Register Number:……..

KUMARAGURU
college of technology
character is life

B.E DEGREE EXAMINATIONS: APRIL / MAY 2023

(Regulation 2018)

Sixth Semester

INFORMATION SCIENCE AND ENGINEERING

U18ISI6204**:** Machine Learning Techniques

COURSE OUTCOMES

CO1: Understanding basic concepts and techniques of Machine Learning

CO2: Analyze the regression and classification techniques

CO3: Be able to create solutions to real world problems using Machine Learning

Time: Three Hours **Maximum Marks: 100**

Answer all the Questions:-

PART A (10 x 2 = 20 Marks)

(Answer not more than 40 words)

1. Outline the importance of Machine Learning. CO1 [K_2]
2. State F-Scores and its purpose. CO1 [K_1]
3. Name any two machine learning algorithms that use gradient descent. CO1 [K_1]
4. Calculate MSE and MAE for the following data CO1 [K_3]

Actual Value (y)	Predicted Value (y hat)
100	130
150	170
200	220
250	260
300	325

5. Distinguish between Classification and Regression algorithms. CO2 [K_2]
6. Recite the basic idea behind the Support Vector Machine. CO2 [K_2]
7. Outline the purpose of decision boundaries. CO2 [K_2]
8. Name the commonly used types of regularized regression methods. CO2 [K_2]
9. Why clustering is called unsupervised learning? CO2 [K_2]
10. List the benefits of dimensionality reduction. CO2 [K_2]

Answer any FIVE Questions:-
PART B (5 x 16 = 80 Marks)
(Answer not more than 400 words)

11. a) Tabulate the similarities and dissimilarities between supervised and unsupervised learning. Also, identify the applications of it. 8 CO1 [K_L]

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 CO3 [K_3]

12. a) Obtain a linear regression for the following data assuming that *y* is the independent variable 8 CO2 [K_3]

x	1.0	2.0	3.0	4.0	5.0
y	1.00	2.00	1.30	3.75	2.25

b) Illustrate non-linear regression and relate it with respect to Linear regression. 8 CO2 [K_2]

13. a) Consider the dataset for prediction of student needs tutor or not. 12 CO3 [K_3]

SNo	Regular	Pass	Needs Tutor
1	Yes	Yes	Yes
2	Yes	Yes	No
3	Yes	No	Yes
4	Yes	No	No
5	No	Yes	Yes
6	No	Yes	No
7	No	No	Yes
8	No	No	No
9	Yes	Yes	No
10	No	Yes	No
11	No	No	No
12	No	No	No

Greedily learn a decision tree using the ID3 algorithm for the above dataset and draw the tree.

b) Write the advantages and disadvantages of Decision tree algorithm 4 CO3 [K_2]

14. a) The following table contains data from the questionnaires survey (to ask people opinion) and objective testing with two attributes (acid durability and strength) to classify whether a special paper tissue is good or not. Here is four training samples. 10 CO3 [K3]

X1 = Acid Durability (seconds)	X2 = Strength (kg/square meter)	Y = Classification
7	7	Bad
7	4	Bad
3	4	Good
1	4	Good

Now the factory produces a new paper tissue that pass laboratory test with X1 = 3 and X2 = 7. Without another expensive survey, can we guess what the classification of this new tissue is using kNN algorithm?

b) Discuss about the evaluation metrics in classification. 6 CO3 [K2]

15. a) Exemplify the One-Vs-All machine learning approach for solving a multiclass classification problem. 8 CO2 [K1]

b) Elaborate how a regularization method is used to solve the problem of overfitting. 8 CO2 [K1]

16. a) Provide a step-by-step logical explanation of how to use Principal Component Analysis (PCA) Algorithm. 8 CO1 [K2]

b) Consider the following eight points (with (x, y) representing locations) for grouping into three clusters: A1(2, 10), A2(2, 5), A3(8, 4), A4(5, 8), A5(7, 5), A6(6, 4), A7(1, 2) and A8(4, 9). Assume, initial cluster centers are: A1(2, 10), A4(5, 8) and A7(1, 2). Use K-Means Algorithm to find the three cluster centers after the first iteration. 8 CO3 [K3]

****