MACHINE LEARNING
(IOT3144)

Time Allotted : 2½ hrs **Full Marks : 60**

Figures out of the right margin indicate full marks.

Candidates are required to answer Group A and any 4 (four) from Group B to E, taking one from each group.

Candidates are required to give answer in their own words as far as practicable.

Group – A

1. Answer any twelve: **12 × 1 = 12**

Choose the correct alternative for the following

(i) Which of the following is a cost function in linear regression?
(a) Cross-entropy loss (b) KL divergence
(c) Hinge loss (d) Mean Squared Error.

(ii) If data is not linearly separable, a common approach is to:
(a) Use nonlinear transformation or kernel methods.
(b) Discard noisy samples.
(c) Increase batch size.
(d) Reduce learning rate.

(iii) A high-variance model usually leads to:
(a) Overfitting. (b) Underfitting.
(c) High bias. (d) Simpler models.

(iv) In machine learning, bias refers to:
(a) Error due to oversensitivity to training data.
(b) Error due to wrong assumptions in the learning algorithm.
(c) Random fluctuations in the dataset.
(d) Noise in the data.

(v) If a model has high precision but low recall, it means:
(a) Many false positives.
(b) Many false negatives.
(c) Few false positives but many false negatives.
(d) Few false negatives but many false positives.

(vi) Maximum likelihood estimation (MLE) finds parameters that:
(a) Minimize variance. (b) Minimize entropy.
(c) Reduce overfitting. (d) Maximize probability of data.

(vii) A very small learning rate in gradient descent may cause:
(a) Slow convergence. (b) Divergence.
(c) No training (d) Overshooting.

(viii) Activation functions in neural networks introduce:
(a) Non-linearity. (b) Linearity. (c) Noise. (d) Bias.

(ix) The "discount factor" (γ) in RL:
(a) Reduces variance. (b) Increases exploration.
(c) Balances immediate and future rewards. (d) Increases exploration.

(x) The solution to the SVM optimization problem is found using:
(a) Gradient Descent. (b) Newton-Raphson.
(c) Lagrange multipliers. (d) Backpropagation.

Fill in the blanks with the correct word

(xi) Logistic regression uses the _____ function to convert linear outputs into probabilities.

(xii) The main operation in CNN is the _____ operation.

(xiii) Logistic regression outputs _____ instead of direct class labels.

(xiv) Kernel methods are particularly useful for handling _____ relationships between features.

(xv) L1 regularization encourages _____ solutions.

Group - B

2. (a) What are the in-sample and out-of-sample errors in the context of classification? *[(CO1)(Remember/LOCQ)]*

(b) Derive the linear regression formula for multiple dependent variables. Also explain how the derived linear regression formula can be used for nonlinear cases. *[(CO2)(Understand/LOCQ)]*

(c) Examine whether the Boolean function XOR is linearly separable or not. *[(CO4)(Analyse/HOCQ)]*

2 + 6 + 4 = 12

3. (a) Consider the following dataset with two independent variables x1 and x2 and a dependent variable y: (1, 2, 3), (2, 0, 2), and (3, 1, 4). The task is to determine the multiple linear regression equation of the form y=w0+w1x1+w2x2 by calculating the regression coefficients. Once the regression equation is obtained, use it to predict the value of y for x1=2 and x2=3. *[(CO2)(Understand/IOCQ)]*

(b) What is the relationship between linear regression and classification models. *[(CO2)(Remember/LOCQ)]*

8 + 4 = 12

Group - C

4. Two classifiers (A and B) produce the following performance metrics on the same dataset: Classifier A: Precision = 0.92, Recall = 0.60

Classifier B: Precision = 0.75, Recall = 0.88

(a) Compute F1-score for each classifier.

(b) For two application scenarios, choose which classifier is preferable and justify with the metrics:

(i) Disease screening where missing a positive (false negative) has high cost (prioritize Recall).

(ii) A spam filter where false positives (legitimate mail flagged) must be minimized (prioritize Precision). *[(CO5)(Apply/IOCQ)]*

(c) Describe the purpose and techniques of regularization used to prevent overfitting in machine learning. *[(CO3)(Understand/LOCQ)]*

(2 + 3 + 3 + 4) = 12

5. (a) Consider the following 2-dimensional dataset:

Point	X	Y
A	1	2
B	2	1
C	1	1
D	8	8
E	9	8
F	8	9
G	5	5
H	6	5

Apply K-Means clustering with k=3 using initial centroids as points A, D, and G. Show cluster assignments and calculate the new centroids. *[(CO2)(Apply/IOCQ)]*

(b) A multiple linear regression model has an R^2 of 0.90 with 8 predictors and 100 observations. Compute the adjusted R^2 and interpret its meaning. *[(CO3)(Understand/IOCQ)]*

8 + 4 = 12

Group - D

6. (a) Describe the Maximum Likelihood Estimation (MLE) approach for estimating the parameters of Logistic Regression. Derive the likelihood function and its log-likelihood form. *[(CO4)(Understand/IOCQ)]*

(b) You observe the following dataset:

x	1	2	3
y	0	1	1

If model predicts P(y=1|x) = [0.2, 0.6, 0.9], compute: (i) Likelihood, (ii) Log-likelihood, (iii) Negative log-likelihood. *[(CO4)(Understand/IOCQ)]*

9 + 3 = 12

7. (a) A company wants to predict whether a customer will buy a product based on their age using logistic regression. The dataset is:

Customer	Age (x)	Buy (y)
1	22	0
2	25	0
3	28	1
4	35	1

Assume initial weights: $w=0.2$, bias $b=-5$, and learning rate $\eta=0.01$.

(i) Compute the probability $P(y=1|x)$ for each customer.

(ii) Calculate the log-likelihood.

(iii) Perform one iteration of gradient descent to update w and b.

[(CO6)(Understand/IOCQ)]

(b) Consider a neural network with:

Input layer: 2 neurons

Hidden layer: 2 neurons with sigmoid activation

Output layer: 1 neuron with sigmoid activation

Weights and biases:

Input → Hidden: $w_{11}=0.1, w_{12}=0.2, w_{21}=0.3, w_{22}=0.4$.

Hidden → Output: $v_1=0.5, v_2=0.6$.

Biases: hidden layer $b_1=b_2=0.1$, output layer $c=0.2$.

Input: $x=[1,0]$ Target: $y=1$.

(i) Compute the hidden layer activations.

(ii) Compute the output of the network.

(iii) Compute the error using squared error loss.

(iv) Perform one step of backpropagation to compute the gradient with respect to output weights v_1 and v_2. *[(CO6)(Apply/IOCQ)]*

6 + 6 = 12

Group - E

8. (a) What is overfitting? Discuss the random and deterministic noises that impact on overfitting. *[(CO6)(Analyse/HOCQ)]*

(b) Write short notes on regularization and validation to deal with the overfitting. *[(CO4)(Remember/LOCQ)]*

6 + 6 = 12

9. (a) Consider a 1-dimensional dataset with three training points:
$x_1=1$, Class Label $y_1=-1$, $x_2=2$, Class Label $y_2=+1$, $x_3=3$, Class Label $y_3=+1$
The goal is to find the optimal separating hyperplane (a point in 1D) that maximizes the margin. The hyperplane equation is $w^Tx+b=0$, which simplifies to $wx+b=0$ in 1D. *[(CO6)(Apply/HOCQ)]*

(b) Given 2D data points:

x_1	x_2	y
1	2	-1
2	3	-1
4	2	1
5	3	1

Find the weight vector w and bias b for the linear SVM hyperplane. *[(CO6)(Apply/HOCQ)]*

6 + 6 = 12

Cognition Level	LOCQ	IOCQ	HOCQ
Percentage distribution	22.92	54.16	22.92