JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

MOOC End Term Examination- 2025

B.Tech-VII Semester (CSE/IT)

COURSE CODE(CREDITS): 20B2WCI601 MAX. MARKS: 70

COURSE NAME: Introduction to Machine Learning

COURSE INSTRUCTORS: Ekta Gandotra MAX. TIME: 3 Hours

Note: *(a)All questions are compulsory.*
(b)Marks are indicated against each question in square brackets.
(c) The candidate is allowed to make Suitable numeric assumptions wherever required for solving problems
(d) Calculator is allowed.

<table>
<tr><th>Q. No</th><th>Question</th><th>Marks</th></tr>
<tr><td>Q1.</td><td>You are given the training and testing errors of two machine learning models trained on the same dataset:
<table>
<tr><th>Model</th><th>Training Error</th><th>Testing Error</th></tr>
<tr><td>Model A</td><td>4%</td><td>22%</td></tr>
<tr><td>Model B</td><td>18%</td><td>20%</td></tr>
</table>
a. Based on the errors, identify which model is overfitting and which is underfitting. Explain your reasoning by relating the errors to bias and variance.
b. Suggest suitable strategies to reduce overfitting and underfitting for the respective models.</td><td>3

3</td></tr>
<tr><td>Q2.</td><td>What is gradient descent? Explain its purpose in optimization. Outline the step-by-step process by which gradient descent is used to find the local minimum of a differentiable cost function when fitting a regression line $y = mx + c$.</td><td>6</td></tr>
<tr><td>Q3.</td><td>Consider the logistic regression hypothesis function
$$h(X) = \frac{1}{1 + e^{-wX}}$$
Plot $h(X)$ versus $X \in R$ for weights $w \in \{0.5, 2, 10\}$. (A qualitative sketch is sufficient). Based on these plots, explain:
a. How increasing the weight w changes the steepness of the sigmoid curve.
b. Why very large weights can cause overfitting in logistic regression.</td><td>6</td></tr>
<tr><td>Q4.</td><td>Compute the first principal component for the following 2-dimensional dataset. Also, find the proportion of total variance explained by the first and second principal components separately.
X= (x_1, x_2) = (1,2), (3,3), (3,5), (5,4)</td><td>6</td></tr>
</table>

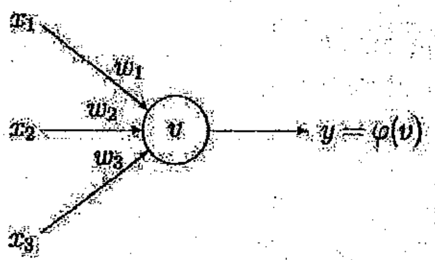
11

Q5.	Describe step-by-step how Linear Discriminant Analysis transforms a labeled dataset into a lower-dimensional space while maximizing class separability. Clearly define the within-class scatter matrix and between-class scatter matrix, and explain how the optimal projection directions are obtained using the generalized eigenvalue problem.	6

Q6.

a. Given a feedforward neural network with an input layer of 3 neurons, one hidden layer with 4 neurons using ReLU activation, and an output layer with 2 neurons using softmax activation, how many weights and biases are there in total? **2**

b. Consider the following diagram of a single artificial neuron. The node has three inputs $x = (x1, x2, x3)$ that receive only binary signals (either 0 or 1). **4**



Suppose that the weights corresponding to the three inputs have the values w1 = 2, w2 = −4 and w3 = 1 and the activation of the unit is given by the following step-function:

$$\varphi(v) = \begin{cases} 1 & \text{if } v \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Calculate the output value y of the unit for each of the following input patterns:

Pattern	P1	P2	P3
x1	1	0	1
x2	0	1	1
X3	0	1	1

Q7.	Explain how bootstrapping is used in bagging. Describe the complete process of training a bagged ensemble model using bootstrapped datasets, and discuss how this approach helps in reducing variance. Provide a simple numerical example to illustrate the concept.	6
Q8.	Explain the Expectation Maximization (EM) algorithm in the context of clustering. Describe its key steps, including the E-step and M-step, and illustrate how the algorithm iteratively improves parameter estimates. Discuss the major advantages and disadvantages of using EM algorithm.	6

Q9. Apply the DBSCAN algorithm to the following dataset and label each data point as Core, Border, or Noise. **6**

A(3, 7), B(4, 6), C(5, 5), D(6, 4), E(7, 3), F(6, 2), G(7, 2), H(8, 4).

Use the parameters Epsilon (ε) = 2, Minimum Points (minPts) = 3, and the following distance matrix.

	A	**B**	**C**	**D**	**E**	**F**	**G**	**H**
A	0	1.41	2.83	4.24	5.66	5.83	6.40	5.83
B		0	1.41	2.83	4.24	4.47	5.00	4.47
C			0	1.41	2.83	3.16	3.61	3.16
D				0	1.41	2.00	2.24	2.00
E					0	1.41	1.00	1.41
F						0	1.00	2.83
G							0	2.24
H								0

Q10. Consider the following dataset.

Student_ID	**Study Hours (H)**	**Attendance (A)**	**Result**
1	High	Good	Pass
2	High	Poor	Pass
3	Medium	Good	Pass
4	Medium	Poor	Fail
5	Low	Good	Fail
6	Low	Poor	Fail
7	High	Good	Pass
8	Medium	Good	Pass
9	Low	Poor	Fail
10	High	Poor	Pass

a. Compute the Information Gain (IG) for both attributes (Study Hours and Attendance) with respect to the class Result. **4**

b. Evaluate why selecting Student_ID as the top feature based on Information Gain is not appropriate for predicting the result. **2**

c. State the major limitation of IG in feature selection and explain how the Gini Index helps to overcome this limitation. **2**

Q11.

a. Describe the architecture of Hidden Markov Model. Explain the role of the Markov assumption and the output independence assumption, and discuss how these assumptions influence the behaviour of the model. **4**

b. Explain geometric interpretation of Support Vector Machine, focussing on the decision boundaries, margins and support vectors using graphical example. **4**

12