COMP1942 Exploring and Visualizing Data (Spring Semester 2019)
Final Examination (Question Paper)
Date: 22 May, 2019 (Wed)
Time: 12:30-15:00
Duration: 2 hours 30 minutes

Student ID:__________________ Student Name:__________________________________

Seat No. :__________________

Instructions:
(1) Please answer **all** questions in the **answer sheet**.
(2) You can use a calculator.

Question Paper

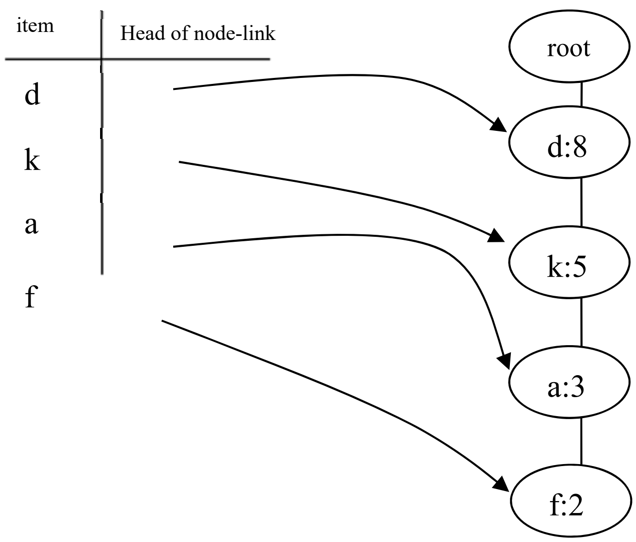
Q1 (20 Marks)

(a) Given a dataset with the following transactions in *binary* format, and the support threshold = 2.

R	A	Y	M	N
0	1	1	0	0
0	1	1	0	1
1	0	0	0	0
1	1	1	0	1
1	0	1	0	1

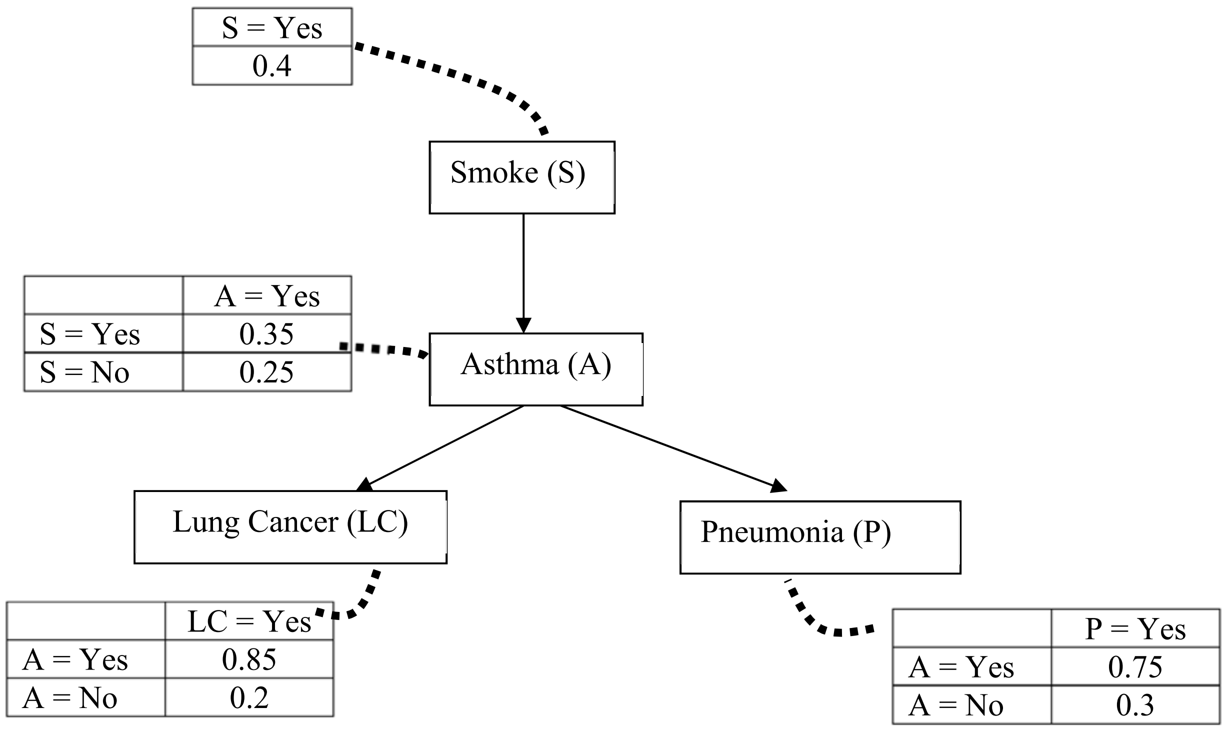
After we perform the join step and the prune step in the Apriori algorithm, we obtain a set C of itemsets. Then, we need to do the counting step for C (i.e., we need to find the frequency of each itemset in C). Finally, we output all itemsets in C with frequency at least a given support threshold as a part of the final output. Why do we need to do the counting step? That is, why can't we simply output C as a part of the final output? You can use the above dataset for illustration.

(b) The following shows an FP-tree. Let the support threshold be 2.
Please list all frequent itemsets with their correspondence frequency counts.



Q2 (20 Marks)

We have the following Bayesian Belief Network.



S = Yes
0.4

	A = Yes
S = Yes	0.35
S = No	0.25

	LC = Yes
A = Yes	0.85
A = No	0.2

	P = Yes
A = Yes	0.75
A = No	0.3

Suppose that there is a new person. We know that

(1) he has no lung cancer
(2) he has pneumonia

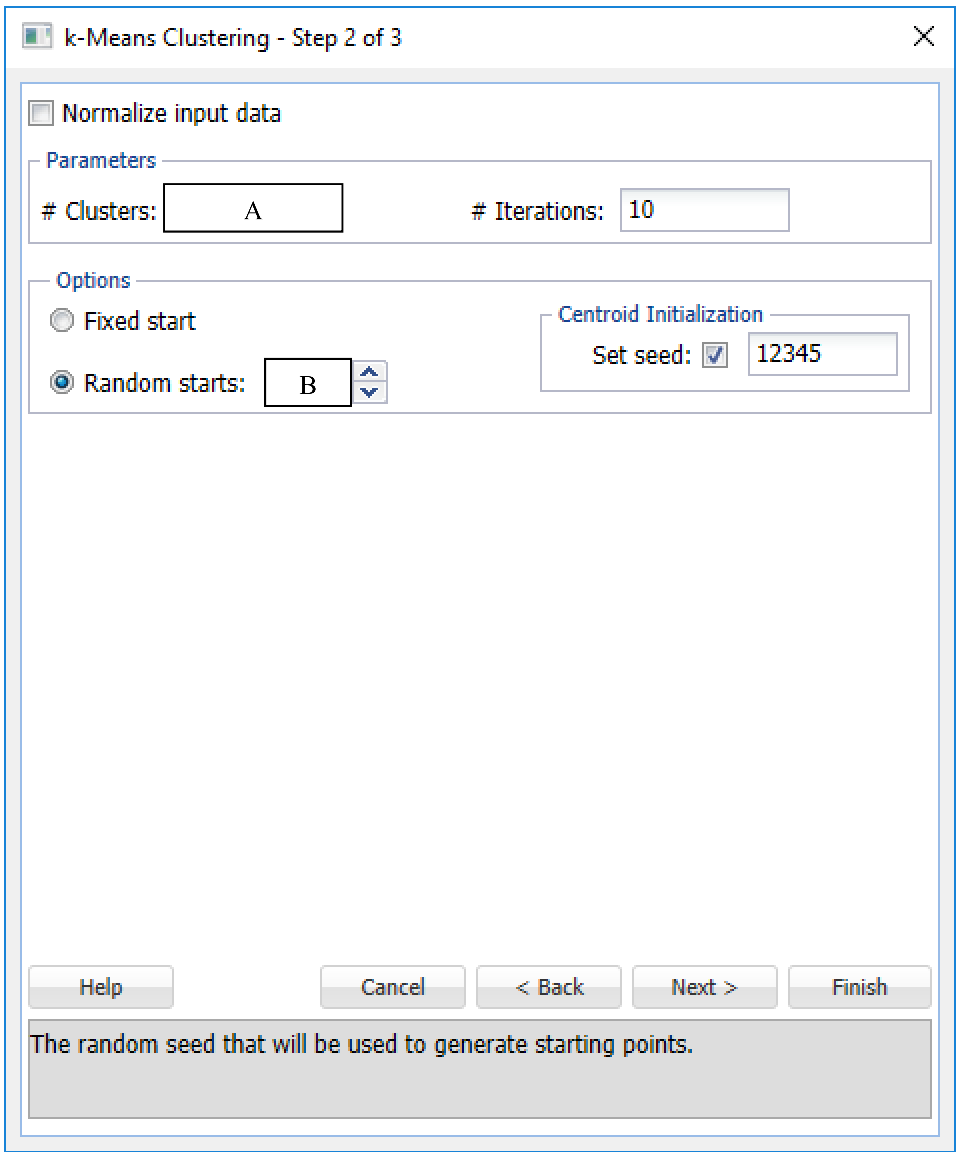
We would like to know whether he is likely to smoke or not.

Please use Bayesian classifier with the use of Bayesian Belief Network to predict whether he is likely to smoke.

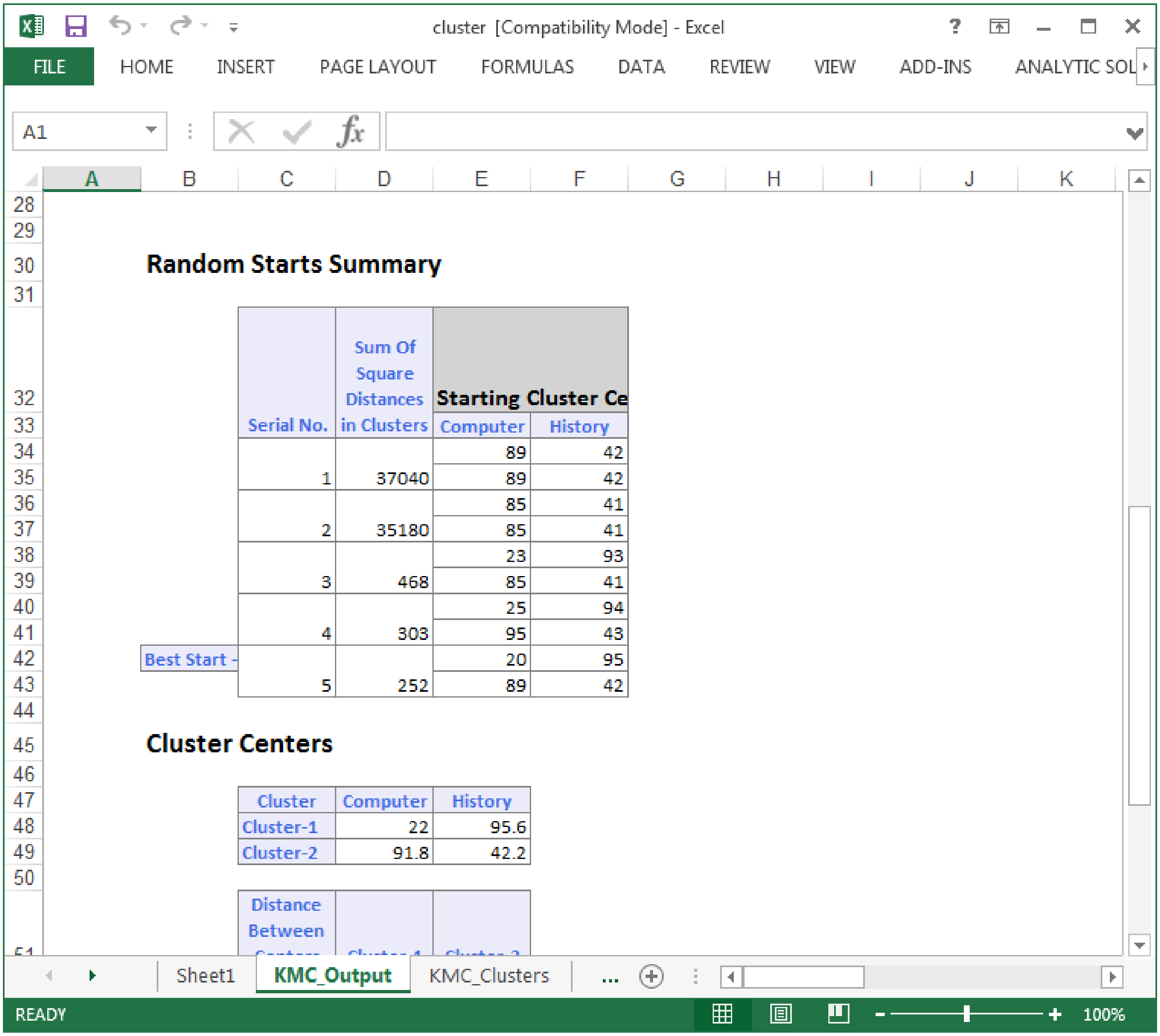
Q3 (20 Marks)

(a) We know how to compute the impurity measurement of an attribute A under the ID3 decision tree, denoted by Imp-ID3(A). We also know how to compute the impurity measurement of an attribute A under the CART decision tree, denoted by Imp-CART(A). Consider two attributes A and B. Is it always true that if Imp-CART(A) > Imp-CART(B), then Imp-ID3(A) > Imp-ID3(B)? If yes, please show that it is true. Otherwise, please give a counter example showing that this is not true and then explain it.

(b) In XLMiner, suppose that we want to perform k-means clustering. We have to specify some parameters in the following input dialog box shown here (and other input dialog boxes not shown here). Note that there is an unknown number "A" and another unknown number "B" in the following input dialog box. Both number "A" and number "B" are inputted by a user.



After that, we execute XLMiner at Raymond's machine and obtain the following output O_1 from XLMiner.



Random Starts Summary

	Serial No.	Sum Of Square Distances in Clusters	Starting Cluster Ce Computer	History
			89	42
	1	37040	89	42
			85	41
	2	35180	85	41
			23	93
	3	468	85	41
			25	94
	4	303	95	43
Best Start -			20	95
	5	252	89	42

Cluster Centers

Cluster	Computer	History
Cluster-1	22	95.6
Cluster-2	91.8	42.2

Please answer the following questions.

(i) What is the value of A (in the input dialog box)?

(ii) What is the value of B (in the input dialog box)?

(iii) What is the mean of each of the clusters finally chosen by XLMiner?

(iv) What is the initial mean of each of the clusters in the k-means clustering such that we have the final clusters with means given in (iii)?

(v) Suppose that this Excel file is sent from Raymond to you. You execute XLMiner at your machine with the same configuration described above (i.e., the same input parameter values given above and the same input parameter values in other input dialog boxes not shown here) and obtain the output O_2. Is it always true that O_1 is equal to O_2? Please give the answer and elaborate it.

Q4 (20 Marks)

(a) In class, we learnt "Sequential K-means Clustering" and "Forgetful Sequential K-means Clustering". What is the scenario or application that "Forgetful Sequential K-means Clustering" is better used compared with "Sequential K-means Clustering"?

(b) Consider eight data points.

The following matrix shows the pairwise distances between any two points.

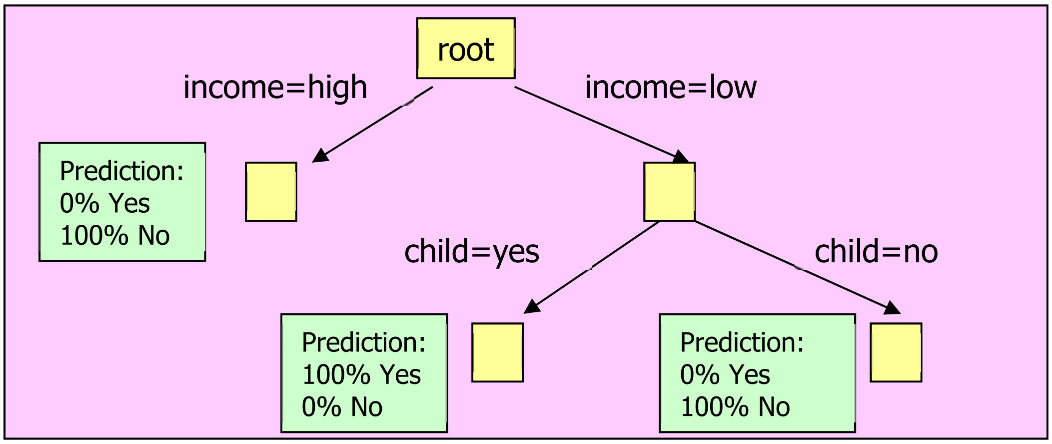
$$
\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \end{array}
\begin{array}{c}
\begin{array}{cccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{array} \\
\begin{pmatrix}
0 & & & & & & & \\
11 & 0 & & & & & & \\
5 & 13 & 0 & & & & & \\
12 & 2 & 14 & 0 & & & & \\
7 & 17 & 1 & 18 & 0 & & & \\
13 & 4 & 15 & 5 & 20 & 0 & & \\
9 & 15 & 12 & 16 & 15 & 19 & 0 & \\
11 & 20 & 12 & 21 & 17 & 22 & 30 & 0
\end{pmatrix}
\end{array}
$$

Please use the agglomerative approach to group these points with distance group average linkage. Draw the corresponding dendrogram for the clustering. You are required to specify the distance metric in the dendrogram.

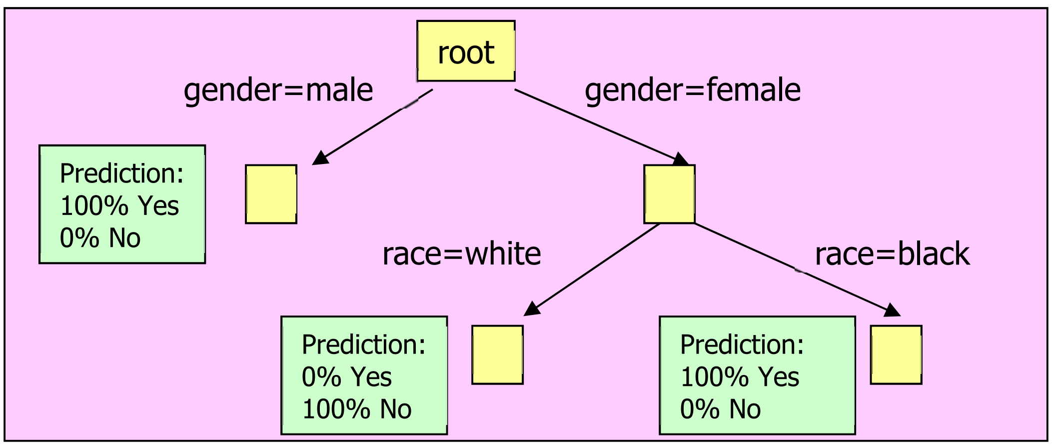
Q5 (20 Marks)

(a) The insurance company is given a table with five input attributes, namely Race, Gender, Married, Income and Child, and one target attribute, namely Insurance. Based on this table, the insurance company constructed three classifiers based on different criteria, namely Classifier 1, Classifier 2 and Classifier 3.

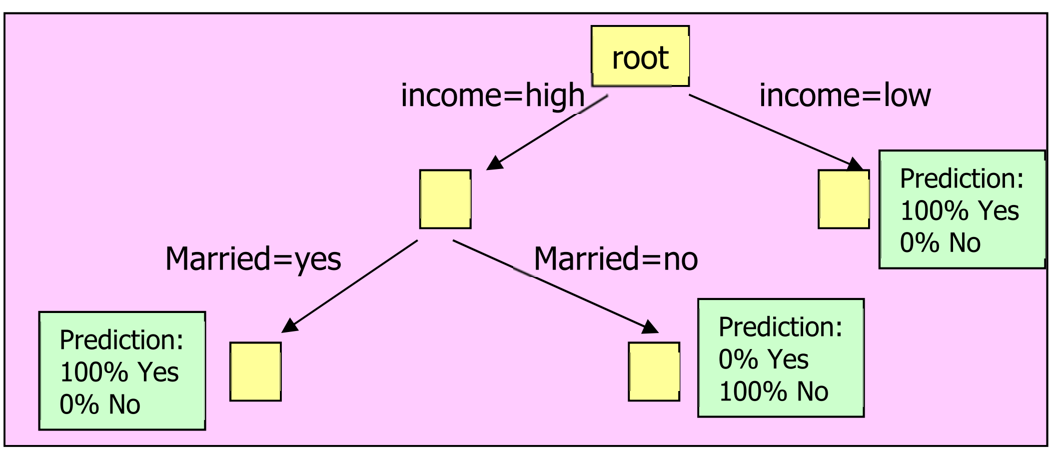
Classifier 1



Classifier 2



Classifier 3



Consider a group of 3 classifiers called an "ensemble" studied in class. Consider a new customer. All input attribute values of this new customer are known to the insurance company. The company uses this "ensemble" to do the prediction and predicts that this new customer will not buy an insurance policy. Suppose that we are very "curious" about the input attribute values of this new customer. What we know about the new customer is that he or she has a low income and the "predicted" result is that this new customer will not buy an insurance policy. We also know the 3 exact classifiers used in the insurance company. Is it possible for us to find the values of some input attribute values of this customer? If yes, please state (1) all these input attribute values and (2) all input attribute(s) that could not be found with their values. Otherwise, please write down the reason why we could not find those values.

(b) Consider that we want to conduct an experiment on a particular chemical. We want to test whether this chemical will have any reaction with another chemical of a fixed amount when the temperature is kept to be a certain value and the weight of this chemical is adjusted to be another certain value. The following table shows the experimental results. This table contains 2 numeric attributes, namely temperature and weight, and one binary attribute, namely react. Each record in the following table corresponds to a chemical test.

Record ID	Temperature	Weight	React
1	2	18	Yes
2	4	20	Yes
3	8	16	No
4	12	14	No
5	16	12	No
6	14	10	No
7	10	12	No
8	8	6	No
9	6	4	Yes
10	8	18	No
11	18	2	Yes
12	20	8	Yes
13	12	4	Yes
14	16	6	Yes

We want to predict whether the chemical will have any reaction when the temperature is equal to 10 and the weight of this chemical is equal to 4. Suppose that we want to use a 3-nearest neighbor classifier and we adopt the Euclidean distance as a distance measurement between two given points/records. What is the prediction? Please write down the prediction (i.e., Yes or No) and the record IDs of the corresponding 3 nearest neighbors.

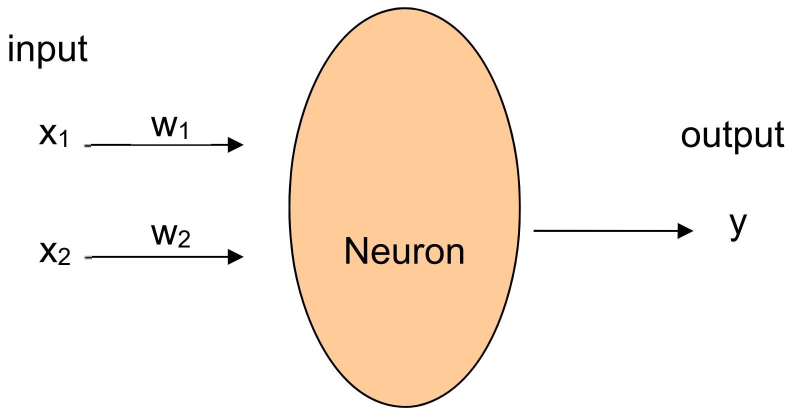
Q6 (20 Marks)

Consider the following table with three attributes where “No. of Phones” and “No. of Laptops” are input attributes and “Buy_NintendoSwtich” is the target attribute. Each tuple corresponds to a customer. An attribute “Record ID” denotes the ID of each record.

Record ID	No. of Phones	No. of Laptops	Buy_ NintendoSwtich
1	0	0	No
2	0	1	No
3	1	0	No
4	1	1	Yes

(a) Rewrite the above table such that values “Yes” and “No” in attribute “Buy_NintendoSwtich” are mapped to values 1 and 0, respectively.

(b) Consider a neural network containing a single neuron where x_1 = “No. of Phones”, x_2 = “No. of Laptops” and y = “Buy_NintendoSwtich”.



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 sigmoid function as an activation function.
Please try to train the neural network with five instances by the following inputs in the given sequence and the answer in Part (a).

1. $(x_1, x_2) = (0, 0)$
2. $(x_1, x_2) = (0, 1)$
3. $(x_1, x_2) = (1, 0)$
4. $(x_1, x_2) = (1, 1)$
5. $(x_1, x_2) = (0, 0)$

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

Q7 (20 Marks)

Suppose that c is a positive real number where we do not know the exact value.

(a) Consider the four 2-dimensional data points:

a:(8, 8), b:(10, 10), c:(7, 11), d:(11, 7)

We can make use of PCA for dimensionality reduction. In dimensionality reduction, given an L-dimensional data point, we want to transform this point to a K-dimensional data point where $K < L$ such that the information loss during the transformation is minimized. Suppose that $L = 2$ and $K = 1$.
Please illustrate with the above example.

(b) Consider the four 2-dimensional data points:

a:(8 + c, 8 + c), b:(10 + c, 10 + c), c:(7 + c, 11 + c), d:(11 + c, 7 + c)

We can make use of PCA for dimensionality reduction. In dimensionality reduction, given an L-dimensional data point, we want to transform this point to a K-dimensional data point where $K < L$ such that the information loss during the transformation is minimized. Suppose that $L = 2$ and $K = 1$.
Can we make use of the answers in part (a) to perform the dimensionality reduction? If yes, please write down each transformed data point. If no, please write down the reasons why we cannot make use of the answers of part (a).

(c) Consider the four 2-dimensional data points:

a:(8c, 8c), b:(10c, 10c), c:(7c, 11c), d:(11c, 7c)

We can make use of PCA for dimensionality reduction. In dimensionality reduction, given an L-dimensional data point, we want to transform this point to a K-dimensional data point where $K < L$ such that the information loss during the transformation is minimized. Suppose that $L = 2$ and $K = 1$.
Can we make use of the answers in part (a) to perform the dimensionality reduction? If yes, please write down each transformed data point. If no, please write down the reasons why we cannot make use of the answers of part (a).

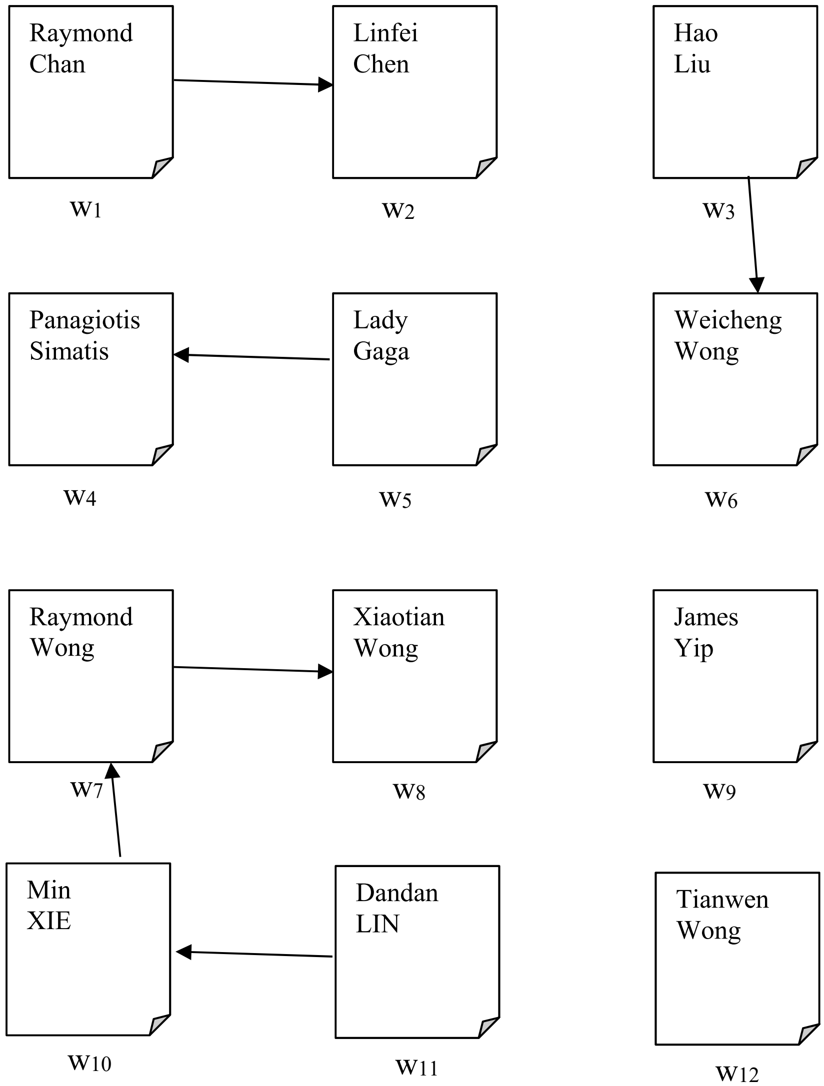
Q8 (20 Marks)

We are given the following adjacency matrix according to four sites, namely p, q, r and s.

$$\begin{array}{c} \\ p \\ q \\ r \\ s \end{array} \begin{array}{c} \begin{array}{cccc} p & q & r & s \end{array} \\ \begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \end{array}$$

(a) Is it possible to find the corresponding stochastic matrix? If yes, write down the stochastic matrix. Otherwise, please explain it.

(b) We are given the following 12 webpages, namely w_1, w_2, …, w_{12}.



The query terms typed by the user are "Raymond" and "Wong".

(i) What is the root set in this query? Please list the webpages in this set.

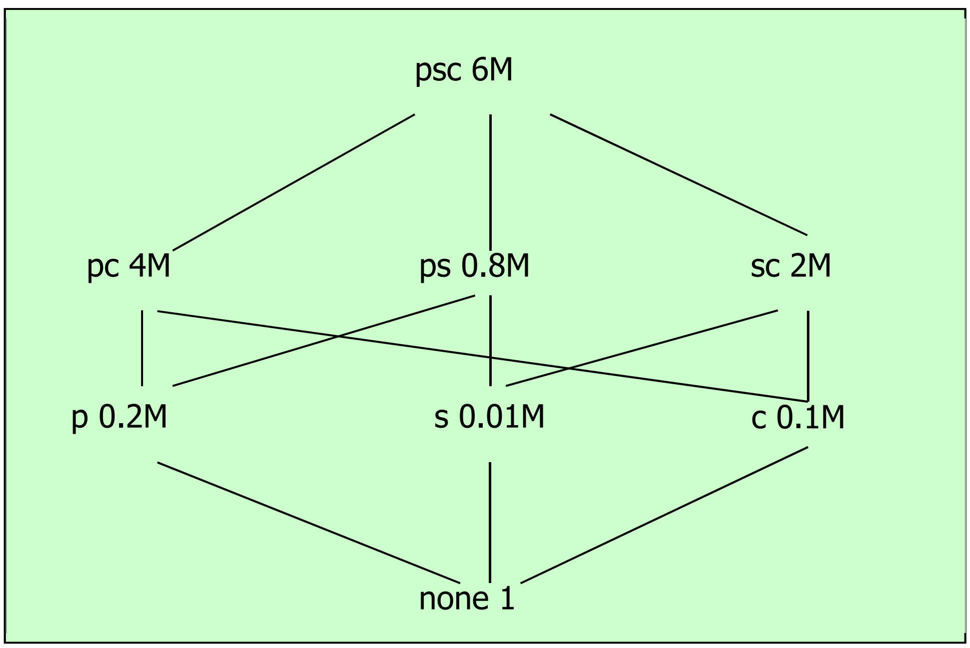
(ii) What is the base set in this query? Please list the webpage in this set.

Q9 (20 Marks)

In class, we learnt that we are given the following table describing the scenario that "parts are bought from suppliers and then sold to customers at a sale price SP".

part	supplier	customer	SP
p1	s1	c1	4
p3	s1	c2	3
p2	s3	c1	7
…	…	…	…

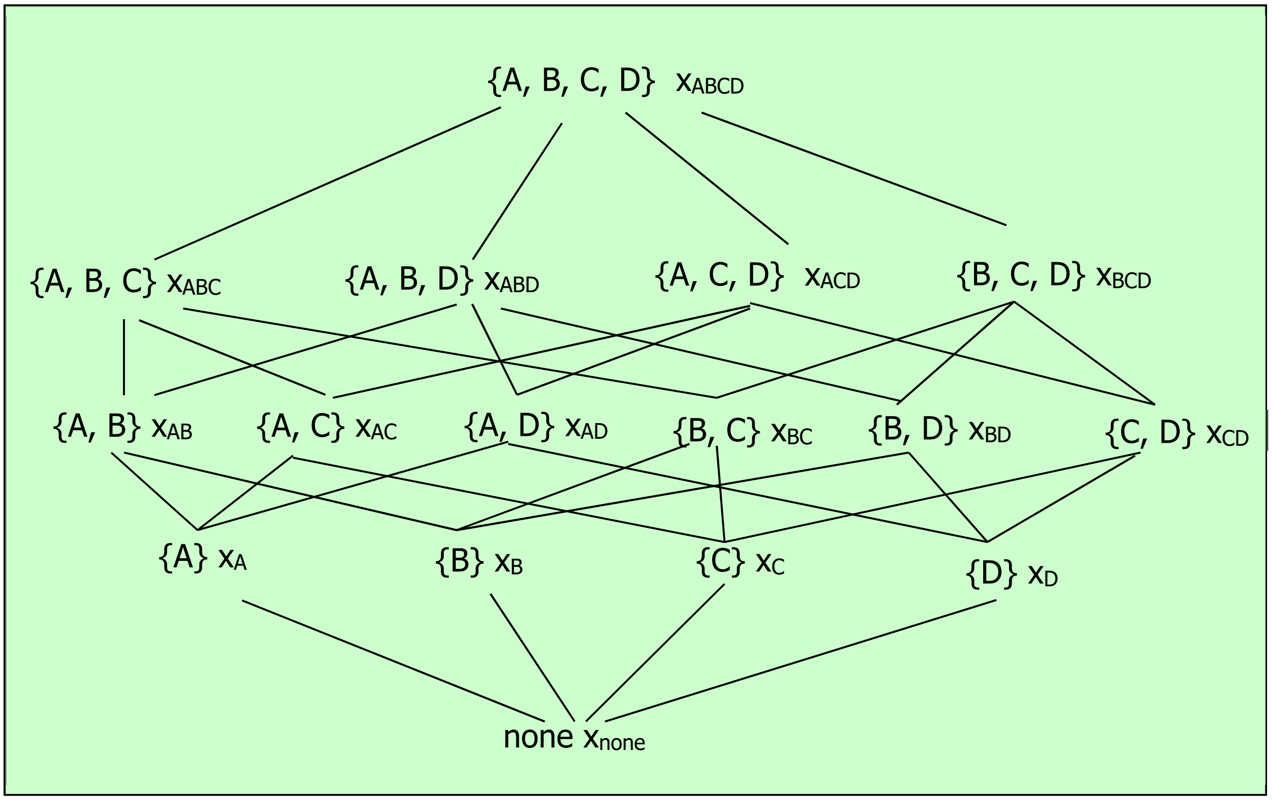
Then, we could answer a query like "for each customer, find the sum of the sale prices (SP) (i.e., find SUM(SP))". Suppose the total number of records in the output of this query is 0.1M. In the class, we learnt that we represent this query and its output by "c 0.1M". We also learnt how to derive (or obtain) the output of query "c" from the output of query "sc". We also learnt how to construct the following graph (or figure) due to this derivation. Suppose that we materialize the outputs of all queries.



(a) In this question, consider another scenario that we are given the following transactions. Suppose that the support threshold is equal to 1.

A	B	C	D
1	1	0	1
1	1	0	1
0	0	1	0
1	1	1	1
1	0	1	0

Now, we would like to answer a query like "finding all frequent itemsets such that each frequent itemset contain at least one item from set {A, B, D}". Two examples of this output are {A, B} and {B, C} but {C} is not in this output. Note that the frequency of each frequent itemset is not needed in the output of this query. Suppose that the total number of frequent itemsets in this output is a number x (to be found by you in this question). Then, similar to what we learnt in the class, we represent this query and its output by "{A, B, D} x". Based on the concept we learnt in class, we could derive (or obtain) the output of query {A, D} from the output of query {A, B, D}. We could construct the following graph due to the derivation. Note that each variable x (with a subscript) in the following corresponds to a number to be found by you in this question. Suppose that we materialize the outputs of all queries.



(i) Please state all frequent itemsets (i.e., query "{A, B, C, D}").
You are not required to give the frequency of each itemset.

(ii) Please find the value of each variable x (e.g., x_{ABCD}, x_{ABC} and x_{none}).

(iii) Assume that we do not consider "none x_{none}". **Now**, suppose that 4 views (instead of all views) are to be materialized (other than the top view). Apply the greedy algorithm and find the resulting views. (Note: For each iteration/selection in the greedy algorithm, if there are ties, just pick the query in the lexicographical order (or alphabetical order) (e.g., {A, B} is ordered before {A, C} in this ordering)).

(b) In Part (a), we did not consider any Apriori property to construct the graph. Now, we want to use the Apriori property learnt in class to "reduce" the total size of the storage.

(i) We know that the Apriori property is in the form of "if <itemset 1> is frequent, then <itemset 2> is frequent." What is the relationship between <itemset 1> and <itemset 2>?

(ii) Consider this transactional dataset **only**. Due to this Apriori property, we do not need to store all frequent itemsets for each query. We could store fewer frequent itemsets for each query. For example, before we use the Apriori property, for a particular query, we store the set S of all frequent itemests for this query, and when we answer this query, we obtain set S and return this as the output. However, after we use this Apriori property, for this query, we store a subset S' of S for this query, and when we answer this query, we obtain set S', derive S'' based on S' and return this derived set S'' as the output (where S'' is equal to S). In the above, x (with a subscript) denotes the total number of all frequent itemsets for each query. Let y (with a subscript) denote the total smallest possible number of all frequent itemsets "stored" for each query. Please give the value of each variable y (e.g., y_{ABCD}, y_{ABC} and y_{none}).

Q10 (20 Marks)

Consider a classification problem where the target attribute contains two possible values, "Yes" and "No".

We are given a training dataset. We generate a classifier based on this dataset. We find that there are exactly 30 tuples which target attribute values are predicted as "Yes", and there are exactly 20 tuples which target attribute values are predicted as "No". We also know that the f-measure of this classifier is 16/22 and the accuracy of this classifier is 0.70 (out of 1.0) (i.e., 70%).

(a) Is it a must that we can find the number of false positives? If yes, please write down the number of false positives. Otherwise, please elaborate why we cannot find it.
(b) Is it a must that we can find the precision of this classifier? If yes, please write down the precision of this classifier. Otherwise, please elaborate why we cannot find it.
(c) Is it a must that we can find the recall of this classifier? If yes, please write down the recall of this classifier. Otherwise, please elaborate why we cannot find it.
(d) Is it a must that we can find the specificity of this classifier? If yes, please write down the specificity of this classifier. Otherwise, please elaborate why we cannot find it.
(e) Is it a must that we can find the decile-wise lift chart of this classifier? If yes, please write down the decile-wise lift chart of this classifier. Otherwise, please elaborate why we cannot find it.

End of Paper

COMP1942 Exploring and Visualizing Data (Spring Semester 2019)
Final Examination (Answer Sheet)
Date: 22 May, 2019 (Wed)
Time: 12:30-15:00
Duration: 2 hours 30 minutes

Student ID:__________________ Student Name:__________________________________

Seat No. :__________________

Instructions:
(1) Please answer **all** questions in this paper.
(2) You can use a calculator.

Answer Sheet

Question	Full Mark	Mark
Q1	20	
Q2	20	
Q3	20	
Q4	20	
Q5	20	
Q6	20	
Q7	20	
Q8	20	
Q9	20	
Q10	20	
Total	200	

Q1 (20 Marks)
(a)

The reason why we cannot simply output C as the final output is that not all itemsets in C are frequent (i.e., not all itemsets in C can be in the final output).

Let us use the size-2 itemset generation for illustration.

Originally, L_1 = {R, A, Y, N}
After the counting step and the pruning step, we have
C_2 = {RA, RY, RN, AY, AN, YN}
Not all itemsets in C_2 have frequency at least 2.
E.g., RA is not frequent since its frequency is equal to 1. Thus, RA is not in the output.

(b)

Itemset	Frequency
{a, d, f, k}	2
{a, d, f}	2
{a, d, k}	3
{a, f, k}	2
{d, f, k}	2
{a, d}	3
{a, f}	2
{a, k}	3
{d, f}	2
{d, k}	5
{f, k}	2
a	3
d	8
f	2
k	5

Q2 (20 Marks)

P(A = Y)
= P(A=Y | S = Y)P(S=Y) + P(A=Y | S = N)P(S=N)
= 0.35 x 0.4 + 0.25 x 0.6
= 0.29

P(LC = N, P = Y)
= P(LC = N, P = Y | A=Y)P(A=Y) + P(LC = N, P = Y | A=N)P(A=N)
= P(LC = N | A=Y) P(P = Y | A=Y)P(A=Y) + P(LC = N | A=N) P(P = Y | A=N)P(A=N)
= 0.15 x 0.75 x 0.29 + 0.8 x 0.3 x 0.71
= 0.2030

P(LC = N, P = Y | S = Y)
= P(LC = N, P = Y | A= Y, S = Y) P(A= Y | S=Y) + P(LC = N, P = Y | A= N, S = Y) P(A= N | S=Y)
= P(LC = N, P = Y | A= Y) x 0.35 + P(LC = N, P = Y | A= N) x 0.65
= P(LC = N | A= Y) P(P = Y | A= Y) x 0.35 + P(LC = N | A= N) P(P = Y | A= N) x 0.65
= 0.15 x 0.75 x 0.35 + 0.8 x 0.3 x 0.65
= 0.1954

P(S = Yes | LC = N, P = Y)

$$= \frac{P(LC = N, P = Y|S = Y)P(S = Y)}{P(LC = N, P = Y)}$$

$$= \frac{0.1954 \times 0.4}{0.2030}$$

= 0.3850

He is less likely to smoke.

Q2 (Continued)

Q3 (20 Marks)
(a)

No.

P(Insurance = Yes) = 1/2
P(Insurance = No) = 1/2

P(Insurance = Yes | A = 0) = 3/4
P(Insurance = No | A = 0) = 1/4

P(Insurance = Yes | A = 1) = 3/4
P(Insurance = No | A = 1) = 1/4

P(A = 0) = 1/2
P(A = 1) = 1/2

P(Insurance = Yes | B = 0) = 1
P(Insurance = No | B = 0) = 0

P(Insurance = Yes | B = 1) = 1/3
P(Insurance = No | B = 1) = 2/3

P(B = 0) = 1/8
P(B = 1) = 7/8

Q3 (Continued)

Consider ID3.
Info(T) = - 1/2 log 1/2 - 1/2 log 1/2 = 1

Consider attribute A.
Info(T_0) = -3/4 log 3/4 - 1/4 log 1/4 = 0.8113
Info(T_1) = -3/4 log 3/4 - 1/4 log 1/4 = 0.8113
Info(A, T) = 1/2 Info(T_0) + 1/2 Info(T_1) = 0.8113
Gain(A, T) = Info(T) - Info(A, T) = 1 - 0.8113 = 0.1887

Consider attribute B.
Info(T_0) = -1 log 1 - 0 log 0 = 0
Info(T_1) = -1/3 log 1/3 - 2/3 log 2/3 = 0.9183
Info(B, T) = 1/8 Info(T_0) + 7/8 Info(T_1) = 0.8035
Gain(b, T) = Info(T) - Info(B, T) = 1 - 0.8035 = 0.1965

Here, Gain(A, T) < Gain(B, T)

Under ID3, Imp-ID3(A) = Gain(A, T)
and Imp-ID3(B) = Gain(B, T).
Thus, we have "Imp-ID3(A) < Imp-ID3(B)".

Consider CART.
Info(T) = 1 - $(1/2)^2$ - $(1/2)^2$ = 1/2

Consider attribute A.
Info(T_0) = 1 - $(3/4)^2$ - $(1/4)^2$ = 0.375
Info(T_1) = 1 - $(3/4)^2$ - $(1/4)^2$ = 0.375
Info(A, T) = 1/2 Info(T_0) + 1/2 Info(T_1) = 0.375
Gain(A, T) = Info(T) - Info(A, T) = 1/2 - 0.375 = 0.125

Consider attirbute B.
Info(T_0) = 1 - 1^2 - 0^2 = 0
Info(T_1) = 1 - $(1/3)^2$ - $(2/3)^2$ = 0.444
Info(B, T) = 1/8 Info(T_0) + 7/8 Info(T_1) = 0.3885
Gain(B, T) = Info(T) - Info(B, T) = 1/2 - 0.3885 = 0.1115

Here, Gain(A, T) > Gain(B, T).

Under CART, Imp-CART(A) = Gain(A, T)
and Imp-CART(B) = Gain(B, T).
Thus, we have "Imp-CART(A) > Imp-CART(B)".

In conclusion, it is possible that
"Imp-CART(A) > Imp-CART(B)" but "Imp-ID3(A) < Imp-ID3(B)".

Q3 (Continued)

Q3 (Continued)

(b) (i)

2

(ii)

5

(iii)

Cluster 1:
mean = (22, 95.6)
Cluster 2:
mean = (91.8, 42.2)

(iv)

Cluster 1:
mean = (20, 95)
Cluster 2:
mean = (89, 42)

(v)

Yes. This is because the only "uncertainty" of the output of the k-means clustering is the randomly chosen means. In the current setting, since we set the seed to 12345, no matter which machine we execute XLMiner, the randomly chosen means obtained from one machine are exactly equal to those obtained from another machine.

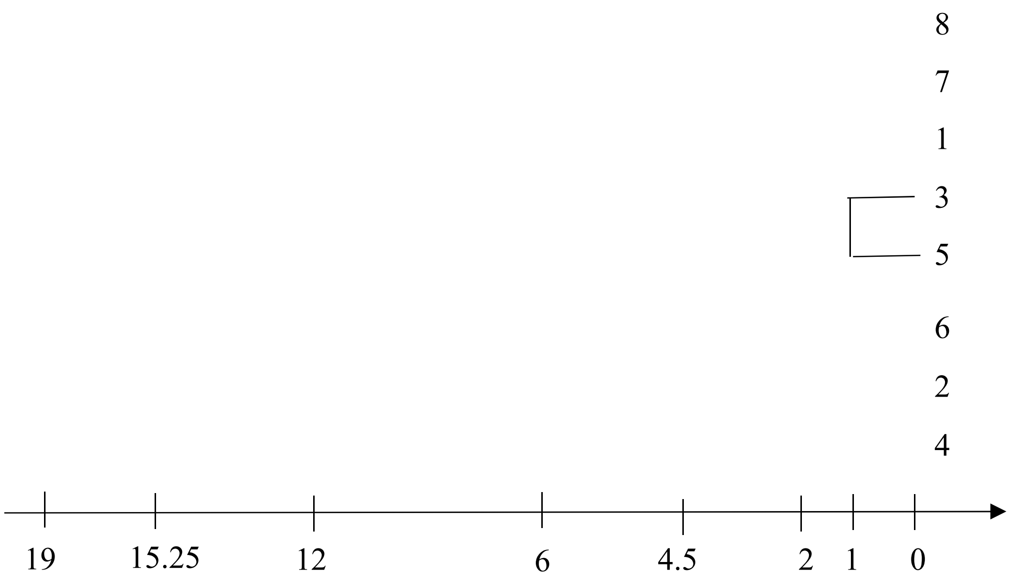
Q4 (20 Marks)

(a)

The scenario is when we are more interested in more recent data points compared with less recent data points.

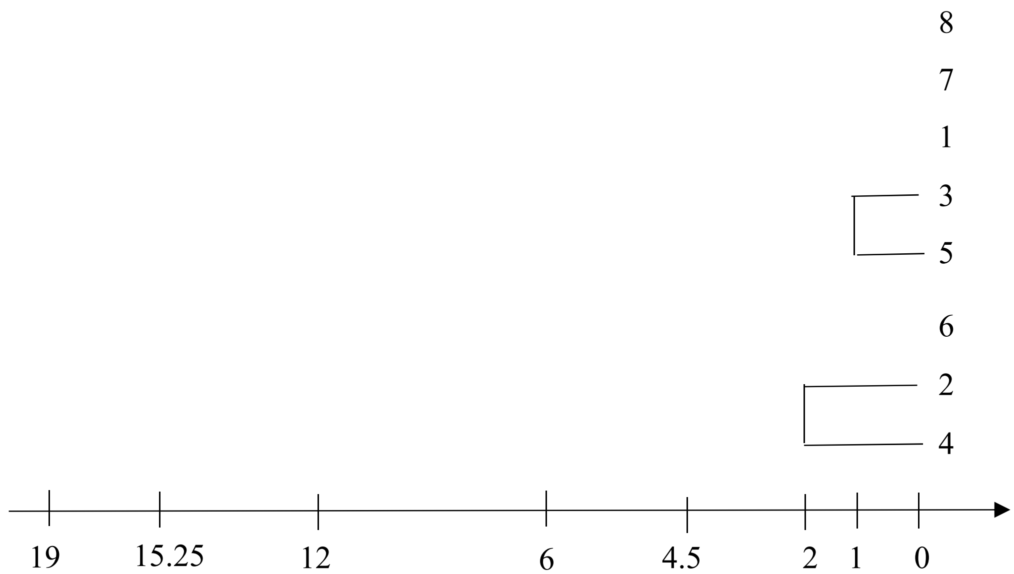
(b)

$$
\begin{array}{c|c}
 & \begin{array}{cccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{array} \\
\begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \end{array} &
\begin{pmatrix}
0 & & & & & & & \\
11 & 0 & & & & & & \\
5 & 13 & 0 & & & & & \\
12 & 2 & 14 & 0 & & & & \\
7 & 17 & 1 & 18 & 0 & & & \\
13 & 4 & 15 & 5 & 20 & 0 & & \\
9 & 15 & 12 & 16 & 15 & 19 & 0 & \\
11 & 20 & 12 & 21 & 17 & 22 & 30 & 0
\end{pmatrix}
\end{array}
$$

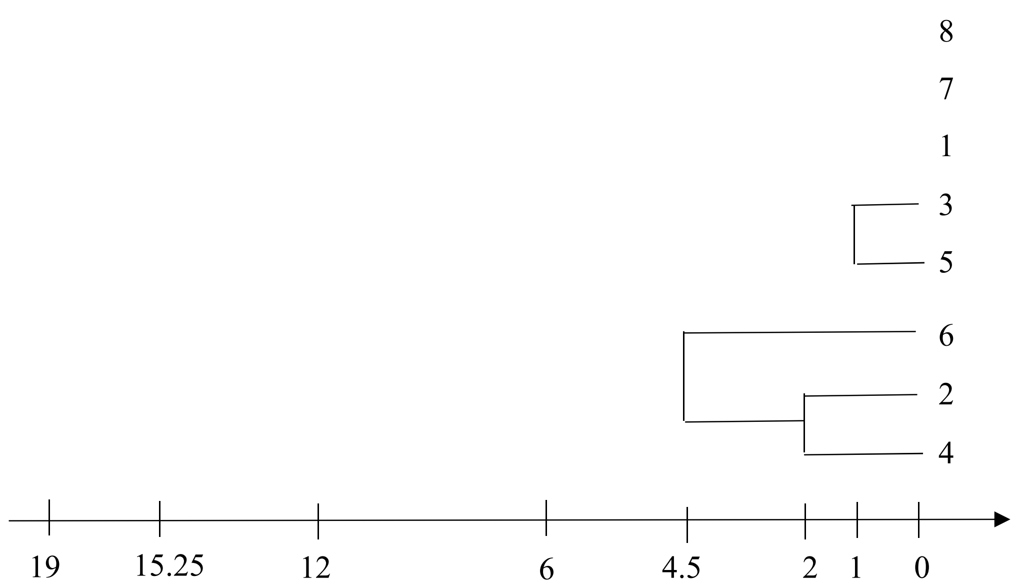


Q4 (Continued)

$$
\begin{array}{c|ccccccc}
 & 1 & 2 & (35) & 4 & 6 & 7 & 8 \\
\hline
1 & 0 & & & & & & \\
2 & 11 & 0 & & & & & \\
(35) & 6 & 15 & 0 & & & & \\
4 & 12 & 2 & 16 & 0 & & & \\
6 & 13 & 4 & 17.5 & 5 & 0 & & \\
7 & 9 & 15 & 13.5 & 16 & 19 & 0 & \\
8 & 11 & 20 & 14.5 & 21 & 22 & 30 & 0
\end{array}
$$

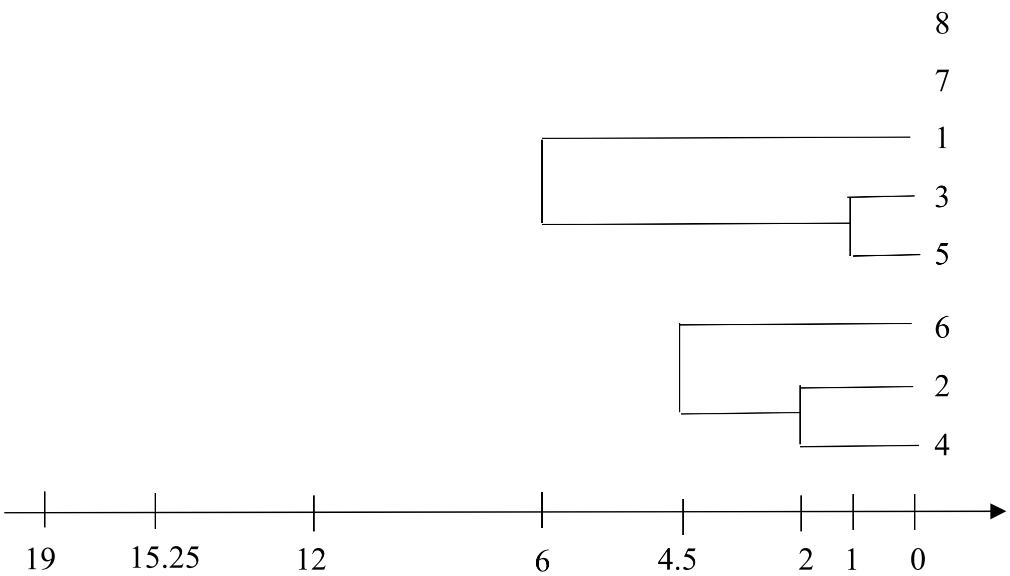


$$
\begin{array}{c|cccccc}
 & 1 & (24) & (35) & 6 & 7 & 8 \\
\hline
1 & 0 & & & & & \\
(24) & 11.5 & 0 & & & & \\
(35) & 6 & 15.5 & 0 & & & \\
6 & 13 & 4.5 & 17.5 & 0 & & \\
7 & 9 & 15.5 & 13.5 & 19 & 0 & \\
8 & 11 & 20.5 & 14.5 & 22 & 30 & 0
\end{array}
$$

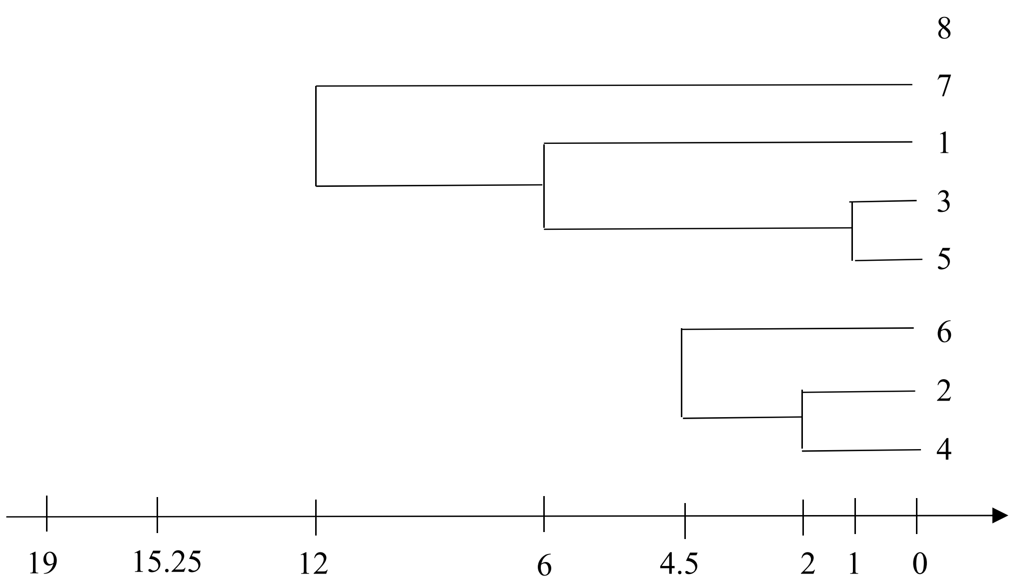


Q4 (Continued)

$$\begin{array}{c|c} & \begin{matrix} 1 & (246) & (35) & 7 & 8 \end{matrix} \\ \begin{matrix} 1 \\ (246) \\ (35) \\ 7 \\ 8 \end{matrix} & \begin{pmatrix} 0 & & & & \\ 12 & 0 & & & \\ 6 & 16.17 & 0 & & \\ 9 & 16.67 & 13.5 & 0 & \\ 11 & 21 & 14.5 & 30 & 0 \end{pmatrix} \end{array}$$

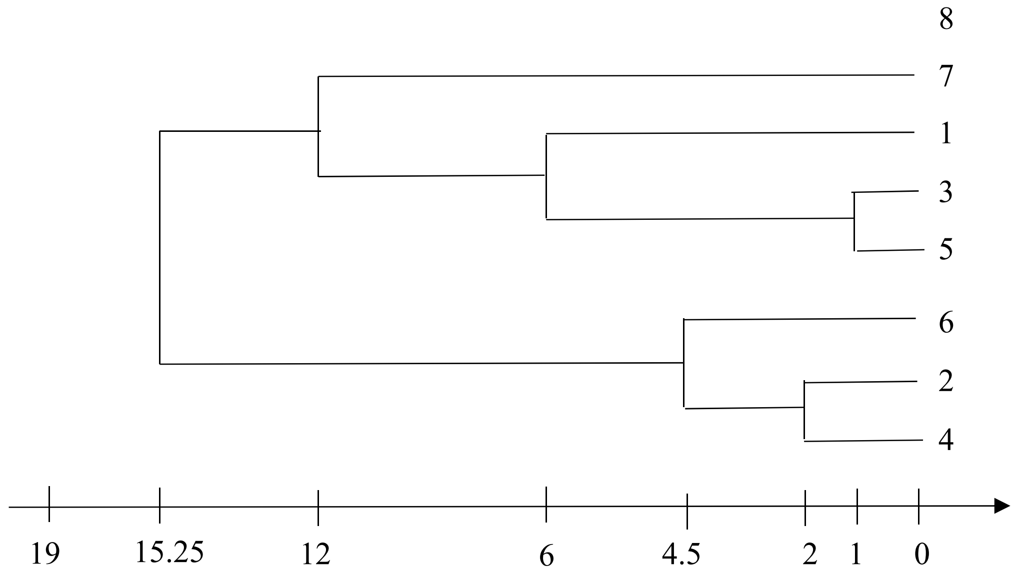


$$\begin{array}{c|c} & \begin{matrix} (135) & (246) & 7 & 8 \end{matrix} \\ \begin{matrix} (135) \\ (246) \\ 7 \\ 8 \end{matrix} & \begin{pmatrix} 0 & & & \\ 14.78 & 0 & & \\ 12 & 16.67 & 0 & \\ 13.33 & 21 & 30 & 0 \end{pmatrix} \end{array}$$

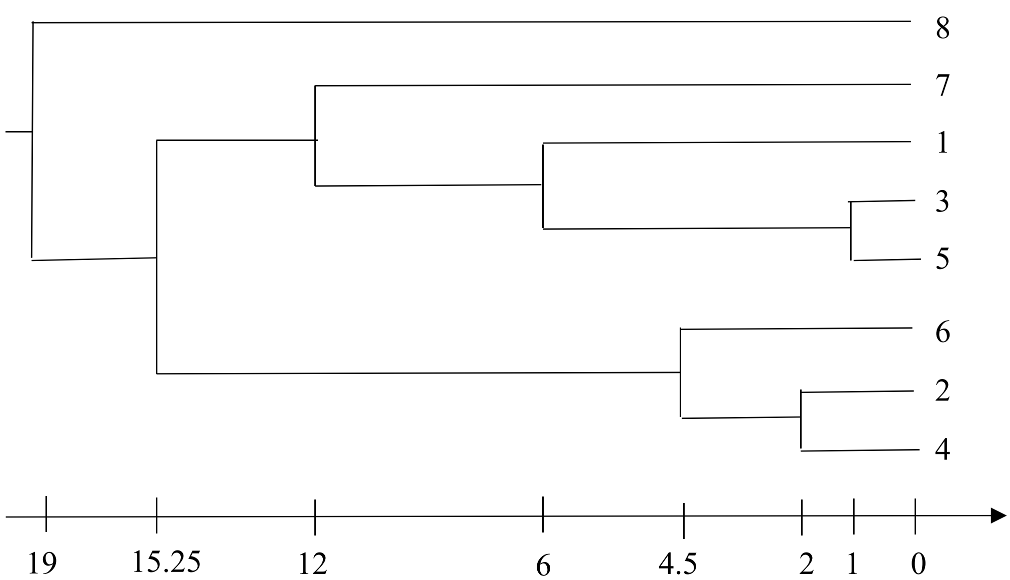


Q4 (Continued)

$$
\begin{array}{cc}
 & \begin{array}{ccc} (1357) & (246) & 8 \end{array} \\
\begin{array}{c} (1357) \\ (246) \\ 8 \end{array} & \begin{pmatrix} 0 & & \\ 15.25 & 0 & \\ 17.5 & 21 & 0 \end{pmatrix}
\end{array}
$$



$$
\begin{array}{cc}
 & \begin{array}{cc} (12345467) & 8 \end{array} \\
\begin{array}{c} (1234567) \\ 8 \end{array} & \begin{pmatrix} 0 & \\ 19 & 0 \end{pmatrix}
\end{array}
$$



Q4 (Continued)

Q5 (20 Marks)
(a)

Yes. It is possible.

gender = female
race = white
child = no

However, we could not know the exact value of the input attribute “married”.

(b)

We predict that the chemical will have a reaction when the temperature is equal to 10 and the weight of this chemical is equal to 4.
The 3 nearest neighbors are 8, 9 and 13.

Q6 (20 Marks)

(a) (i)

No. of Phones	No. of Laptops	Buy_NitendoSwtich
0	0	0
0	1	0
1	0	0
1	1	1

(ii)

Iteration 1:

$(x_1, x_2, y) = (0, 0, 0)$

net = $x_1 w_1 + x_2 w_2 + b$

= 0*0.1 + 0*0.1 + 0.1 = 0.1

b	w_1	w_2
0.1	0.1	0.1

y = 0.5250

$w_1 = w_1 + \alpha(d - y)x_1$

= 0.1 + 0.5*(0 – 0.5250) * 0

= 0.1

$w_2 = w_2 + \alpha(d - y)x_2$

= 0.1 + 0.5*(0 – 0.5250) * 0

= 0.1

$b = b + \alpha(d - y)$

= 0.1 + 0.5*(0 – 0.5250)

= -0.1625

Iteration 2:

$(x_1, x_2, y) = (0, 1, 0)$

net = $x_1 w_1 + x_2 w_2 + b$

= 0*0.1 + 1*0.1 + (-0.1625)

= -0.0625

b	w_1	w_2
-0.1625	0.1	0.1

y = 0.4844

$w_1 = w_1 + \alpha(d - y)x_1$

= 0.1 + 0.5*(0 – 0.4844) * 0

= 0.1

$w_2 = w_2 + \alpha(d - y)x_2$

= 0.1 + 0.5*(0 – 0.4844) * 1

= -0.1422

$b = b + \alpha(d - y)$

= -0.1625 + 0.5*(0 – 0.4844)

= -0.4047

Q6 (Continued)

Iteration 3:

$(x_1, x_2, y) = (1, 0, 0)$

net = $x_1w_1 + x_2w_2 + b$

= 1*0.1 + 0*(-0.1422) + (-0.4047)

= -0.3047

y = 0.4244

b	w_1	w_2
-0.4047	0.1	-0.1422

$w_1 = w_1 + \alpha(d - y)x_1$

= 0.1 + 0.5*(0 – 0.4244) * 1

= -0.1122

$w_2 = w_2 + \alpha(d - y)x_2$

= -0.1422 + 0.5*(0 – 0.4244) * 0

= -0.1422

$b = b + \alpha(d - y)$

= -0.4047 + 0.5*(0 – 0.4244)

= -0.6169

Iteration 4:

$(x_1, x_2, y) = (1, 1, 1)$

net = $x_1w_1 + x_2w_2 + b$

= 1*(-0.1122) + 1*(-0.1422)+(-0.6169)

= -0.8713

y = 0.2950

b	w_1	w_2
-0.6169	-0.1122	-0.1422

$w_1 = w_1 + \alpha(d - y)x_1$

= -0.1122 + 0.5*(1 – 0.2950) * 1

= 0.2403

$w_2 = w_2 + \alpha(d - y)x_2$

= -0.1422 + 0.5*(1 – 0.2950) * 1

= 0.2103

$b = b + \alpha(d - y)$

= -0.6169 + 0.5*(1 – 0.2950)

= -0.2644

Iteration 5:

$(x_1, x_2, y) = (0, 0, 0)$

net = $x_1w_1 + x_2w_2 + b$

= 0*0.2403 + 0*0.2103 +(-0.2644)

= -0.2644

y = 0.4343

b	w_1	w_2
-0.2644	0.2403	0.2103

Q6 (Continued)

$w_1 = w_1 + \alpha(d - y)x_1$
$= 0.2403 + 0.5*(0 – 0.4343) * 0$
$= 0.2403$

$w_2 = w_2 + \alpha(d - y)x_2$
$= 0.2103 + 0.5*(0 – 0.4343) * 0$
$= 0.2103$

$b = b + \alpha(d - y)$
$= -0.2644 + 0.5*(0 – 0.4343)$
$= -0.4816$

b	w_1	w_2
-0.4816	0.2403	0.2103

Q6 (Continued)

Q7 (20 Marks)
(a)

mean vector = $\begin{pmatrix} \frac{8+10+7+11}{4} \\ \frac{8+10+11+7}{4} \end{pmatrix} = \begin{pmatrix} 9 \\ 9 \end{pmatrix}$

For data (8, 8), difference from mean vector = $\begin{pmatrix} 8-9 \\ 8-9 \end{pmatrix} = \begin{pmatrix} -1 \\ -1 \end{pmatrix}$

For data (10, 10), difference from mean vector $\begin{pmatrix} 10-9 \\ 10-9 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$

For data (7, 11), difference from mean vector $\begin{pmatrix} 7-9 \\ 11-9 \end{pmatrix} = \begin{pmatrix} -2 \\ 2 \end{pmatrix}$

For data (11, 7), difference from mean vector =$\begin{pmatrix} 11-9 \\ 7-9 \end{pmatrix} = \begin{pmatrix} 2 \\ -2 \end{pmatrix}$

$$Y = \begin{pmatrix} -1 & 1 & -2 & 2 \\ -1 & 1 & 2 & -2 \end{pmatrix}$$

$$\Sigma = \frac{1}{4} YY^T = \frac{1}{4}\begin{pmatrix} -1 & 1 & -2 & 2 \\ -1 & 1 & 2 & -2 \end{pmatrix}\begin{pmatrix} -1 & -1 \\ 1 & 1 \\ -2 & 2 \\ 2 & -2 \end{pmatrix}$$

$$= \frac{1}{4}\begin{pmatrix} 10 & -6 \\ -6 & 10 \end{pmatrix}$$

$$= \begin{pmatrix} \frac{5}{2} & -\frac{3}{2} \\ -\frac{3}{2} & \frac{5}{2} \end{pmatrix}$$

$$\begin{vmatrix} \frac{5}{2}-\lambda & -\frac{3}{2} \\ -\frac{3}{2} & \frac{5}{2}-\lambda \end{vmatrix} = 0 \quad \Longrightarrow \quad (\frac{5}{2}-\lambda)^2 - (-\frac{3}{2})^2 = 0 \quad \Longrightarrow \quad \lambda = 4 \quad or \quad \lambda = 1$$

when $\lambda = 4$,

$$\begin{pmatrix} \frac{5}{2}-4 & -\frac{3}{2} \\ -\frac{3}{2} & \frac{5}{2}-4 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \Rightarrow \begin{pmatrix} -\frac{3}{2} & -\frac{3}{2} \\ -\frac{3}{2} & -\frac{3}{2} \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \Rightarrow \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \Rightarrow x_1 + x_2 = 0$$

Q7(a) (Continued)

We choose the eigenvector of unit length: $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} \end{pmatrix}$.

When $\lambda = 1$,

$$\begin{pmatrix} \frac{5}{2}-1 & -\frac{3}{2} \\ -\frac{3}{2} & \frac{5}{2}-1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \Rightarrow \begin{pmatrix} \frac{3}{2} & -\frac{3}{2} \\ -\frac{3}{2} & \frac{3}{2} \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \Rightarrow \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \Rightarrow x_1 - x_2 = 0$$

We choose the eigenvector of unit length: $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{pmatrix}$.

Thus, $\Phi = \begin{pmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix}$, $Y = \Phi^T X = \begin{pmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix} X$.

For data (8, 8), $Y = \begin{pmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix}\begin{pmatrix} 8 \\ 8 \end{pmatrix} = \begin{pmatrix} 0 \\ 11.31 \end{pmatrix}$

For data (10, 10), $Y = \begin{pmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix}\begin{pmatrix} 10 \\ 10 \end{pmatrix} = \begin{pmatrix} 0 \\ 14.14 \end{pmatrix}$

For data (7, 11), $Y = \begin{pmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix}\begin{pmatrix} 7 \\ 11 \end{pmatrix} = \begin{pmatrix} -2.83 \\ 12.73 \end{pmatrix}$

For data (11, 7), $Y = \begin{pmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix}\begin{pmatrix} 11 \\ 7 \end{pmatrix} = \begin{pmatrix} 2.83 \\ 12.73 \end{pmatrix}$

The mean vector of the above transformed data points is $\begin{pmatrix} \frac{0+0+(-2.83)+2.83}{4} \\ \frac{11.31+14.14+12.73+12.73}{4} \end{pmatrix} = \begin{pmatrix} 0 \\ 12.73 \end{pmatrix}$

The final transformed data points are:

Q7(a) (Continued)

For data (8, 8), final transformed vector = $\begin{pmatrix} 0-0 \\ 11.31-12.73 \end{pmatrix} = \begin{pmatrix} 0 \\ -1.42 \end{pmatrix}$

For data (10, 10), final transformed vector =$\begin{pmatrix} 0-0 \\ 14.14-12.73 \end{pmatrix} = \begin{pmatrix} 0 \\ 1.41 \end{pmatrix}$

For data (7, 11), final transformed vector =$\begin{pmatrix} -2.83-0 \\ 12.73-12.73 \end{pmatrix} = \begin{pmatrix} -2.83 \\ 0 \end{pmatrix}$

For data (11, 7), final transformed vector =$\begin{pmatrix} 2.83-0 \\ 12.73-12.73 \end{pmatrix} = \begin{pmatrix} 2.83 \\ 0 \end{pmatrix}$

Thus, (8, 8) is reduced to (0);
(10, 10) is reduced to (0);
(7, 11) is reduced to (-2.83);
(11, 7) is reduced to (2.83).

(Note: Another possible answer is
(8, 8) is reduced to (0);
(10, 10) is reduced to (0);
(7, 11) is reduced to (2.83);
(11, 7) is reduced to (-2.83).
This is because the eigenvectors used in this case are:

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{pmatrix}.$$
)

Q7(a) (Continued)

Q7(a) (Continued)

Q7 (Continued)

(b)

Yes. The transformed data points are:

(8 – d, 8 – d) is reduced to (0);
(10 – d, 10 – d) is reduced to (0);
(7 – d, 11 – d) is reduced to (-2.83);
(11 – d, 7 – d) is reduced to (2.83).

(c)

Yes. The transformed data points are:

(8c, 8c) is reduced to (0);
(10c, 10c) is reduced to (0);
(7c, 11c) is reduced to (-2.83c);
(11c, 7c) is reduced to (2.83c).

Q8 (20 Marks)

(a)

Yes.

$$\begin{array}{c c}
 & \begin{array}{cccc} p & q & r & s \end{array} \\
\begin{array}{c} p \\ q \\ r \\ s \end{array} & \begin{pmatrix} 0.33 & 0.5 & 0 & 0 \\ 0.33 & 0 & 0.5 & 1 \\ 0.33 & 0 & 0.5 & 0 \\ 0 & 0.5 & 0 & 0 \end{pmatrix}
\end{array}$$

(b) (i)

w_1, w_6, w_7, w_8, w_{12}

(ii)

w_1, w_2, w_3, w_6, w_7, w_8, w_{10}, w_{12}

Q9 (20 Marks)

(a) (i)

{A, B, C, D,
AB, AC, AD, BC, BD, CD,
ABC, ABD, ACD, BCD,
ABCD}

(ii)

Variable	Value	Variable	Value
x_{ABCD}	15	x_{BC}	12
x_{ABC}	14	x_{BD}	12
x_{ABD}	14	x_{CD}	12
x_{ACD}	14	x_A	8
x_{BCD}	14	x_B	8
x_{AB}	12	x_C	8
x_{AC}	12	x_D	8
x_{AD}	12	x_{none}	0

Q9 (Continued)

(iii)

	First Choice	Second Choice	Third Choice	Fourth Choice
ABC	1x7 = 7	1x4 = 4	1x3 = 3	1x3 = 3
ABD	1x7 = 7	1x4 = 4	1x3 = 3	1x3 = 3
ACD	1x7 = 7	1x6 = 6	1x3 = 3	1x3 = 3
BCD	1x7 = 7	1x6 = 6	1x3 = 3	1x3 = 3
AB	3x3 = 9			
AC	3x3 = 9	3x2 = 6	3x1 = 3	3x1 = 3
AD	3x3 = 9	3x2 = 6	3x1 = 3	3x1 = 3
BC	3x3 = 9	3x2 = 6	3x1 = 3	3x1 = 3
BD	3x3 = 9	3x2 = 6	3x1 = 3	3x1 = 3
CD	3x3 = 9	3x3 = 9		
A	7x1 = 7	4x1 = 4	4x1 = 4	
B	7x1 = 7	4x1 = 4	4x1 = 4	4x1 = 4
C	7x1 = 7	7x1 = 7	4x1 = 4	4x1 = 4
D	7x1 = 7	7x1 = 7	4x1 = 4	4x1 = 4

First choice = AB
Second choice = CD
Third choice = A
Fourth choice = B

Resulting views = {AB, CD, A, B}

Q9 (Continued)

(b) (i)

<itemset 2> is a subset of <itemset 1>

(ii)

Variable	Value	Variable	Value
y_{ABCD}	1	y_{BC}	1
y_{ABC}	1	y_{BD}	1
y_{ABD}	1	y_{CD}	1
y_{ACD}	1	y_{A}	1
y_{BCD}	1	y_{B}	1
y_{AB}	1	y_{C}	1
y_{AC}	1	y_{D}	1
y_{AD}	1	y_{none}	0

Q10 (20 Marks)

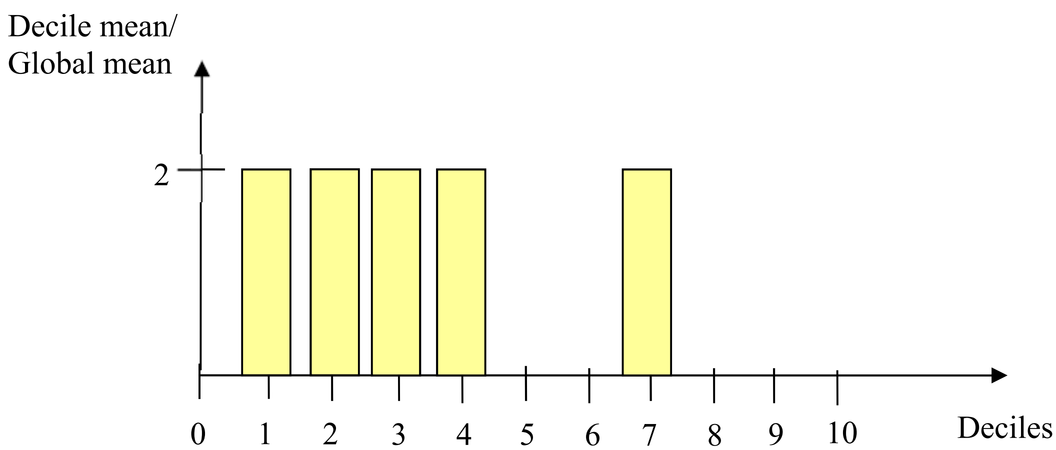
(a) Yes
no. of false positives = 10

(b) Yes
precision = 2/3 (= 20/30)

(c) Yes
recall = 4/5 (=20/25)

(d) Yes
specificity = 3/5 (=15/25)

(e)
Yes.
Decile-wise Lift Chart:



End of Answer Sheet