



Daffodil International University

Faculty of Science & Information Technology

Department of Computer Science & Engineering

Final Semester Examination, Spring 2025

Course Code: CSE325, Course Title: Data Mining and Machine Learning

Level: 3 Term: 2 Batch: 62

Time: 2:00 Hrs

Marks: 40

Answer ALL Questions

[The figures in the right margin indicate the full marks and corresponding course outcomes. All portions of each question must be answered sequentially.]

1.	The following dataset contains shopping transactions from an online fashion store. Your task is to apply the FP-Growth algorithm to extract all frequent itemsets that satisfy a minimum support threshold of 50%. <table><tr><th>Transaction ID</th><th>Items Purchased</th></tr><tr><td>T1</td><td>T-Shirt, Sneakers, Jeans, Cap, Sunglasses</td></tr><tr><td>T2</td><td>Jacket, Sneakers, T-Shirt, Jeans</td></tr><tr><td>T3</td><td>Sunglasses, Watch, Sneakers, Jeans</td></tr><tr><td>T4</td><td>T-Shirt, Sneakers, Belt, Cap</td></tr><tr><td>T5</td><td>Jeans, Jacket, Watch, Sneakers, T-Shirt</td></tr><tr><td>T6</td><td>T-Shirt, Sneakers, Watch, Sunglasses</td></tr><tr><td>T7</td><td>Cap, Sneakers, Jeans, T-Shirt, Jacket</td></tr></table>	Transaction ID	Items Purchased	T1	T-Shirt, Sneakers, Jeans, Cap, Sunglasses	T2	Jacket, Sneakers, T-Shirt, Jeans	T3	Sunglasses, Watch, Sneakers, Jeans	T4	T-Shirt, Sneakers, Belt, Cap	T5	Jeans, Jacket, Watch, Sneakers, T-Shirt	T6	T-Shirt, Sneakers, Watch, Sunglasses	T7	Cap, Sneakers, Jeans, T-Shirt, Jacket	CO1
Transaction ID	Items Purchased																	
T1	T-Shirt, Sneakers, Jeans, Cap, Sunglasses																	
T2	Jacket, Sneakers, T-Shirt, Jeans																	
T3	Sunglasses, Watch, Sneakers, Jeans																	
T4	T-Shirt, Sneakers, Belt, Cap																	
T5	Jeans, Jacket, Watch, Sneakers, T-Shirt																	
T6	T-Shirt, Sneakers, Watch, Sunglasses																	
T7	Cap, Sneakers, Jeans, T-Shirt, Jacket																	
	a) Construct the FP-Tree based on item frequency and item ordering. [5] b) Explain how conditional pattern bases and conditional FP-Trees are formed. [3] c) Discuss the efficiency of FP-Growth over Apriori, especially for dense datasets. [2]																	
2.	You are given a set of 15 two-dimensional points representing spatial coordinates of tracked devices in a smart campus. Apply the DBSCAN algorithm using the following parameters: minPts = 4, eps (ϵ) = 2.0. $P_1(1,6), P_2(2,5), P_3(3,7), P_4(4,6), P_5(5,5), P_6(5,3), P_7(4,2), P_8(6,2), P_9(2,2), P_{10}(1,3), P_{11}(3,3), P_{12}(6,5), P_{13}(7,4), P_{14}(8,5), P_{15}(7,6)$	CO1																
	a) Identify all core points and border points based on the parameters provided. [2.5] b) Use the DBSCAN algorithm to form clusters – i) List all clusters and ii) Indicate which points are noise (if any) [2.5]																	
3.	You are given the following dataset consisting of 10 samples with 3 features:	CO1																

		<table><tr><th>Sample</th><th>Feature A</th><th>Feature B</th><th>Feature C</th></tr><tr><td>A</td><td>2</td><td>0</td><td>1</td></tr><tr><td>B</td><td>0</td><td>1</td><td>4</td></tr><tr><td>C</td><td>4</td><td>3</td><td>2</td></tr><tr><td>D</td><td>3</td><td>5</td><td>3</td></tr><tr><td>E</td><td>1</td><td>2</td><td>2</td></tr><tr><td>F</td><td>5</td><td>1</td><td>0</td></tr><tr><td>G</td><td>2</td><td>4</td><td>1</td></tr><tr><td>H</td><td>3</td><td>2</td><td>4</td></tr><tr><td>I</td><td>1</td><td>0</td><td>3</td></tr><tr><td>J</td><td>0</td><td>3</td><td>2</td></tr></table>	Sample	Feature A	Feature B	Feature C	A	2	0	1	B	0	1	4	C	4	3	2	D	3	5	3	E	1	2	2	F	5	1	0	G	2	4	1	H	3	2	4	I	1	0	3	J	0	3	2		
Sample	Feature A	Feature B	Feature C																																													
A	2	0	1																																													
B	0	1	4																																													
C	4	3	2																																													
D	3	5	3																																													
E	1	2	2																																													
F	5	1	0																																													
G	2	4	1																																													
H	3	2	4																																													
I	1	0	3																																													
J	0	3	2																																													
a)	Standardize the dataset so that each feature has zero mean and unit variance.			[2.5]																																												
b)	Compute the covariance matrix of the standardized dataset.			[2.5]																																												
c)	Find the eigenvalues and eigenvectors of the covariance matrix.			[2.5]																																												
d)	Identify the principal components in descending order of variance explained.			[2.5]																																												
4.	An artificial neural network has been trained on a binary classification dataset where the independent attributes are Feature A and Feature B. The target classes are encoded as Class A = 0 and Class B = 1 before training. In this network: (i) All hidden nodes use the ReLU activation function, and (ii) The output node uses the Sigmoid activation function The trained network is shown below. Classify the data samples in Table 1 using this network. Table 1: Test Data <table><tr><th>Sample</th><th>Feature A</th><th>Feature B</th></tr><tr><td>S1</td><td>1.2</td><td>2.5</td></tr><tr><td>S2</td><td>2.0</td><td>1.0</td></tr></table> Neural Network Architecture - Input Layer: Feature A, Feature B, and Bias - Hidden Layer: 3 neurons (H1, H2, H3) - Output Layer: 1 neuron (O) Input to Hidden Weights: <table><tr><th>From → To</th><th>H1</th><th>H2</th><th>H3</th></tr><tr><td>Feature A</td><td>-1.5</td><td>1.0</td><td>2.1</td></tr><tr><td>Feature B</td><td>2.3</td><td>-0.7</td><td>1.8</td></tr><tr><td>Bias</td><td>-1.0</td><td>-1.2</td><td>-0.5</td></tr></table> Hidden to Output Weights: <table><tr><th>From → To</th><th>Output</th></tr><tr><td>H1</td><td>2.0</td></tr><tr><td>H2</td><td>-1.5</td></tr><tr><td>H3</td><td>1.3</td></tr><tr><td>Bias</td><td>-2.0</td></tr></table>			Sample	Feature A	Feature B	S1	1.2	2.5	S2	2.0	1.0	From → To	H1	H2	H3	Feature A	-1.5	1.0	2.1	Feature B	2.3	-0.7	1.8	Bias	-1.0	-1.2	-0.5	From → To	Output	H1	2.0	H2	-1.5	H3	1.3	Bias	-2.0		CO2								
Sample	Feature A	Feature B																																														
S1	1.2	2.5																																														
S2	2.0	1.0																																														
From → To	H1	H2	H3																																													
Feature A	-1.5	1.0	2.1																																													
Feature B	2.3	-0.7	1.8																																													
Bias	-1.0	-1.2	-0.5																																													
From → To	Output																																															
H1	2.0																																															
H2	-1.5																																															
H3	1.3																																															
Bias	-2.0																																															
a)	Compute the input to each hidden neuron using the given weights and bias.			[2.5]																																												
b)	Apply the ReLU activation to hidden neurons.			[2.5]																																												
c)	Compute the input to the output neuron and apply the Sigmoid activation.			[5]																																												
d)	Classify the result as Class A (0) or Class B (1) based on threshold 0.5.			[5]																																												