CS6190: Probabilistic Modeling Homework 2 Bayesian Networks

Gopal Menon

06 March, 2018

Written Part

1. The total probability of the network is given by

$$\begin{aligned} p(Beysian\ Network) = & p(income)p(smoke|income)p(exercise|income)p(bmi|income, exercise) \\ & \times p(blood\ pressure|exercise, income, smoking)p(cholesterol|exercise, income, smoking) \\ & \times p(diabetes|bmi)p(stroke|bmi, bp, cholesterol)p(attack|bmi, bp, cholesterol) \\ & \times p(angina|bmi, bp, cholesterol) \end{aligned}$$

The total probability distribution requires the product of 10 random variables. If the probabilities are multiplied as shown above, the probability table will have 10 variables. Number of allowed states for the random variables are:

Table 1: CDC Survey Variable states

Survey Variable	Number of states
income	8
exercise	2
smoke	2
bmi	4
bp	4
cholesterol	2
angina	2
stroke	2
attack	2
diabetes	4

The maximum number of rows in the probability distribution table for the above variables will be the product of the number of allowed states - $8 \times 2 \times 2 \times 4 \times 4 \times 2 \times 2 \times 2 \times 2 \times 4 = 32768$. This is the number of probabilities that are needed to store the full joint distribution.

2. For each of the four health outcomes (diabetes, stroke, heart attack, angina)
   (a) 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)?

Table 2: Probability for diabetes with smoking and no exercise

Probability	diabetes
0.1593304	Diabetic
0.0078815	Diabetes only during pregnancy

Probability	diabetes
0.8126324	Not diabetic
0.0201557	Pre-diabetic

Table 3: Probability for stroke with smoking and no exercise

Probability	stroke
0.0501267	Will have stroke
0.9498733	Will not have stroke

Table 4: Probability for attack with smoking and no exercise

Probability	attack
0.0724154	Will have heart attack
0.9275846	Will not have heart attack

Table 5: Probability for angina with smoking and no exercise

Probability	angina
0.0777561	Will have angina
0.9222439	Will not have angina

Table 6: Probability for diabetes with no smoking and exercise

Probability	diabetes
0.1350624	Diabetic
0.0076735	Diabetes only during pregnancy
0.8393916	Not diabetic
0.0178726	Pre-diabetic

Table 7: Probability for stroke with no smoking and exercise

Probability	stroke
0.0368082	Will have stroke
0.9631918	Will not have stroke

Table 8: Probability for attack with no smoking and exercise

Probability	attack
0.0510272	Will have heart attack
0.9489728	Will not have heart attack

Table 9: Probability for angina with no smoking and exercise

Probability	angina
0.0523189	Will have angina
0.9476811	Will not have angina

(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)?

Table 10: Probability for diabetes with bad health

Probability	diabetes
0.1222794	Diabetic
0.0067179	Diabetes only during pregnancy
0.8540029	Not diabetic
0.0169998	Pre-diabetic

Table 11: Probability for stroke with bad health

Probability	stroke
0.0839749	Will have stroke
0.9160251	Will not have stroke

Table 12: Probability for attack with bad health

Probability	attack
0.13433	Will have heart attack
0.86567	Will not have heart attack

Table 13: Probability for angina with bad health

Probability	angina
0.1531853	Will have angina
0.8468147	Will not have angina

Table 14: Probability for diabetes with good health

Probability	diabetes
0.0616345	Diabetic
0.0078003	Diabetes only during pregnancy
0.9198968	Not diabetic
0.0106684	Pre-diabetic

Table 15: Probability for stroke with good health

Probability	stroke
0.0138704	Will have stroke
0.9861296	Will not have stroke

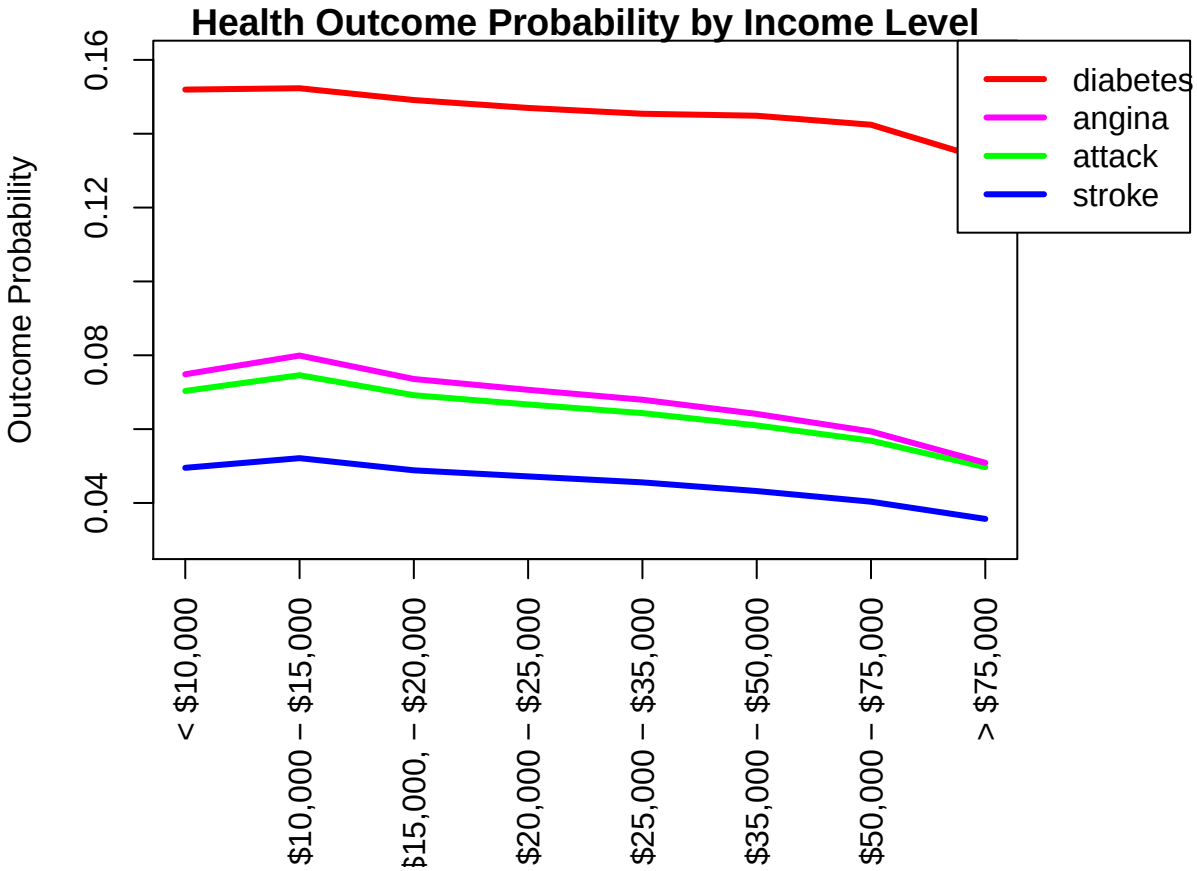
Table 16: Probability for attack with good health

Probability	attack
0.0158879	Will have heart attack
0.9841121	Will not have heart attack

Table 17: Probability for angina with good health

Probability	angina
0.0128387	Will have angina
0.9871613	Will not have angina

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



It looks like overall the probability of a bad health outcome reduces with increasing income. However, it looks like the probability first increases from the lowest to the next highest income and then reduces.

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?

   I think the assumption made in the graph is that the habits (smoking and exercise) have an impact on the health (bmi, bp and cholesterol) of the person and thus indirectly impact the outcomes (diabetes, stroke, attack and angina).

Table 18: Probability for diabetes with smoking and no exercise

Probability	diabetes
0.2267429	Diabetic
0.0062083	Diabetes only during pregnancy
0.7429210	Not diabetic
0.0241279	Pre-diabetic

Table 19: Probability for stroke with smoking and no exercise

Probability	stroke
0.0790277	Will have stroke
0.9209723	Will not have stroke

Table 20: Probability for attack with smoking and no exercise

Probability	attack
0.1175417	Will have heart attack
0.8824583	Will not have heart attack

Table 21: Probability for angina with smoking and no exercise

Probability	angina
0.1142626	Will have angina
0.8857374	Will not have angina

Table 22: Probability for diabetes with no smoking and exercise

Probability	diabetes
0.1025276	Diabetic
0.0088837	Diabetes only during pregnancy
0.8736715	Not diabetic
0.0149172	Pre-diabetic

Table 23: Probability for stroke with no smoking and exercise

Probability	stroke
0.0253386	Will have stroke
0.9746614	Will not have stroke

Table 24: Probability for attack with no smoking and exercise

Probability	attack
0.0303845	Will have heart attack
0.9696155	Will not have heart attack

Table 25: Probability for angina with no smoking and exercise

Probability	angina
0.0359644	Will have angina
0.9640356	Will not have angina

Table 26: Probability for diabetes with bad health

Probability	diabetes
0.1310737	Diabetic
0.0065005	Diabetes only during pregnancy
0.8445742	Not diabetic
0.0178516	Pre-diabetic

Table 27: Probability for stroke with bad health

Probability	stroke
0.085754	Will have stroke
0.914246	Will not have stroke

Table 28: Probability for attack with bad health

Probability	attack
0.1361617	Will have heart attack
0.8638383	Will not have heart attack

Table 29: Probability for angina with bad health

Probability	angina
0.1548693	Will have angina
0.8451307	Will not have angina

Table 30: Probability for diabetes with good health

Probability	diabetes
0.0576544	Diabetic
0.0079690	Diabetes only during pregnancy
0.9240094	Not diabetic
0.0103672	Pre-diabetic

Table 31: Probability for stroke with good health

Probability	stroke
0.0133544	Will have stroke
0.9866456	Will not have stroke

Table 32: Probability for attack with good health

Probability	attack
0.0152652	Will have heart attack
0.9847348	Will not have heart attack

Table 33: Probability for angina with good health

Probability	angina
0.0124368	Will have angina
0.9875632	Will not have angina

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(stroke = 1|diabetes = 1) and P(stroke = 1|diabetes = 3)

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

Table 34: Probability for stroke when diabetes is present

Probability	stroke
0.0764278	Will have stroke
0.9235722	Will not have stroke

Table 35: Probability for stroke when diabetes is not present

Probability	stroke
0.0358626	Will have stroke
0.9641374	Will not have stroke

6. Finally,make sure that your code runs correctly on all of the examples in BayesNetExamples.r. Your code will be graded for correctness on these also.

Table 36: Bishop 8.30 - p(G = 0)

probs	gauge
0.685	1
0.315	0

Table 37: Bishop 8.31 - p(G = 0|F = 0)

probs	gauge	fuel
0.19	1	0
0.81	0	0

Table 38: Bishop 8.32 - p(F = 0|G = 0)

probs	gauge	fuel
0.7428571	0	1
0.2571429	0	0

Table 39: Bishop 8.33 - p(F = 0|G = 0, B = 0)

probs	gauge	battery	fuel
0.8888889	0	0	1
0.1111111	0	0	0

Table 40: Kevin Murphy - Pr(S|W=1)

probs	wet	sprinkler
0.5702364	T	F
0.4297636	T	T

Table 41: Kevin Murphy - Pr(R|W=1)

probs	wet	rain
0.2920723	T	F
0.7079277	T	T

Table 42: Kevin Murphy - Pr(W)

probs	wet
0.3529	F

probs	wet
0.6471	T

Table 43: Kevin Murphy - Pr(S|W=1,R=1)

probs	wet	sprinkler	rain
0.805501	T	F	T
0.194499	T	T	T