III B. TECH II SEMESTER REGULAR EXAMINATIONS MAY - 2023
MACHINE LEARNING
(INFORMATION TECHNOLOGY)

Time: 3 hours   Max. Marks: 70

Note: Answer **ONE** question from each unit **(5 × 14 = 70 Marks)**

~~~~~~

UNIT-I

1. a) Briefly describe the PAC learning model.   [7M]

    b) Discuss Vapnik-Chervonenkis dimension.   [7M]

                              (OR)

2. a) How do probabilistic models use statistical inference to   [7M]
        make predictions and decisions in machine learning, and
        what are some of the main advantages of this approach?

    b) What are some common techniques used for feature   [7M]
        selection and feature engineering in machine learning, and
        how do these impact the performance of learning
        algorithms?

UNIT-II

3. a) Present the mathematical intuition behind linear   [7M]
        regression.

    b) What is the process of learning a class from examples, and [7M]
        how do supervised learning algorithms use this approach
        to classify data?

                              (OR)

4. Construct a decision tree using CART algorithm for the   [14M]
    following data:

Outlook	Temp.	Humidity	Wind	Decision
Sunny	Hot	High	Weak	No
Sunny	Hot	High	Strong	No
Overcast	Hot	High	Weak	Yes
Rain	Mild	High	Weak	Yes
Rain	Cool	Normal	Weak	Yes
Rain	Cool	Normal	Strong	No
Overcast	Cool	Normal	Strong	Yes
Sunny	Mild	High	Weak	No
Sunny	Cool	Normal	Weak	Yes
Rain	Mild	Normal	Weak	Yes
Sunny	Mild	Normal	Strong	Yes
Overcast	Mild	High	Strong	Yes
Overcast	Hot	Normal	Weak	Yes
Rain	Mild	High	Strong	No

UNIT-III

5. a) What is the naïve Bayes classifier, and how does this simple probabilistic model use Bayes' theorem to estimate the probability of different outcomes based on the features of input data? [7M]

    b) Briefly discuss kernel functions. [7M]

(OR)

6. a) Explain support vector machine. [7M]

    b) Describe K-Nearest neighbors. [7M]

UNIT-IV

7. a) How do agglomerative and divisive hierarchical clustering algorithms differ, and what are some of the key considerations when choosing between these two approaches? [7M]

    b) How does clustering around medoids work, and what are some of the key considerations when using this approach to group data points into clusters based on their similarity? [7M]

(OR)

8. a) Can you explain the concept of hierarchical clustering, and how this approach uses a tree-like structure to group data points into nested clusters based on their similarity? [7M]

    b) What is distance-based clustering, and how do algorithms use distance metrics to group data points together based on their proximity to each other? [7M]

UNIT-V

9. a) What is ensemble learning, and how do machine learning algorithms use this approach to improve the accuracy and robustness of their predictions by combining the outputs of multiple models? [7M]

    b) What are the different types of ensemble learning model combination schemes, and how do voting methods work to aggregate the predictions of multiple models? [7M]

(OR)

10. a) What is bagging, and how does the random forest algorithm use this approach to improve the accuracy and generalization performance of decision tree models? [7M]

    b) Describe the Boosting and Adaboost. [7M]

* * * * *