COMP1942 Exploring and Visualizing Data (Spring Semester 2024)
Homework 2 Solution
Full Mark: 100 Marks

Q1 [20 Marks]

(a) No. This new customer will not go to Disneyland.

(b) (i)

No. of Children	No. of Siblings	Go_Disneyland
2	0	0
0	2	0
4	2	1
2	4	1

(ii)

Iteration 1

x_1 x_2 y
(2, 0, 0)

b	w_1	w_2
0.3	0.3	0.3

net $= x_1w_1 + x_2w_2 + b$
$= 2 \times 0.3 + 0 \times 0.3 + 0.3$
$= 0.9$

y $= 1$ (Incorrect)

w_1 $= w_1 + \alpha\,(d - y)\,x_1$
$= 0.3 + 0.6 \times (0 - 1) \times 2$
$= -0.9$

w_2 $= w_2 + \alpha\,(d - y)\,x_2$
$= 0.3 + 0.6 \times (0 - 1) \times 0$
$= 0.3$

b $= b + \alpha\,(d - y)$
$= 0.3 + 0.6 \times (0 - 1)$
$= -0.3$

Iteration 2

x_1 x_2 y
(0, 2, 0)

b	w_1	w_2
-0.3	-0.9	0.3

net $= x_1w_1 + x_2w_2 + b = 0.3$

y $= 1$ (Incorrect)

w_1 $= w_1 + \alpha (d - y) x_1$
$= -0.9 + 0.6 \times (0 - 1) \times 0$
$= -0.9$

w_2 $= w_2 + \alpha (d - y) x_2$
$= 0.3 + 0.6\times (0-1) \times 2$
$= -0.9$

b $= b + \alpha (d - y)$
$= -0.3 + 0.6 \times (0-1)$
$= -0.9$

Iteration 3

x_1 x_2 y
(4, 2, 1)

b	w_1	w_2
-0.9	-0.9	-0.9

net $= x_1w_1 + x_2w_2 + b$
$= -6.3$

y $= 0$ (Incorrect)

w_1 $= w_1 + \alpha (d - y) x_1$
$= -0.9 + 0.6 \times (1 - 0) \times 4$
$= 1.5$

w_2 $= w_2 + \alpha (d - y) x_2$
$= -0.9 + 0.6 \times (1 - 0) \times 2$
$= 0.3$

b $= b + \alpha (d - y)$
$= -0.9 + 0.6 \times (1 - 0)$
$= -0.3$

Iteration 4

x_1 x_2 y
(2, 4, 1)

b	w_1	w_2
-0.3	1.5	0.3

net $= x_1w_1 + x_2w_2 + b$
$= 3.9$

y $= 1$ (Correct)

w_1 $= w_1 + \alpha\,(d - y)\,x_1$
$= 1.5$

w_2 $= w_2 + \alpha\,(d - y)\,x_2$
$= 0.3$

b $= b + \alpha\,(d - y)$
$= -0.3$

Iteration 5

x_1 x_2 y
(2, 0, 0)

b	w_1	w_2
-0.3	1.5	0.3

net $= x_1w_1 + x_2w_2 + b$
$= 2.7$

y $= 1$ (Incorrect)

w_1 $= w_1 + \alpha\,(d - y)\,x_1$
$= 1.5 + 0.6 \times (0 - 1) \times 2$
$= 0.3$

w_2 $= w_2 + \alpha\,(d - y)\,x_2$
$= 0.3 + 0.6 \times (0 - 1) \times 0$
$= 0.3$

b $= b + \alpha\,(d - y)$
$= -0.3 + 0.6 \times (0 - 1)$
$= -0.9$

b	w_1	w_2
-0.9	0.3	0.3

Q2 [20 Marks]

(a)

mean vector = $\begin{pmatrix} \frac{6+8+5+9}{4} \\ \frac{6+8+9+5}{4} \end{pmatrix} = \begin{pmatrix} 7 \\ 7 \end{pmatrix}$

For data (6, 6), difference from mean vector = $\begin{pmatrix} 6-7 \\ 6-7 \end{pmatrix} = \begin{pmatrix} -1 \\ -1 \end{pmatrix}$

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

For data (5, 9), difference from mean vector = $\begin{pmatrix} 5-7 \\ 9-7 \end{pmatrix} = \begin{pmatrix} -2 \\ 2 \end{pmatrix}$

For data (9, 5), difference from mean vector = $\begin{pmatrix} 9-7 \\ 5-7 \end{pmatrix} = \begin{pmatrix} 2 \\ -2 \end{pmatrix}$

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

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

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

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

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

when $\lambda = 4$,

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

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 (6, 6), $Y=\begin{pmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix}\begin{pmatrix} 6 \\ 6 \end{pmatrix}=\begin{pmatrix} 0 \\ 8.49 \end{pmatrix}$

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 (5, 9), $Y=\begin{pmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix}\begin{pmatrix} 5 \\ 9 \end{pmatrix}=\begin{pmatrix} -2.83 \\ 9.90 \end{pmatrix}$

For data (9, 5), $Y=\begin{pmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix}\begin{pmatrix} 9 \\ 5 \end{pmatrix}=\begin{pmatrix} 2.83 \\ 9.90 \end{pmatrix}$

Thus, (6, 6) is reduced to (0);
(8, 8) is reduced to (0);
(5, 9) is reduced to (-2.83);
(9, 5) is reduced to (2.83).

(Note: Another possible answer is
(6, 6) is reduced to (0);
(8, 8) is reduced to (0);
(5, 9) is reduced to (2.83);
(9, 5) 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 $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}=\begin{pmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{pmatrix}$.)

(b)

(2, 6) is reduced to (2.83);
(3, 3) is reduced to (0);
(5, 5) is reduced to (0);
(6, 2) is reduced to (-2.83).

Q3 [20 Marks]

The greedy algorithm discussed in class can be modified be changing the heuristics function from the computation of the benefit of a view to the computation of the benefit of a view per "unit space".

i.e.

Let C(v) be the cost of view v(the number of rows in v)

Algorithm:

$S \leftarrow \{top\ \ view\}$;

$X \leftarrow X - C(v)$ where v is the top view ;

While there exists a view $v \in S\ s.t.\ C(v) \le X$

  Select the view $v \in S\ s.t.$

    $C(v) \le X$

    $B(v,S)/C(v)$ is maximized

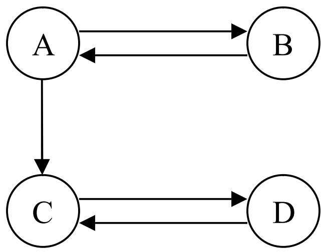
  $S \leftarrow S \cup \{v\}$

  $X \leftarrow X - C(v)$

output S.

Q4 [20 Marks]

(a)



Stochastic matrix:

$$\begin{array}{c} \\ A \\ B \\ C \\ D \end{array} \begin{array}{c} \begin{array}{cccc} A & B & C & D \end{array} \\ \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0.5 & 0 & 0 & 0 \\ 0.5 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \end{array}$$

(b) Equation to be solved:

$$\begin{pmatrix} A \\ B \\ C \\ D \end{pmatrix} = 0.8 \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0.5 & 0 & 0 & 0 \\ 0.5 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} A \\ B \\ C \\ D \end{pmatrix} + \begin{pmatrix} 0.2 \\ 0.2 \\ 0.2 \\ 0.2 \end{pmatrix}$$

(c)

	1	2	3	4	5	6	7	…	13
A	1	1.0	0.68	0.68	0.58	0.58	0.54	…	0.53
B	1	0.6	0.6	0.47	0.47	0.43	0.43	…	0.41
C	1	1.4	1.4	1.53	1.53	1.57	1.57	…	1.59
D	1	1.0	1.32	1.32	1.42	1.42	1.46	…	1.47

So the ranking is C, D, A, B.

Q5 [20 Marks]

(a)

P(SIR = Yes | AP = Yes, P = Yes, WBC = High)

$$= \frac{P(WBC=High|AP=Yes,P=Yes,SIR=Yes)}{P(WBC=High|AP=Y\quad,P=Yes)} P(SIR = Yes|AP = Yes, P = Yes)$$

$$= \frac{P(WBC=High|SIR=Yes)P(SIR=Yes|AP=Yes,P=Y\quad)}{\sum_{x\in\{Yes,No\}} P(WBC=High|SIR\quad)P(SIR\quad|AP=Y\quad,P=Y\quad)}$$

$$= \frac{0.6*0.7}{0.6*0.7+0.3*0.3}$$

= 0.8235

P(SIR = No | AP = Yes, P = Yes, WBC = High) = 1- 0.8235 =0.1765

Since P(SIR = Yes | AP = Yes, P = Yes, WBC = High) > P(SIR = No | AP = Yes, P = Yes, WBC = High), it is more likely that the person has systemic inflammation reaction.

(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.