BIRLA INSTITUTE OF TECHNOLOGY, MESRA, RANCHI
(END SEMESTER EXAMINATION)

CLASS: I MSc. **SEMESTER : VII**
BRANCH: QEDS **SESSION : MO/2025**

SUBJECT: ED407 STATISTICAL MACHINE LEARNING II

TIME: 3 Hours **FULL MARKS: 50**

INSTRUCTIONS:
1. The question paper contains 5 questions each of 10 marks and total 50 marks.
2. Attempt all questions.
3. The missing data, if any, may be assumed suitably.
4. Before attempting the question paper, be sure that you have got the correct question paper.
5. Tables/Data hand book/Graph paper etc. to be supplied to the candidates in the examination hall.

			CO	**BL**
Q.1(a)	Consider the following distance matrix between six points. Draw the dendrogram of these points obtained by the complete linkage clustering algorithm.	[5]	CO1	2,3,4

	P1	P2	P3	P4	P5	P6
P1	0.00	0.24	0.22	0.37	0.34	0.23
P2		0.00	0.15	0.20	0.14	0.25
P3			0.00	0.15	0.28	0.11
P4				0.00	0.29	0.22
P5					0.00	0.39
P6						0.00

Q.1(b) Two different real numbers, x and y, are picked (their distributions are unknown). You get to see one of them, say x, and it is known that x is the larger (smaller) with probability 1/2. Construct an algorithm that can decide if x is larger, with error probability less than 1 / 2, no matter what the distributions of x and y are. Using the Bayesian framework, prove that your algorithm is correct. [5] CO1 1,2,3

Q.2(a) A Genetic Algorithm (GA) is used to **maximize** the function $f(x) = x^2$. The chromosome representation is **5-bit binary**, and the search space is $0 \leq x \leq 31$. You are given the following initial population of 4 chromosomes: [5] **CO2** 2,3,4

Chromosome	Binary
C1	01001
C2	10110
C3	00111
C4	11100

Single-point crossover with **Crossover probability** $p_c = 0.8$ and, **Mutation method is** Bit-flip mutation with **Mutation probability** $p_m = 0.05$ is given. Perform **one full GA generation.**

Q.2(b) A Particle Swarm Optimization (PSO) algorithm is used to minimize the function $f(x) = x^2$. Two particles are initialized with the following positions and velocities; Particle 1: $x_1 = 4$, $v_1 = 0.5$ and Particle 2: $x_2 = -2$, $v_2 = -0.5$. The PSO parameters are given as Inertia weight: $w = 0.6$, Cognitive coefficient: $c_1 = 1.5$, Social coefficient: $c_2 = 1.5$. Random numbers for this iteration are given as for Particle 1: $r_1 = 0.4$, $r_2 = 0.7$ and Particle 2: $r_1 = 0.3$, $r_2 = 0.5$. He Initial personal bests and global bests are given as $pbest_1 = 4, pbest_2 = -2, gbest = -2$.
Perform **one full PSO iteration** and compute updated velocities of both particles, updated positions of both particles, updated personal bests and updated global best. [5] CO2 2,3,5

Q.3(a)	Compute the final stacking probability $\hat{y}$ and the class label (use threshold 0.5) for three base classification models, which are trained on the dataset: • **Model A** outputs: $P(y = 1 \mid x) = 0.6$ • **Model B** outputs: $P(y = 1 \mid x) = 0.7$ • **Model C** outputs: $P(y = 1 \mid x) = 0.4$	[5]	CO3	3,4
Q.3(b)	Consider the function $F(\theta) = \sum_{i=1}^{n} \log(\theta + x_i)$ for data $x = \{2, 4, 5\}$. Assume $\theta > 0$. Apply **Minorization-Maximization (MM)** algorithm to maximise $F(\theta)$.	[5]	CO3	3,4

Q.4(a) Perform the Hidden markov modeling for a speech recognition system that uses the HMM **States: T = Talking, S = Silence with following details.** [5] CO4 3,4,5

Initial Probabilities are $P(T) = 0.7, P(S) = 0.3.$

Transition Probabilities

From/To	**T**	**S**
T	0.9	0.1
S	0.4	0.6

Emission (audio readings):

State	**high**	**low**
T	0.8	0.2
S	0.3	0.7

The observation sequence is given as (high, low, high, high).

Q.4(b) A simple **Recurrent Neural Network** has **Input vector:** $x_t \in \mathbb{R}^2$, **Hidden state:** $h_t \in \mathbb{R}^2$ and **Output:** $y_t \in \mathbb{R}^1$. The RNN update equations are: [5] CO4 4,5

$$h_t = \tanh(W_{xh}x_t + W_{hh}h_{t-1} + b_h)$$
$$y_t = W_{hy}h_t + b_y$$

The **Weights, Biases and Initial Hidden States and Inputs are given below:**

$$W_{xh} = \begin{bmatrix} 0.5 & -0.3 \\ 0.8 & 0.2 \end{bmatrix}, W_{hh} = \begin{bmatrix} 0.4 & 0.1 \\ -0.2 & 0.3 \end{bmatrix}; W_{hy} = [1.2 \quad -0.5];\ b_h = \begin{bmatrix} 0.1 \\ -0.2 \end{bmatrix}, b_y = 0.3$$
$$;h_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix};x_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, x_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Compute the **hidden state** h_1 using input x_1 followed by a **hidden state** h_2 using input x_2 and the **output** y_2 at time step 2.

Q.5(a) A company records its **daily sales** for several days and wants to forecast using the **Prophet** model. The Prophet model decomposes the forecast as: [5] CO5 3,4,5

$$y(t) = g(t) + s(t) + h(t)$$

Where $g(t)$: **trend;** (t): **seasonality;** $h(t)$: **holiday effect.** Evaluate **Prophet model forecast** for the following set of inputs.

$$k = 2.5, m = 100, a_1 = 5, b_1 = -3 \ and \ h(t) = \begin{cases} 20 & \text{if } t \text{ is holiday} \\ 0 & \text{otherwise} \end{cases}$$

Q.5(b) Describe LIME (Local Interpretable Model-agnostic Explanations) and SHAP (SHapley Addictive exPlanations) algorithms. A regression model predicts house prices. Instance x_0 has two features: Bedrooms are 3, and the Distance to city are 5 km. LIME samples 5 perturbations as [5] CO5 5,6

Sample	**Bedrooms**	**Distance**	**Model Output**	**Distance**	**Kernel Width σ**
z_1	3	7	220	0.25	0.75
z_2	2	5	180	0.10	0.75
z_3	4	6	260	0.20	0.75
z_4	1	20	120	1.00	0.75
z_5	3	4	250	0.05	0.75

Compute kernel weights and fit a weighted linear model. Provide the LIME feature contributions and interpret the result.

:::::26/11/2025:::::E