K. J. Somaiya Institute of Technology, Sion, Mumbai-22
(Autonomous College Affiliated to University of Mumbai)

May- 2024

Program: **B.Tech** Scheme: **II**
Supplr Examination: **TY** Semester: **VI**
Course Code: **EXC602** Course Name: **Machine Learning**
Date of Exam: 29/05/24 Duration: **02.5 Hours** Max. Marks: **60**

Instructions:
(1) All questions are compulsory.
(2) Draw neat diagrams wherever applicable.
(3)Assume suitable data, if necessary.

		M	CO	BT
Q 1	**Solve any six questions out of eight:**	**12**		
i)	Define Machine Learning.	2	1	U
ii)	Write significance of learning rate(α) in training of linear regression.	2	2	U
iii)	Define Kernel in SVM.	2	3	U
iv)	What is overfitting of a machine learning model?	2	4	U
v)	What is clustering ?	2	5	U
vi)	Differentiate between Anomaly Detection and Supervised Learning	2	6	A
vii)	What is the importance of scaling in machine learning.	2	2	U
viii)	What are advantages of DB Scan over other clustering algorithms?	2	5	A
Q.2	**Solve any four questions out of six.**	**16**		
i)	Differentiate between supervised learning and unsupervised learning.	4	1	A
ii)	Write expression for hypothesis, cost function and parameter using gradient descent, for logistic regression. Also, write meaning of all symbols used in the equations.	4	2	U
iii)	Write a short note on online learning.	4	6	U
iv)	Why the dataset need to be split into training set, cross validation and test set. Write the ratio of division of the dataset.	4	4	U
v)	A1(2,4), A2(3,1), A3(0,1), A4(1,2) apply PCA to the given dataset.	4	5	A
vi)	The quality of a paper tissue can be determined by its acid durability and strength. Based on the following dataset predict quality a paper tissue of 4 sec acid durability and 8 kg/mtr strength, using KNN. Consider K =3. Use Euclidean distance.	4	3	A

Acid durability (sec)	Strength (Kg/mtr)	Quality
5	7.5	Good
4	3	Bad
8	7	Good
2.5	9	Good
3	6	Bad
2.5	7.5	Bad
6	9	Good

Q.3 Solve any two questions out of three. **16**

i) What are different types of machine learning problems? Write and explain at-least two applications of each type. 8 1 U

ii) An online preowned car selling website uses machine learning for predicting the resale value based on its age and purchased price (in lakh). The dataset used training the model is as follow: 8 2 A

Age	Purchase Price	Resale Price
3.5	12	8.4
3	8.5	4.6
6	5.6	1.4
2.5	18	14.5
4	7.8	3.8

After training the parameters are $\theta_0 = 2.54$, $\theta_1 = -0.4$, and $\theta_2 = 0.54$. Find out the cost. Predict resale price of 3.5 years old car with purchase price of 8 lakh.
Also find out the new values of parameters after an iteration. Assume $\alpha = 0.01$.

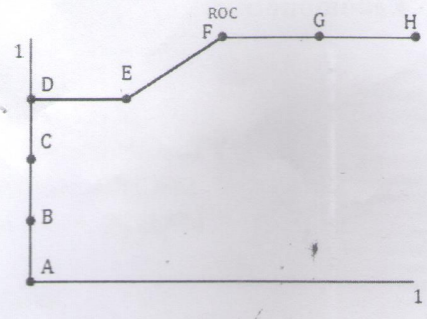
iii) Explain a recommender system in details. 8 6 U

Q.4 Solve any two questions out of three. **16**

i) Use K means clustering algorithm to form two clusters for given data based on Euclidean distance calculation. Take first 2 data points as centroids. Do maximum 3 iterations. 8 5 A

X1	185	170	168	179	182	188	180	180	183	180	180	177
X2	72	56	60	68	72	77	71	70	84	88	67	76

ii) Define ROC. Which of the following point on ROC gives the best threshold for the application that predicts if the patient can be discharged after Covid treatment? Why? (Fig. on the next page) 8 4 A



Explain the significance of ROC-AUC with an example

iii) Construct a decision tree using following dataset 8 3 A

Weekend	Weather	Parents	Money	Decision
W1	Sunny	Yes	Rich	Cinema
W2	Sunny	No	Rich	Tennis
W3	Windy	Yes	Rich	Cinema
W4	Rainy	Yes	Poor	Cinema
W5	Rainy	No	Rich	Stay-in
W6	Rainy	Yes	Poor	Cinema
W7	Windy	No	Poor	Cinema
W8	Windy	No	Rich	Shopping
W9	Windy	Yes	Rich	Cinema
W10	Sunny	No	Rich	Tennis

****