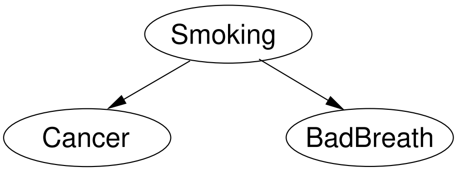
6CCS3AIN, 2019, Tutorial 03 Answers (Version 1.0)

1. The Bayesian network is:



We are told that:

$$\begin{aligned} Psmoking) &= 0.2 \\ P(cancer|smoking) &= 0.6 \\ P(badBreath|smoking) &= 0.95 \end{aligned}$$

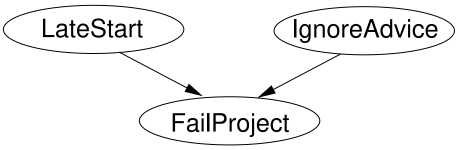
Now, the Naive Bayes model tells us that:

$$\mathbf{P}(Cause, Effect_1, \ldots, Effect_n) = \mathbf{P}(Cause) \prod_i \mathbf{P}(Effect_i|Cause)$$

so:

$$\begin{aligned} P(smoking, cancer, badBreath) &= P(smoking) \times P((cancer|smoking) \times P(badBreath|smoking) \\ &= 0.2 \times 0.6 \times 0.95 \\ &= 0.114 \end{aligned}$$

2. The Bayesian network is:



We are told that:

$$\begin{aligned} P(f|l) &= 0.7 \\ P(f|i) &= 0.8 \end{aligned}$$

The Noisy-Or model tells us that, since $LateStart$ and $IgnoreAdvice$ are the only causes of $FailProject$:

$$\begin{aligned} P(f|\neg l, \neg i) &= 0 \\ P(\neg f|\neg l, \neg i) &= 1 \end{aligned}$$

Then, from $P(f|l)$ we can infer that:

$$P(f|l, \neg i) = 0.7$$

and so;

$$P(\neg f|l, \neg i) = 0.3$$

and similarly:

$$\begin{aligned} P(f|\neg l, i) &= 0.8 \\ P(\neg f|\neg l, i) &= 0.2 \end{aligned}$$

Then:

$$\begin{aligned}P(\neg f|l,i) &= P(\neg f|i) \times P(\neg f|l)\\ &= 0.2 \times 0.3\\ &= 0.06\end{aligned}$$

Thus:

$$P(f|l,i) = 0.94$$

In summary:

$$\begin{aligned}P(f|l,i) &= 0.94 & P(\neg f|l,i) &= 0.06\\ P(f|\neg l,i) &= 0.8 & P(\neg f|\neg l,i) &= 0.2\\ P(f|l,\neg i) &= 0.7 & P(\neg f|l,\neg i) &= 0.3\\ P(f|\neg l,\neg i) &= 0 & P(\neg f|\neg l,\neg i) &= 1\end{aligned}$$

3. With no network, 5 variables need:

$$2^5 - 1 = 31$$

numbers. With the network, we need:

$$1 + 2 + 2 + 4 + 2 = 11$$

4. To get the joint probability over the set of variables in the network we apply the chain rule as on page 23 of the slides. We have:

$$\begin{aligned}P(m,\neg t,h,s,\neg c) &= P(h|\neg t).P(\neg c|s,\neg t).P(\neg t|m).P(s|m).P(m)\\ &= 0.7 \times 0.15 \times 0.3 \times 0.8 \times 0.1\\ &= 0.00252\end{aligned}$$

Note that the 0.15 in the second line above is computed from $1 - P(c|s,\neg t)$. Similarly 0.3 is computed from $1 - P(t|m)$.

5. From page 37 of the slides we have:

$$\begin{aligned}\mathbf{P}(M|h,s) &= \frac{\mathbf{P}(M,h,s)}{P(h,s)}\\ &= \alpha\mathbf{P}(M,h,s)\\ &= \alpha\sum_t\sum_c \mathbf{P}(M,h,s,t,c)\end{aligned}$$

Factorising as in Q4, we then have:

$$\begin{aligned}\mathbf{P}(M|h,s) &= \alpha\sum_t\sum_c P(h|t).P(c|s,t).\mathbf{P}(t|M).\mathbf{P}(s|M).\mathbf{P}(M)\\ &= \alpha\mathbf{P}(M).\mathbf{P}(s|M)\sum_t\sum_c P(h|t).P(c|s,t).\mathbf{P}(t|M)\\ &= \alpha\left\langle P(m).P(s|m)\sum_t\sum_c P(h|t).P(c|s,t).P(t|m),\right.\\ &\qquad \left. P(\neg m).P(s|\neg m)\sum_t\sum_c P(h|t).P(c|s,t).P(t|\neg m)\right\rangle\end{aligned}$$

Let's say that:

$$
\begin{aligned}
pm &= P(m).P(s|m)\sum_t\sum_c P(h|t).P(c|s,t).P(t|m) \\
&= P(m).P(s|m)\,(P(h|t).P(t|m).P(c|s,t) + P(h|t).P(t|m).P(\neg c|s,t) \\
&\qquad\qquad P(h|\neg t).P(\neg t|m).P(c|s,\neg t) + P(h|\neg t).P(\neg t|m).P(\neg c|s,\neg t)) \\
&= 0.1 \times 0.8(0.9 \times 0.7 \times 0.95 + 0.9 \times 0.7 \times 0.05 + 0.7 \times 0.3 \times 0.85 + 0.7 \times 0.3 \times 0.15) \\
&= 0.0672
\end{aligned}
$$

Similarly, let:

$$
\begin{aligned}
pm' &= P(\neg m).P(s|\neg m)\sum_t\sum_c P(h|t).P(c|s,t).P(t|\neg m) \\
&= 0.9 \times 0.2(0.9 \times 0.1 \times 0.95 + 0.9 \times 0.1 \times 0.05 + 0.7 \times 0.9 \times 0.85 + 0.7 \times 0.9 \times 0.15) \\
&= 0.1296
\end{aligned}
$$

Now:

$$
\begin{aligned}
\mathbf{P}(M|h,s) &= \alpha\langle pm, pm'\rangle \\
&= \alpha\langle 0.0672, 0.1296\rangle \\
&= \langle 0.34, 0.66\rangle
\end{aligned}
$$

6. To compute the first sample:

   - Sample m.

     $P(m)$ is 0.1, so we want to generate m one time in 10 and $\neg m$ 9 times in 10. If we generate a random number with equal probability of picking any number between 0 and 1, then that number will be less than or equal to 0.1 1 time in 10, and greater than 0.1 9 times in 10.

     So, our method for picking whether m is true or false is to compare a random number (picked between 0 and 1, inclusive, with equal probability for everry number in that range) with $P(m)$. If the random number is less than or equal to $P(m)$, then m, is true. Otherwise m is false.

     Our first random number is 0.14, so m is false.

   - Sample s.

     Given than we have $\neg m$, we now sample s given $\neg m$. $P(s|\neg m)$ is 0.2. Our random number is 0.57. So s is false.

   - Sample t.

     Similarly to the previous case, $P(t|\neg m = 0.1$, and our random number is 0.01, less than 0.1 So t is true.

   - Sample h.

     Since t is true, we need to sample given t. $P(h|t) = 0.9$, our random number is 0.43, so h is true.

   - Sample c.

     We sample c given $\neg s$ and t. $P(c|\neg s, t) = 0.85$ and the next random number is 0.59, so c is true.

   So we have a sample $\langle \neg m, \neg s, t, h, c\rangle$.

   Following the same procedure again we get, in turn, the samples:

$$
\begin{gathered}
\langle \neg m, s, \neg t, \neg h, c\rangle \\
\langle \neg m, \neg s, \neg t, h, \neg c\rangle \\
\langle \neg m, \neg s, \neg t, h, \neg c\rangle \\
\langle \neg m, \neg s, \neg t, h, \neg c\rangle
\end{gathered}
$$

Joint probabilities are then the proportion of samples with the relevant variables, so we have:

$$\begin{aligned} P(\neg m, \neg s, t, h, c) &\approx 0.2 \\ P(\neg m, s, \neg t, \neg h, c) &\approx 0.2 \\ P(\neg m, \neg s, \neg t, h, \neg c) &\approx 0.6 \end{aligned}$$

and all other probabilities are zero.

Note that these results are approximate. The probabilities only become accurate after many iterations (as the number of iterations approaches infinity, the probability approaches the correct value).

7. Since we are computing $P(m|h, s)$ we will only use samples in which h and s are true. Using rejection sampling we proceed as before, but stop as soon as we get a value for h or s which is false.

   So, starting from the begining of the list of random numbers, we have:

   - Sample m
     As in the previous question, the first number is 0.14, and so m is false.
   - Sample s.
     As in the previous question, s is false.

   This means that the sample will not contribute to computing $P(m|h, s)$, so we reject it and start over.

   - Sample m
     Our random number is 0.01, so m is true.
   - Sample s.
     The random number is 0.43, $P(s|m)$ is 0.8, so s is true, and we continue.
   - Sample t.
     t is true. (At this point we will stop saying how to determine if the variable is true or false — I assume you get that by now.)
   - Sample h.
     If h were false, we would reject this sample, but it is true, so we continue.
   - Sample c.
     c is true.

   So we have a sample $\langle m, s, t, h, c \rangle$.

   The next sample involves 5 false starts where we establish that m is false, but then find that s is also false.

   Eventually, we have:

   - Sample m
     Our random number is 0.34, so m is false.
   - Sample s.
     The random number is 0.14, $P(s|\neg m)$ is 0.2, so s is true, and we continue.
   - Sample t.
     t is false.
   - Sample h.
     h is true.
   - Sample c.
     c is true.

So we have a sample $\langle \neg m, s, \neg t, h, c \rangle$.

Again there are a series of false starts before we get the third sample, which, by my calculation is also $\langle \neg m, s, \neg t, h, c \rangle$.

At this point we have three samples where s and t are true:

$$\langle m, s, t, h, c \rangle$$
$$\langle \neg m, s, \neg t, h, c \rangle$$
$$\langle \neg m, s, \neg t, h, c \rangle$$

and we compute the probability $P(m|h, s)$ as the proportion of the samples where h and s are ture, where m is also true, so:

$$P(m, s, t, h, c) \approx 0.33$$
$$P(\neg m, s, \neg t, h, c) \approx 0.66$$

Again, this is approximate, and will improve with more samples. Indeed, if you complete the 5 samples that the question asks for, you may get a better approximation.

8. For importance sampling, we start by picking an order in which we will evaluate the variables. We will use the same order as before.

   Then, starting at the beginnning of the list of random numbers

   - Sample m
     m is false. (We use the 0.14 to get this — we are starting over with the random numbers again.).
   - s is true by definition.
     w is set to the value of $P(s|\neg m)$, so w is 0.2.
   - Sample t
     t is false.(We use 0.57 to get this.)
   - h is true by definition.
     Thus we update w with $P(h|\neg t) = 0.7$. $w = 0.2 \times 0.7 = 0.14$.
   - Sample for c.
     c is true.

   So we have the sample $\langle \neg m, s, \neg t, h, c \rangle$ and the weight is 0.14.

   For the second sample:

   - Sample m
     m is false. (We use the 0.43 to get this).
   - s is true by definition.
     w is set to the value of $P(s|\neg m)$, so w is 0.2.
   - Sample t
     t is false.(We use 0.59 to get this.)
   - h is true by definition.
     Thus we update w with $P(h|\neg t) = 0.7$. $w = 0.2 \times 0.7 = 0.14$.
   - Sample for c.
     c is true.

   So again we have the sample $\langle \neg m, s, \neg t, h, c \rangle$ and the weight is 0.14.

   My calculation says that the next 3 samples are all $\langle \neg m, s, \neg t, h, c \rangle$ with weight 0.14.

   Since we only have samples with m false, we have $P(m|h, s) \approx 0$.

   However this doesn't show us how the weights work. Let's look at a 6th sample, and pretend that the next random number is 0.05 rather than 0.21 as it is in the list. That would mean we have:

- Sample m
  m is true. (Since we have 0.05 as the random number).
- s is true by definition.
  w is set to the value of $P(s|m)$, so w is 0.8.
- Sample t
  t is true. (We go back to using the values in the table, 0.34 in this case.)
- h is true by definition.
  Thus we update w with $P(h|t) = 0.9$. $w = 0.8 \times 0.9 = 0.72$.
- Sample for c.
  c is true.

So we have $\langle m, s, t, h, c\rangle$ with $w = 0.72$.

So now we have a weight for $P(m|h, s)$ of 0.72 from that last sample, and weight of $5 \times 0.14 = 0.7$ from the previous 5 samples. So:

$$P(M|h, s) \approx \alpha\langle 0.72, 0.7\rangle \qquad\qquad \approx \langle 0.507, 0.493\rangle$$

and $P(m|h, s) = 0.507$ even though we only have one sample.

Again I stress that these values are very approximate with so few samples and as you can see the two values for $P(m|h, s)$ are not particularly close. I would not expect them to be. However, if the values were very different after several thousand samples, that would be a concern (and a suggestion that something was wrong).

9. For Gibbs sampling, we start by picking a state. We know that $S = s$ and $H = h$ from the evidence, but we need to pick values for M, T and C. We assume both values are equally likely, and then sample, using the first three numbers from the sequence.

   - Sample M. The random number is 0.14, which is less than 0.5, so $M = m$.
   - Sample T. The random number is 0.57, which is greater than 0.5, so $T = \neg t$.
   - Sample C. The random number is 0.01, which is less than 0.5, so $C = c$.

   Thus we start with a state $\langle m, s, \neg t, c, h\rangle$. We will modify the variables in order.

   - First we sample M, given its Markov blanket. The Markov blanket of M is S and T, so we need $P(M|s, \neg t)$. The slides (page 79) tells us that:

     $$\mathbf{P}(X|mb(X)) = \alpha\mathbf{P}(X|parents(X)) \prod_{Y\in Children(X)} \mathbf{P}(Y|parents(Y))$$

     which in this case, since S and T are the children of M, becomes:

     $$\begin{aligned}\mathbf{P}(M|s, \neg t) &= \alpha\mathbf{P}(M)\mathbf{P}(s|M)\mathbf{P}(\neg t|M)\\ &= \alpha\langle 0.1, 0.9\rangle\langle 0.8, 0.2\rangle\langle 0.3, 0.9\rangle\\ &= \alpha\langle 0.1 \times 0.8 \times 0.3, 0.9 \times 0.2 \times 0.9\rangle\\ &= \alpha\langle 0.024, 0.162\rangle\\ &\langle 0.13, 0.87\rangle\end{aligned}$$

     Now we have this distribution, we can sample M. The next number on the list is 0.43. This is greater than 0.13, so $M = \neg m$, and that state is $\langle \neg m, s, \neg t, c, h\rangle$
   - The next variable is S, but that is evidence, so it doesn't change.

- The next variable is T, so we need to sample T given its Markov blanket. The Markov blanket of T is all the other variables, so we will write it as $mb(T)$ as shorthand:

$$\begin{aligned}\mathbf{P}(T|mb(T)) &= \alpha\mathbf{P}(T|\neg m)\mathbf{P}(h|T)\mathbf{P}(c|s,T)\\ &= \langle 0.1, 0.9\rangle\langle 0.9, 0.7\rangle\langle 0.95, 0.85\rangle\\ &= \alpha\langle 0.0855, 0.5355\rangle\\ &= \langle 0.14, 0.86\rangle\end{aligned}$$

  Now we sample T from this distribution. The next number on the list os 0.59, this is greater than 0.14, so $T = \neg t$, and the state is $\langle \neg m, s, \neg t, c, h\rangle$.

- The next variable is C, so we need to sample C given its Markov banket. The Markov blanket of C is S and T. Since C has no children, this is very simple — there is no computation:

$$\begin{aligned}\mathbf{P}(C|s,\neg t) &= \alpha\mathbf{P}(C|s,\neg t)\\ &= \langle 0.85, 0.15\rangle\end{aligned}$$

  Sampleing C from this we get $C = c$. That is because the next random numberis 0.5 which is less than 0.85. The state (continues to be) $\langle \neg m, s, \neg t, c, h\rangle$.

- The next variable, H is evidence, so there is no chnage of state.
- We cycle back to M. The state of the Markov blanket of M is still $S = s$ and $T = \neg t$, so the probability of M given its Markov blanket is:

$$\mathbf{P}(M|s,\neg t) = \langle 0.13, 0.87\rangle$$

  just aswe calculated before. Note that if S or T had changed value, we would have to recompute $\mathbf{P}(M|mb(M))$.

  Sampling from $\langle 0.13, 0.87\rangle$, the next random number is 0.12, so $M = m$, and the state has changed to $\langle m, s, \neg t, c, h\rangle$.

- The next variable is S, which is evidence.
- For our fifth (and, for this example, final) sample, the variable is T. The state of the Markov blanket of T has changed (because M has changed state), so we have to recompute:

$$\begin{aligned}\mathbf{P}(T|mb(T)) &= \alpha\mathbf{P}(T|m)\mathbf{P}(h|T)\mathbf{P}(c|s,T)\\ &= \alpha\langle 0.7, 0.3\rangle\langle 0.9, 0.7\rangle\langle 0.95, 0.85\rangle\\ &= \alpha\langle 0.5985, 0.1785\rangle\\ &= \langle 0.77, 0.23\rangle\end{aligned}$$

  We then sample T using this distribution. The next random number is 0.54, which gives us $T = t$ and the state changes to $\langle m, s, t, c, h\rangle$.

Now, to get the value of $P(m|h,s)$ we count how many states had $M = m$ and our estimate of $P(m|h,s)$ is thsi number divided by the total number of states. We have six states in total, the initial state, and one created by every sample. The algorithm (page 78 of the slides), tells us that the only states to use are the ones that follow a sample.

Thus we have 5 states, and $M = m$ in two of them (the last two), so:

$$P(m|h,s) \approx 0.4$$

10. Again, I'm not going to post a solution for this optional bit of the tutorial, but if you did it, you can check the correctness of your solution against the value for $P(m|h,s)$ in question 5.