

K. J. Somaiya Institute of Engineering and Information Technology, Sion, Mumbai-22

(Autonomous College Affiliated to University of Mumbai)

End Semester Exam

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

Examination: TY Semester: VI

Course Code: 01UEXC602 and Course Name: Machine Learning

Duration: 03 Hours

Max. Marks: 60

Instructions:

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

Nov/Dec 2022, 14/12/22

		Max Marks	CO	BT level
Q 1	Solve any six questions out of eight:	12		
i)	Write applications of machine learning in different domain.	2	CO1	R
ii)	Write expression for hypothesis, cost function and for parameter using gradient descent for univariate linear regression. Explain each term in short.	2	CO2	U
iii)	Identify the type of the following machine learning problem, a) Predicting the survival of a passenger in the Titanic disaster b) Recognizing handwritten digit c) Forecasting sales for next 6 months for D-Mart d) Suggesting songs on Spotify	2	CO3	U
iv)	Define ROC.	2	CO4	U
v)	Which of the following is the correct split of the dataset of size 1000000? Justify. a) Training dataset = 700000, Cross Validation = 150000, Test dataset = 150000 b) Training dataset = 333333, Cross Validation = 333333, Test dataset = 333333 c) Training dataset = 900000, Cross Validation = 50000, Test dataset = 50000	2	CO5	E
vi)	What is the significance of regularization parameter (λ)?	2	CO6	U
vii)	What are different types of clustering in unsupervised learning?	2	CO6	U
viii)	What is online machine learning?	2	CO1	R
Q.2	Solve any four questions out of six.	16		
1	Explain anomaly detection with an example.	4	CO1	U

2	Illustrate process of learning with the gradient descent for a univariate linear regression, using a bell shaped error curve. Explain how a step size is modulated on every iteration.	4	CO2	U																																																																											
3	Differentiate between Supervised and Unsupervised Learning.	4	CO3	U																																																																											
4	A machine learning model predicted 930 non-sick and 30 sick patients correctly. Also predicted 25 non-sick patients to be sick and 15 sick patient to be non-sick. Draw a confusion matrix for the machine learning model, considering sick patients to be belonged to positive class. Calculate the precision and recall for the same.	4	CO4	U																																																																											
5	<table border="1" data-bbox="220 526 1118 599"> <tr> <td>X</td> <td>4.51</td> <td>3.58</td> <td>4.31</td> <td>5.06</td> <td>5.64</td> <td>4.99</td> <td>5.29</td> <td>5.83</td> <td>4.7</td> <td>5.61</td> <td>4.9</td> </tr> <tr> <td>Y</td> <td>2.48</td> <td>2.26</td> <td>2.47</td> <td>2.77</td> <td>2.99</td> <td>3.05</td> <td>3.18</td> <td>3.46</td> <td>0.03</td> <td>3.26</td> <td>2.67</td> </tr> </table> <p>In the above dataset x represents the production in Cr Rupees and y represents the energy consumption in lakh kWh. Find the squared error function if parameters for linear regressions are $\theta_0 = 0.4$ and $\theta_1 = 0.49$</p>	X	4.51	3.58	4.31	5.06	5.64	4.99	5.29	5.83	4.7	5.61	4.9	Y	2.48	2.26	2.47	2.77	2.99	3.05	3.18	3.46	0.03	3.26	2.67	4	CO5	A																																																			
X	4.51	3.58	4.31	5.06	5.64	4.99	5.29	5.83	4.7	5.61	4.9																																																																				
Y	2.48	2.26	2.47	2.77	2.99	3.05	3.18	3.46	0.03	3.26	2.67																																																																				
6	What are the different types of machine learning problems? Explain each in brief.	4	CO6	U																																																																											
Q.3	Solve any two questions out of three.	16																																																																													
i)	<table border="1" data-bbox="220 940 951 1050"> <tr> <td>Height</td> <td>158</td> <td>158</td> <td>158</td> <td>160</td> <td>160</td> <td>163</td> <td>163</td> <td>160</td> <td>163</td> </tr> <tr> <td>Weight</td> <td>58</td> <td>59</td> <td>63</td> <td>59</td> <td>60</td> <td>60</td> <td>61</td> <td>64</td> <td>64</td> </tr> <tr> <td>Size</td> <td>M</td> <td>M</td> <td>M</td> <td>M</td> <td>M</td> <td>M</td> <td>M</td> <td>L</td> <td>L</td> </tr> </table> <table border="1" data-bbox="220 1090 951 1201"> <tr> <td>Height</td> <td>165</td> <td>165</td> <td>165</td> <td>168</td> <td>168</td> <td>168</td> <td>170</td> <td>170</td> <td>170</td> </tr> <tr> <td>Weight</td> <td>61</td> <td>62</td> <td>65</td> <td>62</td> <td>63</td> <td>66</td> <td>63</td> <td>64</td> <td>68</td> </tr> <tr> <td>Size</td> <td>L</td> <td>L</td> <td>L</td> <td>L</td> <td>L</td> <td>L</td> <td>L</td> <td>L</td> <td>L</td> </tr> </table> <p>The above dataset (both table carry the entire dataset together) shows the height and weight of a person and the T-Shirt size required for the person. Using KNN algorithm find the T-Shirt size for a person with height 161 cm and weight 67 Kg. Consider K=5 for prediction.</p>	Height	158	158	158	160	160	163	163	160	163	Weight	58	59	63	59	60	60	61	64	64	Size	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	Size	L	L	L	L	L	L	L	L	L	8	CO5	A															
Height	158	158	158	160	160	163	163	160	163																																																																						
Weight	58	59	63	59	60	60	61	64	64																																																																						
Size	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																																																																						
Size	L	L	L	L	L	L	L	L	L																																																																						
ii)	<table border="1" data-bbox="220 1382 1107 1884"> <thead> <tr> <th>Outlook</th> <th>Temperature</th> <th>Humidity</th> <th>Wind</th> <th>Played</th> </tr> </thead> <tbody> <tr><td>Sunny</td><td>Hot</td><td>High</td><td>Weak</td><td>No</td></tr> <tr><td>Sunny</td><td>Hot</td><td>High</td><td>Strong</td><td>No</td></tr> <tr><td>Overcast</td><td>Hot</td><td>High</td><td>Weak</td><td>Yes</td></tr> <tr><td>Rain</td><td>Mild</td><td>High</td><td>Weak</td><td>Yes</td></tr> <tr><td>Rain</td><td>Cool</td><td>Normal</td><td>Weak</td><td>Yes</td></tr> <tr><td>Rain</td><td>Cool</td><td>Normal</td><td>Strong</td><td>No</td></tr> <tr><td>Overcast</td><td>Cool</td><td>Normal</td><td>Strong</td><td>Yes</td></tr> <tr><td>Sunny</td><td>Mild</td><td>High</td><td>Weak</td><td>No</td></tr> <tr><td>Sunny</td><td>Cool</td><td>Normal</td><td>Weak</td><td>Yes</td></tr> <tr><td>Rain</td><td>Mild</td><td>Normal</td><td>Weak</td><td>Yes</td></tr> <tr><td>Sunny</td><td>Mild</td><td>Normal</td><td>Strong</td><td>Yes</td></tr> <tr><td>Overcast</td><td>Mild</td><td>High</td><td>Strong</td><td>Yes</td></tr> <tr><td>Overcast</td><td>Hot</td><td>Normal</td><td>Weak</td><td>Yes</td></tr> <tr><td>Rain</td><td>Mild</td><td>High</td><td>Strong</td><td>No</td></tr> </tbody> </table> <p>The above dataset contains whether or not a football match is played based on the weather conditions. Construct a decision tree for the dataset.</p>	Outlook	Temperature	Humidity	Wind	Played	Sunny	Hot	High	Weak	No	Sunny	Hot	High	Strong	No	Overcast	Hot	High	Weak	Yes	Rain	Mild	High	Weak	Yes	Rain	Cool	Normal	Weak	Yes	Rain	Cool	Normal	Strong	No	Overcast	Cool	Normal	Strong	Yes	Sunny	Mild	High	Weak	No	Sunny	Cool	Normal	Weak	Yes	Rain	Mild	Normal	Weak	Yes	Sunny	Mild	Normal	Strong	Yes	Overcast	Mild	High	Strong	Yes	Overcast	Hot	Normal	Weak	Yes	Rain	Mild	High	Strong	No	8	CO2	A
Outlook	Temperature	Humidity	Wind	Played																																																																											
Sunny	Hot	High	Weak	No																																																																											
Sunny	Hot	High	Strong	No																																																																											
Overcast	Hot	High	Weak	Yes																																																																											
Rain	Mild	High	Weak	Yes																																																																											
Rain	Cool	Normal	Weak	Yes																																																																											
Rain	Cool	Normal	Strong	No																																																																											
Overcast	Cool	Normal	Strong	Yes																																																																											
Sunny	Mild	High	Weak	No																																																																											
Sunny	Cool	Normal	Weak	Yes																																																																											
Rain	Mild	Normal	Weak	Yes																																																																											
Sunny	Mild	Normal	Strong	Yes																																																																											
Overcast	Mild	High	Strong	Yes																																																																											
Overcast	Hot	Normal	Weak	Yes																																																																											
Rain	Mild	High	Strong	No																																																																											

iii)	<p>Explain following concepts with an example,</p> <p>a) Margin b) Maximal Margin Classifier c) Soft Margin Classifier d) Hyperplane e) Kernel</p>	8	CO1	U																		
Q.4	Solve any two questions out of three.	16																				
i)	<table border="1" data-bbox="210 437 513 645"> <thead> <tr> <th></th> <th>X_1</th> <th>X_2</th> </tr> </thead> <tbody> <tr> <td>P1</td> <td>0.07</td> <td>0.83</td> </tr> <tr> <td>P2</td> <td>0.85</td> <td>0.14</td> </tr> <tr> <td>P3</td> <td>0.66</td> <td>0.89</td> </tr> <tr> <td>P4</td> <td>0.49</td> <td>0.64</td> </tr> <tr> <td>P5</td> <td>0.8</td> <td>0.46</td> </tr> </tbody> </table> <p>The above dataset contains the five examples with two features. Calculate the Euclidean distance between the points. Form a distance matrix. Perform Agglomerative Clustering on it.</p>		X_1	X_2	P1	0.07	0.83	P2	0.85	0.14	P3	0.66	0.89	P4	0.49	0.64	P5	0.8	0.46	8	CO4	A
	X_1	X_2																				
P1	0.07	0.83																				
P2	0.85	0.14																				
P3	0.66	0.89																				
P4	0.49	0.64																				
P5	0.8	0.46																				
ii)	<p>Explain how large dataset is useful in machine learning? What is the problem with large dataset in gradient descent? How large datasets are dealt with in gradient descent.</p>	8	CO1	U																		
iii)	<p>What are hyperparameters? What are different methods of hyperparameter tuning? Explain any two hyperparameters in brief.</p>	8	CO6	U																		