MACHINE LEARNING
(MCA2131)

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) What is the primary objective of linear regression?
- (a) To find a hyperplane that maximizes the margin between data points of different classes
- (b) To estimate the relationship between a dependent variable and one or more independent variables
- (c) To classify data into different categories based on training data
- (d) To reduce the dimensionality of data while retaining most of the variance.

(ii) Which of the following is not a supervised machine learning algorithm?
(a) Naïve Bayes (b) K-means (c) SVM (d) Decision tree.

(iii) In Supervised Learning, the algorithm learns from which of the following?
- (a) Data with no labels
- (b) Data that is organized into groups but has no target outcome
- (c) Data that is labelled and includes the correct answer
- (d) Data that is used only for clustering.

(iv) What is the purpose of the training set in machine learning?
(a) To evaluate model performance
(b) To teach the model patterns and relationships
(c) To validate the model
(d) To tune hyperparameters.

(v) In a confusion matrix, which metric is calculated as TP / (TP + FN)?
(a) Precision (b) Recall (Sensitivity) (c) Specificity (d) F1 Score.

(vi) In a Naive Bayes classifier, what is the role of the prior probability?
(a) To measure the likelihood of a feature given a class
(b) To represent the overall probability of a class
(c) To determine the conditional independence of features
(d) To adjust the feature weights in the model.

(vii) What is the significance of the hyperplane in a linear SVM?
(a) It determines the direction of the data points
(b) It maximizes the distance between the closest points of two classes
(c) It minimizes the distance between all data points
(d) It represents the boundary between the training and testing data.

(viii) What is a primary characteristic of deep learning compared to traditional machine learning algorithms?
(a) Requires less training data
(b) Doesn't use neural networks
(c) Learns hierarchical features from data
(d) Relies on rule-based decision-making.

(ix) What is the vanishing gradient problem in deep learning?
(a) When the loss function has multiple local minima
(b) When gradients become too small during backpropagation, hampering training
(c) When the neural network architecture is too complex to converge
(d) When the learning rate is too high, causing divergence.

(x) What is the primary function of a convolutional layer in a convolutional neural network (CNN)?
(a) Reducing the spatial dimensions of the input data
(b) Extracting features from the input using filters
(c) Connecting each neuron to every neuron in the previous layer
(d) Introducing non-linearity to the model.

Fill in the blanks with the correct word

(xi) If a classification model correctly predicts 80 true positives, 10 false positives, 15 false negatives, and 495 true negatives, then the value of recall will be _____.

(xii) In a neural network, the process of adjusting the weights based on the calculated gradients is known as _____.

(xiii) Application of machine learning methods to large databases is called _____.

(xiv) Naive Bayes is a probabilistic classifier based on applying Bayes' theorem with strong _____ assumptions between the features.

(xv) The _____ is a technique used in deep learning to prevent overfitting by randomly dropping units from the neural network during training, which helps the model generalize better to new data.

Group - B

2. (a) Define Machine Learning in Mathematical Paradigm. *[(CO1)(Understand/LOCQ)]*
(b) Explain overfitting and underfitting in regression problem with suitable diagram. *[(CO2)(Analyse/IOCQ)]*
(c) Derive the slope and intercept of single linear regression model. *[(CO2)(Apply/IOCQ)]*

2 + 2 + 8 = 12

3. (a) Given the following data points and the predicted values from a linear regression model:

Actual values Y: [3, 5, 2, 8]
Predicted values Y': [2.5, 5.5, 2.0, 7.5]
Compute the Mean Squared Error (MSE) for this model based on the given data.
[(CO2)(Apply/IOCQ)]

(b) Given the following data points and the predicted probabilities from a logistic regression model:
Actual class labels Y: [1, 0, 1, 1, 0]
Predicted probabilities p: [0.8, 0.3, 0.7, 0.6, 0.4]
Compute the log-likelihood value for this model based on the given data.
[(CO2)(Apply/IOCQ)]

4 + 8 = 12

Group - C

4. (a) You have a binary classification model that predicts whether an email is spam or not. You are given the following predicted probabilities and the actual class labels:
Actual labels Y: [1, 0, 1, 1, 0, 1, 0, 0, 1, 0]
Predicted probabilities p: [0.9, 0.4, 0.8, 0.7, 0.3, 0.85, 0.2, 0.5, 0.75, 0.3]
Calculate the True Positive Rate (TPR) and False Positive Rate (FPR) for the following thresholds: 0.6, 0.7, 0.8, and 0.9. *[(CO3)(Apply/IOCQ)]*

(b) Plot the approximate ROC curve using the TPR and FPR values computed in the previous section. *[(CO3)(Apply/IOCQ)]*

8 + 4 = 12

5. (a) Explain the process of training and testing in machine learning. How do you evaluate the performance of a model using these phases? Include the significance of splitting data, common evaluation metrics, and the potential issues that can arise during this process. *[(CO3)(Understand/LOCQ)]*

(b) Given a linear regression model $h(X) = w^TX + b$, the regularized cost function is defined as:

$$J(\mathbf{w}) = \frac{1}{2m}\sum_{i=1}^{m}\left(h(\mathbf{X}^{(i)}) - y^{(i)}\right)^2 + \frac{\lambda}{2}\|\mathbf{w}\|^2$$

Derive the gradient of the cost function with respect to the weights w.
[(CO3)(Apply/IOCQ)]

(c) In context to the previous regularization, explain how regularization helps prevent overfitting and how the value of λ. *[(CO3)(Apply/IOCQ)]*

5 + 4 + 3 = 12

Group - D

6. (a) Derive the primal form of the SVM optimization problem for a linear classifier. Given a dataset with linearly separable classes, express the optimization problem and discuss how it can be solved. *[(CO5)(Understand/LOCQ)]*

(b) A diagnostic test for a certain disease is 95% accurate, meaning it correctly identifies the disease 95% of the time when the person has it (true positive) and correctly identifies a healthy person 95% of the time (true negative). The prevalence of the disease in the population is 1%. If a person tests positive,

calculate the probability that they actually have the disease using Bayes' Theorem. *[(CO4)(Apply/IOCQ)]*

6 + 6 = 12

7. (a) Compare and contrast the polynomial kernel and the Radial Basis Function (RBF) kernel in the context of SVMs. Discuss their advantages and disadvantages, and provide scenarios where one might be preferred over the other. *[(CO5)(Understand/LOCQ)]*

   (b) Consider a Bayesian Network with two nodes: A and B. The conditional probability table (CPT) for B given A is provided as follows:
   P(B=1|A=0)=0.2,
   P(B=1|A=1)=0.8
   If P(A=0)=0.3 and P(A=1) = 0.7, calculate the marginal probability P(B=1). *[(CO4)(Apply/IOCQ)]*

6 + 6 = 12

Group - E

8. (a) Implement the Hebb network for OR Logic gate. Use binary/bipolar inputs and targets suitably. *[(CO6)(Apply/IOCQ)]*

   (b) Describe the key components of a Convolutional Neural Network (CNN) and their functions. Specifically, explain the roles of convolutional layers, pooling layers, and fully connected layers in a CNN. *[(CO6)(Remember/LOCQ)]*

6 + 6 = 12

9. (a) Explain the concepts of vanishing gradients and exploding gradients in the context of training deep neural networks. How do these issues affect the training process, and what are some common strategies used to mitigate these problems? Provide examples of techniques or approaches that address these challenges. *[(CO6)(Understand/LOCQ)]*

   (b) Consider a grayscale image with a size of 5×5 pixels and a convolutional kernel of size 3×3. The image and kernel are as follows:
   Image:

$$\begin{pmatrix} 1 & 2 & 3 & 0 & 1 \\ 4 & 5 & 6 & 1 & 2 \\ 7 & 8 & 9 & 2 & 3 \\ 0 & 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 & 5 \end{pmatrix}$$

   Kernel:

$$\begin{pmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{pmatrix}$$

   Perform a convolution operation using the kernel on the top-left 3×3 region of the image with a stride of 1 and no padding. Compute the value of the output feature map at the top-left position. *[(CO6)(Apply/IOCQ)]*

4 + 8 = 12

Cognition Level	LOCQ	IOCQ	HOCQ
Percentage distribution	30.21	69.79	0