COMP1942 Exploring and Visualizing Data (Spring Semester 2017)
Final Examination (Answer Sheet)
Date: 27 May, 2017 (Sat)
Time: 16:30-19:30
Duration: 3 hours

Student ID:__________________ Student Name:________________________________

Seat No. :__________________

Instructions:
(1) Please answer **all** questions in **Part A** in this paper.
(2) You can **optionally** answer the bonus question in **Part B** in this paper. You can obtain additional marks for the bonus question if you answer it correctly.
(3) The total marks in Part A are 200.
(4) The total marks in Part B are 20.
(5) The total marks you can obtain in this exam are 200 only.
If you answer the bonus question in Part B correctly, you can obtain additional marks.
But, if the total marks you obtain from Part A and Part B are over 200, your marks will be truncated to 100 only.
(6) You can use a calculator.

# Answer Sheet

| Part | Question | Full Mark | Mark |
|---|---|---|---|
| A | Q1 | 20 | |
| | Q2 | 20 | |
| | Q3 | 20 | |
| | Q4 | 20 | |
| | Q5 | 20 | |
| | Q6 | 20 | |
| | Q7 | 20 | |
| | Q8 | 20 | |
| | Q9 | 20 | |
| | Q10 | 20 | |
| Total (Part A) | | 200 | |
| B | Q11 (OPTIONAL) | 20 | |
| Total (Parts A and B) | | 200 | |

# Part A (Compulsory Short Questions)

**Q1 (20 Marks)**

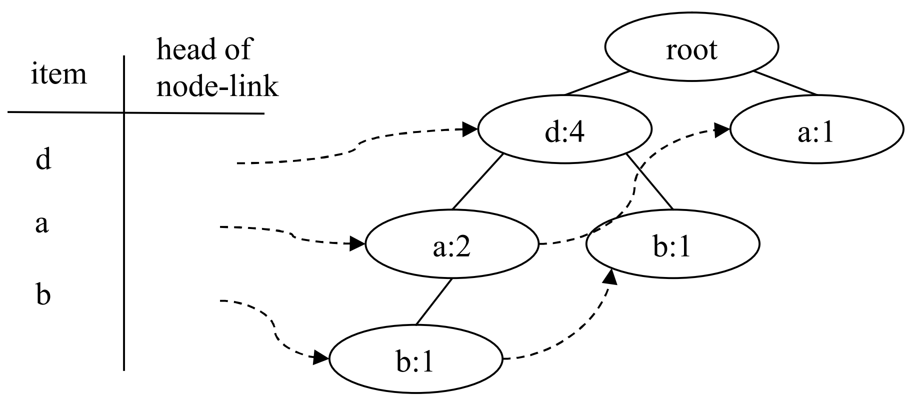
(a) (i)

Confidence of "d→a" = 2/4 = 50%

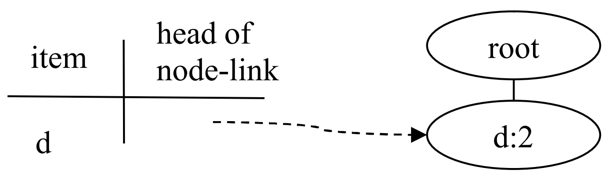
(ii)

Expected confidence of the consequent of the rule = 3/5 = 60%
Lift ratio of the rule = 50/60 = 0.833

(b) (i)




(ii)




(c)

(1) It is costly to handle a large number of candidate sets
(2) It is tedious to repeatedly scan the database and check the candidate patterns.

**Q2 (20 Marks)**
(a) (i)

C, D, E

(ii)

A, B

(iii)

The total number of tuples which have the number of children at most 1.

(iv)

Yes.

(b)
Yes.
The impurity measurement in C4.5 is shown as follows.

Gain(A, T) = (Info(T) – Info(A, T))/SplitInfo(A)

where A is an attribute and T is a set of records.

Let X be the target attribute
Let $n_A$ be the total number of values in attribute A.
Let $A_i$ be the i-th value in attribute A where i = 1, 2, …, $n_A$.

Note that

$\text{Info}(T) = -\Sigma_{v\in X}\, p(v) \log p(v)$
$\text{Info}(A_i) = -\Sigma_{v\in X}\, p(v|A_i) \log p(v|A_i)$ where i = 1, 2, …, $n_A$.
$\text{Info}(A, T) = \Sigma_{i\in[1, n_A]}\, p(A_i)\, \text{Info}(A_i)$
$\text{SplitInfo}(A) = -\Sigma_{v\in A}\, p(v) \log p(v)$

**Q2 (Continued)**

Consider base = 2.

Gain(A, T)
$= (\text{Info}(T) - \text{Info}(A, T))/\text{SplitInfo}(A)$
$= (-\Sigma_{v\in X}\, p(v) \log_2 p(v) - \Sigma_{i\in[1, n_A]}\, p(A_i)\, \text{Info}(A_i) )/[ -\Sigma_{v\in A}\, p(v) \log_2 p(v)]$
$= (-\Sigma_{v\in X}\, p(v) \log_2 p(v) - \Sigma_{i\in[1, n_A]}\, p(A_i)\, [-\Sigma_{v\in X}\, p(v|A_i) \log_2 p(v|A_i) ]\ )/[ -\Sigma_{v\in A}\, p(v) \log_2 p(v)]$ …….(*)

Consider base = b.
Similarly, we have the following derivation.

Gain(A, T)
$= (-\Sigma_{v\in X}\, p(v) \log_b p(v) - \Sigma_{i\in[1, n_A]}\, p(A_i)\, [-\Sigma_{v\in X}\, p(v|A_i) \log_b p(v|A_i) ]\ )/[ -\Sigma_{v\in A}\, p(v) \log_b p(v)]$

Since $\log_b x = \log_2 x/\log_2 b$ where x is a real number, we have the following.

Gain(A, T)
$=(-\Sigma_{v\in X}\, p(v) \log_2 p(v)/\log_2 b - \Sigma_{i\in[1, n_A]}\, p(A_i)\, [-\Sigma_{v\in X}\, p(v|A_i) \log_2 p(v|A_i)/\log_2 b\ ])/[ -\Sigma_{v\in A}\, p(v) \log_2 p(v)/\log_2 b\ ]$
$=(-\Sigma_{v\in X}\, p(v) \log_2 p(v) - \Sigma_{i\in[1, n_A]}\, p(A_i)\, [-\Sigma_{v\in X}\, p(v|A_i) \log_2 p(v|A_i) ])/[ -\Sigma_{v\in A}\, p(v) \log_2 p(v) ]$……(**)

Thus, equation (*) is equal to equation (**).
In other words, no matter what base we use, we have the same formulae at the end.
In C4.5, we must obtain the same tree. That is, $T_2$ is equal to $T_b$.

**Q3 (20 Marks)**

(a) (i)

Classification has a target attribute for prediction.
But, clustering has no target attribute.

(ii)

Both classification and clustering are used for data analysis.

(b)

$$
\begin{array}{c c}
 & \begin{array}{ccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{array} \\
\begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{array} &
\begin{pmatrix}
0 & & & & & & \\
10 & 0 & & & & & \\
7 & 7 & 0 & & & & \\
30 & 23 & 21 & 0 & & & \\
29 & 25 & 22 & 7 & 0 & & \\
38 & 34 & 31 & 10 & 11 & 0 & \\
42 & 36 & 36 & 13 & 17 & 9 & 0
\end{pmatrix}
\end{array}
$$

D(1, *) = 26.0

D(2, *) = 22.5

D(3, *) = 20.7

D(4, *) = 17.3

D(5, *) = 18.5

D(6, *) = 22.2

D(7, *) = 25.5

A = {1 }

B = {2, 3, 4, 5, 6, 7}

## Q3 (Continued)

$$
\begin{array}{c|c}
 & \begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{matrix} \\
\begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{matrix} &
\begin{pmatrix}
0 & & & & & & \\
10 & 0 & & & & & \\
7 & 7 & 0 & & & & \\
30 & 23 & 21 & 0 & & & \\
29 & 25 & 22 & 7 & 0 & & \\
38 & 34 & 31 & 10 & 11 & 0 & \\
42 & 36 & 36 & 13 & 17 & 9 & 0
\end{pmatrix}
\end{array}
$$

| | | |
|---|---|---|
| D(2, A) = 10 | D(2, B) = 25.0 | $\Delta_2$ = 15.0 |
| D(3, A) = 7 | D(3, B) = 23.4 | $\Delta_3$ = 16.4 |
| D(4, A) = 30 | D(4, B) = 14.8 | $\Delta_4$ = -15.2 |
| D(5, A) = 29 | D(5, B) = 16.4 | $\Delta_5$ = -12.6 |
| D(6, A) = 38 | D(6, B) = 19.0 | $\Delta_6$ = -19.0 |
| D(7, A) = 42 | D(7, B) = 22.2 | $\Delta_7$ = -19.8 |

A = {1, 3 }

B = {2, 4, 5, 6, 7}

$$
\begin{array}{c|c}
 & \begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{matrix} \\
\begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{matrix} &
\begin{pmatrix}
0 & & & & & & \\
10 & 0 & & & & & \\
7 & 7 & 0 & & & & \\
30 & 23 & 21 & 0 & & & \\
29 & 25 & 22 & 7 & 0 & & \\
38 & 34 & 31 & 10 & 11 & 0 & \\
42 & 36 & 36 & 13 & 17 & 9 & 0
\end{pmatrix}
\end{array}
$$

| | | |
|---|---|---|
| D(2, A) = 8.5 | D(2, B) = 29.5 | $\Delta_2$ = 21.0 |
| D(4, A) = 25.5 | D(4, B) = 13.2 | $\Delta_4$ = -12.3 |
| D(5, A) = 25.5 | D(5, B) = 15.0 | $\Delta_5$ = -10.5 |
| D(6, A) = 34.5 | D(6, B) = 16.0 | $\Delta_6$ = -18.5 |
| D(7, A) = 39.0 | D(7, B) = 18.6 | $\Delta_7$ = -20.4 |

A = {1, 3, 2 }

B = { 4, 5, 6, 7}

## Q3 (Continued)

$$\begin{array}{c|c} & \begin{array}{ccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{array} & \begin{pmatrix} 0 & & & & & & \\ 10 & 0 & & & & & \\ 7 & 7 & 0 & & & & \\ 30 & 23 & 21 & 0 & & & \\ 29 & 25 & 22 & 7 & 0 & & \\ 38 & 34 & 31 & 10 & 11 & 0 & \\ 42 & 36 & 36 & 13 & 17 & 9 & 0 \end{pmatrix} \end{array}$$

D(4, A) = 24.7 | D(4, B) = 10.0 | $\Delta_4$ = -14.7

D(5, A) = 25.3 | D(5, B) = 11.7 | $\Delta_5$ = -13.6

D(6, A) = 34.3 | D(6, B) = 10.0 | $\Delta_6$ = -24.3

D(7, A) = 38.0 | D(7, B) = 13.0 | $\Delta_7$ = -25.0

A = {1, 3, 2 }

B = {4, 5, 6, 7}

**Q4 (20 Marks)**

(a) P(LC=Yes) = $\sum_{x\in\{Yes,No\}}\sum_{y\in\{Yes,No\}} P(LC=Yes \mid FH=x, S=y)P(FH=x, S=y)$

$= 0.7\times 0.3\times 0.6 + 0.45\times 0.3\times 0.4 + 0.55\times 0.7\times 0.6 + 0.2\times 0.7\times 0.4$

$= 0.467$

$$P(LC=Yes \mid FH=Yes, Smoker=No, PR=Yes)$$
$$= \frac{P(PR=Yes \mid FH=Yes, Smoker=No, LC=Yes)}{P(PR=Yes \mid FH=Yes, Smoker=No)} P(LC=Yes \mid FH=Yes, Smoker=No)$$
$$= \frac{P(PR=Yes \mid LC=Yes)\times P(LC=Yes \mid FH=Yes, Smoker=No)}{\sum_{x\in\{Yes,No\}} P(PR=Yes \mid LC=x)P(LC=x \mid FH=Yes, Smoker=No)}$$
$$= \frac{0.85\times 0.45}{0.85\times 0.45 + 0.45\times 0.55}$$
$$= 0.607143$$
$$P(LC=No \mid FH=Yes, Smoker=No, PR=Yes) = 1 - 0.601743 = 0.392857$$

$\because P(LC=Yes \mid FH=Yes, Smoker=No, PR=Yes) > P(LC=No \mid FH=Yes, Smoker=No, PR=Yes)$ $\therefore$ It is more likely that the person is likely to have Lung Cancer.

(b) Disadvantages:

The Bayesian Belief network classifier requires a predefined knowledge about the network.

The Bayesian Belief Network classifier cannot work directly when the network contains cycles.

**Q5 (20 Marks)**
(a)

We need to transform the first three categorical attributes to the other corresponding three numeric attributes.

(b) (i)

Yes. The number is 3.

(ii)

Yes. The number is 7.

(iii)

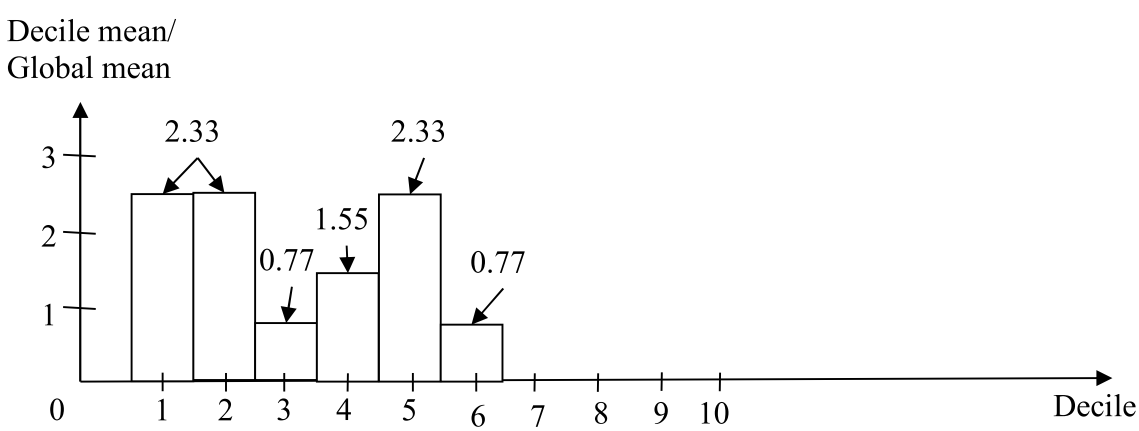
Yes. The number is 6.

(iv)

Yes. The number is 14.

(v)

Yes. The chart is shown as follows.



**Q6 (20 Marks)**

(a)

(1) Training Set
It is used to train or build a model

(2) Validation Set
It is used to fine-tune the model

(3) Test Set
It is used to evaluate the accuracy of the model on completely unseen data

(b) (i)

Yes.
The following table shows an example where the recall is 50% and the specificity is 100%.

| Target | Predicted |
|---|---|
| Yes | Yes |
| Yes | No |
| No | No |
| No | No |

**Q6 (Continued)**

(b)(ii)

No.
If the precision is equal to 50%, then there exists a tuple which has the actual value equal to “No” and the predicted value equal to “Yes”. This means that the specificity must be strictly smaller than 100%.

(iii)
No.
If the precision is equal to 100%, then there exists a tuple which has the actual value equal to "Yes" and the predicted value equal to "Yes". This means that the recall must not be equal to 0%.

(iv)

No.
If the recall is equal to 100%, then there exists a tuple which has the actual value equal to "Yes" and the predicted value equal to "Yes". This means that the precision must not be equal to 0%.

**Q7 (20 Marks)**
(a)

Solving this problem with the four data points, a:(7 + c, 7 + c), b:(9 + c, 9 + c), c:(6 + c, 10 + c), d:(10 + c, 6 + c), is equivalent to solving the problem with the other four data points, a:(7, 7), b:(9, 9), c:(6, 10), d:(10, 6).

mean vector = $\begin{pmatrix} \frac{7+9+6+10}{4} \\ \frac{7+9+10+6}{4} \end{pmatrix} = \begin{pmatrix} 8 \\ 8 \end{pmatrix}$

For data (7, 7), difference from mean vector = $\begin{pmatrix} 7-8 \\ 7-8 \end{pmatrix} = \begin{pmatrix} -1 \\ -1 \end{pmatrix}$

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

For data (6, 10), difference from mean vector = $\begin{pmatrix} 6-8 \\ 10-8 \end{pmatrix} = \begin{pmatrix} -2 \\ 2 \end{pmatrix}$

For data (10, 6), difference from mean vector = $\begin{pmatrix} 10-8 \\ 6-8 \end{pmatrix} = \begin{pmatrix} 2 \\ -2 \end{pmatrix}$

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

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

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

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

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

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 (7, 7), $Y = \begin{pmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix}\begin{pmatrix} 7 \\ 7 \end{pmatrix} = \begin{pmatrix} 0 \\ 9.90 \end{pmatrix}$

For data (9, 9), $Y = \begin{pmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix}\begin{pmatrix} 9 \\ 9 \end{pmatrix} = \begin{pmatrix} 0 \\ 12.73 \end{pmatrix}$

For data (6, 10), $Y = \begin{pmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix}\begin{pmatrix} 6 \\ 10 \end{pmatrix} = \begin{pmatrix} -2.83 \\ 11.31 \end{pmatrix}$

For data (10, 6), $Y = \begin{pmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix}\begin{pmatrix} 10 \\ 6 \end{pmatrix} = \begin{pmatrix} 2.83 \\ 11.31 \end{pmatrix}$

The mean vector of the above transformed data points is $\begin{pmatrix} \frac{0+0+(-2.83)+2.83}{4} \\ \frac{9.90+12.73+11.31+11.31}{4} \end{pmatrix} = \begin{pmatrix} 0 \\ 11.31 \end{pmatrix}$

The final transformed data points are:

**Q7(a) (Continued)**

For data (7, 7), final transformed vector = $\begin{pmatrix} 0-0 \\ 9.90-11.31 \end{pmatrix} = \begin{pmatrix} 0 \\ -1.41 \end{pmatrix}$

For data (9, 9), final transformed vector = $\begin{pmatrix} 0-0 \\ 12.73-11.31 \end{pmatrix} = \begin{pmatrix} 0 \\ 1.41 \end{pmatrix}$

For data (6, 10), final transformed vector = $\begin{pmatrix} -2.83-0 \\ 11.31-11.31 \end{pmatrix} = \begin{pmatrix} -2.83 \\ 0 \end{pmatrix}$

For data (10, 6), final transformed vector = $\begin{pmatrix} 2.83-0 \\ 11.31-11.31 \end{pmatrix} = \begin{pmatrix} 2.83 \\ 0 \end{pmatrix}$

Thus, (7, 7) is reduced to (0);
(9, 9) is reduced to (0);
(6, 10) is reduced to (-2.83);
(10, 6) is reduced to (2.83).

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

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

In conclusion,
(7 + c, 7 + c) is reduced to (0);
(9 + c, 9 + c) is reduced to (0);
(6 + c, 10 + c) is reduced to (-2.83);
(10 + c, 6 + c) is reduced to (2.83).

**Q7(a) (Continued)**

## Q7(a) (Continued)

**Q7 (Continued)**

(b)

Yes. The transformed data points are:

(7 – d, 7 – d) is reduced to (0);
(9 – d, 9 – d) is reduced to (0);
(6 – d, 10 – d) is reduced to (-2.83);
(10 – d, 6 – d) is reduced to (2.83).

(c)

Yes. The transformed data points are:

(7c, 7c) is reduced to (0);
(9c, 9c) is reduced to (0);
(6c, 10c) is reduced to (-2.83c);
(10c, 6c) is reduced to (2.83c).

**Q8 (20 Marks)**

| | First Choice | Second Choice | Third Choice |
|---|---|---|---|
| b | 50 x 5 = 250 | | |
| c | 25 x 5 = 125 | 25 x 2 = 50 | 25 x 1 = 25 |
| d | 80 x 2 = 160 | 30 x 2 = 60 | 30 x 2 = 60 |
| e | 70 x 3 = 210 | 20 x 3 = 60 | 20 + 20 + 10 = 50 |
| f | 60 x 2 = 120 | 60 + 10 = 70 | |
| g | 99 x 1 = 99 | 49 x 1 = 49 | 49 x 1 = 49 |
| h | 90 x 1 = 90 | 40 x 1 = 40 | 30 x 1 = 30 |

First choice = b
Second choice = f
Third choice = d

Resulting views = {b, f, d}

**Q9 (20 Marks)**

(a)

$P(X, Y \mid Z)$

$= \dfrac{P(X,Y,Z)}{P(Z)}$

$= \dfrac{P(X,Y,Z)}{P(Y,Z)} \times \dfrac{P(Y,Z)}{P(Z)}$

$= P(X|Y, Z) \times P(Y|Z)$

$= P(X|Z) \times P(Y|Z)$

(b)

<u>Iteration 1</u>:

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

$\text{net} = x_1 w_1 + x_2 w_2 + b$

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

y = 1 $\boxed{\text{Incorrect!}}$

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

$= 0.1 + 0.5*(0 - 1) * 0$

$= 0.1$

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

$= 0.1 + 0.5*(0 - 1) * 0$

$= 0.1$

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

$= 0.1 + 0.5*(0 - 1)$

$= -0.4$

| b | $w_1$ | $w_2$ |
|---|---|---|
| 0.1 | 0.1 | 0.1 |

**Q9 (Continued)**

Iteration 2:

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

$\text{net} = x_1 w_1 + x_2 w_2 + b = -0.3$

y = 0 $\boxed{\text{Correct!}}$

| b | $w_1$ | $w_2$ |
|---|---|---|
| -0.4 | 0.1 | 0.1 |

$w_1 = w_1 + \alpha(d - y)x_1$
$= 0.1 + 0.5*(0 - 0) * 0$
$= 0.1$
$w_2 = w_2 + \alpha(d - y)x_2$
$= 0.1 + 0.5*(0 - 0) * 1$
$= 0.1$
$b = b + \alpha(d - y)$
$= -0.4 + 0.5*(0 - 0)$
$= -0.4$

Iteration 3:

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

$\text{net} = x_1 w_1 + x_2 w_2 + b = -0.3$

y = 0 $\boxed{\text{Correct!}}$

| b | $w_1$ | $w_2$ |
|---|---|---|
| -0.4 | 0.1 | 0.1 |

$w_1 = w_1 + \alpha(d - y)x_1$
$= 0.1 + 0.5*(0 - 0) * 1$
$= 0.1$
$w_2 = w_2 + \alpha(d - y)x_2$
$= 0.1 + 0.5*(0 - 0) * 0$
$= 0.1$
$b = b + \alpha(d - y)$
$= -0.4 + 0.5*(0 - 0)$
$= -0.4$

Iteration 4:

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

$\text{net} = x_1 w_1 + x_2 w_2 + b = -0.2$

y = 0 $\boxed{\text{Incorrect!}}$

| b | $w_1$ | $w_2$ |
|---|---|---|
| -0.4 | 0.1 | 0.1 |

$w_1 = w_1 + \alpha(d - y)x_1$
$= 0.1 + 0.5*(1 - 0) * 1$
$= 0.6$
$w_2 = w_2 + \alpha(d - y)x_2$
$= 0.1 + 0.5*(1 - 0) * 1$
$= 0.6$
$b = b + \alpha(d - y)$
$= -0.4 + 0.5*(1 - 0)$
$= 0.1$

**Q9 (Continued)**

Iteration 5:

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

net $= x_1 w_1 + x_2 w_2 + b = 0.1$

y = 1 Incorrect!

| b | $w_1$ | $w_2$ |
|---|---|---|
| 0.1 | 0.6 | 0.6 |

$w_1 = w_1 + \alpha(d - y)x_1$
$= 0.6 + 0.5*(0 - 1) * 0$
$= 0.6$

$w_2 = w_2 + \alpha(d - y)x_2$
$= 0.6 + 0.5*(0 - 1) * 0$
$= 0.6$

$b = b + \alpha(d - y)$
$= 0.1 + 0.5*(0 - 1)$
$= -0.4$

| b | $w_1$ | $w_2$ |
|---|---|---|
| -0.4 | 0.6 | 0.6 |

## Q9 (Continued)

**Q10 (20 Marks)**

(a)

Adjacency matrix

$$\begin{array}{cc} & \begin{array}{ccc} x & y & z \end{array} \\ \begin{array}{c} x \\ y \\ z \end{array} & \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \end{pmatrix} \end{array}$$

(b)

Stochastic matrix

$$\begin{array}{cc} & \begin{array}{ccc} x & y & z \end{array} \\ \begin{array}{c} x \\ y \\ z \end{array} & \begin{pmatrix} 0 & 1 & 0.5 \\ 0.5 & 0 & 0.5 \\ 0.5 & 0 & 0 \end{pmatrix} \end{array}$$

(c)

1. Site x has to remove the link from site x to site y
2. Site x has to remove the link from site x to z
3. [Optional] Site x has to create a link from site x to itself

**Q10 (Continued)**

(d)

$r_n = 0.8\ M\ r_0 + c$

$$\begin{pmatrix} r_{n,1} \\ r_{n,2} \\ r_{n,3} \end{pmatrix} = 0.8 \begin{pmatrix} r_{0,1} \\ r_{0,2} \\ r_{0,3} \end{pmatrix} \begin{pmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{pmatrix} + \begin{pmatrix} 0.2 \\ 0.2 \\ 0.2 \end{pmatrix}$$

$$= \begin{pmatrix} 0.8(m_{11}r_{0,1} + m_{12}r_{0,2} + m_{13}r_{0,3}) + 0.2 \\ 0.8(m_{21}r_{0,1} + m_{22}r_{0,2} + m_{23}r_{0,3}) + 0.2 \\ 0.8(m_{31}r_{0,1} + m_{32}r_{0,2} + m_{33}r_{0,3}) + 0.2 \end{pmatrix}$$

Sum of the values in $r_n$

$= 0.8(m_{11}r_{0,1} + m_{12}r_{0,2} + m_{13}r_{0,3} + m_{21}r_{0,1} + m_{22}r_{0,2} + m_{23}r_{0,3} + m_{31}r_{0,1} + m_{32}r_{0,2} + m_{33}r_{0,3}) + (0.2+0.2+0.2)$

$= 0.8[(m_{11} + m_{21} + m_{31})r_{0,1} + (m_{12} + m_{22} + m_{32})r_{0,2} + (m_{13} + m_{23} + m_{33})r_{0,3}] + 0.2 \times 3$

$= 0.8(1 \cdot r_{0,1} + 1 \cdot r_{0,2} + 1 \cdot r_{0,3}) + 0.2 \times 3$

$= 0.8(r_{0,1} + r_{0,2} + r_{0,3}) + 0.2 \times 3$

$= 0.8 \times 3 + 0.2 \times 3$

$= 3$

# Part B (Bonus Question)

**Note:** The following bonus question is an **OPTIONAL** question. You can decide whether you will answer it or not.

**Q11 (20 Additional Marks)**

(a)

The Apriori property is:
If a cell S for set Y of k dimensions is dense, then for any proper subset Y' of Y, a cell S' for set Y' is dense where S' is a set of all grids from S each of which comes from one dimension of Y'.

Next, we show the correctness of this property.
We want to show that
"If a cell S for set Y of k dimensions is dense, then for any proper subset Y' of Y, a cell S' for set Y' is dense where S' is a set of all grids from S each of which comes from one dimension of Y'."

The proof is as follows.

Since S is dense, we know that d(S) ≥ 0.2.

Since Y' is a proper subset of Y, we deduce that S' involving a subset of dimensions compared with S includes the same set of points originally involved in S or more points not originally involved in S. Thus, d(S') ≥ d(S). Since d(S) ≥ 0.2, we deduce that d(S') ≥ 0.2.

Algorithm:

The idea is similar to the original Apriori Algorithm learnt in class.
Algorithm:

1. $L_1 \leftarrow$ a set of all possible dense cells
2. $k \leftarrow 2$
3. while $L_{k-1} \neq \varnothing$
   - $C_k \leftarrow$ Generate candidates from $L_{k-1}$ by Join Step and Prune Step discussed in class
   - Perform a counting step on $C_k$ (i.e., counting the density of each element in $C_k$) and obtain $L_k$ (i.e., keeping each element in $C_k$ with the density value at least 0.2)
4. Output $\bigcup_i L_i$

## Q11 (Continued)

**Q11 (Continued)**

**Q11 (Continued)**
(b)
The Apriori property is:
If a set Y of dimensions has good clustering, then any proper subset of S must have good clustering.
(i.e., For any two sets of dimensions, namely Y and Z, where $Z \subset Y$, if $H(Y) < \omega$, $H(Z) < \omega$.)

Next, we show the correctness of this property.
We want to show that
"for any two sets of dimensions, namely Y and Z, where $Z \subset Y$, if $H(Y) < \omega$, $H(Z) < \omega$."

Consider two cases.
**Case 1:** $|Y – Z| = 1$
In this case, let $X_i = Y\text{-}Z$ (where $X_i$ is a dimension). Note that $Y = Z \cup \{X_i\}$.

Given $c' \in A(Z)$ and $x \in [1, 4]$, we define $\alpha(c', x)$ to be the set of all grids in c' together with the grid "$X_i$: $[10(x-1)+1, 10(x-1)+10]$".

We know that $A(Z \cup \{X_i\}) = \{\alpha(c', x) \mid c' \in A(Z) \text{ and } x \in [1, 4]\}$

Consider

$$
\begin{aligned}
& H(Y) \\
&= H(Z \cup \{X_i\}) \\
&= -\sum_{c \in A(Z \cup \{X_i\})} d(c) \log d(c) \\
&= -\sum_{c' \in A(Z)} \sum_{x \in [1,4]} d(\alpha(c', x)) \log d(\alpha(c', x)) \\
&= -\sum_{c' \in A(Z)} \sum_{x \in [1,4]} d(\alpha(c', x)) \log[d(c') \cdot \frac{d(\alpha(c', x))}{d(c')}] \\
&= -\sum_{c' \in A(Z)} \sum_{x \in [1,4]} d(\alpha(c', x)) \log d(c') - \sum_{c' \in A(Z)} \sum_{x \in [1,4]} d(\alpha(c', x)) \log \frac{d(\alpha(c', x))}{d(c')} \\
&= -\sum_{c' \in A(Z)} d(c') \log d(c') - \sum_{c' \in A(Z)} \sum_{x \in [1,4]} d(\alpha(c', x)) \log \frac{d(\alpha(c', x))}{d(c')} \\
&= H(Z) - \sum_{c' \in A(Z)} \sum_{x \in [1,4]} d(\alpha(c', x)) \log \frac{d(\alpha(c', x))}{d(c')} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots(*)
\end{aligned}
$$

We know that $d(\alpha(c', x)) \le d(c')$ and thus $\log \frac{d(\alpha(c', x))}{d(c')} \le 0$.

Since $d(\alpha(c', x)) \ge 0$, we know that $\sum_{c' \in A(Z)} \sum_{x \in [1,4]} d(\alpha(c', x)) \log \frac{d(\alpha(c', x))}{d(c')} \le 0$

**Q11 (Continued)**

Thus, from (*), we deduce that H(Y) ≥ H(Z).

We know that H(Y) < ω. We deduce that H(Z) < ω.

**Case 2:** |Y – Z| > 1

In Case 1, we know that H(Z ∪ {$X_i$}) ≥ H(Z) (since Y = Z ∪ {$X_i$}).
By similar derivations, we could deduce that for any set X of dimensions, H(Z ∪ X) ≥ H(Z).
Thus, we know that if H(Y) < ω, then H(Z) < ω.

Algorithm:

The idea is similar to the original Apriori Algorithm learnt in class.
Algorithm:

5. $L_1 \leftarrow$ a set of dimensions where each dimension has good clustering
6. $k \leftarrow 2$
7. while $L_{k-1} \neq \varnothing$
   - $C_k \leftarrow$ Generate candidates from $L_{k-1}$ by Join Step and Prune Step discussed in class
   - Perform a counting step on $C_k$ (i.e., computing the H value of each element in $C_k$) and obtain $L_k$ (i.e., keeping each element in $C_k$ with the H value smaller than ω)
8. Output $\bigcup_i L_i$

**Q11 (Continued)**

## Q11 (Continued)

## Q11 (Continued)

**End of Answer Sheet**