A Robot to Do Your Groceries: Assignment 2

September 23, 2014

2 Bayesian Grocery Prediction

The next step in making our robot more intelligent, is to infuse it with reason!

A grocery robot is nice, but it would be cumbersome to have to tell it what to get from the store every time he goes. Better to have it decide for itself! We want our robot to use logic to predict a grocery list, given a certain customer profile. For example, if the customer is a vegetarian, the robot should be able to conclude that it should probably not buy steak.

To make your robot even smarter, it should profile its owner by analyzing its past shopping behavior. Then, based on this profile, it should extrapolate a list of products the customer will most likely want to acquire next.

We will be using a 'Bayesian Network' to implement such reasoning schemes. The purpose of this assignment is to gain a solid understanding of Bayesian networks and to implement it into the brain of your robot.

During the Grand Challenge, your bot will use these reasonings to come up with a suitable list of products.

WHAT YOU NEED TO HAND-IN

This assignment consists of 33 exercises. You should answer each of these questions and hand-in everything within the same document as a PDF file. In this assignment we will not be doing any programming ourselves, so no code needs to be handed in.

Most of the answers may be given by an equation + the numerical answer to it: You need not explain the obvious. For example:
"$P(K) = P(K|C)P(C) + P(K|\neg C)P(\neg C) = \sum_{\forall i} P(K|C_i)P(C_i)$, thus $P(K) = 0.2$" would be a sufficient answer to most of the questions. (You need not give both the 'sum' as the 'long' notation: choose what you prefer.)

Please limit your documentation to roughly 5 pages.

2.1 A SIMPLE NETWORK I

In As. 2.1, we will be using the Bayesian network shown in Fig. 2.1. 'V' means 'vegetarian', 'H' means 'healthy eating habbits', 'K' means 'customer has kids', 'F' means 'customer wants to purchase fruit' and 'C' means 'customer wants to purchase candy'.

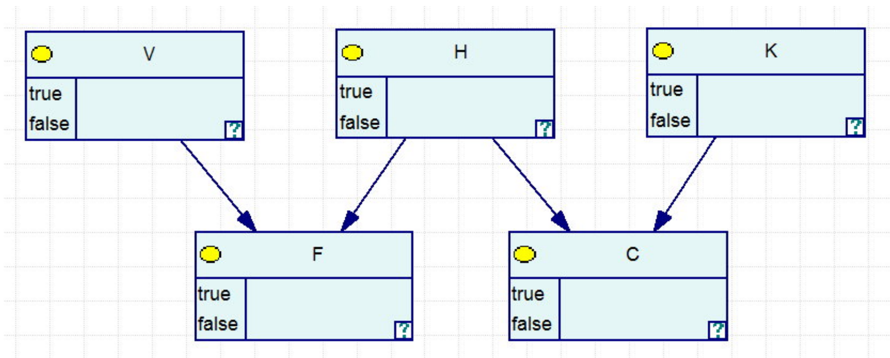

Figure 2.1: *Bayesian network for As. 2.1*

You may use the following estimates of the odds:

$$P(C|KH) = 0.3 \tag{2.1}$$
$$P(C|K\neg H) = 1 \tag{2.2}$$
$$P(C|\neg KH) = 0.01 \tag{2.3}$$
$$P(C|\neg K\neg H) = 0.5 \tag{2.4}$$
$$P(K) = 0.15 \tag{2.5}$$
$$P(H) = 0.3 \tag{2.6}$$
$$P(V) = 0.04 \tag{2.7}$$
$$P(F|VH) = 0.99 \tag{2.8}$$
$$P(F|V\neg H) = 0.7 \tag{2.9}$$
$$P(F|\neg VH) = 0.9 \tag{2.10}$$
$$P(F|\neg V\neg H) = 0.001 \tag{2.11}$$

For clarity, e.g. '$P(F|V\neg H)$' means 'the chance that F is true, given that (V is true and H is false)'; Also known as '$P(F|V = 1, H = 0)$'.

We will first assume that nothing is known about the customer. In other words, Eqs. 2.5, 2.6 and 2.7 do hold.

1. Compute the chance that the customer wants to purchase fruit (i.e., '$P(F)$').
   Please write down an exact equation for $P(F)$ before filling it in. This will help you a lot for the later questions.
2. Compute the chance that the customer wants to purchase candy.
   Again, start with the exact equation for $P(C)$.

Now assume that our customer is a vegetarian: "evidence is found that our customer is a vegetarian".

3. Compute the chance that this customer wants to purchase fruit.
4. And what is the chance that the vegetarian customer wants to purchase candy?
5. Does the fact that the customer is a vegetarian change our belief in the customer having healthy eating habits? I.e., does $P(H)$ change, now that we known that V is true?
   Please give your answer using a maximum of 30 words.

Evidence is found that our customer is a vegetarian with unhealthy eating habits.

6. What is the chance that this customer wants to purchase fruit?
7. And what is the chance for this customer to purchase candy?

Evidence is found that our customer is not a vegetarian, has healthy eating habits and does not have children.

8. What is the chance that this customer wants to purchase fruit?

9. And what is the chance for this customer to purchase candy?

2.2 A SIMPLE NETWORK II

So far, we have computed consequences given the cause, which makes sense as for the purpose of our robot. However, the real magic of Bayesian networks is the ease with which we can compute which of the causes is most likely to blame for an observed consequence. For example, instead of computing the chance that the customer wants to purchase fruit, given that he is a vegetarian, we will compute the chance that the customer is a vegetarian (cause) given that he wants to purchase fruit (consequence). This is the "inverse computation" with respect to As. 2.1: "back-propagation of the evidence".

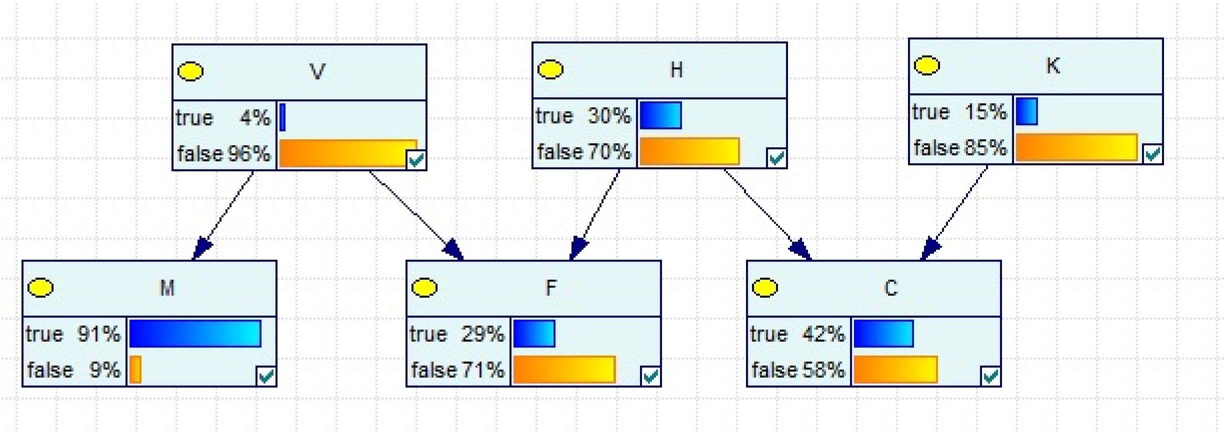

Figure 2.2: *Bayesian network for As. 2.2*

For this assignment, we use the same network as in As. 2.1, using the same odds. However, we will add "meat" just below "vegetarian", as in Fig. 2.2. The odds are extended with:

$$P(M|V) = 0.001 \tag{2.12}$$

$$P(M|\neg V) = 0.95 \tag{2.13}$$

10. In the lecture you have seen Bayes' rule, e.g. in the "cold" example. Use Bayes' rule to derive an expression for $P(V|M)$. Express your answer using only known odds (Eqs. 2.1 - 2.11, Eqs. 2.12 - 2.13). If done correctly, your equation should be of the form $P(V|M) = \frac{a}{a+b}$.
11. Compute the chance that the customer is a vegetarian, given that he does not want to purchase any meat ($P(V|\neg M)$).
12. A similar, but yet more complicated, equation may be written down to compute $P(V|F)$. The only real difference is the fact that you will need an extra elimination of variables.

Derive an expression for $P(V|F)$, using only known odds. If done correctly, your equation should be of the form $P(V|F) = \frac{a+c}{a+c+b+d}$.
(Hint: Since $P(F|V)$ is unknown, you will need to do an extra elimination of the variable H.)

13. Compute the chance that the customer is a vegetarian, given that he does want to purchase fruit ($P(V|F)$).

14. Now, write down an equation for $P(V|MF)$. If done correctly, your equation should be of the form $P(V|MF) = \frac{a+c}{a+c+b+d}$.

15. Compute the chance that the customer is a vegetarian, given that he does want to purchase fruit, but no meat ($P(V|F\neg M)$).

2.3 A NON-BOOLEAN NETWORK

A Bayesian network may be far more complex than we have seen in the last 2 assignments. For example, each node of the network needs not be a boolean (i.e., true/false). We may for example have a network as in Fig. 2.3. Here, 'S' refers to how actively the customer is involved in any sport. Instead of 'active' and 'inactive', we add a third possibility: 'passive'. This is somewhere in between 'active' and 'inactive'. Denote $S_1 \equiv$ 'S =inactive', $S_2 \equiv$ 'S =passive' and $S_3 \equiv$ 'S =active'.

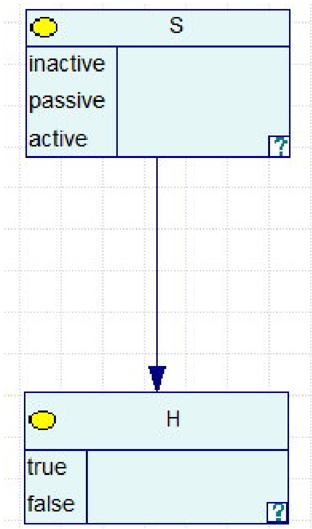

Figure 2.3: *Bayesian network for As. 2.3*

It is given that

$$P(S_1) = 0.05 \quad (2.14)$$

$$P(S_2) = 0.35 \quad (2.15)$$

$$P(H|S_1) = 0.2 \quad (2.16)$$

$$P(H|S_2) = 0.5 \quad (2.17)$$

$$P(H|S_3) = 0.9 \quad (2.18)$$

16. Write down an equation for $P(H)$ using the terms listed in Eqs. 2.14 - 2.18.
17. Compute $P(H)$ and $P(\neg H)$.
18. Use the knowledge acquired in As. 2.2 to compute the chance that the customer is an inactive sporter, given that the customer has healthy eating habbits.

2.4 A NETWORK FOR OUR ROBOT

"84% of the adults are married. Research has shown that married people are less likely to buy flowers then non-married, probably because unwed people are still trying to impress their counterpart. Of all the people that are married, only 34% buys flowers on regular basis, while of the people with no marriage this number is 64%. Fast food consumption is also linked to marriage: 9 out of 10 unhealthy and unmarried people like to eat junk food. On the other hand, almost the same ratio (86%) of healthy and married persons practically never eat junk food. This number shrinks to 57% for healthy persons that are not married. Finally, of the married people that do not have healthy eating habit, 56% buys junk food often. 3 out of 10 of all the people have healthy eating habits. Healthy citizens are more keen to buy something fresh: 88% puts salads in their baskets everyday. Surprisingly, also unhealthy citizens eat salads, but not as much: 64%."

19. Give a schematic diagram of the network relevant to the above text. Show all the proper variables and causalities. *(Hint: There are 5 variables in total)*
20. write down all the conditional- and a-priori probabilities.

Now you should expand the network. The products that a person can buy are divided into 7 categories. From now on, junk food is class 1, flowers is denoted as class 2 and salads are grouped in class 3. You can use the values from the previous question except for class 3, because it already had dependencies. You can forget those and exchange them for the ones listed in the table. The following tables summarize the probabilities involved with this network.

Table 2.1: *Probability Tables*

$P(3\|H,C,V)$	0.88
$P(3\|H,C,\neg V)$	0.78
$P(3\|H,\neg C,V)$	0.96
$P(3\|H,\neg C,\neg V)$	0.72
$P(3\|\neg H,C,V)$	0.64
$P(3\|\neg H,C,\neg V)$	0.28
$P(3\|\neg H,\neg C,V)$	0.46
$P(3\|\neg H,\neg C,\neg V)$	0.09

$P(5\|V,H)$	0.97
$P(5\|V,\neg H)$	0.77
$P(5\|\neg V,H)$	0.89
$P(5\|\neg V,\neg H)$	0.32

$P(6\|C,M)$	0.74
$P(6\|C,\neg M)$	0.81
$P(6\|\neg C,M)$	0.07
$P(6\|\neg C,\neg M)$	0.45

$P(4\|C,F)$	0.86
$P(4\|C,\neg F)$	0.64
$P(4\|\neg C,F)$	0.68
$P(4\|\neg C,\neg F)$	0.32

$P(7\|F)$	0.81
$P(7\|\neg F)$	0.33

$P(C)$	0.15
$P(V)$	0.04
$P(F)$	0.46

21. Expand the network with the given probabilities. Make a diagram of the nodes with arrows indicating the direction of influence
(*H = Healthy, M = Married, C = Children, V = Vegetarian, F = Female*).

Now, you should have all the odds you need! Let's download a toolkit for simulating Bayesian networks. It is recommended that you use "GeNIe 2.0" from the university of Pittsburgh: `https://dslpitt.org/genie/`

In "GeNIe", you may make variables by pressing the yellow oval in the toolbar and then clicking in the workspace. By double-clicking the variable-node you may insert the odds in the 'definition' tab. Using the black arrow icon in the toolbar, you may implement dependencies between the variables. First click the parent node, then click the child node. The chance tables of the 'definition' tab will automatically adapt to the dependencies you set. Consequently, it is recommended that you first implement all dependencies, before inserting all odds.

The Bayesian network may be ran by pressing the lightning bolt icon in your toolbar. The results are most easily visualized in the 'bar chart' representation of the network (press F8 or go to 'node-viewAs-BarChart').

You may set a variable to a value by double-clicking the value in the 'bar chart' perspective (the probability will be set to 100%).

22. Make a screenshot of your network for in the report.

We will now grant our robot evidence for several case studies. The robot should then predict whether the customer wants to purchase fruit and/or candy.

23. What's the probability the customer will want to buy products from class 5?

24. What's the probability the customer will not want anything from class 2?
25. Given that the customer is a female. What is the probability of having to buy something in class 4?
26. This same woman turns out to have two children. What is the chance of her wanting something from class 4 now?
27. A different customer profiles as a healthy vegetarian. What's the probability of buying a product from class 3?
28. A customer was seen buying something out of class 2. What is the probability that he/she is married?
29. What's the chance this customer is female?
30. A wealthy man was observed to buy products from all categories. What is the probability that he's is married?
31. Another customer appeared to not buy anything from the listed classes. What can you say about this customer?
32. We know that a certain person bought stuff from class 1 and 2, but we did not observe the rest of his basket. Is this person likely to have any children?
33. We do not know what a given person bought, but we are certain he didn't buy anything from class 2, 3 or 5. What can you say about this customer?

Congratulations! If you have worked your way through these questions correctly, your robot will now be infused with the power of reason! For the Grand Challenge your robot will be expected to profile a customer based on its purchasing history (back-propagating the evidence). After it has settled on a certain profile, it will predict its owners grocery list and make its way to the supermarket!