AU 332 Artificial Intelligence: Principles and Techniques

By: ZheHao Huang (518021910660)

HW#: 4

November 23, 2020

1 Bayesian Networks

In this assignment we need to implement code for computing exact inferences in Bayesian networks of discrete random variables using variable elimination. Two functions **readFactorTable** and **readFactorTableFromData** are already written, which used to build conditional probability tables, represented as factors. To successfully finish this assignment, we first need to implement the code part to make our inferences in the Bayesian network correctly. In the following part, I will briefly introduce how I implement the functions to use **Pandas** to build up the process of the exact inferences. More details can be found in my **BayesianNetworks.py**.

1.1 joinFactors(Factor1, Factor2)

Instruction: Should return a factor table that is the join of factor 1 and 2. The join of two factors is a valid operation.

Implementation: Actually, there are two situations we will meet with when joining two factor tables.

1. The intersection of the columns of factor 1 and 2 is empty. We need to join each row in factor table 1 with each row in factor table 2.

2. The intersection of the columns of factor 1 and 2 is not empty. We need to join the row in factor table 1 with the row in factor table 2 that have the same value of the intersection columns.

To combine two situations, I use a simple trick that I add a column named **common** with the same value **1** to both factor tables. And then compute the intersection of the columns of factor tablt 1 and 2 which will at least has the column of **common**. Using the function **merge()** we can easily get the expected result and join two factor tables successfully. Finally we just need to drop the redundant columns to return the joined factor table.

1.2 marginalizeFactor(factorTable, hiddenVar)

Instruction: Should return a factor table that marginalizes margVar out of it. Assume that **hiddenVar** is on the left side of the conditional.

Implementation: To marginalize the hidden variable, we just need to sum the rows with the same value of the columns without hidden variable. And the function **groupby()** and its **sum()** can help us do the grouping and summation.

1.3 marginalizeNetworkVariables(bayesNet, hiddenVar)

Instruction: Should return a bayesNet and marginalizes out a list of variables **hiddenVar** from each table in the bayesNet.

Implementation: In this function, we need to marginalize a list of hidden variables in a bayesNet without considering the most efficient order of the variable elimination. For each hidden variable we first join all tables in the bayesNet that has that hidden variable, and then use the function **marginalizeFactor()** above to eliminate the hidden variable from the table.

1.4 evidenceUpdateNet(bayesNet, evidenceVars, evidenceVals)

Instruction: Should return a bayesNet without the values that not in the **evidenceVals**. We do not need to normalize the factors to make them proper probabilities.

Implementation: In this function, I use a simple method that I iterate each row of each table in the bayesNet to check whether it shares some columns in the **evidenceVars** and has the same values of the **evidenceVals**. In the implementation, there also has a trick that the data type of the evidenceVal and the value in the table may be different, so we need to tranform the data type of the evidenceVal to match the value in the table.

1.5 inference(bayesNet, hiddenVar, evidenceVars, evidenceVals)

Instruction: This function takes in a Bayesian network and should return a single joint probability table resulting from the given set of evidence variables and marginalizing a set of hidden variables. The final table should be a proper probability table (entries sum to 1). The hidden variables shown in **hiddenVar** should not be in the returned table.

Implementation: I first implement a helper function to normalize the factor table to make the entries sum to 1, which will be used in the final step of the inference. And the inference can be easliy build up by the function **marginalizeNetworkVariables()** and **evidenceUpdateNet()** written above. Finally return a normalized factor table to make the exact inference.

1.6 The screen-shot of BayesNetworkTestScript.py

In order to make sure that the code of the inference of bayesian network part is correct. I run the test script of the given bayes network and achieve exactly the same result as following(Figure 1).

```
inference starts
   gauge  probs
0      0  0.315
1      1  0.685
   fuel  gauge  probs
0     0      0   0.81
1     0      1   0.19
   fuel  gauge     probs
0     1      0  0.742857
1     0      0  0.257143
   battery  fuel  gauge     probs
0        0     1      0  0.888889
1        0     0      0  0.111111
inference ends
income dataframe is
      probs  income
0  0.050848       1
1  0.059429       2
2  0.074042       3
3  0.094414       4
4  0.116356       5
5  0.150725       6
6  0.164430       7
7  0.289755       8
   smoke  diabetes  long_sit  exercise     probs
0      1         1         1         2  0.136815
1      1         2         1         2  0.008916
2      1         3         1         2  0.837218
3      1         4         1         2  0.017052
```

Figure 1: Screen-shot of the BayesNetworkTestScripy.py output

2 Analyzing risk factors for certain health problems

In this part we need to analyzing risk factors for certain health problems (heart disease, stroke, heart attack, diabetes). The data is from the 2015 Behavioral Risk Factor Surveillance System (BRFSS) survey, which is run by the Centers for Disease Control (CDC). The distilled data is in the spreadsheet **RiskFactorData.csv**. The bayesian network can be visualized as follows:

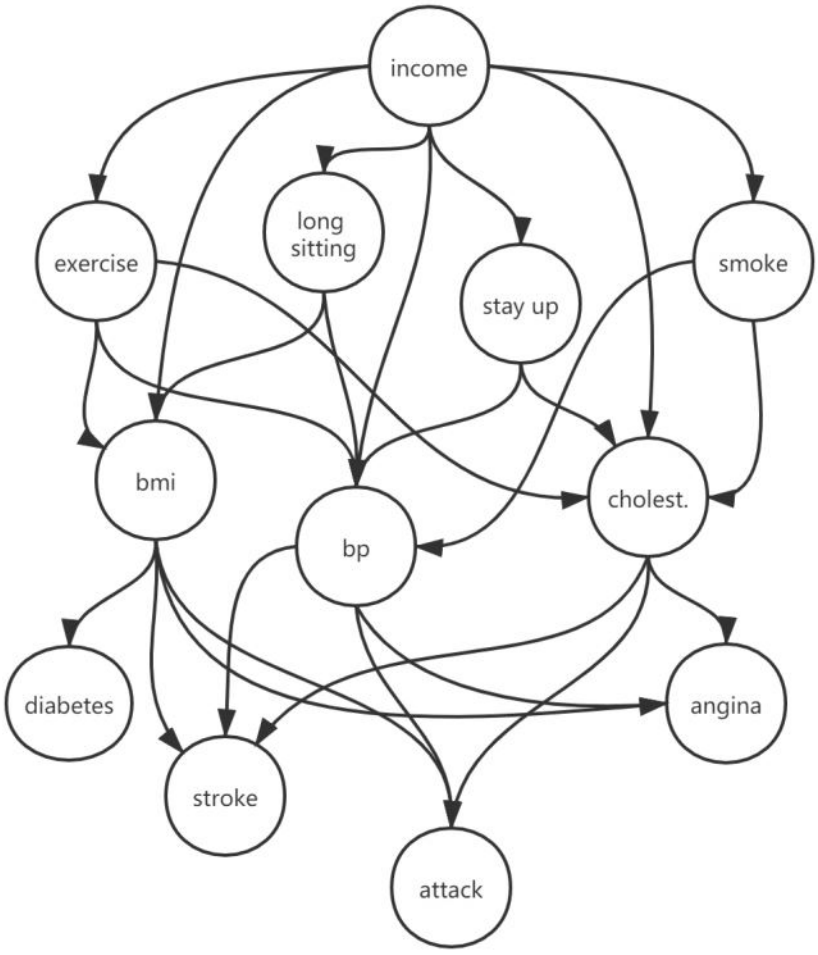

Figure 2: The bayesian network of the risk factors and the health problems

2.1

Instruction: Create the following Bayesian network to analyze the survey results. You will want to use the provided function **readFactorTablefromData**. What is the size (in terms of the number of probabilities needed) of this network? Alternatively, what is the total number of probabilities needed to store the full joint distribution?

Solution:

We can derive 12 CPT tables from the bayesian network of risk factors and health problems(Figure 2):

$$P(income) \tag{1}$$
$$P(exercise|income) \tag{2}$$
$$P(longsitting|income) \tag{3}$$
$$P(stayup|income) \tag{4}$$
$$P(smoke|income) \tag{5}$$
$$P(bmi|exercise, income, longsitting) \tag{6}$$
$$P(bp|exercise, longsitting, stayup, smoke, income) \tag{7}$$
$$P(cholest.|exercise, stayup, income, smoke) \tag{8}$$
$$P(diabetes|bmi) \tag{9}$$
$$P(stroke|bmi, bp, cholest.) \tag{10}$$
$$P(attack|bmi, bp, cholest.) \tag{11}$$
$$P(angina|bmi, bp, cholest.) \tag{12}$$

And the total number of probabilities needed to store the full joint distribution is the summation of the number of these 12 CPT tables entries, which is 1048.(Figure 3)

```
My risk factor net is of 1048 probabilities.
```

Figure 3: The numebr of probabilities

2.2

Instruction: For each of the four health outcomes (diabetes, stroke, heart attack, angina), answer the following by querying your network (using your infer function):
(a) What is the probability of the outcome if I have bad habits (smoke , don't exercise,long sitting and stay up)? How about if I have good habits?
(b) What is the probability of the outcome if I have poor health (high blood pressure,high cholesterol, and overweight)? What if I have good health (low blood pressure, low cholesterol, and normal weight)?

Solution:
(a)The probability of the health outcomes (diabetes, stroke, heart attack, angina) if I have bad and good habits are as follows(Figure 4). (Back up by Figure 15 16)

The probability of the outcomes if I have bad habits						The probability of the outcomes if I have good habits					
diabetes	exercise	smoke	long_sit	stay_up	probability	diabetes	exercise	smoke	long_sit	stay_up	probability
yes	no	yes	yes	yes	0.179597	yes	yes	no	no	no	0.075195
only during pregnancy	no	yes	yes	yes	0.008754	only during pregnancy	yes	no	no	no	0.009409
no	no	yes	yes	yes	0.79116	no	yes	no	no	no	0.903426
pre-diabetic	no	yes	yes	yes	0.020489	pre-diabetic	yes	no	no	no	0.01197
stroke	exercise	smoke	long_sit	stay_up	probability	stroke	exercise	smoke	long_sit	stay_up	probability
yes	no	yes	yes	yes	0.053214	yes	yes	no	no	no	0.029202
no	no	yes	yes	yes	0.946786	no	yes	no	no	no	0.970798
attack	exercise	smoke	long_sit	stay_up	probability	attack	exercise	smoke	long_sit	stay_up	probability
yes	no	yes	yes	yes	0.085704	yes	yes	no	no	no	0.036655
no	no	yes	yes	yes	0.914296	no	yes	no	no	no	0.963345
angina	exercise	smoke	long_sit	stay_up	probability	angina	exercise	smoke	long_sit	stay_up	probability
yes	no	yes	yes	yes	0.09542	yes	yes	no	no	no	0.03551
no	no	yes	yes	yes	0.90458	no	yes	no	no	no	0.96449

Figure 4: The probability of the health outcomes based on habits

Comparing the probability of the health problems in the condition of having bad or good habits, it can be concluded that those who have bad habits are more likely developing health problems than those who have good habits.

(b)The probability of the health outcomes (diabetes, stroke, heart attack, angina) if I have bad and good health are as follows(Figure 5). (Back up by Figure 17 18).

The probability of the outcomes if I have poor health				
diabetes	bp	cholesterol	bmi	probability
yes	high	yes	overweight	0.115423
only during pregnancy	high	yes	overweight	0.007662
no	high	yes	overweight	0.860873
pre-diabetic	high	yes	overweight	0.016043
stroke	bp	cholesterol	bmi	probability
yes	high	yes	overweight	0.082686
no	high	yes	overweight	0.917314
attack	bp	cholesterol	bmi	probability
yes	high	yes	overweight	0.140784
no	high	yes	overweight	0.859216
angina	bp	cholesterol	bmi	probability
yes	high	yes	overweight	0.161608
no	high	yes	overweight	0.838392

The probability of the outcomes if I have good health				
diabetes	bp	cholesterol	bmi	probability
yes	low	no	noraml weight	0.05771
only during pregnancy	low	no	noraml weight	0.009543
no	low	no	noraml weight	0.922194
pre-diabetic	low	no	noraml weight	0.010553
stroke	bp	cholesterol	bmi	probability
yes	low	no	noraml weight	0.01446
no	low	no	noraml weight	0.98554
attack	bp	cholesterol	bmi	probability
yes	low	no	noraml weight	0.016161
no	low	no	noraml weight	0.983839
angina	bp	cholesterol	bmi	probability
yes	low	no	noraml weight	0.013326
no	low	no	noraml weight	0.986674

Figure 5: The probability of the health outcomes based on health

Comparing the probability of the health problems in the condition of having bad or good health, it can be concluded that those who have bad health are more likely developing health problems than those who have good health.

Vertically comparing the probability of the health problems in different habits and health conditions, the health condition exert more influence on health problems than the habits condition because the difference between having good and bad health are more significant that bad and good habits.

2.3

Instruction: Evaluate the effect a person's income has on their probability of having one of the four health outcomes (diabetes, stroke, heart attack, angina). For each of these four outcomes, plot their probability given income status (your horizontal axis should be i = 1, 2, . . . , 8, and your vertical axis should be P(y = 1|income = i), where y is the outcome). What can you conclude?

Solution:
The probability of the four health outcomes (diabetes, stroke, heart attack, angina) on different income conditions can be plotted as follows(Figure 6 7 8 9).(Back up by Figure 19 20 21 22 23)

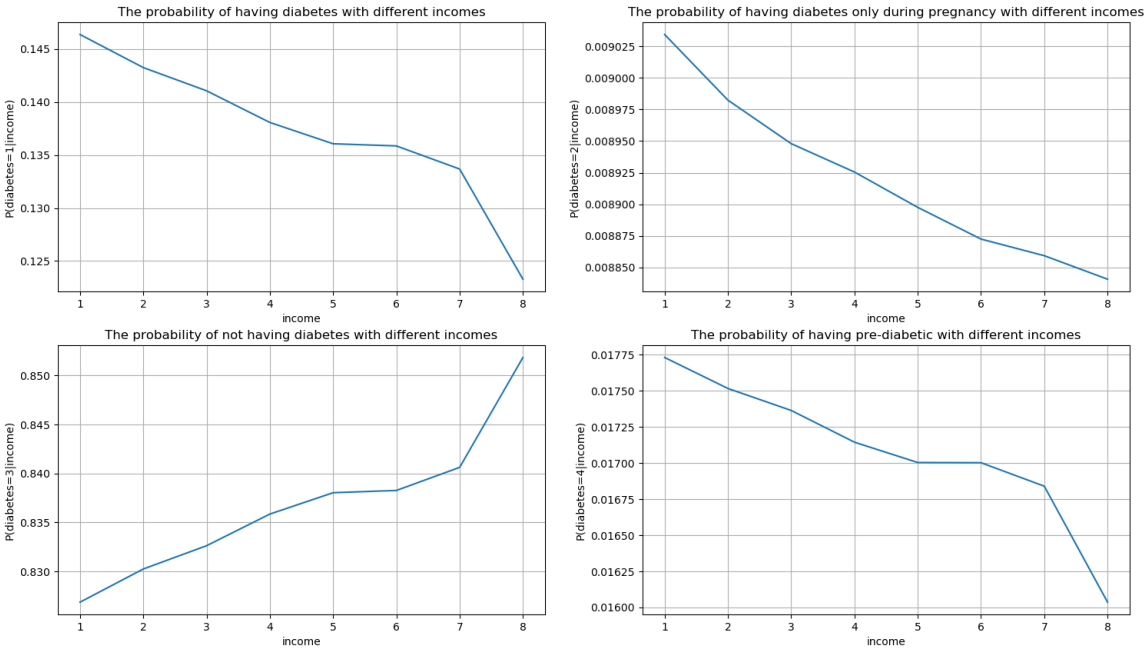

Figure 6: The probability of diabetes based on income

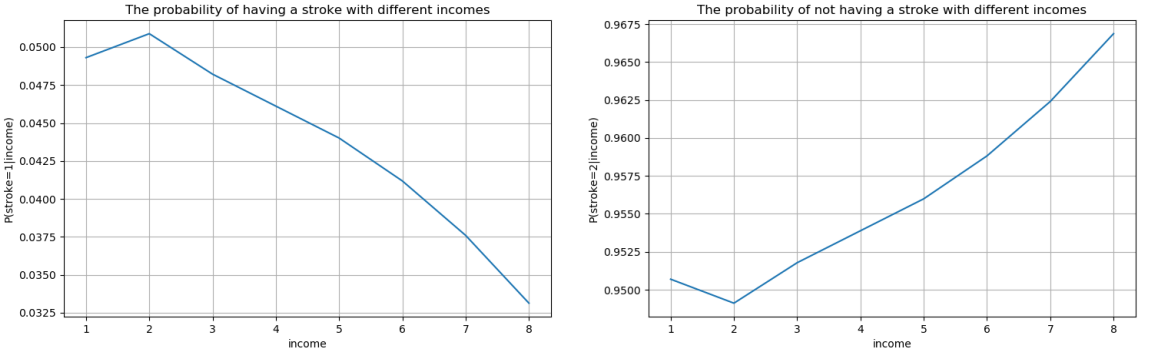

Figure 7: The probability of stroke based on income

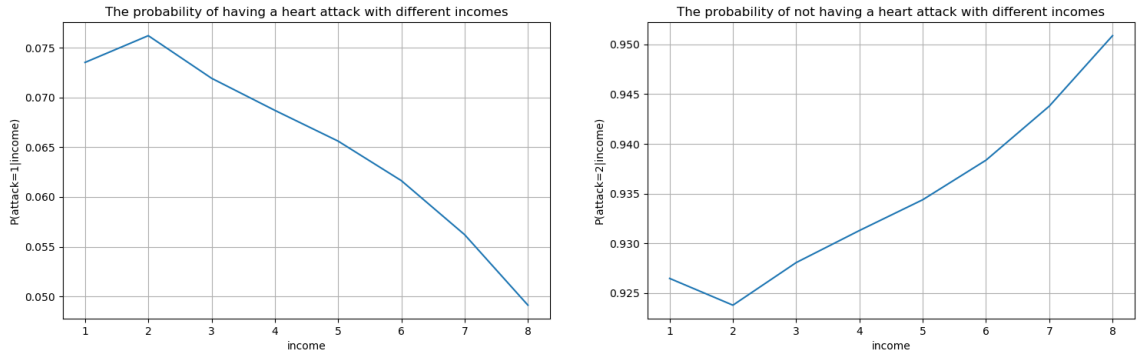

Figure 8: The probability of heart attack based on income

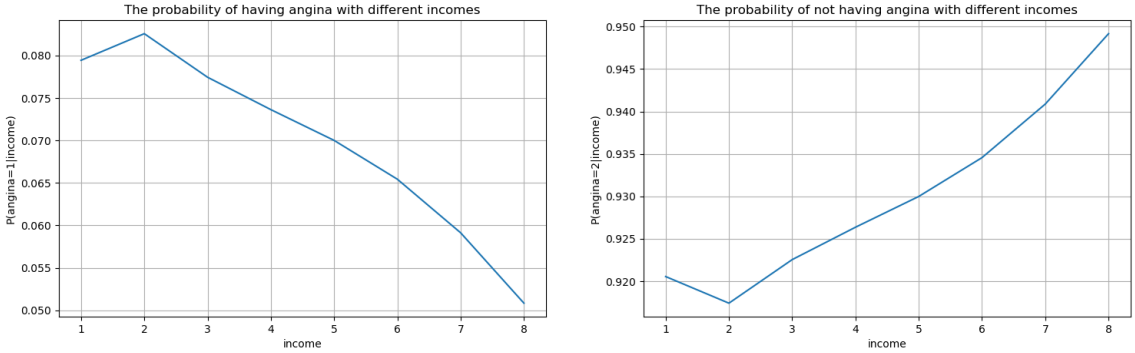

Figure 9: The probability of angina based on income

We can conclude that:

1. Looking at the curve of the probability of diabetes with different incomes, it's obvious that the more income you gain, the less likely you will develop diabetes whether only during pregnancy or not and the pre-diabetes, and the highest income level 8($>$ \$75,000) has a quite lower probability of having the diabetes than other less income levels.

2. Looking at the curve of the probability of the health problems of stroke, heart attack and angina, they share a common factor that those who are the income level 2(\$10, 000-\$15, 000) have the highest probabilities of having stroke, heart attack and angina. And the income level 1 ($<$ \$10, 000) has the second highest probabilities of these health problems. From income level 3 to 8, the probabilities of those three health outcomes decrease , and those who are the income level still have the lowest probabilities of developing those health problems.

2.4

Instruction: Notice there are no links in the graph between the habits and the outcomes. What assumption is this making about the effects of smoking and exercise on health problems? Let's test the validity of these assumptions. Create a second Bayesian network as above, but add edges from smoking to each of the four outcomes and edges from exercise to each of the four outcomes. Now redo the queries in Question 2. What was the effect, and do you think the assumptions of the first graph were valid or not?

Solution:

The assumption is that the smoking and exercise do not directly affect the health problems, so whether having bad habits of smoking and not exercising or good habits will not significantly influence the probabilities of the health outcomes. But if there are the edges from smoking and exercise to each of the four outcomes, that means the habits of smoking and exercise are very relevant to the health problems.

The following tables(Figure 10 11 12 13) respectively show the probabilities of the four health problems both in the condition that there are edges from smoking and exercise to the health outcomes and there are no links between them. (Back up by Figure 24 25 26 27)

The probability of the outcomes if I have bad habits in two net(without edges and adding edges)						
diabetes	exercise	smoke	long_sit	stay_up	probability(without edges)	probability(adding edges)
yes	no	yes	yes	yes	0.179597	0.245992
only during pregnancy	no	yes	yes	yes	0.008754	0.006928
no	no	yes	yes	yes	0.79116	0.723721
pre-diabetic	no	yes	yes	yes	0.020489	0.023359
stroke	exercise	smoke	long_sit	stay_up	probability(without edges)	probability(adding edges)
yes	no	yes	yes	yes	0.053214	0.080488
no	no	yes	yes	yes	0.946786	0.919512
attack	exercise	smoke	long_sit	stay_up	probability(without edges)	probability(adding edges)
yes	no	yes	yes	yes	0.085704	0.135301
no	no	yes	yes	yes	0.914296	0.864699
angina	exercise	smoke	long_sit	stay_up	probability(without edges)	probability(adding edges)
yes	no	yes	yes	yes	0.09542	0.138072
no	no	yes	yes	yes	0.90458	0.861928

Figure 10: The probability of the health outcome based on bad habits with and without links

The probability of the outcomes if I have good habits in two net(without edges and adding edges)						
diabetes	exercise	smoke	long_sit	stay_up	probability(without edges)	probability(adding edges)
yes	yes	no	no	no	0.075195	0.056227
only during pregnancy	yes	no	no	no	0.009409	0.01016
no	yes	no	no	no	0.903426	0.92371
pre-diabetic	yes	no	no	no	0.01197	0.009903
stroke	exercise	smoke	long_sit	stay_up	probability(without edges)	probability(adding edges)
yes	yes	no	no	no	0.029202	0.019464
no	yes	no	no	no	0.970798	0.980536
attack	exercise	smoke	long_sit	stay_up	probability(without edges)	probability(adding edges)
yes	yes	no	no	no	0.036655	0.021213
no	yes	no	no	no	0.963345	0.978787
angina	exercise	smoke	long_sit	stay_up	probability(without edges)	probability(adding edges)
yes	yes	no	no	no	0.03551	0.023948
no	yes	no	no	no	0.96449	0.976052

Figure 11: The probability of the health outcome based on good habits with and without links

The probability of the outcomes if I have poor health in two net(without edges and adding edges)					
diabetes	bp	cholesterol	bmi	probability(without edges)	probability(adding edges)
yes	high	yes	overweight	0.115423	0.121241
only during pregnancy	high	yes	overweight	0.007662	0.007492
no	high	yes	overweight	0.860873	0.854769
pre-diabetic	high	yes	overweight	0.016043	0.016498
stroke	bp	cholesterol	bmi	probability(without edges)	probability(adding edges)
yes	high	yes	overweight	0.082686	0.082697
no	high	yes	overweight	0.917314	0.917303
attack	bp	cholesterol	bmi	probability(without edges)	probability(adding edges)
yes	high	yes	overweight	0.140784	0.140083
no	high	yes	overweight	0.859216	0.859917
angina	bp	cholesterol	bmi	probability(without edges)	probability(adding edges)
yes	high	yes	overweight	0.161608	0.161096
no	high	yes	overweight	0.838392	0.838904

Figure 12: The probability of the health outcome based on bad health with and without links

The probability of the outcomes if I have good health in two net(without edges and adding edges)					
diabetes	bp	cholesterol	bmi	probability(without edges)	probability(adding edges)
yes	low	no	noraml weight	0.05771	0.055937
only during pregnancy	low	no	noraml weight	0.009543	0.009697
no	low	no	noraml weight	0.922194	0.924042
pre-diabetic	low	no	noraml weight	0.010553	0.010323
stroke	bp	cholesterol	bmi	probability(without edges)	probability(adding edges)
yes	low	no	noraml weight	0.01446	0.014544
no	low	no	noraml weight	0.98554	0.985456
attack	bp	cholesterol	bmi	probability(without edges)	probability(adding edges)
yes	low	no	noraml weight	0.016161	0.016183
no	low	no	noraml weight	0.983839	0.983817
angina	bp	cholesterol	bmi	probability(without edges)	probability(adding edges)
yes	low	no	noraml weight	0.013326	0.013328
no	low	no	noraml weight	0.986674	0.986672

Figure 13: The probability of the health outcome based on good health with and without links

From the table of the probability based on habits, we can see that the edges from smoking and exercise to the four health outcomes make a big difference on the divergence of the health outcomes with good habits or bad habits. It's obvious that these edges enhance the correlation between smoking and exercise and the four health problems, so with these edges, those who smoking and not exercise are more likely to developing the harmful health diseases than without these edges. So the assumptions are valid.

But if given the evidence of the health condition like blood pressure, bmi or cholesterol, whether there are edges from smoking and exercise to the four health outcomes makes almost no difference to their probabilities.

2.5

Instruction: Also notice there are no edges between the four outcomes. What assumption is this making about the interactions between health problems? Make a third network, starting from the network in Question 4, but adding an edge from diabetes to stroke. For both networks, evaluate the following probabilities:

$$P(stroke = 1|diabetes = 1) and P(stroke = 1|diabetes = 3) \tag{13}$$

Again, what was the effect, and was the assumption about the interaction between diabetes and stroke valid?

Solution:

Noticing that there are no edges between the four outcomes, we can assume that there is little inter-connection between the four health problems, which means whether a person have any health problems or not, it will little influence the probabilities of having other health outcomes.

The probability table of (equation 13) is as follow(Figure 14). (Back up by Figure **??**)

The probability of diabetes with or without an edge from diabetes to stroke			
stroke	diabetes	probability without an edge frome diabetes to stroke	probability with an edge from diabetes to stroke
yes	yes	0.044417	0.076542
yes	no	0.039955	0.034456

Figure 14: The probability of diabetes based on stroke

From the table and comparing the probabilities of having diabetes condition on having a stroke or not having a stroke, we can see that when there is not a link from stroke to diabetes, the probabilities of having diabetes in two conditions are almost the same. But if we add an edge from stroke to diabetes, having a stroke obviously increase the probability of having diabetes. So the assumption above is valid. And in the original bayesian network, we actually do not consider the interaction between the four health problems.

3 Conclusion

In this lab, we first implement the exact inference of bayesian networks by skillfully using **Pandas** to run the process of variable elimination and get the final normalized probability table. That step is the most challenging and most important part of this assignment because first I have never used the data structure of **Pandas** like **Dataframe** and then I need to use the **inference()** function to do the second part of analyzing the risk factors for certain health problems. So I take a quiet long time to make sure I correctly implement the inference process in the first part. But that's really a rewarding task since I learned a new powerful data structure different from matrix in **Numpy** and improved my programming skills.

And in the second part of the this assignment, I realize that we not only need to understand how to successfully implement the algorithm, but also use these methods to actually solve and analyze some real problems. And according to the characteristics the target mission, we need to learn to design sophisticated algorithm to help get the better result.

4 Appendix

```
The probability of the diabetes with the bad habits
    exercise  long_sit  smoke  stay_up  diabetes     probs
0         2         1      1        1         1  0.179597
1         2         1      1        1         2  0.008754
2         2         1      1        1         3  0.791160
3         2         1      1        1         4  0.020489
The probability of the stroke with the bad habits
    stroke  exercise  long_sit  smoke  stay_up     probs
0       1         2         1      1        1  0.053214
1       2         2         1      1        1  0.946786
The probability of the attack with the bad habits
    attack  exercise  long_sit  smoke  stay_up     probs
0       1         2         1      1        1  0.085704
1       2         2         1      1        1  0.914296
The probability of the angina with the bad habits
    angina  exercise  long_sit  smoke  stay_up    probs
0       1         2         1      1        1  0.09542
1       2         2         1      1        1  0.90458
```

Figure 15: 2.1.1

```
The probability of the diabetes with the good habits
    exercise  long_sit  smoke  stay_up  diabetes     probs
0         1         2      2        2         1  0.075195
1         1         2      2        2         2  0.009409
2         1         2      2        2         3  0.903426
3         1         2      2        2         4  0.011970
The probability of the stroke with the good habits
    stroke  exercise  long_sit  smoke  stay_up     probs
0       1         1         2      2        2  0.029202
1       2         1         2      2        2  0.970798
The probability of the attack with the good habits
    attack  exercise  long_sit  smoke  stay_up     probs
0       1         1         2      2        2  0.036655
1       2         1         2      2        2  0.963345
The probability of the angina with the good habits
    angina  exercise  long_sit  smoke  stay_up    probs
0       1         1         2      2        2  0.03551
1       2         1         2      2        2  0.96449
```

Figure 16: 2.1.2

```
The probability of the diabetes with the bad health
    bmi  diabetes  cholesterol  bp     probs
0    3         1            1   1  0.115423
1    3         2            1   1  0.007662
2    3         3            1   1  0.860873
3    3         4            1   1  0.016043
The probability of the stroke with the bad health
    cholesterol  bp  bmi  stroke     probs
0            1   1    3       1  0.082686
1            1   1    3       2  0.917314
The probability of the attack with the bad health
    cholesterol  bp  bmi  attack     probs
0            1   1    3       1  0.140784
1            1   1    3       2  0.859216
The probability of the angina with the bad health
    cholesterol  bp  bmi  angina     probs
0            1   1    3       1  0.161608
1            1   1    3       2  0.838392
```

Figure 17: 2.2.1

```
The probability of the diabetes with the good health
    bmi  diabetes  cholesterol  bp     probs
0    2         1            2   3  0.057710
1    2         2            2   3  0.009543
2    2         3            2   3  0.922194
3    2         4            2   3  0.010553
The probability of the stroke with the good health
    cholesterol  bp  bmi  stroke    probs
0            2   3    2       1  0.01446
1            2   3    2       2  0.98554
The probability of the attack with the good health
    cholesterol  bp  bmi  attack     probs
0            2   3    2       1  0.016161
1            2   3    2       2  0.983839
The probability of the angina with the good health
    cholesterol  bp  bmi  angina     probs
0            2   3    2       1  0.013326
1            2   3    2       2  0.986674
```

Figure 18: 2.2.2

```
The probability of the diabetes with income[1]
    income  diabetes     probs
0       1         1  0.146381
1       1         2  0.009034
2       1         3  0.826854
3       1         4  0.017731
The probability of the diabetes with income[2]
    income  diabetes     probs
0       2         1  0.143259
1       2         2  0.008982
2       2         3  0.830243
3       2         4  0.017516
The probability of the diabetes with income[3]
    income  diabetes     probs
0       3         1  0.141068
1       3         2  0.008948
2       3         3  0.832620
3       3         4  0.017364
The probability of the diabetes with income[4]
    income  diabetes     probs
0       4         1  0.138079
1       4         2  0.008925
2       4         3  0.835851
3       4         4  0.017144
```

Figure 19: 3.1.1

```
The probability of the diabetes with income[5]
    income  diabetes     probs
0       5         1  0.136066
1       5         2  0.008897
2       5         3  0.838033
3       5         4  0.017004
The probability of the diabetes with income[6]
    income  diabetes     probs
0       6         1  0.135862
1       6         2  0.008872
2       6         3  0.838263
3       6         4  0.017002
The probability of the diabetes with income[7]
    income  diabetes     probs
0       7         1  0.133681
1       7         2  0.008859
2       7         3  0.840620
3       7         4  0.016840
The probability of the diabetes with income[8]
    income  diabetes     probs
0       8         1  0.123280
1       8         2  0.008841
2       8         3  0.851843
3       8         4  0.016037
```

Figure 20: 3.1.2

```
The probability of the stroke with income[1]
    income  stroke     probs
0       1       1  0.049305
1       1       2  0.950695
The probability of the stroke with income[2]
    income  stroke     probs
0       2       1  0.050885
1       2       2  0.949115
The probability of the stroke with income[3]
    income  stroke     probs
0       3       1  0.048218
1       3       2  0.951782
The probability of the stroke with income[4]
    income  stroke     probs
0       4       1  0.046113
1       4       2  0.953887
The probability of the stroke with income[5]
    income  stroke    probs
0       5       1  0.04401
1       5       2  0.95599
The probability of the stroke with income[6]
    income  stroke     probs
0       6       1  0.041183
1       6       2  0.958817
The probability of the stroke with income[7]
    income  stroke     probs
0       7       1  0.037597
1       7       2  0.962403
The probability of the stroke with income[8]
    income  stroke     probs
0       8       1  0.033132
1       8       2  0.966868
```

Figure 21: 3.2

```
The probability of the attack with income[1]
    income  attack     probs
0       1       1  0.073515
1       1       2  0.926485
The probability of the attack with income[2]
    income  attack     probs
0       2       1  0.076203
1       2       2  0.923797
The probability of the attack with income[3]
    income  attack     probs
0       3       1  0.071924
1       3       2  0.928076
The probability of the attack with income[4]
    income  attack   probs
0       4       1  0.0687
1       4       2  0.9313
The probability of the attack with income[5]
    income  attack    probs
0       5       1  0.06561
1       5       2  0.93439
The probability of the attack with income[6]
    income  attack     probs
0       6       1  0.061636
1       6       2  0.938364
The probability of the attack with income[7]
    income  attack     probs
0       7       1  0.056209
1       7       2  0.943791
The probability of the attack with income[8]
    income  attack     probs
0       8       1  0.049129
1       8       2  0.950871
```

Figure 22: 3.3

```
The probability of the angina with income[1]
    income  angina     probs
0       1       1  0.079443
1       1       2  0.920557
The probability of the angina with income[2]
    income  angina     probs
0       2       1  0.082579
1       2       2  0.917421
The probability of the angina with income[3]
    income  angina     probs
0       3       1  0.077446
1       3       2  0.922554
The probability of the angina with income[4]
    income  angina     probs
0       4       1  0.073632
1       4       2  0.926368
The probability of the angina with income[5]
    income  angina     probs
0       5       1  0.070007
1       5       2  0.929993
The probability of the angina with income[6]
    income  angina     probs
0       6       1  0.065431
1       6       2  0.934569
The probability of the angina with income[7]
    income  angina    probs
0       7       1  0.05912
1       7       2  0.94088
The probability of the angina with income[8]
    income  angina     probs
0       8       1  0.050829
1       8       2  0.949171
```

Figure 23: 3.4

```
The probability of the diabetes with the bad habits after adding edges on smoke and exercise
    exercise  long_sit  smoke  stay_up  diabetes     probs
0         2         1      1        1         1  0.245992
1         2         1      1        1         2  0.006928
2         2         1      1        1         3  0.723721
3         2         1      1        1         4  0.023359
The probability of the stroke with the bad habits after adding edges on smoke and exercise
    stroke  exercise  long_sit  smoke  stay_up     probs
0       1         2         1      1        1  0.080488
1       2         2         1      1        1  0.919512
The probability of the attack with the bad habits after adding edges on smoke and exercise
    attack  exercise  long_sit  smoke  stay_up     probs
0       1         2         1      1        1  0.135301
1       2         2         1      1        1  0.864699
The probability of the angina with the bad habits after adding edges on smoke and exercise
    angina  exercise  long_sit  smoke  stay_up     probs
0       1         2         1      1        1  0.138072
1       2         2         1      1        1  0.861928
```

Figure 24: 4.1

```
The probability of the diabetes with the good habits after adding edges on smoke and exercise
    exercise  long_sit  smoke  stay_up  diabetes     probs
0         1         2      2        2         1  0.056227
1         1         2      2        2         2  0.010160
2         1         2      2        2         3  0.923710
3         1         2      2        2         4  0.009903
The probability of the stroke with the good habits after adding edges on smoke and exercise
    stroke  exercise  long_sit  smoke  stay_up     probs
0       1         1         2      2        2  0.019464
1       2         1         2      2        2  0.980536
The probability of the attack with the good habits after adding edges on smoke and exercise
    attack  exercise  long_sit  smoke  stay_up     probs
0       1         1         2      2        2  0.021213
1       2         1         2      2        2  0.978787
The probability of the angina with the good habits after adding edges on smoke and exercise
    angina  exercise  long_sit  smoke  stay_up     probs
0       1         1         2      2        2  0.023948
1       2         1         2      2        2  0.976052
```

Figure 25: 4.2

```
The probability of the diabetes with the bad health after adding edges on smoke and exercise
    diabetes  bmi  cholesterol  bp     probs
0         1    3            1   1  0.121241
1         2    3            1   1  0.007492
2         3    3            1   1  0.854769
3         4    3            1   1  0.016498
The probability of the stroke with the bad health after adding edges on smoke and exercise
    bmi  stroke  cholesterol  bp     probs
0    3       1            1   1  0.082697
1    3       2            1   1  0.917303
The probability of the attack with the bad health after adding edges on smoke and exercise
    bmi  cholesterol  attack  bp     probs
0    3            1       1   1  0.140083
1    3            1       2   1  0.859917
The probability of the angina with the bad health after adding edges on smoke and exercise
    bmi  angina  cholesterol  bp     probs
0    3       1            1   1  0.161096
1    3       2            1   1  0.838904
```

Figure 26: 4.3

```
The probability of the diabetes with the good health after adding edges on smoke and exercise
    diabetes  bmi  cholesterol  bp     probs
0         1    2            2   3  0.055937
1         2    2            2   3  0.009697
2         3    2            2   3  0.924042
3         4    2            2   3  0.010323
The probability of the stroke with the good health after adding edges on smoke and exercise
    bmi  stroke  cholesterol  bp     probs
0    2       1            2   3  0.014544
1    2       2            2   3  0.985456
The probability of the attack with the good health after adding edges on smoke and exercise
    bmi  cholesterol  attack  bp     probs
0    2            2       1   3  0.016183
1    2            2       2   3  0.983817
The probability of the angina with the good health after after adding edges on smoke and exercise
    bmi  angina  cholesterol  bp     probs
0    2       1            2   3  0.013328
1    2       2            2   3  0.986672
```

Figure 27: 4.4

```
The probability of the stroke with diabetes in the net4
    stroke  diabetes     probs
0       1         1  0.044417
1       2         1  0.955583
The probability of the stroke with diabetes in the net5
    stroke  diabetes     probs
0       1         1  0.076542
1       2         1  0.923458
The probability of the stroke without diabetes in the net4
    stroke  diabetes     probs
0       1         3  0.039955
1       2         3  0.960045
The probability of the stroke without diabetes in the net5
    stroke  diabetes     probs
0       1         3  0.034456
1       2         3  0.965544
```

Figure 28: 5