

**VIT**

Vellore Institute of Technology

(Approved by the Council for Higher Education, Government of Tamil Nadu)

School of Computer Science and Engineering**Winter Semester 2023-24****Continuous Assessment Test -- 1****SLOT: C1+TC1****Programme Name & Branch : M.Tech (Integrated) Data Science****Course Name & code: Machine Learning for Data Science (MDI4001)****Class Number : 2477, 2478****Faculty Name : Dr. Anisha M. Lal, Dr. Jyotismita Chaki****Exam Duration: 90 Mins****Maximum Marks: 50**

Q. No:	Question	Max. Marks																
1.	Discuss about how to handle the challenges in machine learning listed below, a) Poor Quality Data b) Irrelevant Features c) Overfitting the Training Data d) Underfitting the Training Data	10																
2.	a) Discuss the methods used for feature scaling with proper example. (6) b) Discuss the need of regularization along with a suitable example. (4)	10																
3.	The following data relate to advertisement expenditure (in lakh of rupees) and their corresponding sales (in lakh of rupees). <table border="1"><tr><td>Advertisement expenditure</td><td>40</td><td>50</td><td>38</td><td>60</td><td>65</td><td>50</td><td>35</td></tr><tr><td>Sales</td><td>36</td><td>60</td><td>55</td><td>50</td><td>60</td><td>48</td><td>30</td></tr></table> Perform the following, a) Estimate the parameter values using linear regression (5) b) Draw the graph and represent the linear regression line fit (3) c) Estimate the sales corresponding to advertising expenditure of Rs. 30 lakhs. (2)	Advertisement expenditure	40	50	38	60	65	50	35	Sales	36	60	55	50	60	48	30	10
Advertisement expenditure	40	50	38	60	65	50	35											
Sales	36	60	55	50	60	48	30											
4.	Explicate the types of gradient descent algorithms (along with suitable examples) which minimizes the cost function in machine learning.	10																

5. a) Interpret the observations from the below graph (4)

10

Figure 1: Plot is generated on non-linear data using linear regression

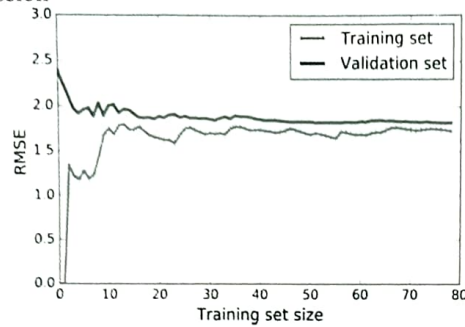
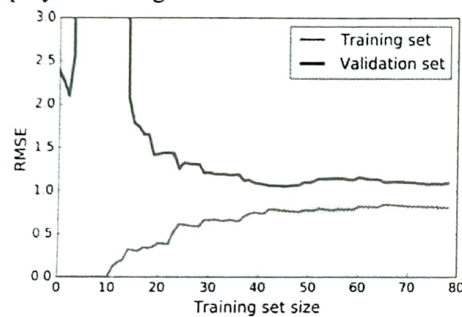


Figure 2: Plot is generated on non-linear data using 10-degree polynomial regression



b) Calculate the precision and recall for the given confusion matrix and plot the ROC curve. (6)

		Predicted		
		Greyhound	Mastiff	Samoyed
Actual	Greyhound	240	15	18
	Mastiff	21	310	14
	Samoyed	22	10	170