



Department of Computer Science and Engineering
Jashore University of Science and Technology

Semester Final Examination

B.Sc. (Engg.) 4th Year 1st Semester, Session: 2018-19

Course Code: CSE-4109

Course Title: Data Warehouse and Mining

Time: 3.00 hours

Marks: 72

[N.B. Answer any 6 set questions from following 8 set questions. The figures in the margin indicate full marks.]

1. (a) Calculate the Euclidean distance among the following two sentences using term-frequency vector. 6
- Sentence1: I am a citizen of Bangladesh
 - Sentence2: I am not a citizen of Bangladesh by birthplace
- (b) Given the physics marks of pupils in a school class are following: 3*2
- 34, 11, 50, 25, 43, 27, 31, 33, 10, 34, 35, 36, 25, 12, 59
- Calculate the five number summary, draw the boxplot and sort out the outlier(s) [if any].

2. Given in the following binary featured dataset of marks of 4 students which they obtained in two class tests. The third column is the target class column which states that whether the students later passed or failed in the final exam: 12

CT#01	CT#02	FinalExam
4	11	Pass
8	4	Fail
13	5	Fail
7	4	Pass

Use Principal Component Analysis (PCA) to reduce the dimension from 2 to 1

3. (a) Consider the following relational schema of an international airport. 3*3
- FLIGHT (IDFlight, Company, DepAirport, ArrAirport, DepTime, ArrTime)
- FLYING (IDFlight, FlightDate)
- AIRPORT (IDAirport, AirName, City, State)
- TICKET (TicketSerialNo, IDFlight, FlightDate, Seat, Rate, Name, Surname, Sex)
- CHECK-IN (TicketSerialNo, CheckInTime, LuggageNr)

Design a Data Warehouse for TICKET:

- Derive the attribute tree
- Derive the fact schema and Identify measures and dimensions
- Design a snowflake schema

- (b) Explain surrogate key with appropriate example 3

4. (a) Suppose following is the set of sales transactions of a baby-item super-shop 6+6

Transaction ID	Items Brought (Itemset)
1	Bread, Butter, Milk
2	Bread, Butter
3	Honey, Cookies, Diapers
4	Milk, Diapers, Bread, Butter
5	Honey, Diapers

- Generate the Candidate itemset and frequent itemset with minimum support 40%
- Generate association rules from the frequent itemset you generated with minimum confidence threshold 70%

5. (a) Write down the steps of K-means clustering. 6
- The following six objects, each with two attributes are to be clustered. A1(4,6), A2(2,5), A3(9,3), A4(6,9), A5(5,9) and A6(4,7). Apply K-means clustering method to build cluster.
- (b) List four desirable features of a cluster analysis method. 2
- (c) Define outlier. Outliers are often discarded as noise but some applications these noisy data can be more interesting than the more regularly occurring ones. Why? Explain your answer. 4

6. (a) Consider the following dataset of weight and height of some neonates

6

Weight	Height
0.5	1.4
2.3	1.9
2.9	3.2

Find a regression line using gradient descent until 6th iteration and then predict height of a neonate whose weight is 3.5.

- (b) Kernel trick in Support Vector Machine (SVM) vs Kernel trick in Neural network: Discuss with necessary illustration
- (c) Propose the ways to build a logistic regression model using a perceptron.

4

2

7. (a) Given the following dataset of income and age of people and the third column is the target class column which states that whether the man owns a house or rented. Construct a decision tree using ID3

12

Income	Age	Own House
Very High	Young	Yes
High	Middle	Yes
Low	Young	Rented
High	Middle	Yes
Very High	Middle	Yes
Medium	Young	Yes
High	Old	Yes
Medium	Middle	Rented
Low	Middle	Rented
Low	Old	Rented
High	Young	Yes
Medium	Old	Rented

12
7+5
4-5

8. (a) Consider the following database containing five transactions with min_sup=60%

6

5*1

TID	Transactions
T100	{M,O,N,O,K,E,Y}
T200	{D,M,O,N,K,E,Y}
T300	{M,O,A,K,E}
T400	{M,U,C,K,Y}
T500	{C,O,O,K,I,E}

Mine frequent item sets using FP growth algorithm.

- (b) The following table shows the height and weight for students in a school.

6

2+2

Height(cm)	Weight(kg)
208	80
205	48
198	75
181	65
172	60

- i. Plot the data. Do Height and Weight seem to have a linear relationship?
- ii. Use the method of least squares to find an equation for the prediction of a student's weight based on the student's height and also calculate the weight of a student who has height 185