

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

April – May 2023

(B.Tech / M.Tech.) Program: B.Tech(Scheme: II)

Examination: **TY**

Semester: **VI**

Course Code: **EXC602**

Course Name: **Machine Learning**

Date of Exam: **15/05/23**

Duration: **2.5 Hours**

Max. Marks: **60**

Instructions:

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

			CO	BT																											
Q 1	Solve any six questions out of eight:	12																													
i)	How is machine learning different from conventional programming?	2	1	U																											
ii)	Write expression for hypothesis, cost function and for parameter using gradient descent for univariate linear regression.	2	2	U																											
iii)	Define following terms related to SVM : a) Soft margin b) Kernel	2	3	R																											
iv)	Define overfitting and underfitting of a ML model.	2	4	U																											
v)	What are the types of unsupervised clustering?	2	5	R																											
vi)	Differentiate between anomaly detection and supervised learning.	2	6	U																											
vii)	Write expression for sigmoid function. Explain utility of Sigmoid function in the logistic regression.	2	3	U																											
viii)	What are the advantages of dimensionality reduction?	2	5	U																											
Q.2	Solve any four questions out of six.	16																													
i)	Differentiate between Supervised and Unsupervised Learning.	4	1	U																											
ii)	What is feature scaling? Explain with appropriate example. Write significance of feature scaling in machine learning.	4	2	U																											
iii)	<table border="1" style="width: 100%; border-collapse: collapse; margin-bottom: 5px;"> <tr> <td style="padding: 2px;"><i>Height</i></td> <td style="padding: 2px;">160</td> <td style="padding: 2px;">160</td> <td style="padding: 2px;">163</td> <td style="padding: 2px;">163</td> <td style="padding: 2px;">160</td> <td style="padding: 2px;">163</td> <td style="padding: 2px;">165</td> <td style="padding: 2px;">165</td> </tr> <tr> <td style="padding: 2px;"><i>Weight</i></td> <td style="padding: 2px;">59</td> <td style="padding: 2px;">60</td> <td style="padding: 2px;">60</td> <td style="padding: 2px;">61</td> <td style="padding: 2px;">63</td> <td style="padding: 2px;">64</td> <td style="padding: 2px;">61</td> <td style="padding: 2px;">62</td> </tr> <tr> <td style="padding: 2px;"><i>T Shirt</i></td> <td style="padding: 2px;">M</td> <td style="padding: 2px;">M</td> <td style="padding: 2px;">M</td> <td style="padding: 2px;">M</td> <td style="padding: 2px;">L</td> <td style="padding: 2px;">L</td> <td style="padding: 2px;">L</td> <td style="padding: 2px;">L</td> </tr> </table>	<i>Height</i>	160	160	163	163	160	163	165	165	<i>Weight</i>	59	60	60	61	63	64	61	62	<i>T Shirt</i>	M	M	M	M	L	L	L	L	4	3	A
<i>Height</i>	160	160	163	163	160	163	165	165																							
<i>Weight</i>	59	60	60	61	63	64	61	62																							
<i>T Shirt</i>	M	M	M	M	L	L	L	L																							
	Predict T Shirt size for a person with height 164 cm and weight 65 Kg using KNN for K=3.																														
iv)	Explain ROC and ROC-AUC in brief.	4	4	U																											

v)

	x_1	x_2
P1	12	25
P2	23	14
P3	32	24
P4	14	23
P5	15	21

4 5 A

The above dataset represents the features of the new species of plants. Group them together using divisive clustering. Use Manhattan distance.

vi) Explain anomaly detection in brief.

4 6 U

Q.3 Solve any two questions out of three.

16

i) Write applications of machine learning in different domain. Elaborate with at least four examples, how machine learning is useful in these applications.

8 1 U

ii) Following dataset represents the height (cm) of the soccer player and the average goal scored in EPL. If $\theta_0 = 65.15$ and $\theta_1 = 0.38$, find the cost with these parameters. Predict the average goals scored by a player with height 179 cm.

8 2 A

Height	180	188	178	191	175	185	183
Goals	25	29	21	20	37	17	36

Perform mean normalization on the features.

iii) The below dataset represents data related to the purchase of cars. Construct a decision tree using information gain.

8 3 A

Buying Price	Maintenance Price	Safety	Evaluation
High	High	High	Unacceptable
High	High	Low	Unacceptable
Medium	High	High	Acceptable
Low	Medium	High	Acceptable
Low	Low	High	Acceptable
Low	Low	Low	Unacceptable
Medium	Low	Low	Acceptable
High	Medium	High	Unacceptable
High	Low	High	Acceptable
Low	Medium	High	Acceptable
High	Medium	Low	Acceptable
Medium	Medium	Low	Acceptable
Medium	High	High	Acceptable
Low	Medium	Low	Unacceptable

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

- i) A ML model has been deployed to predict whether or not a pre-owned car will be sold based on the features. The model could predict the 436 sold and 368 unsold cars correctly, while it predicted 96 sold and 90 unsold cars incorrectly. Hence, 8 4 An
- a) Draw a confusion matrix.
 - b) Define and calculate accuracy, precision and recall.

ii) 8 5 A

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

Use K means clustering algorithm to form two clusters for given data based on Euclidean distance calculation. Take first 2 data points as centroids.

- iii) Write short note on 8 6 U
- a) Online learning
 - b) Recommender system
