: 15%;">2009</td> </tr> <tr> <td>y(sales)</td> <td>12</td> <td>19</td> <td>29</td> <td>37</td> <td>45</td> </tr> </table>	x(year)	2005	2006	2007	2008	2009	y(sales)	12	19	29	37	45																
x(year)	2005	2006	2007	2008	2009																								
y(sales)	12	19	29	37	45																								
	<p>a. Find the least squares regression line $y=ax+b$.</p> <p>b. use the least squares regression line as a model to estimate the sales of the company in 2012.</p>																												
ii)	<p>Find the clusters using single link technique. Use Euclidean distance and draw the dendrogram.</p> <table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="width: 15%;">Sample No.</th> <th style="width: 15%;">P1</th> <th style="width: 15%;">P2</th> <th style="width: 15%;">P3</th> <th style="width: 15%;">P4</th> <th style="width: 15%;">P5</th> <th style="width: 15%;">P6</th> </tr> </thead> <tbody> <tr> <td>X</td> <td>0.40</td> <td>0.22</td> <td>0.35</td> <td>0.26</td> <td>0.08</td> <td>0.45</td> </tr> <tr> <td>Y</td> <td>0.53</td> <td>0.38</td> <td>0.32</td> <td>0.19</td> <td>0.41</td> <td>0.30</td> </tr> </tbody> </table>					Sample No.	P1	P2	P3	P4	P5	P6	X	0.40	0.22	0.35	0.26	0.08	0.45	Y	0.53	0.38	0.32	0.19	0.41	0.30	8	CO3	Ap
Sample No.	P1	P2	P3	P4	P5	P6																							
X	0.40	0.22	0.35	0.26	0.08	0.45																							
Y	0.53	0.38	0.32	0.19	0.41	0.30																							
iii)	<p>Positively labeled data points (2,2)(2,-2)(-2,-2)(-2,2) and Negatively labeled data points (1,1)(1,-1)(-1,-1)(-1,1). Find a hyper-plane for dividing data into two different classes using SVM.</p> $\Phi_1 \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{cases} \begin{pmatrix} 4 - x_2 + x_1 - x_2 \\ 4 - x_1 + x_1 - x_2 \end{pmatrix} & \text{if } \sqrt{x_1^2 + x_2^2} > 2 \\ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} & \text{otherwise} \end{cases}$					8	CO3	Ap																					
Q.4	Solve any two questions out of three.					16																							
i)	<p>What type of machine learning problem is,</p> <p>a) Predicting the survival of a passenger in the Titanic disaster</p> <p>b) Recognizing handwritten digit</p> <p>c) Forecasting sales for next 6 months for D-Mart</p> <p>d) Suggesting songs on Spotify</p> <p>e) Identifying a fraudulent,</p> <p>Explain each in detail.</p>					8	CO1	Un																					
ii)	<p>Describe derivative free methods. Why is it called the Derivative free method?</p>					8	CO4	Un																					
iii)	<p>Given the data in the table, reduce the dimension from 2 to 1 using principle component analysis algorithm.</p> <table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 20%;">Systolic BP</td> <td style="width: 10%;">126</td> <td style="width: 10%;">128</td> <td style="width: 10%;">128</td> <td style="width: 10%;">130</td> <td style="width: 10%;">130</td> <td style="width: 10%;">132</td> </tr> <tr> <td>Diastolic BP</td> <td>78</td> <td>80</td> <td>82</td> <td>82</td> <td>84</td> <td>86</td> </tr> </table>					Systolic BP	126	128	128	130	130	132	Diastolic BP	78	80	82	82	84	86	8	CO5	Ap							
Systolic BP	126	128	128	130	130	132																							
Diastolic BP	78	80	82	82	84	86																							
