Register Number:……..

KUMARAGURU
college of technology
character is life

B.E DEGREE EXAMINATIONS: NOV/DEC 2023

(Regulation 2018)

Sixth Semester

INFORMATION SCIENCE AND ENGINEERING

U18ISI6204T**:** 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.	List few applications of machine learning	CO1	[K_2]
2.	Define regression and state few evaluation metrics for it	CO1	[K_2]
3.	"Convergence depends on the learning rate " - Justify	CO2	[K_4]
4.	Determine the pearson co-efficient value given SS regression as 12 and SS total as 20	CO3	[K_4]
5.	How to avoid overfitting in decision tree ?	CO2	[K_2]
6.	State the advantages of KNN	CO2	[K_2]
7.	Compare Linear and Logistic Regression	CO2	[K_4]
8.	Write the formula for regularized linear regression	CO2	[K_2]
9.	What are the benefits of dimensionality reduction technique?	CO2	[K_2]
10.	Differentiate unsupervised and supervised techniques.	CO1	[K_2]

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

11.	a) i)	State few performance measures used in machine learning.	2	CO1	[K_3]
	ii)	Evaluate the accuracy, Precision, Recall, F1 Measure and construct the confusion matrix for a dataset with 45000 images where 5 images of cat and 5 images of dog are considered for testing and given TP=2, FP=0, TN=5, FN=3.	4	CO1	[K_3]
	iii)	How much data will be considered for training and testing for cross validation testing?	4	CO1	[K_3]

	b)	Write the step by step process involved in Machine Learning.	6	CO1	[K_2]
12.	a)	A survey was conducted among a group of people where the number of people and network usage is calculated based on router count. Apply regression model which determines the best slope equation for network usage. Draw the graph for the same and determine the SSE value.	10	CO3	[K_3]

S. No	1	2	3	4	5	6	7	8
Router Count	4	6	6	7	8	7	8	10
No. of Persons	2	2	4	4	5	5	6	6

	b)	How Convergence is achieved in linear regression? Discuss about gradient descent with appropriate equations and diagrams.	6	CO2	[K_2]
13.	a)	A Student is planning to buy a computer. Based on his age, bank balance consistency, income and credit rate conclude whether the student buys or not. Construct a decision tree for this problem using cross entropy measure.	10	CO2	[K_3]

Category of Age	Income	Student	Credit rate	Buys Computer
Youth	high	no	fair	no
youth	high	no	excellent	no
middle_aged	high	no	fair	yes
senior	medium	no	fair	yes
senior	low	yes	fair	yes
senior	low	yes	excellent	no
middle_aged	low	yes	excellent	yes
youth	medium	no	fair	no
youth	low	yes	fair	yes
senior	medium	yes	fair	yes
youth	medium	yes	excellent	yes
middle_aged	medium	no	excellent	yes
middle_aged	high	yes	fair	yes
senior	medium	no	excellent	no

	b) i)	The Margin in SVM is to be maximum. Justify.	2	CO2	[K_4]
	ii)	Discuss how a hyperplane is constructed in SVM	4	CO3	[K_2]
14.	a)	What are the methods to overcome overfitting problem? How regularization is used in overfitting problem with cost function?	10	CO2	[K_2]
	b)	Discuss how high bias and variance is used in regularization with regression and classification problems.	6	CO2	[K_2]
15.	a)	A set of dimension values of ears and eyes of animal species is specified and it is	16	CO3	[K_4]

to be grouped according to the category. Use K-means clustering algorithm for the same and categorise the following into three clusters and proceed for 2 iterations.

S.No.	X Coordinate	Y Coordinate
1	2	10
2	2	5
3	8	4
4	5	8
5	7	5
6	6	4
7	1	2
8	4	9

Initial cluster centers are: Point1(2, 10), Point4(5, 8) and Point7(1, 2).

16. a) Illustrate about linear regression with single and multi-variables 8 CO2 [K_2]

b) Define PCA. How to reduce n features to k features using Principal Component Analysis? Write the algorithm for the same. 8 CO2 [K_2]