Register Number:……..

KUMARAGURU college of technology
character is life

B.E/B.TECH DEGREE EXAMINATIONS: APRIL /MAY 2024

(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. Differentiate between Classification and Regression. CO1 [K_1]

2. Construct a confusion matrix with the datset given below. CO1 [K_3]

ID	Target	Prediction
1	False	False
2	False	False
3	False	True
4	False	False
5	True	True
6	False	False
7	True	True
8	True	True
9	False	True
10	False	False
11	False	False
12	True	True
13	False	False
14	True	True
15	False	False
16	False	False
17	True	False
18	True	True
19	True	True
20	True	True

Find Accuracy and Misclassification rate.

3.	State the role of learning rate in convergence without overshooting		CO1	[K_1]
4.	Define cost function.		CO2	[K_1]
5.	What is Information Gain in decision tree construction		CO2	[K_1]
6.	List the problems of over fitting in machine learning models		CO2	[K_1]
7.	Recall Multiclass Classification.		CO3	[K_1]
8.	Explain the role of parameters in hypothesis representation.		CO3	[K_2]
9.	Define data compression in the context of dimensionality reduction.		CO3	[K_1]
10.	List the benefits of dimensionality reduction in Machine Learning		CO3	[K_1]

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

11.	a)	Identify the categories of Machine Learning and illustrate each with a real-time application	8	CO1	[K_1]
	b)	Distinguish between over fitting and under fitting. How it can affect model Generalization?	8	CO1	[K_1]
12.	a)	Formulate gradient descent algorithm with suitable illustration.	8	CO1	[K_3]
	b)	Consider the example below where the mass, y(grams),of a chemical is related to time x(Seconds),for which the chemical reaction has been taking place according to the table:	8	CO1	[K_3]

Time, x(seconds)	5	7	12	16	20
Mass, y(grams)	40	120	180	210	240

a) Find the equation of the regression line

b) By how much does the chemical increase in weight in five seconds?

13.	a)	Explain the K-Nearest Neighbor (KNN) algorithm in the context of classification. Provide a step-by-step overview of how KNN works and discuss its key parameters.	8	CO2	[K_2]
	b)	Explain the working of ID3 decision tree construction algorithm	8	CO2	[K_2]
14.	a)	Formulate the logistic regression training process with suitable illustration	8	CO2	[K_3]
	b)	State the primary objective of PCA in reducing the dimensionality of data, and how is it mathematically formulated? Give an outline of Principal Component	8	CO2	[K_1]

Analysis (PCA) algorithm, from data preprocessing to computing principal components?

15.	a)	Illustrate the kernel trick in SVM and its significance in handling non-linearly separable data.	8	CO3	[K_2]
	b)	Explain how regularization techniques such as L1 and L2 regularization are applied to mitigate over fitting in linear regression and logistic regression models.	8	CO3	[K_2]
16.	a)	Use k-means algorithm to cluster the following eight points with (x , y) representing locations (Manhattan Distance) into 3 clusters A1(2,10),A2(2,5),A3(8,4),A4(5,8),A5(7,5),A6(6,4),A7(1,2),A8(4,9) Initialize A1, A4 and A7 as cluster centroids.	8	CO3	[K_3]
	b)	Discuss Elbow method for determining the optimal number of clusters in K-Means clustering.	8	CO3	[K_2]

****