

2

JK LAKSHMIPAT UNIVERSITY JAIPUR
INSTITUTE OF ENGINEERING AND TECHNOLOGY
B.Tech., Theory Exam-III Examination, May 2026
Machine Learning

Course Code: CS1138

Time: 2.5 Hrs.

Roll No.....

Max. Marks: 30

Instructions to students:

Do not write anything other than your roll number on question paper.

Assume suitable data wherever required and mention it clearly.

Students are allowed to use calculators.

Writing appropriate units, nomenclature, and drawing neat sketches/schematics/ flow chart wherever, required are an integral part of the answer.

- Q.1** A startup wants to predict the monthly sales (in ₹1000s) based on the amount spent on digital advertising (in ₹1000s) using a simple linear regression model.

4×1

CO1

The observed data is:

Advertising (x)	Sales (y)
1	3
2	5
4	9
5	11

Assume the linear model: $h_{\theta}(x) = \theta_0 + \theta_1 x$

with initial parameters: $\theta_0 = 0, \theta_1 = 1$

Tasks:

- Compute the predicted sales for each observation.
- Calculate the error for each data point.
- Write the gradient expressions for the cost function $J(\theta)$ with respect to θ_0 and θ_1 .
- Perform one iteration of gradient descent with learning rate $\alpha = 0.1$, and compute the updated values of θ_0 and θ_1 .

- Q.2** Suppose a specific classifier classifies images into three categories: Dog, Cat, or Bird. For a given image, the final layer produces logits: $z = [2.0, 1.0, 0.1]$

4×1

**CO1,
CO3**

Compute the following:

- The Softmax output from the logits
- Compute the cross-entropy loss for the true level obtained from the above function.

OR

Consider a soft-margin SVM with regularization parameter $C = 10$.

4×1

- A. Explain how slack variables help in handling non-separable data by illustrating the meaning of $\xi_i = 0$, $0 < \xi_i < 1$, and $\xi_i > 1$.
- B. Derive the mathematical optimization problem of soft-margin SVM from hard-margin SVM.
- C. Discuss what happens when $C \rightarrow \infty$ and when $C \rightarrow 0$.
- D. Discuss the limitations of hard-margin SVM and explain why soft-margin SVM is generally preferred. Support your answer with examples from real-world domains such as healthcare, fraud detection, or text classification.

CO3

Q.3 You are given a small dataset of 8 instances for predicting whether a student will Pass (Yes/No) an exam based on the following attributes:

5×1

- Study Hours: Low / High
- Attendance: Poor / Good
- Assignments Completed: Yes / No

CO3,

CO4

Student	Study Hours	Attendance	Assignments	Pass
1	High	Good	Yes	Yes
2	Low	Poor	No	No
3	High	Poor	Yes	Yes
4	Low	Good	No	No
5	High	Good	No	Yes
6	Low	Poor	Yes	No
7	High	Poor	No	No
8	Low	Good	Yes	Yes

Tasks:

- A. Compute the entropy of the dataset.
- B. Calculate the Information Gain for the attribute “Study Hours”.
- C. Compute the Gini Index of the parent node.
- D. Identify the best attribute for splitting using either IG or Gini (no need to compute all—justify briefly).
- E. For the chosen attribute, state the majority class for each split and count misclassification errors.

Q.4 Use the k-means algorithm and Euclidean distance to cluster the following 8 examples into $k = 3$ clusters: $A_1 = (2,10)$, $A_2 = (2,5)$, $A_3 = (8,4)$, $A_4 = (5,8)$, $A_5 = (7,5)$, $A_6 = (6,4)$, $A_7 = (1,2)$, $A_8 = (4,9)$. Suppose that the initial seeds (centers of each cluster) are A_1 , A_4 and A_7 . Run the k-means algorithm for 1 epoch only. At the end of this epoch show:

2+1+2

CO3,

CO4

- A. The new clusters (i.e. the examples belonging to each cluster)
- B. The centers of the new clusters
- C. How many more iterations are needed to converge? Draw the result for each epoch.

Q.5 A fraud detection system models transactions using a 1-D Gaussian Mixture Model **5×1**
with two components. **CO3**

Component 1: $\pi_1 = 0.7$, $\mu_1 = 50$, $\sigma_1^2 = 25$

Component 2: $\pi_2 = 0.3$, $\mu_2 = 120$, $\sigma_2^2 = 36$

For a transaction amount $x = 90$:

- A. Compute likelihood under each Gaussian component.
- B. Find posterior probabilities of belonging to each component.
- C. Classify the transaction into a component.
- D. Determine an unknown transaction amount $x = ?$ where both posterior probabilities become equal.
- E. Discuss whether this point should be treated as suspicious.

Q. 6 A neural network has the following architecture: **2+2+1**

- Input layer: 2 features (x_1, x_2)
- One hidden layer with 2 neurons
- Output layer with 1 neuron
- Activation (hidden): ReLU
- Activation (output): Sigmoid

CO3,
CO4

Parameters are:

Hidden Layer:

$$w_{11}^{[1]} = 0.6, \quad w_{12}^{[1]} = -0.2, \quad b_1^{[1]} = -0.1$$

$$w_{21}^{[1]} = 0.4, \quad w_{22}^{[1]} = 0.9, \quad b_2^{[1]} = 0.2$$

Output Layer:

$$w_1^{[2]} = -0.5, \quad w_2^{[2]} = 0.8, \quad b^{[2]} = -0.05$$

Input:

$$x_1 = 2, \quad x_2 = 1$$

- A. Compute $z_1^{[1]}$ and $z_2^{[1]}$
- B. Compute $a_1^{[1]}$ and $a_2^{[1]}$
- C. Compute final output $a^{[2]}$

Q.7 A company develops an ML model to shortlist candidates for job interviews. After deployment, it is observed that candidates from certain backgrounds are consistently rejected. What could be the possible sources of bias in this model? **02**
CO4

