

AU 311 INTRODUCTION TO PATTERN RECOGNITION

By: Harry Lording (717030990012)

HW#: 1

November 29, 2019

Coding Part

Code included in zip file. For the most part the implementation of the this part of the homework was easy however very tedious, especially in respect of debugging the code.

Writing Part

1. What is the size (in terms of the number of probabilities needed) of this network? And what is the total number of probabilities needed to store the full joint distribution?

a) The size of the network is the total number of rows across all the factor tables in the network.

```
Table has 8 rows
Table has 64 rows
Table has 16 rows
Table has 16 rows
Table has 128 rows
Table has 64 rows
Table has 16 rows
Table has 64 rows
Table has 64 rows
Table has 64 rows
Total rows = 504
```

b) The number of probabilities in the joint distribution is the the multiplication of the number of values each random variable can take. Therefore the number of probabilities is $8 * 2 * 2 * 4 * 4 * 2 * 2 * 2 * 2 * 4 = 32,768$

2. For each of the four health outcomes (diabetes, stroke, heart attack, angina), answer the following by querying and inferring from your Bayesian network. Please write the results in a table format.

(a) What is the probability of the outcome if I have bad habits (smoke and don't exercise)? How about if I have good habits (don't smoke and do exercise)?

Question 2a

Diabetes outcomes with bad habits:

	Diabetes	Probability
0	yes	0.150516
1	only during pregnancy	0.008965
2	no	0.822423
3	pre-diabetic	0.018096

Stroke outcomes with bad habits:

	Stroke	Probability
0	yes	0.049264
1	no	0.950736

Heart attack outcomes with bad habits:

	Heart Attack	Probability
0	yes	0.07433
1	no	0.92567

Angina outcomes with bad habits:

	Angina	Probability
0	yes	0.080448
1	no	0.919552

Diabetes outcomes with good habits:

	Diabetes	Probability
0	yes	0.127119
1	only during pregnancy	0.008865
2	no	0.847693
3	pre-diabetic	0.016323

Stroke outcomes with good habits:

	Stroke	Probability
0	yes	0.03611
1	no	0.96389

Heart attack outcomes with good habits:

	Heart Attack	Probability
0	yes	0.052798
1	no	0.947202

Angina outcomes with good habits:

	Angina	Probability
0	yes	0.054755
1	no	0.945245

(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)?

Question 2b

Diabetes outcomes with poor health:

	Diabetes	Probability
0	yes	0.115423
1	only during pregnancy	0.007662
2	no	0.860873
3	pre-diabetic	0.016043

Stroke outcomes with poor health:

	Stroke	Probability
0	yes	0.082686
1	no	0.917314

Heart attack outcomes with poor health:

	Heart Attack	Probability
0	yes	0.140784
1	no	0.859216

Angina outcomes with poor health:

	Angina	Probability
0	yes	0.161608
1	no	0.838392

Diabetes outcomes with good health:

	Diabetes	Probability
0	yes	0.057710
1	only during pregnancy	0.009543
2	no	0.922194
3	pre-diabetic	0.010553

Stroke outcomes with good health:

	Stroke	Probability
0	yes	0.01446
1	no	0.98554

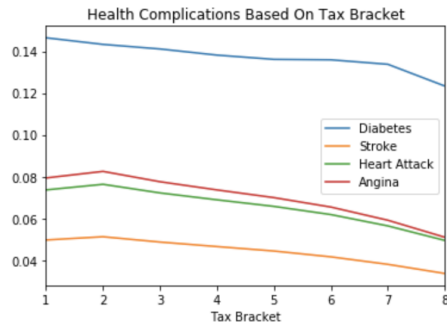
Heart attack outcomes with good health:

	Heart Attack	Probability
0	yes	0.016161
1	no	0.983839

Angina outcomes with good health:

	Angina	Probability
0	yes	0.013326
1	no	0.986674

3. 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 $i = 1; 2; \dots; 8$; and your vertical axis should be $P(y = 1 | \text{income} = i)$, where y is the outcome). What can you conclude?



From these results it is evident that there is a negative correlation between a persons income and the outcomes of the Bayes Network. As a persons income increases they are less likely the have suffer from health problems. This is most likely due to people of higher income being able to afford more and better quality health services than people in lower income brackets

4. Notice there are no links in the graph between the habits (smoking and exercise) 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?

It makes the assumption that smoking and exercise do not directly effect the chances of a person having or not having a health outcome. Instead they only effect a persons bmi, bp and cholestrol. However it is likely that smoking and exercise impact other hidden factors that could effect the chances of someone having a health problem or not. By creating links between smoking and exercise towards the health outcomes in the Bayes Net we can see if and how smoking and exercise effect the chances of a person having an outcome.

```

Question 4

Diabetes outcomes with bad habits:
Diabetes Old Probability New Probability Change
0 yes 0.158516 0.218945 0.060429
1 only during pregnancy 0.088965 0.086715 -0.002250
2 no 0.822423 0.768693 -0.053731
3 pre-diabetic 0.818896 0.821447 0.003552

Stroke outcomes with bad habits:
Stroke Old Probability New Probability Change
0 yes 0.849264 0.878835 0.028571
1 no 0.958736 0.921965 -0.028771

Heart attack outcomes with bad habits:
Heart Attack Old Probability New Probability Change
0 yes 0.07433 0.121166 0.046836
1 no 0.92567 0.878834 -0.046836

Angina outcomes with bad habits:
Angina Old Probability New Probability Change
0 yes 0.088448 0.119887 0.038559
1 no 0.919552 0.880993 -0.038559

Diabetes outcomes with good habits:
Diabetes Old Probability New Probability Change
0 yes 0.227119 0.098552 -0.028567
1 only during pregnancy 0.088865 0.089884 0.001019
2 no 0.847693 0.877576 0.029882
3 pre-diabetic 0.816323 0.813988 -0.002334

Stroke outcomes with good habits:
Stroke Old Probability New Probability Change
0 yes 0.83611 0.824311 -0.0118
1 no 0.96389 0.975689 0.0118

Heart attack outcomes with good habits:
Heart Attack Old Probability New Probability Change
0 yes 0.852798 0.831815 -0.021782
1 no 0.947202 0.968985 0.021782

Angina outcomes with good habits:
Angina Old Probability New Probability Change
0 yes 0.854755 0.8368 -0.017955
1 no 0.945245 0.9632 0.017955

```

As we can see we see a positive correlation between having bad habits and and the probability of having any of the outcomes. That is if a person has bad habits they are more likely to have any of the outcomes. As expected there is a person has good health habits they are less likely to have any of the outcomes.

```

Diabetes outcomes with poor health:
Diabetes Old Probability New Probability Change
0 yes 0.115423 0.123481 0.008058
1 only during pregnancy 0.087662 0.087468 -0.000202
2 no 0.868873 0.852416 -0.008457
3 pre-diabetic 0.816843 0.816643 0.000600

Stroke outcomes with poor health:
Stroke Old Probability New Probability Change
0 yes 0.882866 0.884257 0.001571
1 no 0.917314 0.915743 -0.001571

Heart attack outcomes with poor health:
Heart Attack Old Probability New Probability Change
0 yes 0.148784 0.142199 -0.001415
1 no 0.859216 0.857801 -0.001415

Angina outcomes with poor health:
Angina Old Probability New Probability Change
0 yes 0.161688 0.162972 0.001364
1 no 0.838392 0.837028 -0.001364

Diabetes outcomes with good health:
Diabetes Old Probability New Probability Change
0 yes 0.057718 0.054173 -0.003537
1 only during pregnancy 0.089543 0.089731 0.000188
2 no 0.922194 0.925952 0.003758
3 pre-diabetic 0.818553 0.818144 -0.000409

Stroke outcomes with good health:
Stroke Old Probability New Probability Change
0 yes 0.01446 0.013997 -0.000463
1 no 0.98554 0.986083 0.000463

Heart attack outcomes with good health:
Heart Attack Old Probability New Probability Change
0 yes 0.816161 0.815469 -0.000692
1 no 0.983839 0.984531 0.000692

Angina outcomes with good health:
Angina Old Probability New Probability Change
0 yes 0.813326 0.812944 -0.000382
1 no 0.986674 0.987056 0.000382

```

Retaking the queries in question 2 we see that it is now more likely for a person with poor health to

have any of the health outcomes and it is less likely for a person with good health to have any of the health outcomes. As stated before, it is most likely that smoking and exercise impacted other hidden variables apart from bmi, bp and cholesterol that would then have a impact of a person having any of the health outcomes. Due to this I think the while the first graph was accurate in representing the impacts of the income, bmi, bp, smoking, exercise and cholesterol on a persons likelihood of having diabetes, stroke, heart attack or angina, I think the second graph provides a more accurate representation.

5. 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:

$P(\text{stroke} = 1 \mid \text{diabetes} = 1)$ and $P(\text{stroke} = 1 \mid \text{diabetes} = 0)$

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

The assumption it makes is that having any of the outcomes would not mean that a person would be any more or less likely to have another outcome.

```
Stroke given you have diabetes without added edge:
Stroke Probability
0 yes 0.044164

Stroke given you have diabetes with added edge:
Stroke Probability
0 yes 0.076198

Stroke given you don't have diabetes without added edge:
Stroke Probability
0 yes 0.040478

Stroke given you don't have diabetes with added edge:
Stroke Probability
0 yes 0.035015
```

As we can see the updated table suggests that if there is a link between the pair of outcomes, the likelihood of someone having a stroke goes up given that you have diabetes. Likewise if there is a link, the probability of a person have a stroke will be less given that they don't have diabetes. In my opinion a link between these outcomes makes sense, as it is probable that due to a person having diabetes that they will be more likely to have a stroke.