Roll No. ......

FIFTH SEMESTER

B. Tech.[CSE]

END SEMESTER EXAMINATION ***Nov-Dec 2025***

CS303/CO327 MACHINE LEARNING

Time: 3:00 Hours ***Max. Marks: 40***

Note: Answer ***ALL*** questions.
Assume suitable missing data, if any.
CO# is course outcome(s) related to the question.
BTL# is associated Blooms' Taxonomy Level

1[a] A bank trains a loan-default prediction model on data from stable economic years, but during an economic downturn the model suddenly misclassifies many customers. Identify the underlying ML issues.
[2] [CO1] [BTL-4]

[b] Identify the main components of a machine learning system (data, model, objective function, optimization algorithm). Which component most strongly influences generalization and why?
[2] [CO1] [BTL-4]

[c] A company wants to detect unusual financial transactions, but only a few examples of known fraud cases exist, while most transactions are normal and unlabeled. The model must infer abnormal patterns without clear labels. Identify the ML technique most suitable for this scenario.
[2] [CO3] [BTL-4]

[d] An e-commerce platform uses user click data (0 for no click, 1 for click) to model the probability that a user will click an ad. Identify the most appropriate probability distribution for modeling this behavior.
[2] [CO1] [BTL-3]

2. Answer *any TWO* of the followings

[a] A bank is developing a logistic regression model to predict whether a loan applicant will default. The model is defined as: $y_i = wx_i + b$, where x_i represents the applicant's income (scaled), and y_i is the default label (0 = no default, 1 = default). The bank collects data from two applicants, shown below:
[2+2] [CO3] [BTL-3,4]

Table I

Applicant	Income (x)	Default (y)
A	2	0
B	5	1

The model parameters are initialized as $w = 0.1$ and $b = 0$. The learning rate for gradient descent is $\alpha = 0.05$. Compute the predicted probabilities $\hat{y}_1$ and $\hat{y}_2$. Also, Compute the batch gradient $\frac{\partial L}{\partial w}$, where L is loss.

[b] A financial analytics firm uses a Support Vector Machine (SVM) to classify stocks as "Buy" or "Sell" based on two technical indicators: X_1: 10-day moving average (MA10), X_2: Volume change (%). After training the model, the SVM produces the following decision boundary: $f(x) = 0.8X_1 - 1.5X_2 - 50 = 0$. Two support vectors are identified: ($X_1 = 75$, $X_2 = 10$) for the "Buy" class, and ($X_1 = 60$, $X_2 = 20$) for the "Sell" class. classify a new stock with $X_1 = 72$, $X_2 = 15$ as "Buy" or "Sell." Explain the importance of these support vectors in determining the width and position of the margin, and why removing non-support-vector points does not change the SVM boundary. **[2+2] [CO3] [BTL-3,4]**

[c] A hospital is developing a decision tree model to predict whether a patient will be readmitted within 30 days. During model training, the hospital notices that validation accuracy increases steadily until the tree reaches a depth of 6, but beyond this depth, validation accuracy begins to decline. Meanwhile, training accuracy continues to increase all the way up to a depth of 15, indicating potential overfitting. To address this, two strategies are tested: pre-pruning by limiting the maximum depth to 6, which results in a validation accuracy of 89%, and post-pruning, where the tree is first grown fully and then pruned back, ultimately giving a depth of 8 with a higher validation accuracy of 91%. Based on this information, evaluate which method, pre-pruning (early stopping) or post-pruning, provides the better model and justify your conclusion using evidence from the results. **[4] [CO2, 4] [BTL-5]**

3. Answer *any TWO* of the followings

[a] A researcher is evaluating three clusters created from customer data and has computed the following metrics for each: the average intra-cluster distance and the closest inter-cluster distance, shown in the table below:

Table II

Metric	C1	C2	C3
Avg Intra Distance	0.6	1.1	0.4
Closest Inter Distance	1.8	1.3	3.5

Which cluster shows the strongest indication of cluster assignment ambiguity? Explain using the given measures. **[4] [CO4] [BTL-4,5]**

[b] A financial analyst applies PCA to four stock indicators and obtains the variance explained by each principal component, shown in Table III. The analyst wishes to reduce the dataset's dimensionality for modelling purposes. [4] [CO2] [BTL-4]

Table III

Metric	PC1	PC2	PC3	PC4
Variance Explained (%)	58%	30%	9%	3%

Based on these values, should the dataset be reduced to two, three, or all four principal components? Justify your recommendation.

[c] A cybersecurity company deploys a machine learning model to detect malicious login attempts. The model outputs a probability score for each login, and the security team must choose an appropriate classification threshold for alerting suspicious activity. The performance of the model at two candidate thresholds is shown below: [4] [CO2] [BTL-4]

Table IV

Threshold	Precision	Recall
0.40	0.72	0.88
0.75	0.93	0.46

The security team's primary goal is to prevent as many real attacks as possible, but they must also ensure analysts are not overwhelmed with unnecessary alerts. Based on the precision–recall trade-off and the operational constraints, determine which threshold should be deployed.

4. Answer *any TWO* of the followings

[a] A neural network is trained on 120,000 samples using different batch sizes, producing the results shown in Table V.

Table V

Method	Batch Size	Training Acc.	Validation Acc.
Full-Batch GD	120,000	72%	70%
Mini-Batch GD	256	89%	87%
Stochastic GD	1	97%	75%

Using the mini-batch method (batch size 256), different type of regularizations are applied and yields the following validation accuracies: No regularization: 87%; L2 regularization: 90%; Dropout ($p = 0.4$): 92%. Which combination of training method + regularization gives the best model performance? [4] [CO4] [BTL-4]

[b] Consider a neural network activation layer that applies the Leaky ReLU function elementwise, where $f(z_i) = z_i$ for $z_i > 0$ and $f(z_i) = 0.1z_i$ for $z_i \leq 0$. The input to this activation layer is the pre-activation vector $z = [3, -4, 2]^T$, and the upstream gradient arriving from the next layer is $\frac{\partial L}{\partial a} = [5, -1, 4]^T$. Compute the local gradient, expressed as the

Jacobian matrix $J_{\text{LeakyReLU}}(z)$. Then, compute the downstream gradient $\frac{\partial L}{\partial z}$. **[4] [CO4] [BTL-3]**

[c] A company predicts employee "Performance Score" based on two features: Experience (E) and Certification (C). The model's predictions (Value Function v) for various subsets of features are given below: $v(\{\}) = 50$ (*Base average score*); $v(\{E\}) = 70$ $v(\{C\}) = 60$; $v(\{E, C\}) = 90$. Calculate the Shapley Value for the feature "Certification" (ϕ_C). **[4] [CO4] [BTL-3]**

5[a] A delivery robot operates in a building with two non-terminal states, Lobby (S_1) and Corridor (S_2), and a terminal Goal (G) with $V(G) = 0$. From S_1, taking the Safe action A always moves to S_2 with reward +2, while from S_2, taking Safe action A always moves to G with reward +3. There is also a Risky action B in S_2 that moves to G with probability 0.7 and reward +7, or back to S_1 with probability 0.3 and reward −2. The discount factor is $\gamma = 0.9$. Under the policy π that always chooses Safe (A) in both S_1 and S_2, first compute $V^{\pi}(S_1)$and $V^{\pi}(S_2)$, and then compute $Q^{\pi}(S_2, B)$, the expected return if the robot takes Risky (B) once in S_2 and then follows policy π thereafter. **[4] [CO5] [BTL-3]**

[b] A food-delivery app uses reinforcement learning to estimate the value of two non-terminal states, Pickup (P) and Transit (T), before reaching Delivered (D), which is terminal with $V(D) = 0$. Under the fixed policy, the driver receives a reward of +2 when moving from P to T and +6 when moving from T to D. The discount factor is $\gamma = 0.5$and the TD(0) learning rate is $\alpha = 0.5$. After the first episode, the value estimates are $V_1(P) = 1$and $V_1(T) = 3$. In the next episode, the driver follows the sequence P → T → D with rewards +2 and +6. Using TD(0) and updating values after each transition, compute the updated estimates $V_2(P)$and $V_2(T)$. **[4] [CO5] [BTL-3]**

----Best of Luck----