R20

QP Code: R20A31AI01

Anurag University

B Tech III Year I-Semester Regular & Supplementary Examinations, Oct/Nov-2023

Course: Essentials of Machine Learning

(Common to AI & AIML)

Time: 3Hours **Max Marks: 60**

Section –A (Short answer type questions)

- **Answer all questions: (10 x 2 =20Marks)**

1. Differentiate between Supervised and Unsupervised learning.
2. Distinguish between overfitting and underfitting. How it can affect model generalization?
3. Compare Classification with regression with an example.
4. What types of problems are best suited for decision tree learning?
5. Compare K means clustering with Hierarchical Clustering Techniques.
6. List out the disadvantages of clustering schemes.
7. Why do we need ensemble methods in machine learning?
8. What are Bayesian Belief nets? Where are they used?
9. How does a multilayer perceptron solve the XOR problem?
10. What is the purpose of the activation function in a neural network? Give some activation functions.

Section—B (Essay answer type questions)

- **Answer all questions: (5 x 8 =40 Marks)**

11. A) i) Define reinforcement learning and provide an example of a real-world application where reinforcement learning is commonly used.
    ii) Discuss the concept of bias-variance trade-off in machine learning. How does it affect model performance?

    OR

    B) i) Describe the concept of dimensionality reduction and its significance in feature transformation.
    ii) Explain situations where feature transformation might be preferred over feature extraction and vice versa.

12. A) Sales data of 10 months for a coffee house situated near a prime location of a city comprising the number of customers (in hundreds) and monthly sales (in Thousand Rupees) are given below

S. No.	No. of Customers (in hundreds)	Monthly Sales (in thousand Rs.)
1	6	1
2	6.1	6
3	6.2	8
4	6.3	10
5	6.5	11
6	7.1	20
7	7.6	21
8	7.8	22
9	8	23
10	8.1	25

Find the simple linear regression equation that fits the given data.

OR

B) Suppose 10000 patients get tested for flu; out of them, 9000 are actually healthy and 1000 are actually sick. For the sick people, a test was positive for 620 and negative for 380. For the healthy people, the same test was positive for 180 and negative for 8820. Construct a confusion matrix for the data and compute the precision and recall for the data.

13. A) Explain about Hierarchical clustering algorithm with a suitable example

OR

B) Use K Means clustering to cluster the following data into two groups. Assume cluster centroid are m1=2 and m2=4. The distance function used is Euclidean distance. {2, 4, 10, 12, 3, 20, 30, 11, 25}.

14. A) How does the random forest model work? How is it different from bagging and boosting in ensemble models?

OR

B) Difference between Bagging and Boosting, write the implementation steps for Bagging.

15. A) What is Artificial Neural Network? Explain architecture of Artificial neural network. How does Artificial neural network works, and how it differs to Biological neural network.

OR

B) Explain in detail about multilayer neural networks and back propagation algorithm.