HKUST COMP 1942 homework 2 submission

COMP1942 Exploring and Visualizing Data (Spring Semester 2024)
Homework 2
Deadline: 8 May, 2024, 9:00am
(Please hand in during lecture.)
Full Mark: 100 Marks

Coupon Instructions:

1. You can use a coupon to waive any question you want and obtain full marks for this question.
2. You can waive at most one question in each assignment.
3. You can also answer the question you will waive. We will also mark it but will give full marks to this question.
4. The coupon is non-transferrable. That is, the coupon with a unique ID can be used only by the student who obtained it in class.
5. Please staple the coupon to the submitted assignment.
6. Please write down the question no. you want to waive on the coupon.
7. [New Requirement compared with HW1] Please print this page (Page 1) of this homework specification as the first page of your HW submission. Fill in the correspondence information for the ease of marking.

- Student ID: 20703125
- Student Name: Suo Xun Xin
- Seat No.: 143
    - (Your seat number is the one used in the midterm exam. You could also check it in the HW page in Canvas.)

Question	Full Mark	Mark
Q1	20	
Q2	20	
Q3	20	
Q4	20	
Q5	20	
Total	100	

Q1

[20 Marks]

Consider the following table with three attributes where "No. of Children" and "No. of Siblings" are input attributes and "Go_Disneyland" is the target attribute. Each tuple corresponds to a customer.

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

Q1.a

Suppose that there is a new customer with "No. of Children" = 0 and "No. of Siblings" = 1. Please use XLMiner to predict whether this customer will go to Disneyland or not by a 3-nearest neighbor classifier. In the XLMiner setting, you do not need to re-scale the data. You do not need to submit any softcopy related to XLMiner. You just need to write down the answer without any explanation.

The customer will not go to Disneyland.

Record ID	Prediction: Go_Disneyland	PostProb: No	PostProb: Yes
Record 1	No	0.666666667	0.333333333

Q1.b

Q1.b.i

Rewrite the above table such that values "Yes" and "No" in attribute "Go_Disneyland" are mapped to values 1 and 0, respectively.

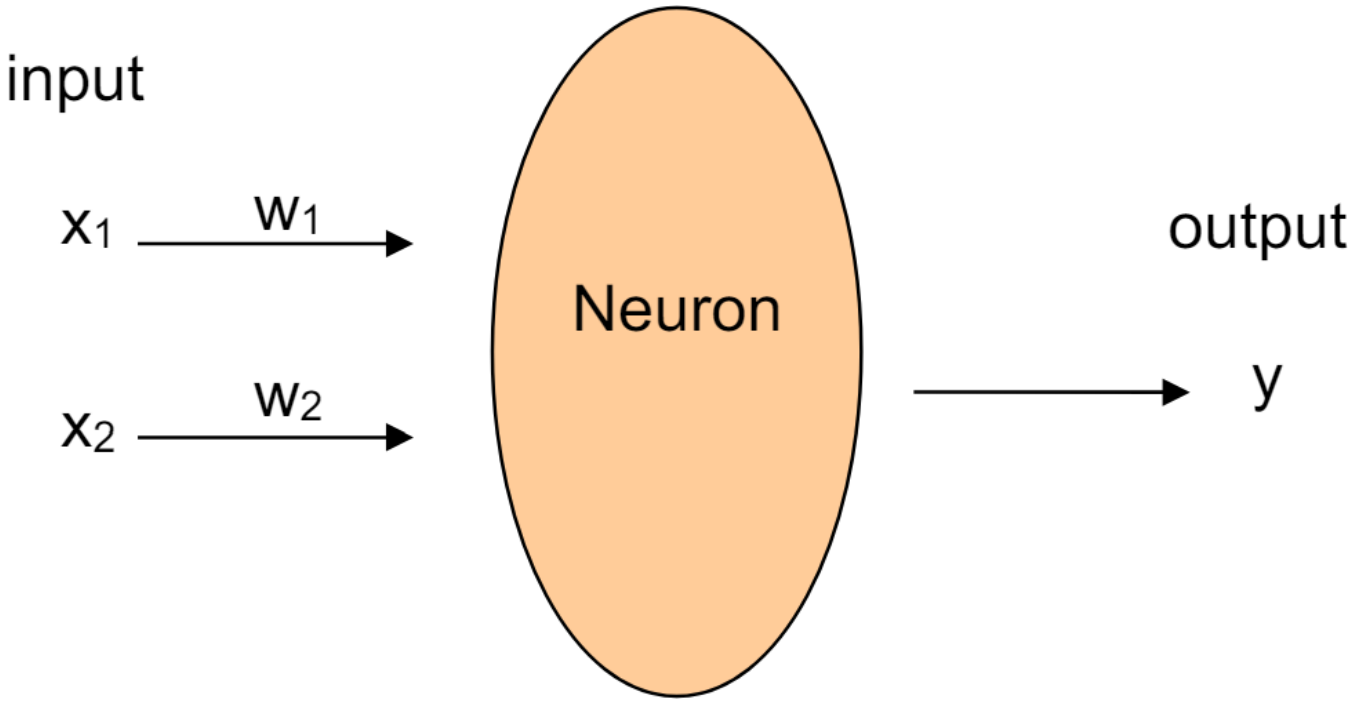
Map "No" to 0 and "Yes" to 1.

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

Q1.b.ii

You cannot use XLMiner in this part.

Consider a neural network containing a single neuron where x_1 = "No. of Children", x_2 = "No. of Siblings" and y = "Go_Disneyland".



Initially, we set the values of w_1, w_2 and b to be 0.3 where b is a bias value in the neuron.

- Suppose the learning rate is denoted by α. Let α = 0.6.
- Suppose we adopt the threshold function as an activation function.
- Please try to train the neural network with five instances by the following inputs in the given sequence.

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

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

- 1st step
  - (w_1, w_2, b) = (0.3, 0.3, 0.3)
  - y' = 0.3 * 2 + 0.3 * 0 + 0.3 = 0.9
  - y = 1 (y' >= 0)
  - diff = d - y = 0 - 1 = -1
  - w_1 = 0.3 + 0.6 * diff * 2 = 0.3 - 1.2 = -0.9
  - w_2 = 0.3 + 0.6 * diff * 0 = 0.3
  - b = 0.3 + 0.6 * diff = -0.3
  - (w_1, w_2, b) = (-0.9, 0.3, -0.3)
- 2nd step
  - (w_1, w_2, b) = (-0.9, 0.3, -0.3)
  - y' = -0.9 * 0 + 0.3 * 2 - 0.3 = 0.3
  - y = 1 (y' >= 0)
  - diff = d - y = 0 - 1 = -1
  - w_1 = -0.9 + 0.6 * diff * 0 = -0.9

    - w_2 = 0.3 + 0.6 * diff * 2 = 0.3 - 1.2 = -0.9
    - b = -0.3 + 0.6 * diff = -0.9
    - (w_1, w_2, b) = (-0.9, -0.9, -0.9)
- 3rd step
    - (w_1, w_2, b) = (-0.9, -0.9, -0.9)
    - y' = -0.9 * 4 + -0.9 * 2 - 0.9 = -6.3
    - y = 0 (y' < 0)
    - diff = d - y = 1 - 0 = 1
    - w_1 = -0.9 + 0.6 * diff * 4 = 1.5
    - w_2 = -0.9 + 0.6 * diff * 2 = 0.3
    - b = -0.9 + 0.6 * diff = -0.3
    - (w_1, w_2, b) = (1.5, 0.3, -0.3)
- 4th step
    - (w_1, w_2, b) = (1.5, 0.3, -0.3)
    - y' = 1.5 * 2 + 0.3 * 4 - 0.3 = 3.9
    - y = 1 (y' >= 0)
    - diff = d - y = 1 - 1 = 0
    - diff is 0, can skip updating, as updating does not change any values
    - (w_1, w_2, b) = (1.5, 0.3, -0.3)
- 5th step
    - (w_1, w_2, b) = (1.5, 0.3, -0.3)
    - y' = 1.5 * 2 + 0.3 * 0 - 0.3 = 2.7
    - y = 1 (y' >= 0)
    - diff = d - y = 0 - 1 = -1
    - w_1 = 1.5 + 0.6 * diff * 2 = 0.3
    - w_2 = 0.3 + 0.6 * diff * 0 = 0.3
    - b = -0.3 + 0.6 * diff = -0.9
    - (w_1, w_2, b) = (0.3, 0.3, -0.9)
- The final values are (w_1, w_2, b) = (0.3, 0.3, -0.9).

Q2

[20 Marks]

Q2.a

Consider the four 2-dimensional data points:

> a: (6, 6), b: (8, 8), c: (5, 9) and d: (9, 5)

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 that the information loss during the transformation is minimized.

Please illustrate with the above example when L=2 and K=1. Please write the numbers up to 2 decimal places. You cannot use XLMiner in this part.

- total = a + b + c + d = (6, 6) + (8, 8) + (5, 9) + (9, 5) = (28, 28)
- mean = total / 4 = (7, 7)
- modify data points
    - a' = a - mean = (6, 6) - (7, 7) = (-1, -1)

- `b' = b - mean = (8, 8) - (7, 7) = (1, 1)`
- `c' = c - mean = (5, 9) - (7, 7) = (-2, 2)`
- `d' = d - mean = (9, 5) - (7, 7) = (2, -2)`

- `covariance matrix:`

$$
\begin{aligned}
&\text{covariance matrix} \\
&= \frac{1}{4}\begin{bmatrix} -1 & 1 & -2 & 2 \\ -1 & 1 & 2 & -2 \end{bmatrix}\begin{bmatrix} -1 & -1 \\ 1 & 1 \\ -2 & 2 \\ 2 & -2 \end{bmatrix} \\
&= \frac{1}{4}\begin{bmatrix} 10 & -6 \\ -6 & 10 \end{bmatrix} \\
&= \begin{bmatrix} 2.5 & -1.5 \\ -1.5 & 2.5 \end{bmatrix}
\end{aligned}
$$

- `eigenvalues` λ`:`

$$
\begin{aligned}
\begin{vmatrix} 2.5-\lambda & -1.5 \\ -1.5 & 2.5-\lambda \end{vmatrix} &= 0 \\
(2.5-\lambda)^2 - (-1.5)^2 &= 0 \\
6.25 - 5\lambda + \lambda^2 - 2.25 &= 0 \\
\lambda^2 - 5\lambda + 4 &= 0 \\
(\lambda - 1)(\lambda - 4) &= 0 \\
\lambda &= 1 \text{ or } 4
\end{aligned}
$$

- `eigenvectors` $\vec{x} = [x_1, x_2]^{\mathrm{T}}$`:`

When $\lambda = 1$,

$$
\begin{aligned}
\begin{bmatrix} 2.5 & -1.5 \\ -1.5 & 2.5 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} &= \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\
\begin{bmatrix} 2.5x_1 - 1.5x_2 \\ -1.5x_1 + 2.5x_2 \end{bmatrix} &= \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\
\begin{bmatrix} 1.5x_1 - 1.5x_2 \\ -1.5x_1 + 1.5x_2 \end{bmatrix} &= \begin{bmatrix} 0 \\ 0 \end{bmatrix} \\
\begin{bmatrix} 1.5x_1 \\ -1.5x_1 \end{bmatrix} &= \begin{bmatrix} 1.5x_2 \\ -1.5x_2 \end{bmatrix} \\
\begin{bmatrix} x_1 \\ x_1 \end{bmatrix} &= \begin{bmatrix} x_2 \\ x_2 \end{bmatrix} \\
x_1 &= x_2
\end{aligned}
$$

Choose $\vec{x}$ such that its norm is 1.

$$
\vec{x} = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}
$$

When $\lambda = 4$,

$$
\begin{aligned}
\begin{bmatrix} 2.5 & -1.5 \\ -1.5 & 2.5 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} &= 4\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\
\begin{bmatrix} 2.5x_1 - 1.5x_2 \\ -1.5x_1 + 2.5x_2 \end{bmatrix} &= \begin{bmatrix} 4x_1 \\ 4x_2 \end{bmatrix} \\
\begin{bmatrix} -1.5x_1 - 1.5x_2 \\ -1.5x_1 - 1.5x_2 \end{bmatrix} &= \begin{bmatrix} 0 \\ 0 \end{bmatrix} \\
\begin{bmatrix} -1.5x_1 \\ -1.5x_1 \end{bmatrix} &= \begin{bmatrix} 1.5x_2 \\ 1.5x_2 \end{bmatrix} \\
\begin{bmatrix} x_1 \\ x_1 \end{bmatrix} &= \begin{bmatrix} -x_2 \\ -x_2 \end{bmatrix} \\
x_1 &= -x_2
\end{aligned}
$$

Choose $\vec{x}$ such that its norm is 1.

$$
\vec{x} = \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}
$$

- sort the eigenvectors in descending eigenvalues:

$$\vec{x}_1 = \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix} \qquad (\lambda = 4)$$
$$\vec{x}_2 = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} \qquad (\lambda = 1)$$

- combine the sorted eigenvectors to make a transformation matrix:

$$\Phi = \begin{bmatrix} \vec{x}_1^{\mathrm{T}} \\ \vec{x}_2^{\mathrm{T}} \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$$

- transform the original data points:

$$a'_{\text{new}} = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} \begin{bmatrix} 6 \\ 6 \end{bmatrix} = \begin{bmatrix} 0 \\ 12\sqrt{2} \end{bmatrix}$$
$$b'_{\text{new}} = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} \begin{bmatrix} 8 \\ 8 \end{bmatrix} = \begin{bmatrix} 0 \\ 16\sqrt{2} \end{bmatrix}$$
$$c'_{\text{new}} = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} \begin{bmatrix} 5 \\ 9 \end{bmatrix} = \begin{bmatrix} -2\sqrt{2} \\ 14\sqrt{2} \end{bmatrix}$$
$$d'_{\text{new}} = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} \begin{bmatrix} 9 \\ 5 \end{bmatrix} = \begin{bmatrix} 2\sqrt{2} \\ 14\sqrt{2} \end{bmatrix}$$

- only keep the first K=1 coordinates of each data point:

$$a_{\text{new}} = [0]$$
$$b_{\text{new}} = [0]$$
$$c_{\text{new}} = [-2\sqrt{2}]$$
$$d_{\text{new}} = [2\sqrt{2}]$$

Therefore, the new data points are:

- a: (0)
- b: (0)
- c: (-2.83)
- d: (2.83)

Q2.b

Consider the four 2-dimensional data points:

> a: (2, 6), b: (3, 3), c: (5, 5) and d: (6, 2)

Similar to (a), we can make use of PCA for dimensionality reduction.

Please use XLMiner to find the transformed data points when L=2 and K=1. Please write the numbers up to 2 decimal places. You do not need to submit any softcopy related to XLMiner. You just need to write down the answer without any explanation.

- a: (2.83)
- b: (0)
- c: (0)

- d: (-2.83)

Record ID	Comp1
Record 1	2.828427125
Record 2	0
Record 3	0
Record 4	-2.828427125

Q3

[20 Marks]

In this question, there is no need to use XLMiner.

In the view materialization method discussed in class, the constraint is given by the number of views that can be materialized. However, different views are of different sizes and the same number of views may not occupy the same amount of storage. In a more realistic problem setting, we are given available memory of size X. We still assume the equal probability of querying of each possible view. Please propose a solution for this problem setting.

Given input:

- X = available memory size

Run the following algorithm:

- Set S to {top view}.
- While the sum of sizes of views in S is less than X:
  - Find the view V with the highest benefit given the views in S is materialized B(V, S).
  - Set S' to the union of S and {V}.
  - If the sum of sizes of views in S' is less than or equal to X:
    - Set S to S'.
  - Else:
    - Stop the while statement.
- The resulting S is the solution and is also a greedy selection.

Q4

[20 Marks]

In this question, there is no need to use XLMiner.

Assume that there are four web sites A, B, C and D. Suppose that A and B point to each other, and C and D point to each other. Furthermore, A points to C.

Q4.a

What is the stochastic matrix created by the page rank method for the four sites?

The row/column order is (A, B, C, D).

$$M = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0.5 & 0 & 0 & 0 \\ 0.5 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

Q4.b

What equation will be solved to decide on the ranking, taking into the consideration of the possibility of the spider trap problem? For the ease of marking, please use the equation taught in the class (though it is possible have other forms of equations).

Let a be a number between 0 and 1 (inclusive). The equation is:

$$r_{\text{next}} = (1-a)Mr_{\text{current}} + \begin{bmatrix} a \\ a \\ a \\ a \end{bmatrix}$$

This equation makes the resulting importance values more reasonable even if there is a spider trap.

For this question, we will use a = 0.2, as taught in the class, so the equation is

$$r_{\text{next}} = 0.8Mr_{\text{current}} + \begin{bmatrix} 0.2 \\ 0.2 \\ 0.2 \\ 0.2 \end{bmatrix}$$

Initialize the initial importance values with

$$r_{\text{initial}} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

.

Q4.c

What will be the resulting ranking? Please use the equation given in (b) for doing this part. Please show your steps where all numbers are shown up to 2 decimal places.

Using a = 0.2 and the stochastic matrix

$$M = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0.5 & 0 & 0 & 0 \\ 0.5 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

:

Iteration No.	**[A B C D]^T**
1	[1.00 1.00 1.00 1.00]^T
2	[1.00 0.60 1.40 1.00]^T
3	[0.68 0.60 1.40 1.32]^T

Iteration No.	$[A\ B\ C\ D]^T$
4	$[0.68\ 0.47\ 1.53\ 1.32]^T$
5	$[0.58\ 0.47\ 1.53\ 1.42]^T$
6	$[0.58\ 0.43\ 1.57\ 1.42]^T$
7	$[0.54\ 0.43\ 1.57\ 1.46]^T$
8	$[0.54\ 0.42\ 1.58\ 1.46]^T$
9	$[0.54\ 0.42\ 1.58\ 1.46]^T$

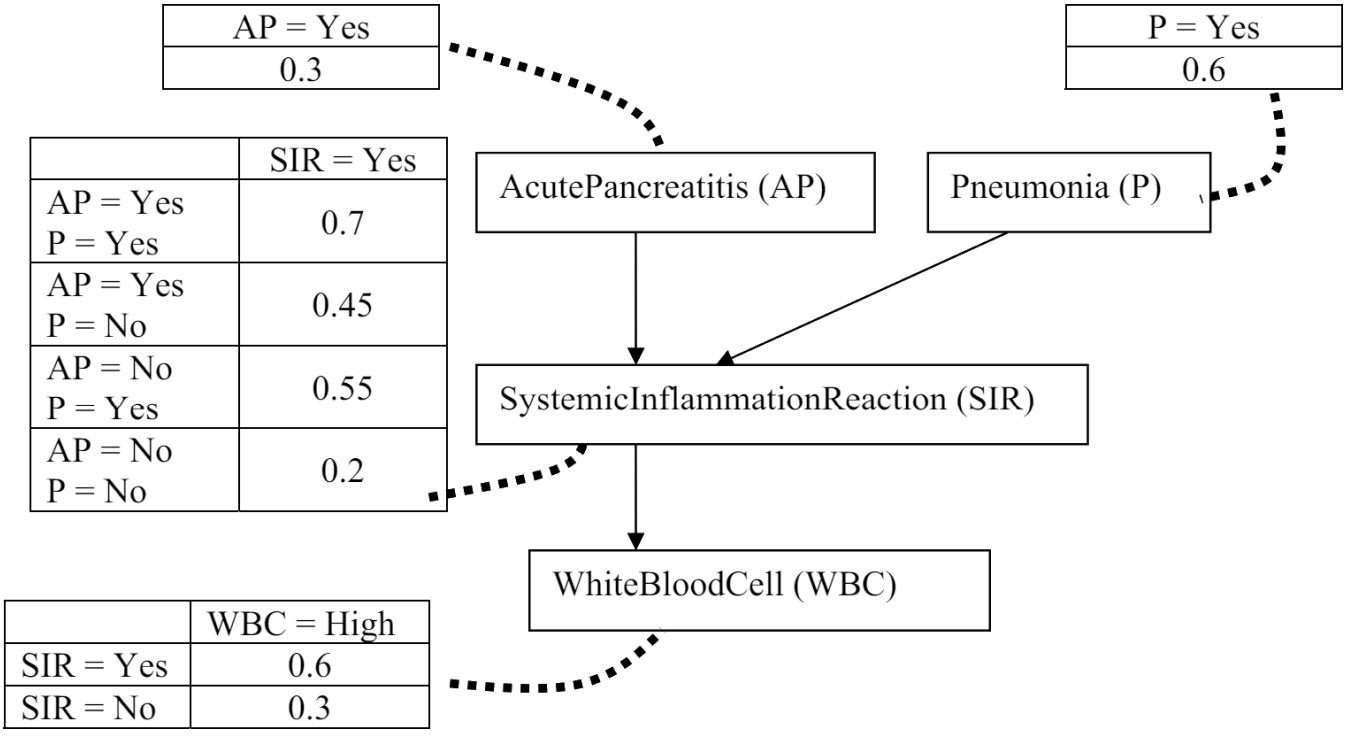
The importance values has stopped changing, so the matrix has converged. Rank the pages by descending importance: C > D > A > B.

Q5

[20 Marks]

In this question, you cannot use XLMiner.

We have the following Bayesian Belief Network.



Suppose that there is a new patient. We know that

1. he has acute pancreatitis
2. he has pneumonia
3. his result of white blood cell is high

We would like to know whether he is likely to have systemic inflammation reaction.

Acute Pancreatitis	Pneumonia	White Blood Cell	Systemic Inflammation Reaction
Yes	Yes	High	?

Q5.a

Please use Bayesian Belief Network classifier with the use of Bayesian Belief Network to predict whether he is likely to have systemic inflammation reaction.

$$
\begin{aligned}
&P(\text{AP} = \text{Yes}, \text{P} = \text{Yes}) \\
&= P(\text{AP} = \text{Yes})P(\text{P} = \text{Yes}) && \text{(independence)} \\
&= 0.3 \cdot 0.6 \\
&= 0.18
\end{aligned}
$$

$$
\begin{aligned}
&P(\text{WBC} = \text{High}|\text{AP} = \text{Yes}, \text{P} = \text{Yes}) \\
&= \sum_{\text{SIR} \in \{\text{Yes}, \text{No}\}} P(\text{WBC} = \text{High}|\text{SIR})P(\text{SIR}|\text{AP} = \text{Yes}, \text{P} = \text{Yes}) \\
&= 0.6 \cdot 0.7 + 0.3 \cdot (1 - 0.7) \\
&= 0.42 + 0.09 \\
&= 0.51
\end{aligned}
$$

$$
\begin{aligned}
&P(\text{WBC} = \text{High}|\text{SIR} = \text{Yes}, \text{AP} = \text{Yes}, \text{P} = \text{Yes}) \\
&= P(\text{WBC} = \text{High}|\text{SIR} = \text{Yes}) && \text{(conditional independence)} \\
&= 0.6
\end{aligned}
$$

$$
\begin{aligned}
&P(\text{SIR} = \text{Yes}, \text{AP} = \text{Yes}, \text{P} = \text{Yes}) \\
&= P(\text{SIR} = \text{Yes}|\text{AP} = \text{Yes}, \text{P} = \text{Yes})P(\text{AP} = \text{Yes}, \text{P} = \text{Yes}) \\
&= 0.7 \cdot 0.18 \\
&= 0.126
\end{aligned}
$$

$$
\begin{aligned}
&P(\text{SIR} = \text{Yes}, \text{AP} = \text{Yes}, \text{P} = \text{Yes}, \text{WBC} = \text{High}) \\
&= P(\text{WBC} = \text{High}|\text{SIR} = \text{Yes}, \text{AP} = \text{Yes}, \text{P} = \text{Yes})P(\text{SIR} = \text{Yes}, \text{AP} = \text{Yes}, \text{P} = \text{Yes}) \\
&= 0.6 \cdot 0.126 \\
&= 0.0756
\end{aligned}
$$

$$
\begin{aligned}
&P(\text{AP} = \text{Yes}, \text{P} = \text{Yes}, \text{WBC} = \text{High}) \\
&= P(\text{WBC} = \text{High}|\text{AP} = \text{Yes}, \text{P} = \text{Yes})P(\text{AP} = \text{Yes}, \text{P} = \text{Yes}) \\
&= 0.51 \cdot 0.18 \\
&= 0.0918
\end{aligned}
$$

$$
\begin{aligned}
&P(\text{SIR} = \text{Yes}|\text{AP} = \text{Yes}, \text{P} = \text{Yes}, \text{WBC} = \text{High}) \\
&= \frac{P(\text{SIR} = \text{Yes}, \text{AP} = \text{Yes}, \text{P} = \text{Yes}, \text{WBC} = \text{High})}{P(\text{AP} = \text{Yes}, \text{P} = \text{Yes}, \text{WBC} = \text{High})} \\
&= \frac{0.0756}{0.0918} \\
&= 0.823529412 \text{ (cor. to 9 sig. fig.)} \\
&= 0.824 \text{ (cor to 3 sig. fig.)}
\end{aligned}
$$

$$
\begin{aligned}
&P(\text{SIR} = \text{No}|\text{AP} = \text{Yes}, \text{P} = \text{Yes}, \text{WBC} = \text{High}) \\
&= 1 - P(\text{SIR} = \text{Yes}|\text{AP} = \text{Yes}, \text{P} = \text{Yes}, \text{WBC} = \text{High}) \\
&= 1 - 0.823529412 \\
&= 0.176470588 \\
&= 0.176 \text{ (cor. to 3 sig. fig.)}
\end{aligned}
$$

$$
\begin{aligned}
&\because P(\text{SIR} = \text{Yes}|\text{AP} = \text{Yes}, \text{P} = \text{Yes}, \text{WBC} = \text{High}) \\
&> P(\text{SIR} = \text{No}|\text{AP} = \text{Yes}, \text{P} = \text{Yes}, \text{WBC} = \text{High}) \\
&\therefore \text{Yes, he is likely to have systemic inflammation reaction.}
\end{aligned}
$$

Q5.b

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

The disadvantage is that the complexity of the network increases very quickly with the number of variables, increasing its space and time complexity quickly, which makes it

impractical if there are lot of variables. Also, the network cannot have cycles, so it cannot model cyclic relationship between variables.