

Vivekanand Education Society's Institute of Technology, Chembur, Mumbai,
Department of Computer Engineering,
Year:2023-24 (ODD Sem)
MID TERM TEST

Class : BE	Division: A/B/C
Semester: VII	Subject: Machine Learning
Date: 04/09/2023	Time: 10.30am-11.30am

Course Outcome	CO1	CO2	CO3	CO6
Percentage %	34	39	13	14

Q.1)		(Attempt any five of the following)	Marks (20)	CO's Mapped												
	a)	What is the effect on the training error and generalization error if the training data is increased? Justify the same.	2M	CO1												
	b)	For a crop yield prediction system, explain the steps required for building a good machine learning algorithm.	2M	CO1, CO6												
	c)	A company with 20 million customers has analyzed 10% of all transactions to classify them as fraudulent or not, while the rest of the data wasn't labeled with "fraud" or "non-fraud" labels: Which type of learning in ML will be suitable without sacrificing accuracy? Justify the answer.	2M	CO1, CO6												
	d)	Apply Kernel trick approach of SVM to any example of your choice and justify	2M	CO3												
	e)	Differentiate between soft and hard margins in the SVM classifier with a suitable illustration.	2M	CO3												
	f)	Elaborate on the bias-variance trade-off.	2M	CO1												
Q.2)	a)	<p>A dietician wants to analyze the relationship between weight and height for a few people. Below is the dataset for the same.</p> <p>a) Plot the regression line</p> <p>b) Determine the weight of a person if his height measures 144 cm.</p> <table><tr><td>Height (cm)</td><td>137</td><td>140</td><td>142</td><td>145</td><td>147</td></tr><tr><td>Weight (kg)</td><td>30</td><td>33</td><td>35</td><td>39</td><td>43</td></tr></table>	Height (cm)	137	140	142	145	147	Weight (kg)	30	33	35	39	43	5M	CO2, CO6
Height (cm)	137	140	142	145	147											
Weight (kg)	30	33	35	39	43											
		OR														

		Determine the root node for the following dataset using the ID3 algorithm.		CO2																																																																																										
	b)	<table><tr><th>Sr.No</th><th>Income</th><th>Defaulting</th><th>Credit Score</th><th>Location</th><th>Give Loan?</th></tr><tr><td>1</td><td>Low</td><td>High</td><td>High</td><td>Bad</td><td>No</td></tr><tr><td>2</td><td>Low</td><td>High</td><td>High</td><td>Good</td><td>No</td></tr><tr><td>3</td><td>High</td><td>High</td><td>High</td><td>Bad</td><td>Yes</td></tr><tr><td>4</td><td>Medium</td><td>Medium</td><td>High</td><td>Bad</td><td>Yes</td></tr><tr><td>5</td><td>Medium</td><td>Low</td><td>Low</td><td>Bad</td><td>No</td></tr><tr><td>6</td><td>Medium</td><td>Low</td><td>Low</td><td>Good</td><td>Yes</td></tr><tr><td>7</td><td>High</td><td>Low</td><td>Low</td><td>Good</td><td>Yes</td></tr><tr><td>8</td><td>Low</td><td>Medium</td><td>High</td><td>Bad</td><td>No</td></tr><tr><td>9</td><td>Low</td><td>Low</td><td>Low</td><td>Bad</td><td>No</td></tr><tr><td>10</td><td>Medium</td><td>Medium</td><td>Low</td><td>Bad</td><td>No</td></tr><tr><td>11</td><td>low</td><td>Medium</td><td>Low</td><td>Good</td><td>Yes</td></tr><tr><td>12</td><td>High</td><td>Medium</td><td>High</td><td>Good</td><td>Yes</td></tr><tr><td>13</td><td>High</td><td>High</td><td>Low</td><td>Bad</td><td>No</td></tr><tr><td>14</td><td>Medium</td><td>Medium</td><td>High</td><td>Good</td><td>Yes</td></tr></table>	Sr.No	Income	Defaulting	Credit Score	Location	Give Loan?	1	Low	High	High	Bad	No	2	Low	High	High	Good	No	3	High	High	High	Bad	Yes	4	Medium	Medium	High	Bad	Yes	5	Medium	Low	Low	Bad	No	6	Medium	Low	Low	Good	Yes	7	High	Low	Low	Good	Yes	8	Low	Medium	High	Bad	No	9	Low	Low	Low	Bad	No	10	Medium	Medium	Low	Bad	No	11	low	Medium	Low	Good	Yes	12	High	Medium	High	Good	Yes	13	High	High	Low	Bad	No	14	Medium	Medium	High	Good	Yes	5M	
Sr.No	Income	Defaulting	Credit Score	Location	Give Loan?																																																																																									
1	Low	High	High	Bad	No																																																																																									
2	Low	High	High	Good	No																																																																																									
3	High	High	High	Bad	Yes																																																																																									
4	Medium	Medium	High	Bad	Yes																																																																																									
5	Medium	Low	Low	Bad	No																																																																																									
6	Medium	Low	Low	Good	Yes																																																																																									
7	High	Low	Low	Good	Yes																																																																																									
8	Low	Medium	High	Bad	No																																																																																									
9	Low	Low	Low	Bad	No																																																																																									
10	Medium	Medium	Low	Bad	No																																																																																									
11	low	Medium	Low	Good	Yes																																																																																									
12	High	Medium	High	Good	Yes																																																																																									
13	High	High	Low	Bad	No																																																																																									
14	Medium	Medium	High	Good	Yes																																																																																									
Q.3)	a)	<p>i. State the splitting criteria used in the CART algorithm. ii. Determine the root node using CART</p> <table><tr><th>Name</th><th>Hair</th><th>Height</th><th>Weight</th><th>Location</th><th>Class</th></tr><tr><td>Sunita</td><td>Blonde</td><td>Average</td><td>Light</td><td>No</td><td>Yes</td></tr><tr><td>Anita</td><td>Blonde</td><td>Tall</td><td>Average</td><td>Yes</td><td>No</td></tr><tr><td>Kavita</td><td>Brown</td><td>Short</td><td>Average</td><td>Yes</td><td>No</td></tr><tr><td>Sushma</td><td>Blonde</td><td>Short</td><td>Average</td><td>No</td><td>Yes</td></tr><tr><td>Xavier</td><td>Red</td><td>Average</td><td>Heavy</td><td>No</td><td>Yes</td></tr><tr><td>Balaji</td><td>Blonde</td><td>Tall</td><td>Heavy</td><td>No</td><td>No</td></tr><tr><td>Ramesh</td><td>Blonde</td><td>Average</td><td>Heavy</td><td>No</td><td>No</td></tr><tr><td>Swetha</td><td>Blonde</td><td>Short</td><td>Light</td><td>Yes</td><td>No</td></tr></table>	Name	Hair	Height	Weight	Location	Class	Sunita	Blonde	Average	Light	No	Yes	Anita	Blonde	Tall	Average	Yes	No	Kavita	Brown	Short	Average	Yes	No	Sushma	Blonde	Short	Average	No	Yes	Xavier	Red	Average	Heavy	No	Yes	Balaji	Blonde	Tall	Heavy	No	No	Ramesh	Blonde	Average	Heavy	No	No	Swetha	Blonde	Short	Light	Yes	No	5M	CO2																																				
Name	Hair	Height	Weight	Location	Class																																																																																									
Sunita	Blonde	Average	Light	No	Yes																																																																																									
Anita	Blonde	Tall	Average	Yes	No																																																																																									
Kavita	Brown	Short	Average	Yes	No																																																																																									
Sushma	Blonde	Short	Average	No	Yes																																																																																									
Xavier	Red	Average	Heavy	No	Yes																																																																																									
Balaji	Blonde	Tall	Heavy	No	No																																																																																									
Ramesh	Blonde	Average	Heavy	No	No																																																																																									
Swetha	Blonde	Short	Light	Yes	No																																																																																									
		OR																																																																																												
	b)	<p>Create a confusion matrix for the following dataset. Calculate accuracy, sensitivity, specificity, and the F1-score.</p> <table><tr><th>Patient ID</th><th>1</th><th>2</th><th>3</th><th>4</th><th>5</th><th>6</th><th>7</th><th>8</th></tr><tr><td>Disease detected (Actual)</td><td>Y</td><td>N</td><td>Y</td><td>N</td><td>N</td><td>N</td><td>Y</td><td>Y</td></tr><tr><td>Disease detected (Predicted)</td><td>Y</td><td>Y</td><td>N</td><td>N</td><td>N</td><td>Y</td><td>Y</td><td>N</td></tr></table>	Patient ID	1	2	3	4	5	6	7	8	Disease detected (Actual)	Y	N	Y	N	N	N	Y	Y	Disease detected (Predicted)	Y	Y	N	N	N	Y	Y	N	5M	CO1																																																															
Patient ID	1	2	3	4	5	6	7	8																																																																																						
Disease detected (Actual)	Y	N	Y	N	N	N	Y	Y																																																																																						
Disease detected (Predicted)	Y	Y	N	N	N	Y	Y	N																																																																																						