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

Part A

Q1 (15 Marks) (Version 1)

Item	Freq
a	3
b	5
c	7
d	7
e	1
f	7
g	1
h	1
i	1
j	1
k	1
l	1
m	1
n	1
o	1
p	1
q	1
r	1
s	1
t	1

Freq items:

Item	Freq
a	3
b	5
c	7
d	7
f	7

Sorted Freq items:

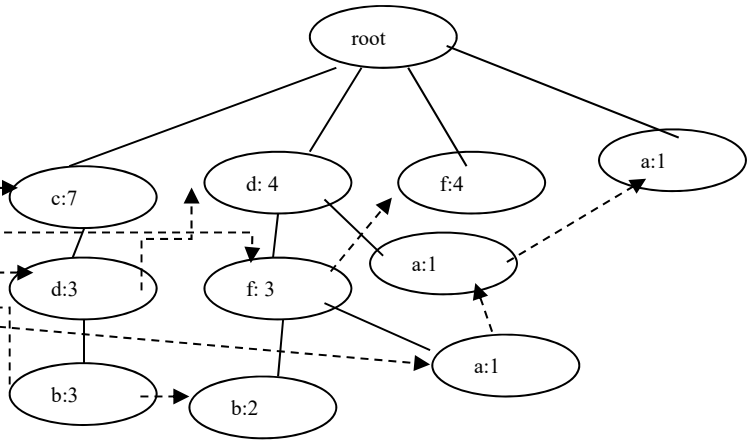
Item	Freq
c	7
d	7
f	7
b	5
a	3

Ordered freq items

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

FP-tree

Item	Head of link
c	
d	
f	
b	
a	



Conditional FP-tree on "a" **count (a)=3**

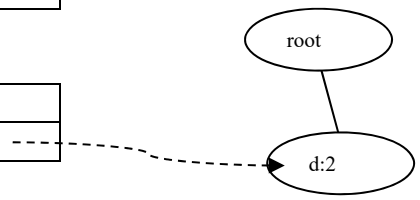
(d:1,f:1,a:1) (a:1,d:1)
(d:1,a:1) ⇒ (a:1,d:1)
(a:1) (a:1)

Item	Freq
c	0
d	2
f	1
b	0
a	3

↓

Item	Freq
a	3
d	2

Item	Head
d	



{a,d}:2

Conditional FP-tree on "b" **count (b)=5**

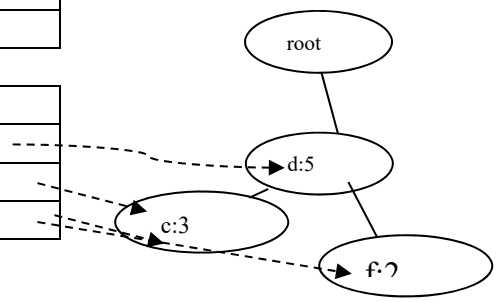
(c:3,d:3,b:3) (b:3,d:3,c:3)
(d:2,f:2,b:2) ⇒ (b:2,d:2,f:2)

Item	Freq
c	3
d	5
f	2
b	5
a	0

↓

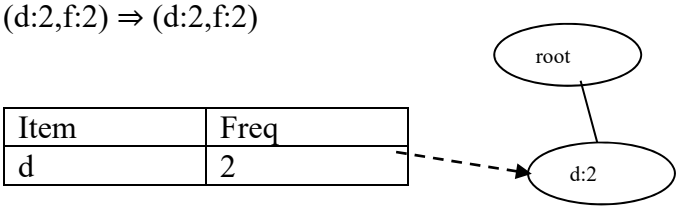
Item	Freq
b	5
d	5
c	3
f	2

Item	Head
d	
c	
f	



Conditional FP-tree on "b,f" count(b,f) = 2

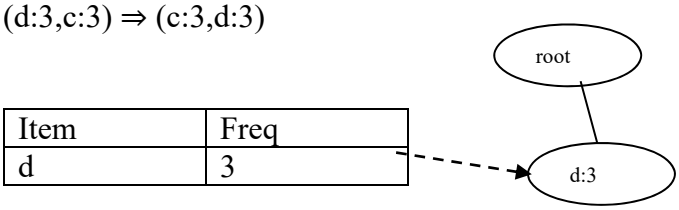
(d:2,f:2) ⇒ (d:2,f:2)



{b,f}=2
{b,d,f}=2

Conditional FP-tree on "b,c" count(b,c) = 3

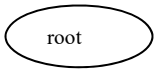
(d:3,c:3) ⇒ (c:3,d:3)



{b,c}=3
{b,c,d}=3

Conditional FP-tree on "b,d" count(b,d) = 5

(d:5)



{b,d}:5

Conditional FP-tree on "f" **count (f)=7**

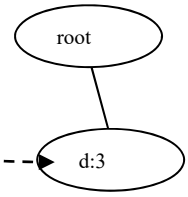
(f:4) ⇒ (f:4)

(d:3,f:3) (f:3,d:3)

Item	freq
f	7
d	3

Item	Head
d	

{d,f}:3



Conditional FP-tree on "d" **count (d)=7**

(c:3,d:3) (d:3,c:3)

⇒

(d:4) (d:4)

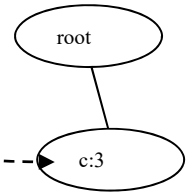
Item	Freq
c	3
d	7
f	0
b	0
a	0

↓

Item	Freq
d	7
c	3

Item	Head
c	

{c,d}:3



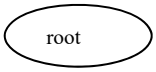
Conditional FP-tree on "c" **count (c)=7**

(c:7) ⇒ (c:7)

Item	freq
c	7

↓

Item	freq
c	7



Freq itemsets

={{c},{b,d}, {c,d}
{d},{a,d}, {b,c,d}
{f},{d,f}, {b, d, f},
{b}, {b,c}
{a},{b,f}}

Q1 (15 Marks) (Version 2)

Item	Freq
a	1
b	1
c	1
d	1
e	1
f	1
g	1
h	1
i	1
j	1
k	1
l	1
r	3
s	5
t	6
u	7
v	1
w	7
x	1
y	1
z	1

Freq items:

Item	Freq
r	3
s	5
t	6
u	7
w	7

Sorted Freq items:

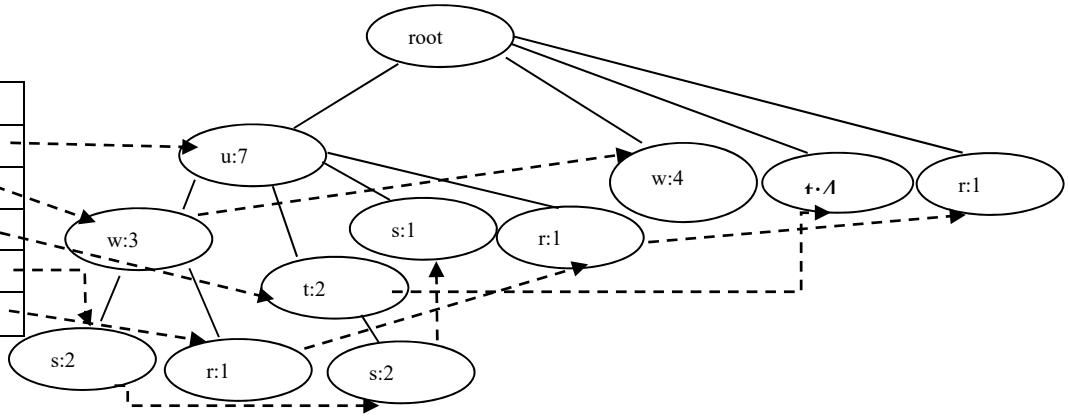
Item	Freq
u	7
w	7
t	6
s	5
r	3

Ordered freq items

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

FP-tree

Item	Head of link
u	
w	
t	
s	
r	



Conditional FP-tree on "r" **count (r)=3**

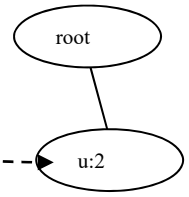
(u:1,r:1) (r:1,u:1)
(u:1,w:1,r:1) ⇒ (r:1,u:1)
(r:1) (r:1)

Item	Freq
u	2
w	0
t	0
s	0
r	3

↓

Item	Freq
r	3
u	2

Item	Head
u	



{r,u}:2

Conditional FP-tree on "s" **count (s)=5**

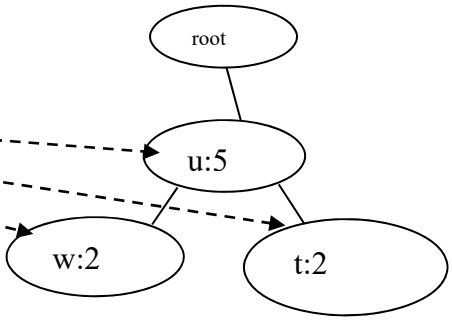
(u:1,s:1) (s:1,u:1)
(u:2,w:2.s:2) ⇒ (s:2,u:2,w:2)
(u:2,t:2,s:2) (s:2,u:2,t:2)

Item	Freq
u	5
w	2
t	2
s	5
r	0

↓

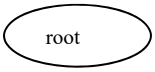
Item	Freq
s	5
u	5
t	2
w	2

Item	Head
u	
t	
w	



Conditional FP-tree on "u,s" **count (u,s)=5**

(u:5)

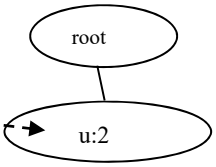


{u,s}:5

Conditional FP-tree on "s,w" **count (s,w)=2**

(u,w:2)

Item	head
u	

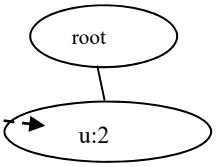


{w,s}:2
{u,s,w}:2

Conditional FP-tree on "s,t" **count (s,t)=2**

(u,t:2)

Item	head
u	



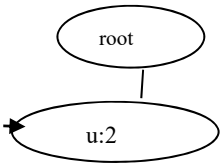
{t,s}:2
{u,s,t}:2

Conditional FP-tree on "t" **count (t)=6**

(t:4) ⇒ (t:4)

(u:2,t:2) (t:2,u:2)

Item	head
U	

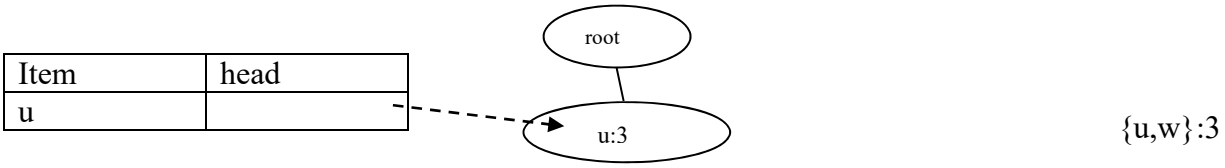


{u,t}:2

Conditional FP-tree on "w" **count (w)=7**

(u:3,w:3) ⇒ (w:3,u:3)

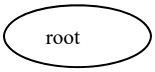
(w:4) (w:4)



{u,w}:3

Conditional FP-tree on "u" **count (u)=7**

(u:7)



Freq itemsets=

{ {u},{w},{t},{s}{r},
{r,u},{u,s},{w,s},{t,s},{u,t},{u,w}
{u,s,w},{u,s,t}
}

Q2 (15 Marks)

Center		(1, 2)	(2, 0)	(-10, -5)	(-5, -2)	(10, 12)	(8, 6)	(-8, -6)	(2, 1)
	Points	1	2	3	4	5	6	7	8
(1, 2)	1	0							
(2, 0)	2	2.24	0						
(-10, -5)	3	13.04	13	0					
(-5, -2)	4	7.21	7.28	5.83	0				
(10, 12)	5	13.45	14.42	26.25	20.52	0			
(8, 6)	6	8.06	8.49	21.1	15.26	6.32	0		
(-8, -6)	7	12.04	11.66	2.24	5	25.46	20	0	
(2, 1)	8	1.41	1	13.42	7.62	13.6	7.81	12.21	0

Center		(1, 2)	(2, 0.5)	(-10, -5)	(-5, -2)	(10, 12)	(8, 6)	(-8, -6)
	Points	1	(28)	3	4	5	6	7
(1, 2)	1	0						
(2, 0.5)	(28)	1.8	0					
(-10, -5)	3	13.04	13.2	0				
(-5, -2)	4	7.21	7.43	5.83	0			
(10, 12)	5	13.45	14.01	26.25	20.52	0		
(8, 6)	6	8.06	8.14	21.1	15.26	6.32	0	
(-8, -6)	7	12.04	11.93	2.24	5	25.46	20	0

Center		(1.5, 1.25)	(-10, -5)	(-5, -2)	(10, 12)	(8, 6)	(-8, -6)
	Points	(128)	3	4	5	6	7
(1.5, 1.25)	(128)	0					
(-10, -5)	3	13.09	0				
(-5, -2)	4	7.27	5.83	0			
(10, 12)	5	13.7	26.25	20.52	0		
(8, 6)	6	8.05	21.1	15.26	6.32	0	
(-8, -6)	7	11.95	2.24	5	25.46	20	0

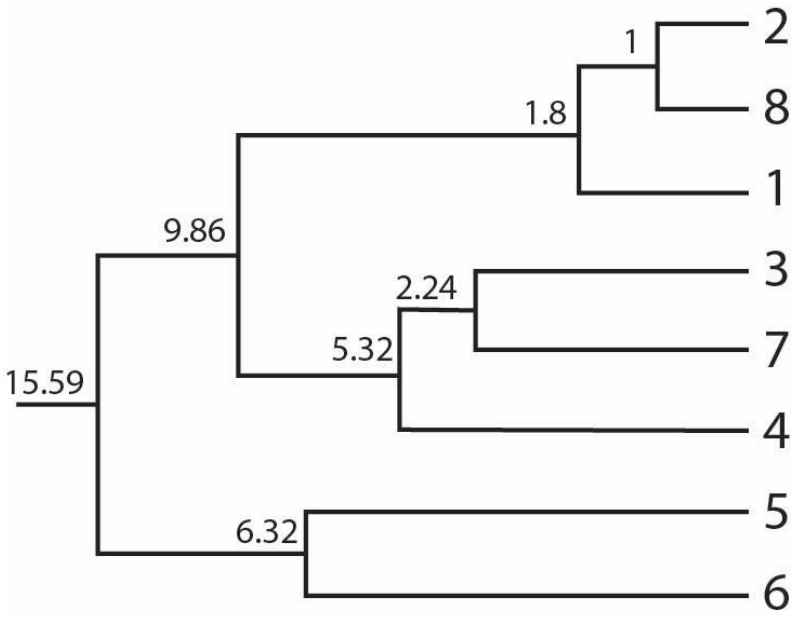
Center		(1.5, 1.25)	(-9, -5.5)	(-5, -2)	(10, 12)	(8, 6)
	Points	(128)	(37)	4	5	6
(1.5, 1.25)	(128)	0				
(-9, -5.5)	(37)	12.48	0			
(-5, -2)	4	7.27	5.32	0		
(10, 12)	5	13.7	25.83	20.52	0	
(8, 6)	6	8.05	20.52	15.26	6.32	0

Center		(1.5, 1.25)	(-7, -3.75)	(10, 12)	(8, 6)
	Points	(128)	(347)	5	6
(1.5, 1.25)	(128)	0			
(-7, -3.75)	(347)	9.86	0		
(10, 12)	5	13.7	23.17	0	
(8, 6)	6	8.05	17.89	6.32	0

Center		(1.5, 1.25)	(-7, -3.75)	(9, 9)
	Points	(128)	(347)	(56)
(1.5, 1.25)	(128)	0		
(-7, -3.75)	(347)	9.86	0	
(9, 9)	(56)	10.78	20.46	0

Center		(-2.75, -1.25)	(9, 9)
	Points	(123478)	(56)
(-2.75, -1.25)	(123478)	0	
(9, 9)	(56)	15.59	0

Dendrogram (not to scale)



Q3 (15 Marks)

(a)
Yes.
P(LC = Yes | FH = Yes, S = yes, PR = No) = 0.4375

$$\frac{P(PR = No \mid LC = Yes, FH = Yes, S = Yes) * P(LC = Yes \mid FH = Yes, S = Yes)}{P(PR = No \mid FH = Yes, S = Yes)} = 0.4375$$

P(PR = No | LC = Yes, FH = Yes, S = Yes)
= P(PR = No | LC = Yes) (conditional independence)
= (1 - x)

P(LC = Yes | FH = Yes, S = Yes) = 0.7

P(PR = No | FH = Yes, S = Yes)
= P(PR = No | LC = Yes) * P(LC = Yes | FH = Yes, S = Yes) +
P(PR = No | LC = No) * P(LC = No | FH = Yes, S = Yes)
= (1 - x) * 0.7 + (1 - y) * 0.3

Thus, (1 - x) * 0.7 / ((1 - x) * 0.7 + (1 - y) * 0.3) = 0.4375
According to x = 2y, we can solve the final result of x = 0.8, y = 0.4

(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.

Q4 (15 Marks) (Version 1)

mean vector = $\begin{pmatrix} \frac{6+8+5+10}{4} \\ \frac{6+8+11+4}{4} \end{pmatrix} = \begin{pmatrix} 29/4 \\ 29/4 \end{pmatrix} = \begin{pmatrix} 7.25 \\ 7.25 \end{pmatrix}$

For data (6, 6), difference from mean vector = $\begin{pmatrix} 6-7.25 \\ 6-7.25 \end{pmatrix} = \begin{pmatrix} -5/4 \\ -5/4 \end{pmatrix} = \begin{pmatrix} -1.25 \\ -1.25 \end{pmatrix}$

For data (8, 8), difference from mean vector = $\begin{pmatrix} 8-7.25 \\ 8-7.25 \end{pmatrix} = \begin{pmatrix} 3/4 \\ 3/4 \end{pmatrix} = \begin{pmatrix} 0.75 \\ 0.75 \end{pmatrix}$

For data (5, 11), difference from mean vector = $\begin{pmatrix} 5-7.25 \\ 11-7.25 \end{pmatrix} = \begin{pmatrix} -9/4 \\ 15/4 \end{pmatrix} = \begin{pmatrix} -2.25 \\ 3.75 \end{pmatrix}$

For data (10, 4), difference from mean vector = $\begin{pmatrix} 10-7.25 \\ 4-7.25 \end{pmatrix} = \begin{pmatrix} 11/4 \\ -13/4 \end{pmatrix} = \begin{pmatrix} 2.75 \\ -3.25 \end{pmatrix}$

$Y = \begin{pmatrix} -1.25 & 0.75 & -2.25 & 2.75 \\ -1.25 & 0.75 & 3.75 & -3.25 \end{pmatrix}$

$$\Sigma = \frac{1}{4}YY^T = \frac{1}{4}\begin{pmatrix} -1.25 & 0.75 & -2.25 & 2.75 \\ -1.25 & 0.75 & 3.75 & -3.25 \end{pmatrix}\begin{pmatrix} -1.25 & -1.25 \\ 0.75 & 0.75 \\ -2.25 & 3.75 \\ 2.75 & -3.25 \end{pmatrix}$$

$$= \frac{1}{4}\begin{pmatrix} 14.75 & -15.25 \\ -15.25 & 26.75 \end{pmatrix}$$

$$= \begin{pmatrix} \frac{59}{16} & -\frac{61}{16} \\ -\frac{61}{16} & \frac{107}{16} \end{pmatrix} = \begin{pmatrix} 3.6875 & -3.8125 \\ -3.8125 & 6.6875 \end{pmatrix}$$

$$\begin{vmatrix} 3.6875-\lambda & -3.8125 \\ -3.8125 & 6.6875-\lambda \end{vmatrix} = 0 \quad \Rightarrow \quad \lambda^2 - 10.375\lambda + 10.125 = 0$$

$$\Rightarrow \lambda = \frac{83+\sqrt{4297}}{16} = 9.2845 \text{ or } \lambda = \frac{83-\sqrt{4297}}{16} = 1.0905$$

When $\lambda = \frac{83+\sqrt{4297}}{16}$,

$$\begin{pmatrix} \frac{59}{16} - \frac{83+\sqrt{4}297}{16} & -\frac{61}{16} \\ -\frac{61}{16} & \frac{107}{16} - \frac{83+\sqrt{4}297}{16} \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \Rightarrow \begin{pmatrix} -24-\sqrt{4}297 & -61 \\ -61 & 24-\sqrt{4}297 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$\Rightarrow x_1 + 0.6812x_2 = 0$$

We choose the eigenvector of unit length: $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0.5630 \\ -0.8265 \end{pmatrix}$.

when $\lambda = \frac{83-\sqrt{4297}}{16}$,

$$\begin{pmatrix} \frac{59}{16} - \frac{83-\sqrt{4}297}{16} & -\frac{61}{16} \\ -\frac{61}{16} & \frac{107}{16} - \frac{83-\sqrt{4}297}{16} \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \Rightarrow \begin{pmatrix} -24+\sqrt{4}297 & -61 \\ -61 & 24+\sqrt{4}297 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$\Rightarrow x_1 - 1.4681x_2 = 0$$

We choose the eigenvector of unit length: $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} -0.8265 \\ -0.5630 \end{pmatrix}$.

Order can be interchangeable from left-to-right

Thus, $\Phi = \begin{pmatrix} 0.5630 & -0.8265 \\ -0.8265 & -0.5630 \end{pmatrix}$, $Y = \Phi^T X = \begin{pmatrix} 0.5630 & -0.8265 \\ -0.8265 & -0.5630 \end{pmatrix} X$.

For data (6, 6), $Y = \begin{pmatrix} 0.5630 & -0.8265 \\ -0.8265 & -0.5630 \end{pmatrix} \begin{pmatrix} 6 \\ 6 \end{pmatrix} = \begin{pmatrix} -1.5810 \\ -8.3367 \end{pmatrix}$

For data (8, 8), $Y = \begin{pmatrix} 0.5630 & -0.8265 \\ -0.8265 & -0.5630 \end{pmatrix} \begin{pmatrix} 8 \\ 8 \end{pmatrix} = \begin{pmatrix} -2.1080 \\ -11.1156 \end{pmatrix}$

For data (5, 11), $Y = \begin{pmatrix} 0.5630 & -0.8265 \\ -0.8265 & -0.5630 \end{pmatrix} \begin{pmatrix} 5 \\ 11 \end{pmatrix} = \begin{pmatrix} -6.2764 \\ -10.3251 \end{pmatrix}$

For data (10, 4), $Y = \begin{pmatrix} 0.5630 & -0.8265 \\ -0.8265 & -0.5630 \end{pmatrix} \begin{pmatrix} 10 \\ 4 \end{pmatrix} = \begin{pmatrix} 2.3238 \\ -10.5166 \end{pmatrix}$

The mean vector of the above transformed data points is $= \begin{pmatrix} -1.9104 \\ -10.0735 \end{pmatrix}$

The final transformed data points are:

For data (6, 6), final transformed vector $= \begin{pmatrix} -1.5810 \\ -8.3367 \end{pmatrix} - \begin{pmatrix} -1.9104 \\ -10.0735 \end{pmatrix} = \begin{pmatrix} 0.3294 \\ 1.7368 \end{pmatrix}$

For data (8, 8), final transformed vector $= \begin{pmatrix} -2.1080 \\ -11.1156 \end{pmatrix} - \begin{pmatrix} -1.9104 \\ -10.0735 \end{pmatrix} = \begin{pmatrix} -0.1976 \\ -1.0421 \end{pmatrix}$

For data (5, 11), final transformed vector $= \begin{pmatrix} -6.2764 \\ -10.3251 \end{pmatrix} - \begin{pmatrix} -1.9104 \\ -10.0735 \end{pmatrix} = \begin{pmatrix} -4.3660 \\ -0.2516 \end{pmatrix}$

For data (10, 4), final transformed vector $= \begin{pmatrix} 2.3238 \\ -10.5166 \end{pmatrix} - \begin{pmatrix} -1.9104 \\ -10.0735 \end{pmatrix} = \begin{pmatrix} 4.2342 \\ -0.4431 \end{pmatrix}$

Thus, (6, 6) is reduced to (0.3294);
(8, 8) is reduced to (-0.1976);
(5, 11) is reduced to (-4.3660);
(10, 4) is reduced to (4.2342).

(Note: Another possible answer is
(6, 6) is reduced to (-0.3294);
(8, 8) is reduced to (0.1976);
(5, 11) is reduced to (4.3660);
(10, 4) is reduced to (-4.2342).
This is because the eigenvectors used in this case are:

$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} -0.5630 \\ 0.8265 \end{pmatrix}$ and $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0.8265 \\ 0.5630 \end{pmatrix}$

Q4 (15 Marks) (Version 2)

mean vector = $\begin{pmatrix} \frac{18+24+15+30}{4} \\ \frac{18+24+33+12}{4} \end{pmatrix} = \begin{pmatrix} 85/4 \\ 85/4 \end{pmatrix} = \begin{pmatrix} 21.25 \\ 21.25 \end{pmatrix}$

For data (18, 18), difference from mean vector = $\begin{pmatrix} 18-21.25 \\ 18-21.25 \end{pmatrix} = \begin{pmatrix} -3.75 \\ -3.75 \end{pmatrix}$

For data (24, 24), difference from mean vector = $\begin{pmatrix} 24-21.25 \\ 24-21.25 \end{pmatrix} = \begin{pmatrix} 2.25 \\ 2.25 \end{pmatrix}$

For data (15, 33), difference from mean vector = $\begin{pmatrix} 15-21.25 \\ 33-21.25 \end{pmatrix} = \begin{pmatrix} -6.75 \\ 11.25 \end{pmatrix}$

For data (30, 12), difference from mean vector = $\begin{pmatrix} 30-21.25 \\ 12-21.25 \end{pmatrix} = \begin{pmatrix} 8.25 \\ -9.75 \end{pmatrix}$

$Y = \begin{pmatrix} -1.25 & 0.75 & -2.25 & 2.75 \\ -1.25 & 0.75 & 3.75 & -3.25 \end{pmatrix}$

$$\Sigma = \frac{1}{4} YY^T = \frac{1}{4} \begin{pmatrix} -3.75 & 2.25 & -6.75 & 8.25 \\ -3.75 & 2.25 & 11.25 & -9.75 \end{pmatrix} \begin{pmatrix} -3.75 & -3.75 \\ 2.25 & 2.25 \\ -6.75 & 11.25 \\ 8.25 & -9.75 \end{pmatrix}$$

$$= \frac{1}{4} \begin{pmatrix} 132.75 & -137.25 \\ -137.25 & 240.75 \end{pmatrix}$$

$$= \begin{pmatrix} 33.1875 & -34.3125 \\ -34.3125 & 60.1875 \end{pmatrix}$$

$$\begin{vmatrix} 33.1875-\lambda & -34.3125 \\ -34.3125 & 60.1875-\lambda \end{vmatrix} = 0$$

$\Rightarrow \lambda = 83.5602$ or $\lambda = 9.8148$

When $\lambda = 83.5602$,

$$\begin{pmatrix} 33.1875-83.5602 & -34.3125 \\ -34.3125 & 60.1875-83.5602 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \Rightarrow \begin{pmatrix} -50.3727 & -34.3125 \\ -34.3125 & -23.3727 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$\Rightarrow x_1 + 0.6812x_2 = 0$$

We choose the eigenvector of unit length: $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0.5630 \\ -0.8265 \end{pmatrix}$.

when $\lambda = 9.8148$,

$$\begin{pmatrix} 33.1875-9.8148 & -34.3125 \\ -34.3125 & 60.1875-9.8148 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \Rightarrow \begin{pmatrix} 23.3727 & -34.3125 \\ -34.3125 & 50.3727 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$\Rightarrow x_1 - 1.4681x_2 = 0$$

We choose the eigenvector of unit length: $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} -0.8265 \\ -0.5630 \end{pmatrix}$.

Thus, $\Phi = \begin{pmatrix} 0.5630 & -0.8265 \\ -0.8265 & -0.5630 \end{pmatrix}$, $Y = \Phi^T X = \begin{pmatrix} 0.5630 & -0.8265 \\ -0.8265 & -0.5630 \end{pmatrix} X$

For data (18, 18), $Y = \begin{pmatrix} 0.5630 & -0.8265 \\ -0.8265 & -0.5630 \end{pmatrix} \begin{pmatrix} 18 \\ 18 \end{pmatrix} = \begin{pmatrix} -4.7431 \\ -25.0101 \end{pmatrix}$

For data (24, 24), $Y = \begin{pmatrix} 0.5630 & -0.8265 \\ -0.8265 & -0.5630 \end{pmatrix} \begin{pmatrix} 24 \\ 24 \end{pmatrix} = \begin{pmatrix} -6.3241 \\ -33.3468 \end{pmatrix}$

For data (15, 33), $Y = \begin{pmatrix} 0.5630 & -0.8265 \\ -0.8265 & -0.5630 \end{pmatrix}\begin{pmatrix} 15 \\ 33 \end{pmatrix} = \begin{pmatrix} -18.8291 \\ -30.9752 \end{pmatrix}$

For data (30, 14), $Y = \begin{pmatrix} 0.5630 & -0.8265 \\ -0.8265 & -0.5630 \end{pmatrix}\begin{pmatrix} 30 \\ 12 \end{pmatrix} = \begin{pmatrix} 6.9715 \\ -31.5499 \end{pmatrix}$

The mean vector of the above transformed data points is $= \begin{pmatrix} -5.7312 \\ -30.2205 \end{pmatrix}$

The final transformed data points are:

For data (18, 18), final transformed vector = $\begin{pmatrix} -4.7431 \\ -25.0101 \end{pmatrix} - \begin{pmatrix} -5.7312 \\ -30.2205 \end{pmatrix} = \begin{pmatrix} 0.9881 \\ 5.2104 \end{pmatrix}$

For data (24, 24), final transformed vector = $\begin{pmatrix} -6.3241 \\ -33.3468 \end{pmatrix} - \begin{pmatrix} -5.7312 \\ -30.2205 \end{pmatrix} = \begin{pmatrix} -0.5929 \\ -3.1263 \end{pmatrix}$

For data (15, 33), final transformed vector = $\begin{pmatrix} -18.8291 \\ -30.9752 \end{pmatrix} - \begin{pmatrix} -5.7312 \\ -30.2205 \end{pmatrix} = \begin{pmatrix} -13.0979 \\ -0.7547 \end{pmatrix}$

For data (30, 14), final transformed vector = $\begin{pmatrix} 6.9715 \\ -31.5499 \end{pmatrix} - \begin{pmatrix} -5.7312 \\ -30.2205 \end{pmatrix} = \begin{pmatrix} 12.7027 \\ -1.3294 \end{pmatrix}$

Thus, (18, 18) is reduced to (0.9881);
(24, 24) is reduced to (-0.5929);
(15, 33) is reduced to (-13.0979);
(30, 14) is reduced to (12.7027).

(Note: Another possible answer is
(18, 18) is reduced to (-0.9881);
(24, 24) is reduced to (0.5929);
(15, 33) is reduced to (13.0979);
(30, 14) is reduced to (-12.7027).
This is because the eigenvectors used in this case are:
$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} -0.5630 \\ 0.8265 \end{pmatrix}$ and $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0.8265 \\ 0.5630 \end{pmatrix}$

Q5 (15 Marks)

(a)

Yes. The number is 3. By looking at the second portion of the chart, we have 10 – 7 = 3

(b)

Yes. The number is 7. By looking at the first portion of the chart, we have 7 – 0 = 7

(c)

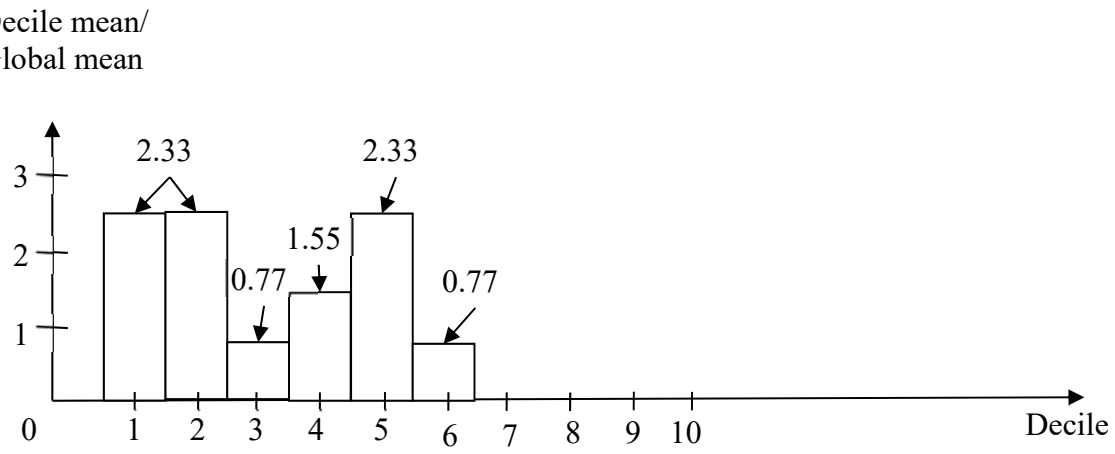
Yes. The number is 6. By looking at the third portion of the chart, we have 16 – 10 = 6

(d)

Yes. The number is 14. By looking at the fourth portion of the chart, we have 30 – 16 = 14

(e)

Yes. The chart is shown as follows.



Global mean = 13/30

By dividing the data into ten deciles, we have the value of each decile to be Decile mean/Global mean.

(f)

Yes.

Precision = True Positive/ (True Positive + False Positive) = 7/(7+3) = 7/10

Recall = True Positive/ (True Positive + False Negative) = 7/(7+6) = 7/13

F1-score = (2 x Precision x Recall)/(Precision + Recall) = (2 x 7/10 x 7/13)/(7/10 + 7/13) = 14/23 = 0.609

Q6 (15 Marks)

	First Choice	Second Choice	Third Choice	Forth Choice
b	**700**			
c	350	50	50	50
d	640	**320**		
e	360	60	60	20
f	390	90	90	30
g	580	180	**180**	
h	330	230	70	60
i	320	320	160	**160**
j	190	90	90	45
k	185	85	85	40
l	180	80	80	35
m	195	195	115	115

Resulting views = {b, d, g, i}

	First Choice	Second Choice	Third Choice	Forth Choice
b	**700**			
c	350	50	50	50
d	640	320	160	**160**
e	360	60	60	20
f	390	90	90	30
g	580	180	**180**	
h	330	230	120	110
i	320	**320**		
j	190	90	90	45
k	185	85	85	40
l	180	80	80	35
m	195	195	35	35

Resulting views = {b, i, g, d}

Q7 (15 Marks)

(a)

ID	x_1	x_2	y
a	13	13	1
b	19	9	1
c	21	15	1
d	15	19	1
e	7	9	-1
f	5	7	-1
g	9	9	-1
h	7	5	-1

(b)

minimize $w_1^2+w_2^2$

subject to

$13w_1 + 13w_2 + b \geq 1$
$19w_1 + 9w_2 + b \geq 1$
$21w_1 + 15w_2 + b \geq 1$
$15w_1 + 19w_2 + b \geq 1$
$-7w_1 - 9w_2 - b \geq 1$
$-5w_1 - 7w_2 - b \geq 1$
$-9w_1 -7w_2 - b \geq 1$
$-7w_1 -5w_2 - b \geq 1$
w_1, w_2 and b are real numbers

(c)

$net=13w_1 +13w_1 +b=2.7$
$y=0.9910$
$w_1 = 0.1+0.5*(1-0.9910)*13=0.1585$
$w_2 = 0.1+0.5*(1-0.9910)*13=0.1585$
$b = 0.1+0.5*(1-0.9910)=0.1045$

$net=19w_1 +9w_2 +b=4.5425$
$y=0.9998$
$w_1 = 0.1585+0.5*(1-0.9998)*19=0.1604$
$w_2 = 0.1585+0.5*(1-0.9998)*9=0.1594$
$b = 0.1045+0.5*(1-0.9998)=0.1046$

net=21w_1 +15w_2 +b=5.864
y=1
w_1 = 0.1604+0.5*(1-1)*21=0.1604
w_2 = 0.1594+0.5*(1-1)*15=0.1594
b = 0.1046+0.5*(1-1)=0.1046

net=15w_1 +19w_2 +b=5.5392
y=1
w_1 = 0.1604+0.5*(1-1)*15=0.1604
w_2 = 0.1594+0.5*(1-1)*19=0.1594
b = 0.1046+0.5*(1-1)=0.1046

net=7w_1 +9w_2 +b=2.662
y=0.9903
w_1 = 0.1604+0.5*(-1-0.9903)*7=-6.8057
w_2 = 0.1594+0.5*(-1-0.9903)*9=-8.7970
b = 0.1046+0.5*(-1-0.9903)=-0.8906

net=5w_1 +7w_2 +b=-96.4981
y=-1
w_1 = -6.8057+0.5*(-1+1)*5=-6.8057
w_2 =-8.7970+0.5*(-1+1)*7=-8.7970
b = -0.8906+0.5*(-1+1)= -0.8906

(d)
The neural network has an assumption that records in the training set are “independent”.
In some cases, records in the training set are related/correlated to (or dependent on) other records in the training set. Thus, the neural network could not capture this “dependent” scenario well if the training set has dependent records.

Q8 (15 Marks)

(a)

Adjacency matrix

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

(b)

Stochastic matrix

$$\begin{array}{c} \\ x \\ y \\ z \end{array} \begin{array}{c} \begin{array}{ccc} x & y & z \end{array} \\ \begin{bmatrix} 0 & ½ & ½ \\ ½ & ½ & ½ \\ ½ & 0 & 0 \end{bmatrix} \end{array}$$

(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

(d)

$r_n = 0.8\ M\ r_0 + c$

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

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

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

Version 1

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

Version 2

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

Version 3

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

Version 4

Question	Your Answer
Q9	B
Q10	E
Q11	C
Q12	B
Q13	D
Q14	A
Q15	A
Q16	E
Q17	B
Q18	B
Q19	A
Q20	B
Q21	C
Q22	A
Q23	B
Q24	E

Version 5

Question	Your Answer
Q9	Q
Q10	T
Q11	R
Q12	Q
Q13	S
Q14	P
Q15	P
Q16	T
Q17	R
Q18	P
Q19	Q
Q20	T
Q21	Q
Q22	Q
Q23	P
Q24	Q

Version 6

Question	Your Answer
Q9	W
Q10	Z
Q11	X
Q12	W
Q13	Y
Q14	V
Q15	V
Q16	Z
Q17	W
Q18	V
Q19	W
Q20	X
Q21	V
Q22	W
Q23	Z
Q24	W

Version 7

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

Version 8

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

Version 9

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

End of Paper