EXIC - Machine Learning

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

April – May 2024
(B.Tech / M.Tech.) Program: B.Tech(Scheme: **II**)

Examination: **TY**   Semester: **VI**
Course Code: **EC602**   Course Name: **Machine Learning**   Max. Marks: **60**
Date of Exam: **17/05/24**   Duration: **2.5 Hours**

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

					CO	BT
				12		
Q 1	**Solve any six questions out of eight:**					
i)	What are the differences between classification and Regression?				2	1	R
ii)	Write the cost function for Linear and Logistic Regression.				2	2	U
iii)	Explain terms hypothesis and hypothesis space?				2	3	U
iv)	What is meant by overfitting, under fitting?				2	4	R
	Define Entropy. What is its importance in creating Decision Tree?				2	5	U
vi)	List the limitations of using a linear regression model for classification problems.				2	6	U
vii)	How is K decided in K-means clustering?				2	3	A
viii)	For univariate dataset given, calculate mean and standard deviation. S = [5, 10, 15, 20, 25, 30]				2	5	U

		CO	BT
		16	
Q.2	**Solve any four questions out of six.**		
i)	How Logistic Regression Model is different from Linear Regression?	4	1	U
ii)	What is feature scaling? Explain with appropriate example. Write significance of feature scaling in machine learning.	4	2	U
iii)	Following table gives height, weight and T-shirt size of some customers.	4	3	A

Height	158	158	158	160	160	163	163	160	163
Weight	58	59	63	59	60	60	61	64	64
T Shirt	M	M	M	M	M	M	M	L	L

Height	165	165	165	168	168	168	170	170	170
Weight	61	62	65	62	63	66	63	64	68
T Shirt	L	L	L	L	L	L	L	L	L

Predict the T-shirt size of a new customer with height 161 cm and weight 61kg with k = 5. Use Euclidean Distance metric. (The dataset split across tables)

		CO	BT
iv)	Find the eigenvalues and eigenvectors of given matrix. $A = \begin{bmatrix} 7 & 3 \\ 3 & -1 \end{bmatrix}$	4	4	A
v)	Apply agglomerative clustering to build the dendogram, for following given dataset. Use Euclidean distance Metric to find the distance. A(1, 1), B(2, 3), C(3, 5), D(4, 5), E(6, 6), F(7, 5)	4	5	
vi)	Write short note on SVM.			

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

i) What are the types of Machine Learning? Explain each type in brief with appropriate examples. 8 1 U

ii) Consider the student performance training data set of 8 instances shown in which describes the performance of individual students in a course and their CGPA obtained in the previous semesters. The Independent attributes are CGPA, Assessment and Project. The target variable is 'Result' which is a discrete valued variable that takes two values 'Pass' or 'Fail'. Based on the performance of a student, classify whether a student will pass or fail in that course. Find what will be the result of a student with CGPA 8.7, Assessment 78, Project submitted 7. 8 2 A

Sr. No.	CGPA	Assessment	Project Submitted	Result
1.	9.2	85	8	Pass
2.	8	80	7	Pass
3.	8.5	81	8	Pass
4.	6	45	5	Fail
5.	6.5	50	4	Fail
6.	8.2	72	7	Pass
7.	5.8	38	5	Fail
8.	8.9	91	9	Fail

iii) Construct a decision tree for following dataset using information gain. 8 3 A

Outlook	Temperature	Wind	Play Tennis
Sunny	Hot	Weak	No
Sunny	Hot	Strong	No
Overcast	Hot	Weak	Yes
Rain	Mild	Weak	Yes
Rain	Cool	Weak	Yes
Rain	Cool	Strong	No
Overcast	Cool	Strong	Yes
Sunny	Mild	Weak	No
Sunny	Cool	Weak	Yes
Rain	Mild	Weak	Yes
Sunny	Mild	Strong	Yes
Overcast	Mild	Strong	Yes
Overcast	Hot	Weak	Yes
Rain	Mild	Strong	No

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

i) Write a note on DBSCAN clustering algorithm. Discuss the parameters required. Differentiate between DBSCAN and k means clustering algorithm. 8 4 U

ii) Write short note on Principal Component Analysis. 8 5 U

iii) Write short note on (a) Recommender system (b) Anomaly detection 8 6 A

****