COMP1942 Exploring and Visualizing Data (Spring Semester 2020)
Online Final Examination (Question Paper)
Date: 24 May, 2020 (Sunday)
Time: 6:30pm-8:30pm
Duration: 2 hours

Instructions:

(1) Guideline
- (a) Please follow **all** instructions about the exam guideline (e.g., your face video capturing) stated in the Canvas website.
- (b) For the sake of space, we do not write them again.

(2) COMP1942 Virtual Barn Log-in Period
- (a) Your allocated period of logging in "COMP1942 Virtual Barn" could be found in the Canvas webpage.
- (b) In the whole exam period, you could use XLMiner installed at your laptop/desktop/browser **at any time** but, you could use XLMiner installed in "COMP1942 Virtual Barn" **in the allocated period only**. If you log in "COMP1942 Virtual Barn" outside the allocated period in the exam period, 40 (out of 200) (which is equivalent to 20 (out of 100)) will be deducted from your exam score.

(3) Question
- (a) Please answer **all** questions. The total scores in this exam are 200.
- (b) There are 2 parts in this exam, Part A (Short/Long Question) and Part B (Multiple-Choice Question).

(4) Answer Sheet
- (a) Please submit your answers in PDF to the Canvas website.
- (b) Please use the cover page stated in the Canvas website as the **first** page of your PDF file. This cover page includes your information and an agreement with your signature.
- (c) Please start to write your answers starting on the **second** page of your PDF file.
- (d) The PDF file should "clearly" show your answers without any blurred images. No marks will be given to any "blurred" parts in the PDF file. Please make sure that the PDF file shows your answers clearly.

(5) Online Exam
- (a) This is an online exam where you could access all online materials.
  However, it is **not** allowed to communicate with other people (except the instructor and the tutors in this course) in any form (including but not limited to orally, electronically and in writing) during the entire exam period.

(6) File Submission
- (a) We allow a 15-minute buffer for your PDF file upload. Remember to upload your file at around 8:30pm. We allow your file uploading time at most 15 minutes. Canvas will terminate any file uploading process at 8:45pm if your file is still being uploaded at 8:45pm.

(7) Zero-Score Regulation
- (a) If your face could not be shown in your video for at least 10 seconds in the exam period, your exam score will be set to 0 (even though you submit your PDF file in Canvas).
- (b) If you do not submit the first cover page, your exam score will be set to 0.
- (c) We only mark your latest PDF file uploaded by 8:45pm. Your exam score will be set to 0 if we could not see any PDF file uploaded by 8:45pm (even though you do the question paper or you "could" upload your PDF file after 8:45pm).

XLMiner Question

In this exam, we have the following questions which need you to use XLMiner.

- Part B – Q9
- Part B – Q10
- Part B – Q11
- Part B – Q12
- Part B – Q13
- Part B – Q14
- Part B – Q15
- Part B – Q16

Part A (Short/Long Question)

In this part, there are 8 short/long questions, namely Q1, Q2, Q3, Q4, Q5, Q6, Q7 and Q8. The total scores in this part are 120 scores (out of 200). Each question weights 15 scores (out of 200).

Q1 (15 Marks) (Version A)
Given the following transactions and the support threshold = 2.

TID	Items bought
1	b, d, f, r
2	b, c, d, s
3	c, m, t
4	b, d, f
5	a, d, f
6	e, f
7	f, h
8	b, d, c
9	a, l
10	c, g
11	c, k
12	f, n, o
13	b, c, d, p
14	f, j, q
15	c, i
16	a, d

Follow the steps of the FP-growth algorithm to find all frequent itemsets. Please draw all (conditional) FP-trees generated in the algorithm (e.g., all conditional FP-trees containing single paths). Please show the steps and list all frequent itemsets.

Q1 (15 Marks) (Version B)

Given the following transactions and the support threshold = 2.

TID	Items bought
1	s, u, w, i
2	s, t, u, j
3	t, d, k
4	s, u, w
5	r, u, w
6	v, w
7	w, y
8	s, u, l
9	r, c
10	t, x
11	t, b
12	w, e, f
13	s, t, u, g
14	w, a, h
15	t, z
16	r, u

Follow the steps of the FP-growth algorithm to find all frequent itemsets. Please draw all (conditional) FP-trees generated in the algorithm (e.g., all conditional FP-trees containing single paths). Please show the steps and list all frequent itemsets.

Q2 (15 Marks)

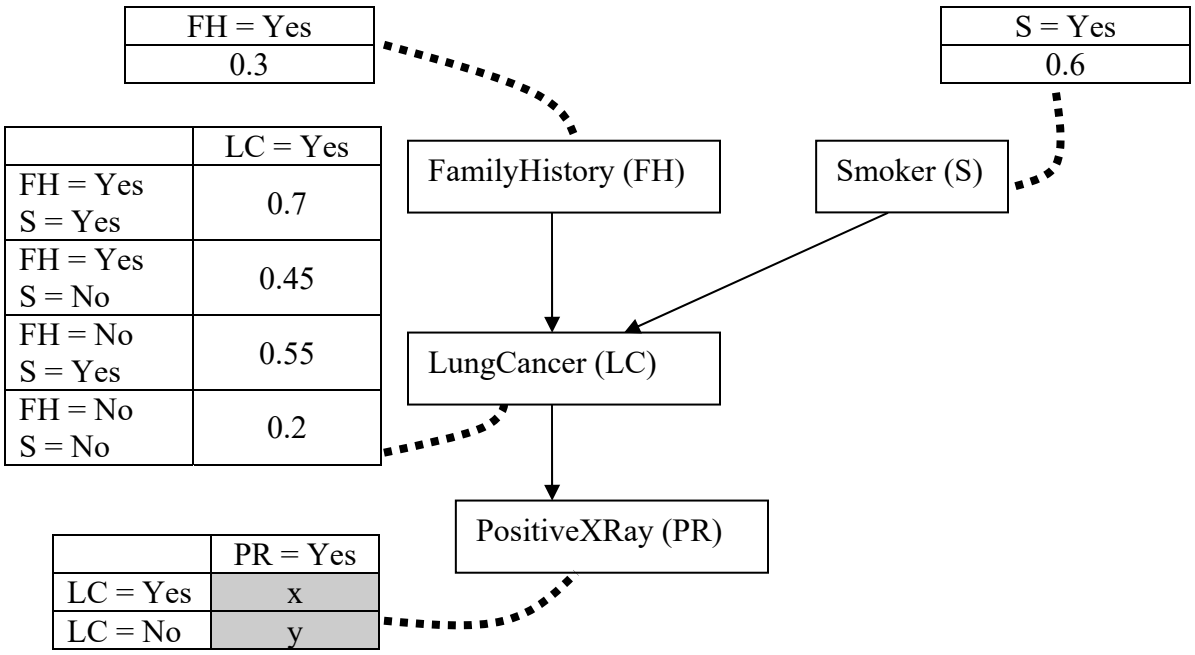
Consider eight data points.

x_1: (1, 2), x_2: (2, 0), x_3: (-10, -5), x_4: (-5, -2), x_5: (10, 12), x_6: (8, 6), x_7: (-8, -6), x_8: (2, 1)

Please use the agglomerative approach to find the dendrogram by using the median linkage as a distance measurement. Please write down the distances (in the x-axis) in the dendrogram. Please show each step of how you do the computation to know that the two (small) clusters are merged into one cluster in the agglomerative approach.

Q3 (15 Marks)

We have the following Bayesian Belief Network.



However, the given network has some missing information marked in grey color. Specifically, we do not know the exact values of variable x and variable y. But, what we know is that x is equal to 2y.

Besides, we know the following patient.

(1) he has his family history
(2) he is a smoker
(3) his result of X-Ray is negative

According to the concept of Bayesian Belief Network (without the missing information), we could compute the probability that he suffers from Lung Cancer is equal to 0.4375.

(a) (13 Marks) Is it possible to know the values of variable x and variable y? If yes, please write down the values of x and y, and give the derivations/explanations. Otherwise, please describe why it is not possible and give the minimal information so that we could derive the values of x and y.

(b) (2 Marks) Although Bayesian Belief Network classifier does not have an independent assumption among all attributes (compared with the naïve Bayesian classifier), what are the disadvantages of this classifier?

Q4 (15 Marks) (Version A)

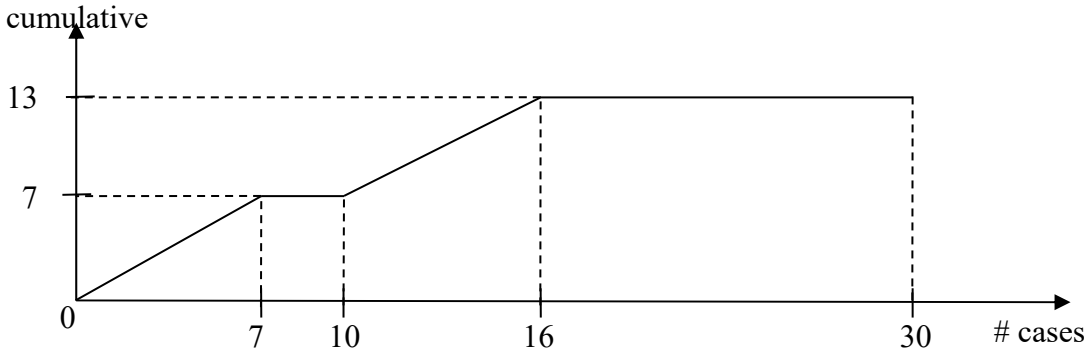
We are given the following 4 data points: (6, 6), (8, 8), (5, 11), (10, 4). Use PCA to reduce from two dimensions to one dimension for each of these 4 data points. In this part, please show your steps.

Q4 (15 Marks) (Version B)

We are given the following 4 data points: (18, 18), (24, 24), (15, 33), (30, 12). Use PCA to reduce from two dimensions to one dimension for each of these 4 data points. In this part, please show your steps.

Q5 (15 Marks)

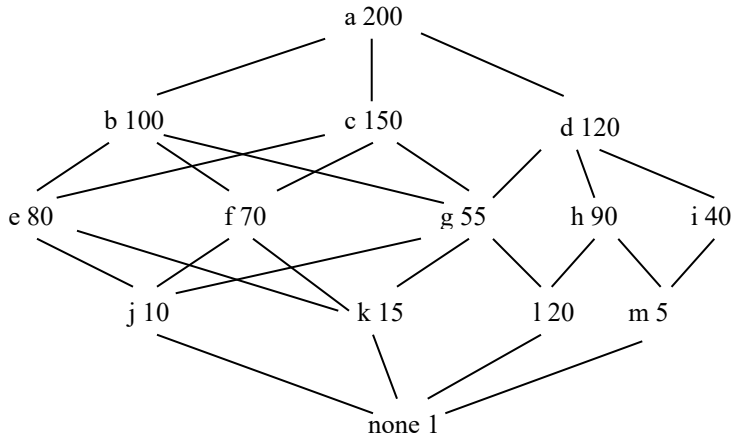
We are given the following lift chart based on a classifier.



(a) (2 Marks) Is it possible to find the number of false positives? If yes, please write down the number and give derivations. Otherwise, please explain it.
(b) (2 Marks) Is it possible to find the number of true positives? If yes, please write down the number and give derivations. Otherwise, please explain it.
(c) (2 Marks) Is it possible to find the number of false negatives? If yes, please write down the number and give derivations. Otherwise, please explain it.
(d) (2 Marks) Is it possible to find the number of true negatives? If yes, please write down the number and give derivations. Otherwise, please explain it.
(e) (4 Marks) Is it possible to find the decile-wise lift chart? If yes, please give the chart and give derivations. Otherwise, please explain it.
(f) (3 Marks) Is it possible to find the f1-score? If yes, please write down the number and give derivations. Otherwise, please explain it.

Q6 (15 Marks)

We are given the following lattice containing nodes where each node corresponds to a possible query. The number associated with each query corresponds to the cost of answering this query.



Assume that we do not consider "none 1". Suppose 4 views are to be materialized (other than the top view). Apply the greedy algorithm and find the resulting views. Please show all steps (e.g., a table showing the benefit of each view for each possible view selection).

Q7 (15 Marks)

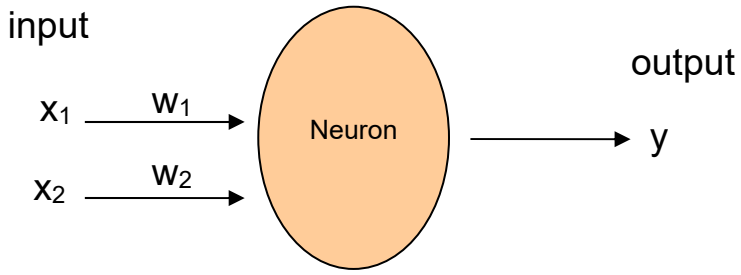
The following shows a table where the input attributes are x_1 and x_2 and y is the target attribute. There are two possible values in the target attribute y, namely "Yes" and "No". Attribute ID corresponds to the ID of each tuple.

ID	x_1	x_2	y
a	13	13	Yes
b	19	9	Yes
c	21	15	Yes
d	15	19	Yes
e	7	9	No
f	5	7	No
g	9	9	No
h	7	5	No

(a) (2 Marks) Consider a classification task by a neural network model or SVM. In order to perform the classification task by a neural network model or SVM, we need to transform the above table T to another table T' such that the target attribute y in the transformed table contains numeric values (i.e., 1 for "Yes" and -1 for "No") instead of categorical values. What is this transformed table T'?

(b) (4 Marks) Consider a classification task by SVM. Formulate this SVM problem by a quadratic programming. Please list the objective function and all constraints in this quadratic programming.

(c) (8 Marks) Consider a classification task by a neural network model containing a single neuron.



Initially, we set the values of w_1, w_2 and b to be 0.1 where b is a bias value in the neuron.

Suppose the learning rate is denoted by α. Let $\alpha = 0.5$.
Suppose we adopt the hyperbolic tangent function as an activation function.
Please try to train the neural network with the following first six instances in the given sequence and the answer in Part (a).

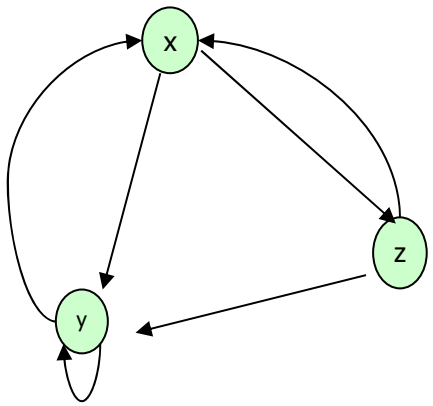
1. $(x_1, x_2) = (13, 13)$
2. $(x_1, x_2) = (19, 9)$
3. $(x_1, x_2) = (21, 15)$
4. $(x_1, x_2) = (15, 19)$
5. $(x_1, x_2) = (7, 9)$
6. $(x_1, x_2) = (5, 7)$

What are the final values of w_1, w_2 and b after these six instances are processed?

(d) (1 Mark) What is the major disadvantage of the traditional neural network model compared with the recurrent neural network model?

Q8 (15 Marks)

The following shows three sites, namely x, y and z, with their linkage.



(a) (2 Marks) What is the adjacency matrix?
(b) (2 Marks) What is the stochastic matrix?
(c) (4 Marks) Suppose that site x would like to become a "spider trap". Please list out all possible operations that site x has to do.
(d) (7 Marks) In the PageRank algorithm, we need to update a ranking vector r by "$0.8 \cdot M \cdot r + c$" iteratively where M is the stochastic matrix and c is a vector $(0.2, 0.2, \ldots 0.2)^T$. Suppose that r_n is the ranking vector after the update and r_0 is the ranking vector before the update. For simplicity, you can assume that there are only three sites in this PageRank algorithm. Prove that if the sum of the values in r_0 is equal to 3, then the sum of the values in r_n is equal to 3. In the proof, please use the following notations.

$$r_0 = \begin{pmatrix} r_{0,1} \\ r_{0,2} \\ r_{0,3} \end{pmatrix}, \qquad r_n = \begin{pmatrix} r_{n,1} \\ r_{n,2} \\ r_{n,3} \end{pmatrix}, \qquad M = \begin{pmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{pmatrix}$$

Part B (Multiple-Choice Question)

In this part, there are 16 multiple-choice questions, namely Q9-Q24. The total scores in this part are 80 scores (out of 200). Each question weighs 5 scores (out of 200). In your answer sheet, please write down the following table on your **last** page of your PDF submission. In the corresponding cell, write down the answer for each question.

Note: Please write the letter **clearly** (i.e., A, B, C, D or E) for each answer so that it could be distinguished from other letters **easily**. In the past, some students wrote the letter unclearly which look like two possible letters. One example is that the hand-written letter "B" (from some students) is similar to the hand-written letter "E". There are more examples which are not included here. In any case, if your letter is judged by us that it is unclear, even though you "thought" that your answer is correct, 0 score will be given to you for that question.

Part B

Question	Your Answer
Q9	
Q10	
Q11	
Q12	
Q13	
Q14	
Q15	
Q16	
Q17	
Q18	
Q19	
Q20	
Q21	
Q22	
Q23	
Q24	

Q9. [XLMiner Question] You are given an Excel file called "Q9-10.xlsx" which is used in Q9 and Q10. This Excel file stores a list of individual records with 6 attributes, namely “AgeRange”, “Gender”, “Education”, “FamilySize”, “Status”, “WorkOutsideHK”. The following shows the header row in this Excel file.

AgeRange	Gender	Education	FamilySize	Status	WorkOutsideHK
…	…	…	…	…	…

Suppose that we want to use the Naïve Bayes classifier to train a model on the given list of individual records. In the XLMiner setting, (1) we do not need to partition the data, (2) we do not need to choose the “Laplace smoothing” option and (3) we use the empirical probability in “prior probability”. We want to predict whether each individual with the following 5 attribute values (i.e., “AgeRange”, “Gender”, “Education”, “FamilySize” and “Status”) will work outside Hong Kong or not. If the predicted probability that an individual will work outside Hong Kong is at least 0.5, we simply say that this individual will work outside Hong Kong. Otherwise, s/he will not. In the following, we include one additional attribute “Name” to describe the name of the individual.

Name	AgeRange	Gender	Education	FamilySize	Status
Raymond	71-80	Male	Master	6	married
Mary	81-90	Female	Secondary School	4	single
Chris	51-60	Male	Doctor	3	married

A. Only Raymond and Mary will work outside Hong Kong but Chris will not.
B. Only Mary and Chris will work outside Hong Kong but Raymond will not.
C. Only Raymond and Chris will work outside Hong Kong but Mary will not.
D. Only Raymond will work outside Hong Kong but Mary and Chris will not.
E. None of the above choices

Q10. [XLMiner Question] Continue from the dataset description of "Q9-10.xlsx" from Q9.

Suppose that we want to use the Naïve Bayes classifier to train a model on the given list of individual records. We want to predict whether each individual with the following 5 attribute values (i.e., “AgeRange”, “Gender”, “Education”, “FamilySize” and “WorkOutsideHK”) is single or not. If the predicted probability that an individual is single is at least 0.5, we say that this individual is single. Otherwise, s/he is not. In the following, we include one additional attribute “Name” to describe the name of the individual.

Name	AgeRange	Gender	Education	FamilySize	WorkOutsideHK
Wong	71-80	Male	Master	6	Yes
Lee	81-90	Female	Secondary School	4	No
Chan	51-60	Male	Doctor	3	Yes

A. Only Wong and Lee are both single but Chan is not.
B. Only Lee and Chan are both single but Wong is not.
C. Only Wong and Chan are both single but Lee is not.
D. Only Wong is single but Lee and Chan are not.
E. None of the above choices

Q11. [XLMiner Question] You are given an Excel file called "Q11-12.xlsx" which is used in Q11 and Q12. This Excel file stores a list of student report score records with 5 attributes, namely “Presentation”, “Organization”, “Grammar”, “Interesting” and “Relevancy”. The following shows the header row in this Excel file.

Presentation	Organization	Grammar	Interesting	Relevancy
…	…	…	…	…

Suppose that we want to perform PCA using the covariance matrix to reduce the dimensionality from 5 to 2. Thus, each record is finally transformed to a 2-dimensional vector. What is the sum of the squares of these 2 dimensional values for the first record?
For example, if the 2-demensional vector is (-0.5, 0.2), then the sum of the squares of these 2 values is equal to $(-0.5)^2 + 0.2^2 = 0.29$.

A. The sum is in the range between 0.00 and 2000.00.
B. The sum is in the range between 2000.01 and 4000.00.
C. The sum is in the range between 4000.01 and 6000.00.
D. The sum is in the range between 6000.00 and 8000.00.
E. None of the above choices

Q12. [XLMiner Question] Continue from the dataset description of "Q11-12.xlsx" from Q11. Please use XLMiner to perform hierarchical clustering to find 2 clusters using the distance measurement as the median linkage. In XLMiner, we do not want to normalize the data. What is the total number of points in each of these 2 clusters.

A. The number of points in one cluster is 98 and the number of points in another cluster is 2.
B. The number of points in one cluster is 97 and the number of points in another cluster is 3.
C. The number of points in one cluster is 75 and the number of points in another cluster is 25.
D. The number of points in one cluster is 77 and the number of points in another cluster is 23.
E. None of the above choices

Q13. [XLMiner Question] You are given an Excel file called "Q13-16.xlsx" which is used in Q13, Q14, Q15 and Q16. This Excel file stores a list of student records with 6 attributes, namely "Gender", "Major", "Attendance", "HW1", "Phase1" and "Midterm". The physical meaning of these attributes could be understood with their attribute names. But, we want to clarify some attributes. Attribute "Attendance" means the score of the students for the "attendance" assessment. Attribute "HW1" and attribute "Phase1" have similar meanings. Attribute "Midterm" has 2 possible values, namely "AboveMean" and "BelowMean". For the sake of simplification, we assume that if the student has his/her midterm score equal to the mean of the midterm exam, s/he has the value of "Midterm" as "AboveMean". The following shows the header row in this Excel file.

Gender	Major	Attendance	HW1	Phase1	Midterm
…	…	…	…	…	…

Please do a transformation on this dataset with the following changes. (Note: Our tutorial notes taught how to perform this kind of transformation before (e.g., "Manual" Option of "Categorical Data – Reduce Categories"). If you don't know how to use this XLMiner feature, you could do this manually by yourself.)

- In attribute "Gender",
  - Change "Male" to 1
  - Change "Female" to 2
- In attribute "Major",
  - Change "COMP" to 1
  - Change "CPEG" to 2
  - Change "MAEC" to 3
  - Change "QFIN" to 4
  - Change "QSA" to 5

We want to use "Decision Tree" to train this transformed dataset. Please use this setting in XLMiner.

- We do not partition the data.
- We do not re-scale the data
- We need to limit the number of levels to 2
- We need to limit the number of records in a terminal node to 50
- Unspecified parameters should be left as default values.

We want to use the trained model to predict whether each of the following students will have value "AboveMean" in attribute "Midterm".

Name	Gender	Major	Attendance	HW1	Phase1
Susan	Female	QFIN	87	55	84
Tom	Male	QSA	55	30	90
Raymond	Male	QSA	82	87	55

A. Only Susan will have a value "AboveMean" in attribute "Midterm" but the other 2 students will not.
B. Only Tom will have a value "AboveMean" in attribute "Midterm" but the other 2 students will not.
C. Only Raymond will have a value "AboveMean" in attribute "Midterm" but the other 2 students will not.
D. None of these 3 students will have a value "AboveMean" in attribute "Midterm"
E. None of the above choices

Q14. [XLMiner Question] Continue from the dataset description of "Q13-16.xlsx" from Q13. According to the XLMiner result of Q13, what are the accuracy, specificity and f1-score of the model used in Q13 (on the given dataset)?

A. Accuracy = 91.6%; Specificity=100%; f1-score=91.80%
B. Accuracy = 91.6%; Specificity=1%; f1-score=0.92%
C. Accuracy = 87.4%; Specificity=92.7%; f1-score=86.2%
D. Accuracy = 87.4%; Specificity=0.93%; f1-score=0.86%
E. None of the above choices

Q15. [XLMiner Question] Continue from the dataset description of "Q13-16.xlsx" from Q13. We want to use "7-nearest neighbor" classifier to train this transformed dataset. Please use this setting in XLMiner.

- We do not partition the data.
- We do not re-scale the data.
- We used the fixed K as 7.
- Unspecified parameters should be left as default values.

We want to use the trained model to predict whether each of the following students will have value "AboveMean" in attribute "Midterm".

Name	Gender	Major	Attendance	HW1	Phase1
Susan	Female	QFIN	87	55	84
Tom	Male	QSA	55	30	90
Raymond	Male	QSA	82	87	55

A. Only Susan will have a value "AboveMean" in attribute "Midterm" but the other 2 students will not.
B. Only Tom will have a value "AboveMean" in attribute "Midterm" but the other 2 students will not.
C. Only Raymond will have a value "AboveMean" in attribute "Midterm" but the other 2 students will not.
D. None of these 3 students will have a value "AboveMean" in attribute "Midterm"
E. None of the above choices

Q16. [XLMiner Question] Continue from the dataset description of "Q13-16.xlsx" from Q13. We want to use "neural network" classifier to train this transformed dataset. Please use this setting in XLMiner.

- We do not partition the data.
- We do not re-scale the data.
- No. of Hidden Layers = 1
- No. of neurons in this hidden layers = 4
- Sigmoid function is used in both the hidden layer and the output layer.
- Unspecified parameters should be left as default values.

We want to use the trained model to predict whether each of the following students will have value "AboveMean" in attribute "Midterm".

Name	Gender	Major	Attendance	HW1	Phase1
Susan	Female	QFIN	87	55	84
Tom	Male	QSA	55	30	90
Raymond	Male	QSA	82	87	55

A. Only Susan will have a value "AboveMean" in attribute "Midterm" but the other 2 students will not.
B. Only Tom will have a value "AboveMean" in attribute "Midterm" but the other 2 students will not.
C. Only Raymond will have a value "AboveMean" in attribute "Midterm" but the other 2 students will not.
D. None of these 3 students will have a value "AboveMean" in attribute "Midterm"
E. None of the above choices

Q17. Which of the following statements are true?

(1) Consider frequent pattern mining. If we set the support threshold smaller, then the number of frequent itemsets is also smaller.
(2) It is a must that the sum of the values in the final authority vector in the HITS algorithm (i.e., one of the final outputs of the HITS algorithm) is the number of elements.
(3) It is a must that the sum of the values in the final vector in the PageRank algorithm (i.e., the final output of the PageRank algorithm) is the number of elements.

A. Statements (1) and (2) only
B. Statements (1) and (3) only
C. Statements (2) and (3) only
D. Statements (1), (2) and (3)
E. None of the above choices

Q18. Consider association rule mining on a table with items B, C and D. Suppose that we know that the lift ratio of "C→B" is at least 2.0. Which of the following statements are true?

(1) It is a must that the lift ratio of "B→C" is at least 2.0.
(2) It is a must that the lift ratio of "{C, D}→B" is at least 2.0.
(3) It is a must that the lift ratio of "C→{B, D}" is at least 2.0.

A. Statement (1) only
B. Statement (2) only
C. Statement (3) only
D. Statements (1), (2) and (3)
E. None of the above choices

Q19. Which of the following statements are true?

(1) Consider a lift chart. It is always true that the y-axis value of the lift curve in the chart does not decrease when the x-axis value of the lift curve increases.
(2) Consider a decile-wise lift chart. It is always true that the y-axis value of the bar in the chart does not increase when the x-axis value of the bar increases.
(3) Consider three random variables A, B and C. It is always true that
$P(A \mid B, C) = P(B \mid A, C) \cdot P(A \mid C) / P(B \mid C)$.

A. Statements (1) and (2) only
B. Statements (1) and (3) only
C. Statements (2) and (3) only
D. Statements (1), (2) and (3)
E. None of the above choices

Q20. Consider the data warehouse technique we learnt in class. Given a set V of views to be materialized, we know how to compute Gain(V ∪ {top view}, {top view}). Which of the following statements are true?

(1) Let A be a view. It is always true that
Gain(V ∪ {top view}, {top view}) ≥ Gain(V ∪ {top view, view A}, {top view, view A}).
(2) Suppose that view P and view C are two nodes where P is a parent node of C (i.e., P is just above C) in the relationship graph. It is always true that
Gain({view P} ∪ {top view}, {top view}) ≥ Gain({view C} ∪ {top view}, {top view}).
(3) Suppose that S and T are two sets of views to be materialized such that "S ⊆ T".
(Note that "S ⊆ T" means that each view in set S can be found in set T.)
It is always true that for any view x,
Gain({x} ∪ S ∪ {top view}, {top view}) - Gain(S ∪ {top view}, {top view})
≥ Gain({x} ∪ T ∪ {top view}, {top view}) - Gain(T ∪ {top view}, {top view}).

A. Statement (1) only
B. Statement (2) only
C. Statement (3) only
D. Statements (1), (2) and (3)
E. None of the above choices

Q21. You learnt some measurements for a decision tree in class. Two of them are represented in the form of charts. They are a lift chart and a decile-wise lift chart. Which of the following statements are correct?

(1) It is always true that the greatest possible value in the y-axis in the lift chart is equal to the greatest possible value in the y-axis in the decile-wise lift chart.
(2) It is always true that we can construct the decile-wise lift chart according to the lift chart without the original training dataset.
(3) It is always true that we can construct the lift chart according to the decile-wise lift chart without the original training dataset.

A. Statement (1) only
B. Statement (2) only
C. Statement (3) only
D. Statements (1), (2) and (3)
E. None of the above choices

Q22. Consider the FP-tree built based on a given support threshold T. Which of the following statements are true?

(1)Consider a parent node N and a child node N'. It is always true that in the FP-tree, the count stored in N' is greater than the count stored in N.

(2)Consider an item A. It is always true that the total count stored in all nodes for item A in the FP-tree is at least T if there exists a node for item A in this FP-tree.

(3)It is always true that the total number of non-root nodes in the FP-tree must be smaller than the total number of occurrences of frequent items in the original data.

A. Statement (1) only
B. Statement (2) only
C. Statement (3) only
D. Statements (2) and (3) only
E. None of the above choices

Q23. Which of the following statements are true?

(1) Consider the Apriori algorithm. It is always true that C_i is larger than or equal to L_i for each i = 1, 2, 3, …

(2) Consider the Apriori algorithm. It is always true that C_{i+1} is larger than or equal to L_i for each i = 1, 2, 3, …

(3) It is always true that C_i must become smaller after the prune step is executed for each i = 1, 2, 3, ….

A. Statement (1) only
B. Statement (2) only
C. Statement (3) only
D. Statements (1) and (2) only
E. None of the above choices

Q24. Which of the following statements are true?

(1) If we know that the current record is dependent on the previous record, it is better to use the recurrent neural network model compared with the traditional neural network model for training.
(2) In practice, usually, we set parameter k of the k-nearest neighbor classifier to an even number.
(3) It is possible that the decision tree generated based on ID3 is equal to the decision tree generated based on CART.

A. Statements (1) and (2) only
B. Statements (1) and (3) only
C. Statements (2) and (3) only
D. Statements (1), (2) and (3)
E. None of the above choices

End of Paper