Printed Pages – 5 Roll No. .....

C109512(022)

B. Tech. (Fifth Semester) Examination

Nov.-Dec. 2023

CSE(AIML)

INTRODUCTION to MACHINE LEARNING

Time Allowed : Three hours

Maximum Marks : 100

Minimum Pass Marks : 35

Note : *Attempt all questions. Part (a) from each question is compulsory and answers any* **two** *of the remaining (b), (c) and (d).*

Unit-I

1. (a) Distinguish between Supervised and Unsupervised Machine Learning methods. 4

(b) Explain Machine Learning Life Cycle. Discuss various applications of Machine Learning in details. 8

(c) Discussed various methods of Supervised and Unsupervised Machine Learning with example. 8

(d) Discuss data preprocessing and data wrangling process of Machine Learning. Explain significance of matching data to an appropriate algorithm. 8

Unit-II

2. (a) Discuss interpretation of Linear Regression coefficients. 4

(b) Find linear regression euqation for the following two sets of data : 8

X	2	4	6	8
Y	3	7	5	10

(c) Explain validation of Simple Regression model, coefficients of determination (R-squared), estimation of parameters using ordinary least square. 8

(d) Discuss Hypothesis test for regression coefficients (t-test), Residual analysis in details. 8

Unit-III

3. (a) Discuss ordinary least square estimation for simple linear regression in brief. 4

(b) Suppose we have the following dataset with one response variable Y and two predictors variables X1 and X2. Compute the multiple regression coefficient for the dataset given below : 8

Y	X1	X2
140	60	22
155	62	25
159	67	24
179	70	20
192	71	15
200	72	14
212	75	14
215	78	11

(c) Explain Multiple Linear Regression in detail. Discuss validation of Multiple Linear Regression model. 8

(d) Discuss statistical significance of individual variables in Multiple Linear Regression (t-test). Explain interpretation of Multiple Linear Regression coefficients. 8

Unit-IV

4. (a) Explain Bagging and Boosting of Ensemble method in brief. 4

(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. 8

(c) Explain Naive Bayes Classifier in details with example. Discuss the advantages, disadavantages and applications of Naive Bayes Classifier. 8

(d) Explain K-Nearest Nighbor Algorithm in detail with example. Discuss its advantages, disadvantages and applications. 8

Unit-V

5. (a) Compare K-Means clustering and Fuzzy C-Means clustering. 4

(b) Illustrate K-Means clustering algorithm with an example. Also explain advantages disadvantages, applications of K-Means clustering. 8

(c) Explain DBSCAN algorithm for density-based clustering. List out its advantages compared to K-Means clustering. 8

(d) Discuss Agglomerative and Divisive Hierarchical clustering methods. Explain the significance of distance measure in clustering analysis. 8