# 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) In the context of machine learning applications in e-commerce, what is the primary purpose of a recommendation system?
(a) To detect fraudulent transactions
(b) To forecast inventory levels
(c) To personalize product suggestions for users
(d) To optimize supply chain logistics

(ii) Gradient Descent is mainly used to
(a) Increase bias
(b) Minimize cost function
(c) Calculate VC Dimension
(d) Perform cross-validation

(iii) What is the purpose of regularization in Machine Learning?
(a) To make the model more complex
(b) To add noise to the data
(c) To decrease training time
(d) To prevent overfitting and improve generalization

(iv) In the context of model evaluation, which scenario is indicative of high variance?
(a) The model performs well on the training data but poorly on the test data
(b) The model performs equally well on both the training and test data
(c) The model performs poorly on both the training and test data
(d) The model performs well on the test data but poorly on the training data

(v) Which of the following best describes a "support vector" in the context of an SVM?
(a) Any data point that falls outside the margin
(b) A data point that lies closest to the decision boundary
(c) A data point that belongs to the majority class
(d) Any data point that is correctly classified

(vi) Which of the following is NOT a typical application of Naive Bayes?
(a) Spam filtering (b) Sentiment analysis
(c) Image classification (d) Document classification

(vii) In an LSTM cell, which of the following gates is responsible for deciding which information to keep and which to discard from the cell state?
(a) Input Gate (b) Forget Gate
(c) Output Gate (d) Update Gate

(viii) In RNNs, the vanishing gradient problem is mainly solved using
(a) CNN (b) LSTM
(c) Logistic Regression (d) Dropout

(ix) Naive Bayes' approach assumes
(a) Attributes are dependent (b) Attributes are independent given class
(c) Prior probability is zero (d) Linear decision boundary always exists

(x) Which of the following is the primary purpose of using logistic regression?
(a) To predict continuous outcomes
(b) To classify categorical outcomes
(c) To find the relationship between two continuous variables
(d) To cluster similar data points

*Fill in the blanks with the correct word*

(xi) Convolutional Neural Networks (CNNs) often use _____ layers to reduce the spatial dimensions of the feature maps and capture important features.

(xii) In the gradient descent algorithm, the step size that determines the magnitude of each update is called the _____.

(xiii) In the calculation of F1 score in a classification problem, it is the _____ of precision and recall.

(xiv) _____ is the machine learning algorithm, which can be used with labelled data.

(xv) In Support Vector Machines, the _____ function is used to find the optimal hyperplane that separates different classes in the feature space.

## Group - B

2. (a) Compare and contrast supervised and unsupervised learning. Provide an example of a real-world application for each type of learning. *[(CO1)(Understand/LOCQ)]*

(b) Explain how reinforcement learning differs from both supervised and unsupervised learning. Describe a scenario where reinforcement learning would be an appropriate choice and why. *[(CO1)(Evaluate/HOCQ)]*

**6 + 6 = 12**

3. (a) Discuss the significance of the cost function for the linear regression model and how it is used to optimize the model. Provide an example scenario where a linear regression model would be appropriate and describe the expected outcome.
*[(CO2)(Understand/LOCQ)]*

(b) Find the least square regression line for the following set of data: {(-1, 0), (0, 2), (1, 4), (2, 5)} *[(CO2)(Apply/IOCQ)]*

**6 + 6 = 12**

## Group - C

4. (a) Define overfitting in the context of machine learning. How can overfitting affect the performance of your model on new, unseen data? *[(CO3)(Understand/LOCQ)]*

(b) Outline the bias-variance tradeoff for error estimation of a model with suitable example. Use suitable diagram to explain the same. *[(CO3)(Remember/LOCQ)]*

**4 + 8 = 12**

5. (a) You are evaluating the performance of a binary classification model, and you have calculated the ROC curve for the model, resulting in the following set of (FPR, TPR) points:
(0.1, 0.3), (0.2, 0.5), (0.4, 0.7), (0.6, 0.8), (0.8, 0.9), (1.0, 1.0).
Calculate the Area under the ROC Curve (ROC-AUC) for this model based on these points. Round your answer to two decimal places. *[(CO3)(Apply/IOCQ)]*

(b) Explain the concept of "probably approximately correct" (PAC) learning in the context of machine learning. *[(CO3)(Remember/LOCQ)]*

(c) Given that the goal is to achieve a hypothesis with an error rate of at most $\epsilon$ with probability at least $1-\delta$, how does the number of training examples m relate to $\epsilon$ and according to PAC theory? *[(CO3)(Understand/LOCQ)]*

**6 + 3 + 3 = 12**

## Group - D

6. (a) Consider the following data set:

| Day | Outlook | Temperature | Humidity | Wind | Play_Tenis |
|---|---|---|---|---|---|
| 1 | Sunny | Hot | High | Weak | No |
| 2 | Sunny | Hot | High | Strong | No |
| 3 | Overcast | Hot | High | Weak | Yes |
| 4 | Rain | Mild | High | Weak | Yes |
| 5 | Rain | Cool | Normal | Weak | Yes |
| 6 | Rain | Cool | Normal | Strong | No |
| 7 | Overcast | Cool | Normal | Strong | Yes |
| 8 | Sunny | Mild | High | Weak | No |
| 9 | Sunny | Cool | Normal | Weak | Yes |
| 10 | Rain | Mild | Normal | Weak | Yes |
| 11 | Sunny | Mild | Normal | Strong | Yes |
| 12 | Overcast | Mild | High | Strong | Yes |
| 13 | Overcast | Hot | Normal | Weak | Yes |
| 14 | Rain | Mild | High | Strong | No |

If a new instance comes as {Sunny, Cool, High, Strong}, Classify the class that it belongs to. *[(CO4)(Apply/IOCQ)]*

(b) Explain the concept of the kernel trick in SVMs. Discuss how it transforms the input space and provides a solution to the problem of non-linearly separable data. Illustrate your explanation with an example of a commonly used kernel function. *[(CO4)(Understand/LOCQ)]*

**7 + 5 = 12**

7. (a) Consider a Bayesian Network with three variables: A, B, and C. Suppose the network structure is as follows: A → B → C. Given the conditional probability tables for each variable, how would you compute the joint probability distribution for the variables A, B, and C? What is the significance of the network structure in this computation? *[(CO4)(Understand/LOCQ)]*

(b) Consider the function $f(x,y)=x^2+y^2$. Find the minimum value of this function subject to the constraint $x+y=1$ using the method of Lagrange multipliers. *[(CO5)(Apply/IOCQ)]*

**7 + 5 = 12**

## Group - E

8. Discuss the impact of the three following neural network architectures: feedforward, convolutional, and recurrent neural networks on their suitability for various tasks. How does the choice of architecture influence the model's performance and its ability to handle specific types of data? *[(CO6)(Evaluate/HOCQ)]*

**(4 × 3) = 12**

9. (a) Find the weights using ADALINE network for XOR function. Use binary/bipolar inputs and targets suitably. *[(CO6)(Apply/IOCQ)]*

(b) Explain the role of activation functions in deep learning models. Compare and contrast the sigmoid, ReLU, and tanh activation functions. Discuss one advantage and one disadvantage of each function in the context of training neural networks. *[(CO6)(Remember/LOCQ)]*

**6 + 6 = 12**

---

| Cognition Level | LOCQ | IOCQ | HOCQ |
|---|---|---|---|
| Percentage distribution | 50 | 31 | 19 |