

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

TEST -1 EXAMINATION- 2026

B.Tech-VI Semester (BI)

COURSE CODE (CREDITS):18B11BI611

MAX. MARKS: 15

COURSE NAME: MACHINE LEARNING FOR BIOINFORMATICS

COURSE INSTRUCTOR: KLK, SKP

MAX. TIME: 1 Hour

Note: (a) All questions are compulsory.

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

(c) Use of scientific calculators is allowed.

Q.No	Question	CO	Marks
Q1	<p>A system is designed to recognize handwritten digits (0–9) from images (28X28).</p> <ol style="list-style-type: none"> Define the Task (T), Performance Measure (P), and Training Experience (E) for this system. Specify the Target Function of the system. Explain Function Approximation and How Empirical Risk Minimization (ERM) is used in this system and why? What is the Inductive Bias, and why is it important for this system? Suggest one suitable model type for handwritten digit recognition and justify your choice. 	1	4
Q2	<p>A teacher network outputs the softened probabilities.</p> $p(T) = (0.60, 0.25, 0.10, 0.05)$ <p>A student network outputs:</p> $p(S) = (0.50, 0.30, 0.15, 0.05)$ <ol style="list-style-type: none"> Compute the KL divergence: $D_{KL}(p(T) p(S))$. And $D_{KL}(p(S) p(T))$. If the true label is Class 1, compute the cross-entropy loss of the student. 	2	3

Q3	<p>A system is designed to detect whether an email is Spam (Yes) or Not Spam (No). Each email is represented using 4 binary and 4 tertiary features.</p> <ol style="list-style-type: none"> 1. Define the Instance Space (X) for this spam detection system and compute its size. 2. Calculate the Total Number of Concepts (C) for this system. 3. Explain Syntactic Hypotheses and describe how they relate to the model used. 4. Explain Semantic Hypotheses and determine their total number. 	3	2																									
Q4	<p>Consider the following training examples for the target concept EnjoySport, described using three attributes: (4) Sky \in {Sunny, Rainy}, AirTemp \in {Warm, Cold}, and Humidity \in {Normal, High}.</p> <table border="1" data-bbox="287 884 1061 1198"> <thead> <tr> <th>Example</th> <th>Sky</th> <th>AirTemp</th> <th>Humidity</th> <th>EnjoySport</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>Sunny</td> <td>Warm</td> <td>Normal</td> <td>Yes</td> </tr> <tr> <td>2</td> <td>Sunny</td> <td>Cold</td> <td>Normal</td> <td>Yes</td> </tr> <tr> <td>3</td> <td>Rainy</td> <td>Warm</td> <td>High</td> <td>No</td> </tr> <tr> <td>4</td> <td>Rainy</td> <td>Cold</td> <td>Normal</td> <td>No</td> </tr> </tbody> </table> <p>Using the Candidate Elimination algorithm:</p> <ol style="list-style-type: none"> 1. Determine the sequence of updates to the specific boundary (S) and generic boundary (G) after each training example. 2. Find the final specific boundary and final generic boundary. 3. List all hypotheses in the version space consistent with the given training data. 	Example	Sky	AirTemp	Humidity	EnjoySport	1	Sunny	Warm	Normal	Yes	2	Sunny	Cold	Normal	Yes	3	Rainy	Warm	High	No	4	Rainy	Cold	Normal	No	2	4
Example	Sky	AirTemp	Humidity	EnjoySport																								
1	Sunny	Warm	Normal	Yes																								
2	Sunny	Cold	Normal	Yes																								
3	Rainy	Warm	High	No																								
4	Rainy	Cold	Normal	No																								
Q5	<p>Why do robust regression loss functions combine properties of squared error and absolute error? Discuss how this helps in handling outliers in real-world datasets</p>	1	2																									