

M.Sc. (A.I. & M.L.) Sem.-2 Examination
Machine Learning

Time : 3.00 Hours]

June-2025

[Max.Marks : 100

Instructions:

- Write both the Sections in the separate answer book.
- Both Sections having equal weightage.
- Draw Diagrams wherever necessary.
- Make Assumptions wherever necessary.

SECTION – I

Q.1 Explain the following terms with example. [16]

- (a) Compare the Max Voting, Averaging, and Weighted Averaging Techniques along with an example.
- (b) How does the Boosting algorithm handle misclassified data points?
- (c) Compare Information Gain and Gini Index as splitting criteria.
- (d) What is the kernel trick, and why is it important in SVM?

Q.2 Attempt the following (Any FOUR) [16]

- (a) How does a decision tree use entropy and information gain to split nodes?
- (b) Consider a dataset where 9 students passed and 5 failed. A feature "Studies Regularly" divides them into:
 - Yes: 6 passed, 1 failed
 - No: 3 passed, 4 failed

Calculate the Information Gain for the feature "Studies Regularly".

- (c) Why is accuracy alone insufficient for evaluating classification models? What alternatives would you use for imbalanced data?
- (d) Describe how Random Forest overcomes the problem of overfitting in Decision Trees.
- (e) Explain the difference between Linear SVM and Non-Linear SVM with proper example.

Q.3 Attempt the following (Any TWO) [18]

- (a) Compare the following Ensemble methods:
 - Bagging
 - Boosting
 - Stacking
 - Blending

Explain the working principle, architecture, and when each should be used. Use diagrams where needed.

- (b) Describe the structure of a decision tree. Also provide the pros and cons of Decision Tree.
- (c) Given the following linearly separable 2D dataset:

Points	Class
(1, 2)	Circle
(2, 3)	Circle
(3, 1)	Circle
(0, 0)	Square
(1, 1)	Square
(2, 0)	Square

Let Support vectors are identified as (1, 2), (2, 0), and (1, 1) then compute:

- a) The weight vector w
- b) The bias term b
- c) The equation of the optimal hyperplane
- d) Identify the label of point (2,2).

SECTION – II

Q.4 **Answer the following question**

[16]

- (a) Define clustering and explain its primary objective.
- (b) Explain the concept of edit distance and list the three allowed operations.
- (c) What is the role of the log-odds (logit) in logistic regression?
- (d) What are the ethical implications of deploying ML models in critical areas like healthcare or criminal justice?

Q.5 **Answer the following question(Any FOUR)**

[16]

- (a) What is the purpose of a confusion matrix? How can it be used to calculate other classification metrics?
- (b) Explain how gradient descent works in optimizing the parameters of a linear regression model. What happens if the learning rate is too high or too low?
- (c) Define the Silhouette Score. How is it used to evaluate clustering performance?
- (d) List and explain the key assumptions of linear regression. What happens if these assumptions are violated?
- (e) Consider the following dataset and find the best fit line.

x	y
2	8
4	14
6	30
8	32

Also Calculate the R^2 and Adjusted R^2 metrics.

Q.6 Attempt the following (Any TWO)

[18]

- (a) You are given the following 2D data points:

Point	X	Y
A	2	2
B	3	2
C	6	5
D	7	6

Perform agglomerative hierarchical clustering using Euclidean distance and the average linkage method to form the cluster and also show the dendrogram.

- (b) Describe the steps involved in the KNN algorithm. What are the advantages and disadvantages of KNN? How does the choice of k affect the KNN model's performance?
- (c) Given the following data points and initial fuzzy membership matrix ($m=2$):
 Data points: $x_1 = (1,1)$, $x_2 = (2,3)$, $x_3 = (5,6)$, $x_4 = (-1,2)$
 Initial centroids: $c_1 = (1.5,2)$, $c_2 = (4,5)$
 Initial memberships:
 $u_{11}=0.7$, $u_{12}=0.3$,
 $u_{21}=0.4$, $u_{22}=0.6$,
 $u_{31}=0.2$, $u_{32}=0.5$
- (a) Compute the updated centroids after one iteration.
- (b) Recalculate the membership degrees for each point using the new centroids.
-