

M.Sc.(AI)I/02.22.004 Reg. No.

--	--	--	--	--	--	--	--



COCHIN UNIVERSITY OF SCIENCE AND TECHNOLOGY
M.Sc. COMPUTER SCIENCE WITH SPECIALISATION IN ARTIFICIAL INTELLIGENCE
FIRST SEMESTER EXAMINATION, FEBRUARY 2022
21-344-0104 DATA SCIENCE AND MACHINE LEARNING
(Regular)

Time: 3 Hrs.

Maximum Marks: 50

Answer any five questions

Each question carries 10 Marks

QUESTIONS										MARKS																			
1	a)	Calculate Karl Pearson Coefficient of correlation from the following data								6																			
	<table><tr><td>X</td><td>9</td><td>8</td><td>7</td><td>6</td><td>5</td><td>4</td><td>3</td><td>2</td><td>1</td></tr><tr><td>Y</td><td>15</td><td>16</td><td>14</td><td>13</td><td>11</td><td>12</td><td>10</td><td>8</td><td>9</td></tr></table>									X	9	8	7	6	5	4	3	2	1	Y	15	16	14	13	11	12	10	8	9
X	9	8	7	6	5	4	3	2	1																				
Y	15	16	14	13	11	12	10	8	9																				
	b)	Write short notes on one tail and two tail tests?								4																			
II	Suppose that the data for analysis includes the attribute <i>age</i> . The <i>age</i> values for the data are (in increasing order) 13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52, 70								4																				
	a)	Use smoothing by bin means, smoothing by bin median, smoothing by bin boundary to smooth the data, using a bin depth of 3?																											
	b)	Describe about different attribute subset selection methods?								6																			
III	a)	What is cross-validation in machine learning? Explain the different types of cross validations.								6																			
	b)	Suppose 10000 patients get tested for flu; out of them, 9000 are actually healthy and 1000 are actually sick. For the sick people, a test was positive for 620 and negative for 380. For the healthy people, the same test was positive for 180 and negative for 8820. Construct a confusion matrix for the data and compute the precision and recall for the data.								4																			

IV	<p>Use ID3 algorithm to find the root node of the decision tree for the data in the following table.</p> <table><tr><th>Day</th><th>outlook</th><th>temperature</th><th>humidity</th><th>wind</th><th>playtennis</th></tr><tr><td>D1</td><td>sunny</td><td>hot</td><td>high</td><td>weak</td><td>no</td></tr><tr><td>D2</td><td>sunny</td><td>hot</td><td>high</td><td>strong</td><td>no</td></tr><tr><td>D3</td><td>overcast</td><td>hot</td><td>high</td><td>weak</td><td>yes</td></tr><tr><td>D4</td><td>rain</td><td>mild</td><td>high</td><td>weak</td><td>yes</td></tr><tr><td>D5</td><td>rain</td><td>cool</td><td>normal</td><td>weak</td><td>yes</td></tr><tr><td>D6</td><td>rain</td><td>cool</td><td>normal</td><td>strong</td><td>no</td></tr><tr><td>D7</td><td>overcast</td><td>cool</td><td>normal</td><td>strong</td><td>yes</td></tr><tr><td>D8</td><td>sunny</td><td>mild</td><td>high</td><td>weak</td><td>no</td></tr><tr><td>D9</td><td>sunny</td><td>cool</td><td>normal</td><td>weak</td><td>yes</td></tr><tr><td>D10</td><td>rain</td><td>mild</td><td>normal</td><td>weak</td><td>yes</td></tr><tr><td>D11</td><td>sunny</td><td>mild</td><td>normal</td><td>strong</td><td>yes</td></tr><tr><td>D12</td><td>overcast</td><td>mild</td><td>high</td><td>strong</td><td>yes</td></tr><tr><td>D13</td><td>overcast</td><td>hot</td><td>normal</td><td>weak</td><td>yes</td></tr><tr><td>D14</td><td>rain</td><td>mild</td><td>high</td><td>strong</td><td>no</td></tr></table>	Day	outlook	temperature	humidity	wind	playtennis	D1	sunny	hot	high	weak	no	D2	sunny	hot	high	strong	no	D3	overcast	hot	high	weak	yes	D4	rain	mild	high	weak	yes	D5	rain	cool	normal	weak	yes	D6	rain	cool	normal	strong	no	D7	overcast	cool	normal	strong	yes	D8	sunny	mild	high	weak	no	D9	sunny	cool	normal	weak	yes	D10	rain	mild	normal	weak	yes	D11	sunny	mild	normal	strong	yes	D12	overcast	mild	high	strong	yes	D13	overcast	hot	normal	weak	yes	D14	rain	mild	high	strong	no	10
Day	outlook	temperature	humidity	wind	playtennis																																																																																							
D1	sunny	hot	high	weak	no																																																																																							
D2	sunny	hot	high	strong	no																																																																																							
D3	overcast	hot	high	weak	yes																																																																																							
D4	rain	mild	high	weak	yes																																																																																							
D5	rain	cool	normal	weak	yes																																																																																							
D6	rain	cool	normal	strong	no																																																																																							
D7	overcast	cool	normal	strong	yes																																																																																							
D8	sunny	mild	high	weak	no																																																																																							
D9	sunny	cool	normal	weak	yes																																																																																							
D10	rain	mild	normal	weak	yes																																																																																							
D11	sunny	mild	normal	strong	yes																																																																																							
D12	overcast	mild	high	strong	yes																																																																																							
D13	overcast	hot	normal	weak	yes																																																																																							
D14	rain	mild	high	strong	no																																																																																							
V	<p>a) Explain any two model combination scheme to improve the accuracy of a classifier?</p> <p>b) Write down Apriori algorithm for frequent itemset generation</p>	<p>5</p> <p>5</p>																																																																																										
VI	<p>a) Applying the k- medoid clustering algorithm, find two clusters in the following data.</p> <table><tr><td>X</td><td>2</td><td>3</td><td>3</td><td>4</td><td>6</td><td>6</td><td>7</td><td>7</td><td>8</td><td>7</td></tr><tr><td>Y</td><td>6</td><td>4</td><td>8</td><td>7</td><td>2</td><td>4</td><td>3</td><td>4</td><td>5</td><td>6</td></tr></table> <p>b) What are the different problems occurs when clustering is applied on real world data.</p>	X	2	3	3	4	6	6	7	7	8	7	Y	6	4	8	7	2	4	3	4	5	6	<p>7</p> <p>3</p>																																																																				
X	2	3	3	4	6	6	7	7	8	7																																																																																		
Y	6	4	8	7	2	4	3	4	5	6																																																																																		
VII	<p>a) Given the following distance matrix, construct the dendrogram using agglomerative clustering with complete linkage</p> <table><tr><td></td><td>A</td><td>B</td><td>C</td><td>D</td><td>E</td></tr><tr><td>A</td><td>0</td><td>1</td><td>2</td><td>2</td><td>3</td></tr><tr><td>B</td><td>1</td><td>0</td><td>2</td><td>4</td><td>3</td></tr><tr><td>C</td><td>2</td><td>2</td><td>0</td><td>1</td><td>5</td></tr><tr><td>D</td><td>2</td><td>4</td><td>1</td><td>0</td><td>3</td></tr><tr><td>E</td><td>3</td><td>3</td><td>5</td><td>3</td><td>0</td></tr></table> <p>b) List out the advantages and disadvantages of agglomerative clustering</p>		A	B	C	D	E	A	0	1	2	2	3	B	1	0	2	4	3	C	2	2	0	1	5	D	2	4	1	0	3	E	3	3	5	3	0	<p>7</p> <p>3</p>																																																						
	A	B	C	D	E																																																																																							
A	0	1	2	2	3																																																																																							
B	1	0	2	4	3																																																																																							
C	2	2	0	1	5																																																																																							
D	2	4	1	0	3																																																																																							
E	3	3	5	3	0																																																																																							