Report of HW4

Huanyu Wang 522030910212

1 Coding Part

For specific implementation, please view ./BayesianNetworks.py

```
def joinFactors(Factor1, Factor2):
```

- If the two factors don't have same columns, then just simply cross join them, otherwise inner join them. The new probs are the product of the original probs.

```
def marginalizeFactor(factorTable, hiddenVar):
```

- First calculate the remaining columns, and then use groupby() and sum() to marginalize the factor. Use reset_index() to reset the indexes and make the form of the factor stay unchanged.

```
def evidenceUpdateNet(bayesnet, evidenceVars, evidenceVals):
```

- Select the factors that contain evidenceVars and return the corresponding values.

```
def inference(bayesnet, hiddenVars, evidenceVars, evidenceVals):
```

- First update the network, then marginalize the hidden variables and repeatedly join the factors. Finally normalize the factor to get the result.

```
def marginalizeNetworkVariables(bayesNet, hiddenVar):
```

- This function is required but not defined in the given code, thus implemented in the 'other function' part. For every hidden variable, choose all the related factors and marginalize them in order to get the new bayesNet.

2 Written Part

Problem 1

The bayesNet is created as bellow, covering all the nodes and corresponding relations between them.

```
income       = readFactorTablefromData(riskFactorNet, ['income'])
exercise     = readFactorTablefromData(riskFactorNet, ['exercise', 'income'])
long_sit     = readFactorTablefromData(riskFactorNet, ['long_sit', 'income'])
stay_up      = readFactorTablefromData(riskFactorNet, ['stay_up', 'income'])
smoke        = readFactorTablefromData(riskFactorNet, ['smoke', 'income'])
bmi          = readFactorTablefromData(riskFactorNet, ['bmi', 'exercise', 'income', 'long_sit'])
bp           = readFactorTablefromData(riskFactorNet, ['bp', 'exercise', 'long_sit', 'income', 'stay_up', 'smoke'])
cholest      = readFactorTablefromData(riskFactorNet, ['cholesterol', 'exercise', 'stay_up', 'income', 'smoke'])
diabetes     = readFactorTablefromData(riskFactorNet, ['diabetes', 'bmi'])
stroke       = readFactorTablefromData(riskFactorNet, ['stroke', 'bmi', 'bp', 'cholesterol'])
attack       = readFactorTablefromData(riskFactorNet, ['attack', 'bmi', 'bp', 'cholesterol'])
angina       = readFactorTablefromData(riskFactorNet, ['angina', 'bmi', 'bp', 'cholesterol'])
risk_net = [income, exercise, long_sit, stay_up, smoke, bmi, bp, cholest, diabetes, stroke, attack, angina]
factors = set(riskFactorNet.columns)
```

The number of probabilities of each factor:

- income: 8
- smoke: 2 * 8 = 16
- exercise: 2 * 8 = 16
- long_sit: 2 * 8 = 16
- stay_up: 2 * 8 = 16
- bmi: 4 * 2 * 8 * 2 = 128
- bp: 4 * 2 * 2 * 8 * 2 * 2 = 512
- cholest: 2 * 2 * 2 * 8 * 2 = 128
- stoke: 2 * 4 * 4 * 2 = 64
- attack: 2 * 4 * 4 * 2 = 64
- angina: 2 * 4 * 4 * 2 = 64
- diabetes: 4 * 4 = 16
- **Sum(probabilities needed)** = 1048
- Total number of probabilities needed to store the **full joint distribution** is $\prod \text{domain number} = 131072$

Problem 2

Bad Habits / good habits inference:

```
obsVars  = ['smoke', 'exercise', 'long_sit', 'stay_up']
obsVals  = [1, 2, 1, 1]
print('Bad habits')
print(inference(risk_net, margVars, obsVars, obsVals))
obsVals  = [2, 1, 2, 2]
print('Good habits')
print(inference(risk_net, margVars, obsVars, obsVals))
```

Poor health / good health inference:

```
obsVars  = ['bp', 'cholesterol','bmi']
obsVals = [1, 1, 3]
print('Poor health')
print(inference(risk_net, margVars, obsVars, obsVals))
obsVals = [3, 2, 2]
print('Good health')
print(inference(risk_net, margVars, obsVars, obsVals))
```

Result

```
For diabetes

Bad habits
   long_sit  stay_up  smoke  exercise  diabetes     probs
0         1        1      1         2         1  0.179597
1         1        1      1         2         2  0.008754
2         1        1      1         2         3  0.791160
3         1        1      1         2         4  0.020489
Good habits
   long_sit  stay_up  smoke  exercise  diabetes     probs
0         2        2      2         1         1  0.075195
1         2        2      2         1         2  0.009409
2         2        2      2         1         3  0.903426
3         2        2      2         1         4  0.011970
Poor 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
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
```

For stroke

Bad habits

	long_sit	stay_up	smoke	exercise	stroke	probs
0	1	1	1	2	1	0.053214
1	1	1	1	2	2	0.946786

Good habits

	long_sit	stay_up	smoke	exercise	stroke	probs
0	2	2	2	1	1	0.029202
1	2	2	2	1	2	0.970798

Poor health

	cholesterol	bp	bmi	stroke	probs
0	1	1	3	1	0.082686
1	1	1	3	2	0.917314

Good health

	cholesterol	bp	bmi	stroke	probs
0	2	3	2	1	0.01446
1	2	3	2	2	0.98554

For attack

Bad habits

	long_sit	stay_up	smoke	exercise	attack	probs
0	1	1	1	2	1	0.085704
1	1	1	1	2	2	0.914296

Good habits

	long_sit	stay_up	smoke	exercise	attack	probs
0	2	2	2	1	1	0.036655
1	2	2	2	1	2	0.963345

Poor health

	cholesterol	bp	bmi	attack	probs
0	1	1	3	1	0.140784
1	1	1	3	2	0.859216

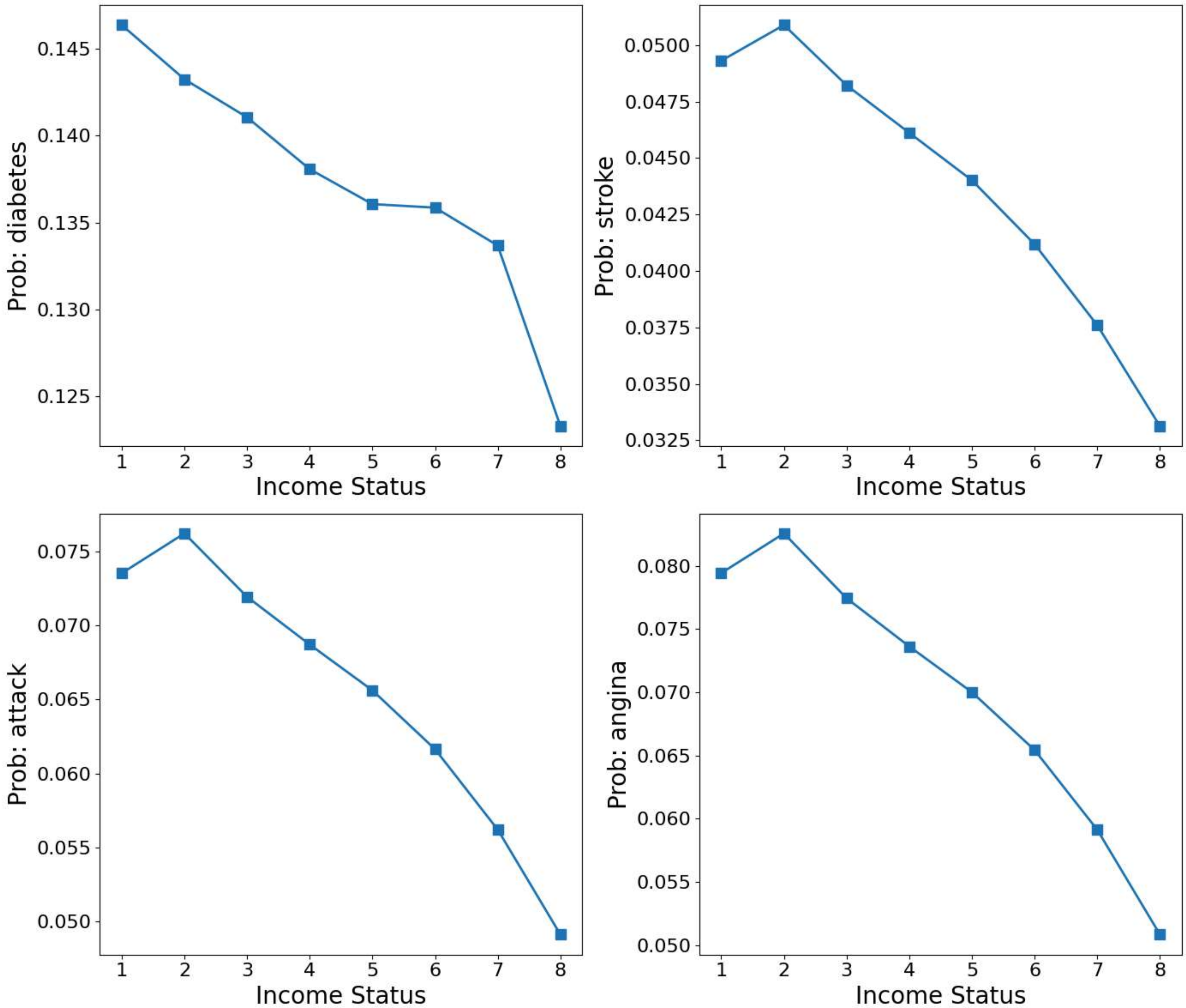
Good health

	cholesterol	bp	bmi	attack	probs
0	2	3	2	1	0.016161
1	2	3	2	2	0.983839

```
For angina

Bad habits
   angina  long_sit  stay_up  smoke  exercise    probs
0       1         1        1      1         2  0.09542
1       2         1        1      1         2  0.90458
Good habits
   angina  long_sit  stay_up  smoke  exercise    probs
0       1         2        2      2         1  0.03551
1       2         2        2      2         1  0.96449
Poor health
   cholesterol  bp  bmi  angina     probs
0            1   1    3       1  0.161608
1            1   1    3       2  0.838392
Good health
   cholesterol  bp  bmi  angina     probs
0            2   3    2       1  0.013326
1            2   3    2       2  0.986674
```

Problem 3



According to the plots, the probability of having diabetes/stroke/heart attack/angina will increase slightly and then decrease when a person's income rises. When one becomes richer, he/she may have an unhealthy life style, thus increasing the probability of having diseases. But after being even more richer, he/she may have adequate money to ensure his/her health, thus not likely to have diseases.

Problem 4

No links between the habits and the outcomes assumes that they are not related directly, but are bridged by body features like bmi, bp and cholest.

Second Bayesian Network

```
income2      = readFactorTablefromData(riskFactorNet, ['income'])
stroke2      = readFactorTablefromData(riskFactorNet, ['stroke', 'bmi', 'bp', 'cholesterol', 'exercise', 'smoke'])
attack2      = readFactorTablefromData(riskFactorNet, ['attack', 'bmi', 'bp', 'cholesterol', 'exercise', 'smoke'])
angina2      = readFactorTablefromData(riskFactorNet, ['angina', 'bmi', 'bp', 'cholesterol', 'exercise', 'smoke'])
diabetes2    = readFactorTablefromData(riskFactorNet, ['diabetes', 'bmi', 'exercise', 'smoke'])
risk_net = [income2, smoke, exercise, long_sit, stay_up, bmi, bp, cholest, stroke2, attack2, angina2, diabetes2]
factors = set(riskFactorNet.columns)
```

Result

```
For diabetes

Bad habits
   stay_up  smoke  long_sit  exercise  diabetes     probs
0        1      1         1         2         1  0.245992
1        1      1         1         2         2  0.006928
2        1      1         1         2         3  0.723721
3        1      1         1         2         4  0.023359
Good habits
   stay_up  smoke  long_sit  exercise  diabetes     probs
0        2      2         2         1         1  0.056227
1        2      2         2         1         2  0.010160
2        2      2         2         1         3  0.923710
3        2      2         2         1         4  0.009903
Poor health
   cholesterol  bmi  bp  diabetes     probs
0            1    3   1         1  0.121241
1            1    3   1         2  0.007492
2            1    3   1         3  0.854769
3            1    3   1         4  0.016498
Good health
   cholesterol  bmi  bp  diabetes     probs
0            2    2   3         1  0.055937
1            2    2   3         2  0.009697
2            2    2   3         3  0.924042
3            2    2   3         4  0.010323
```

```
For stroke

Bad habits
   stay_up  smoke  stroke  long_sit  exercise     probs
0        1      1       1         1         2  0.080488
1        1      1       2         1         2  0.919512
Good habits
   stay_up  smoke  stroke  long_sit  exercise     probs
0        2      2       1         2         1  0.019464
1        2      2       2         2         1  0.980536
Poor health
   cholesterol  bmi  bp  stroke     probs
0            1    3   1       1  0.082697
1            1    3   1       2  0.917303
Good health
   cholesterol  bmi  bp  stroke     probs
0            2    2   3       1  0.014544
1            2    2   3       2  0.985456
```

```
For attack

Bad habits
   stay_up  attack  smoke  long_sit  exercise     probs
0        1       1      1         1         2  0.135301
1        1       2      1         1         2  0.864699
Good habits
   stay_up  attack  smoke  long_sit  exercise     probs
0        2       1      2         2         1  0.021213
1        2       2      2         2         1  0.978787
Poor health
   cholesterol  bmi  attack  bp     probs
0            1    3       1   1  0.140083
1            1    3       2   1  0.859917
Good health
   cholesterol  bmi  attack  bp     probs
0            2    2       1   3  0.016183
1            2    2       2   3  0.983817
```

```
For angina

Bad habits
   stay_up  angina  smoke  long_sit  exercise     probs
0        1       1      1         1         2  0.138072
1        1       2      1         1         2  0.861928
Good habits
   stay_up  angina  smoke  long_sit  exercise     probs
0        2       1      2         2         1  0.023948
1        2       2      2         2         1  0.976052
Poor health
   angina  cholesterol  bmi  bp     probs
0       1            1    3   1  0.161096
1       2            1    3   1  0.838904
Good health
   angina  cholesterol  bmi  bp     probs
0       1            2    2   3  0.013328
1       2            2    2   3  0.986672
```

As is shown above, the possibility of health outcomes conditioned on having bad habits rises, while other conditions remain the similar distribution. However, the assumptions of the first graph considered scientific reasoning as all the diseases should only be directly caused by body features like bp, bmi and cholest. Thus, I think the assumptions of the first graph were valid still.

Problem 5

No edges between the four outcomes assumes that they are not related to each other, i.e. no interactions between different diseases.

Third Bayesian Network

```
income3          = readFactorTablefromData(riskFactorNet, ['income'])
stroke3          = readFactorTablefromData(riskFactorNet,
                ['stroke', 'bmi', 'bp', 'cholesterol', 'exercise', 'smoke', 'diabetes'])
risk_net = [income3, smoke, exercise, long_sit, stay_up, bmi, bp, cholest, stroke3, attack2, angina2, diabetes2]
```

Result

```
Second Network
   diabetes  stroke     probs
0         1       1  0.044417
1         1       2  0.955583
   diabetes  stroke     probs
0         3       1  0.039955
1         3       2  0.960045
```

```
Third Network
   diabetes  stroke     probs
0         1       1  0.076542
1         1       2  0.923458
   diabetes  stroke     probs
0         3       1  0.034456
1         3       2  0.965544
```

As is shown in the figures above, in the second graph, P(*|1) and P(*|3) are approximately equal, which indicates that having stroke is independent with diabetes. However, when adding an edge between diabetes and stroke, difference between P(*|1) and P(*|3) increases significantly, indicating that there exists relation between them. From my perspective, the assumption about the interaction between diabetes and stroke are not valid because diseases should only interact with each other through body features, as the second graph shows.

2.6

The output is the same as the correct answer.

```
Windows PowerShell
(ai3603) PS C:\Users\17717\Desktop\人工智能理论及应用\作业\4\AI3603_HW4> python .\BayesNetworkTestScript.py
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
   long_sit  exercise  smoke  diabetes     probs
0         1         2      1         1  0.136815
1         1         2      1         2  0.008916
2         1         2      1         3  0.837218
3         1         2      1         4  0.017052
```