Total No. of Printed Pages: 3

PRN No.	

PAPER CODE	U313-293(RE)

December 2023 (REEXAM)

TY (SEMESTER - I)

COURSE NAME: MACHINE LEARNING **Branch: ELECRONICS AND TELECOMMUNICATION** **COURSE CODE: ETUA 31203**

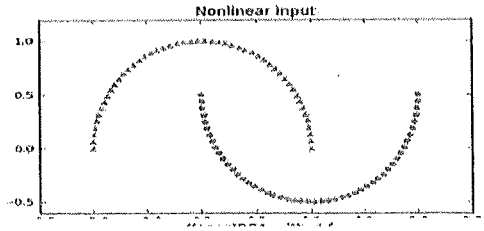
(PATTERN 2020)

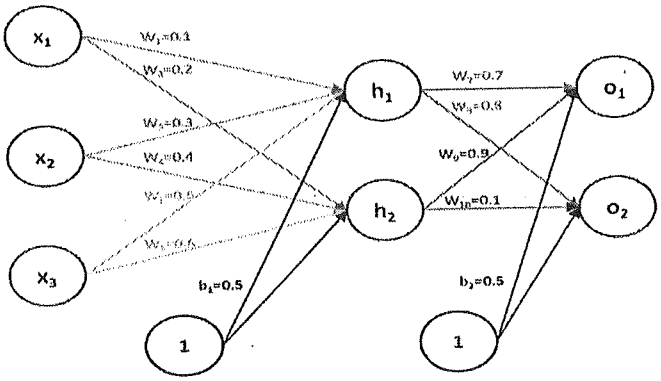
Time: [2 Hrs] **[Max. Marks: 60]**

(*) Instructions to candidates:

1) **Figures to the right indicate full marks.**
2) **Use of scientific calculator is allowed**
3) **Use suitable data wherever required**
4) **All questions are compulsory. Solve any two sub questions each from each Question 1 ,2, 3,4,5,and 6 respectively**

Q. No.		Question Description	Max. Marks	CO mapped	BT Level
Q.1	a)	Calculate the z-score for array [20,15,10,5,0].	[5]	CO1	Apply
	b)	Suppose the p.m.f.(probability mass function) of the discrete random variable X is: x | 0 1 2 3 $f(x)$ | 0.2 0.1 0.4 0.3 Calculate expectations $\mathbb{E}$ (2), $\mathbb{E}$ (x) and $\mathbb{E}$ (2x).	[5]	CO1	Apply
	c)	Overfitting of model reduces accuracy and its said to be a big problem in Machine learning. Explain different techniques to reduce /avoid overfitting.	[5]	CO1	Under-Stand
Q2	a)	In a regression model ,the cost function $J(\theta) = \theta x^3$, the old $\theta = 0.5$, learning rate $\alpha = 0.05$ and input x=2,calculate the updated θ using gradient descent. Explain all steps of gradient descent algorithm.	[5]	CO2	Apply
	b)	In a univariate linear regression model with θ_0=1, θ_1=2 and x= 1.Calculate the output of model when passed through tanh activation.	[5]	CO2	Apply
	c)	The sigmoid activation function is given by $f(z) = \frac{1}{1+e^{-z}}$, prove that $\frac{\delta f(z)}{\delta z} = f(z)(1 - f(z))$	[5]	CO2	Under-Stand

Q3.	a)	In Naïve Bayes Classification, predicting the possibility of playing Golf, the frequency table for the weather outlook is as shown below **Frequency Table** / **Play Golf**: Yes / No Outlook: Sunny 3 2; Overcast 4 0; Rainy 2 3 Compute the posterior probability P (No| Sunny) and P(Yes| Rainy).	[5]	CO3	Apply
	b)	Given the following confusion matrix for a classification task with three classes (A, B and C), Actual /Predicted | A | B | C A | 6 | 2 | 0 B | 1 | 6 | 0 C | 1 | 1 | 8 Calculate Overall accuracy, Precision, Recall and F1 for all classes.	[5]	CO3	Apply
	c)	In context with Support Vector Machines (SVM) ,explain a)Hyperplane b) Margin and c) Support Vectors. Why SVM are called as "Kernal Machines".	[5]	CO3	Under-Stand
Q.4	a)	In a certain PCA based application, the data is arranged in 2x2 matrix and is as follows $A=\begin{bmatrix}2 & 1\\ 1 & 2\end{bmatrix}$ Calculate the co-variance matrix and Eigen Values.	[5]	CO4	Apply
	b)	With the help of suitable example, illustrate the process of agglomerative hierarchical clustering. Draw the dendrogram for the example used in illustration.	[5]	CO4	Under-Stand
	c)	You are working on some classification problem where the input data is non-linear and as shown in figure below  Suggest and demonstrate a technique through which you can convert this nonlinear data into linear one for smooth partition for robust classification.	[5]	CO4	Apply

Q.5	a)	Let v = [2.1 , 4.8, 3.5],be the inputs to the softmax function, calculate the output for all inputs.	[5]	CO5	Apply
	b)	You are training a multilayer perceptron for certain classification application. What will be the effect on training if use 1) Stochastic gradient 2) Batch gradient and 3) Mini batch gradient.	[5]	CO5	Under Stand
	c)	In the perceptron shown below, output O_1=0.8896, the output of hidden layer h_1=0.9866, target t_1=0.1. The activation function used in output layer is sigmoid. The present value of weight connecting h1 to O1(weight W_7) is 0.7. Calculate the updated value of W_7 if learning rate is 0.01. 	[5]	CO5	Apply
Q.6)	a)	Thirty two filters of size 11x11 is applied to an image of size 327x327 with zero padding and stride of 2, The image is RGB. Calculate the volume of final image?	[5]	CO6	Apply
	b)	What is convolution? Explain 2D convolution process with suitable example? How 2D convolution is used in 3D filtering in Convolutional Neural networks?	[5]	CO6	Under Stand
	c)	"Deep learning networks are more susceptible to overfitting". Justify. What are the techniques used to avoid overfitting in deep networks ?	[5]	CO6	Under Stand