6CCS3AIN, Tutorial 02 (Version 1.0)

1. From the joint distribution (see slides), compute:

   (a) $P(cavity)$

   (b) $\mathbf{P}(Toothache)$.

   (c) $\mathbf{P}(Toothache|cavity)$.

   (d) $P(catch \vee cavity)$

   (e) $\mathbf{P}(Cavity|toothache \vee catch)$

   Recall what the difference is between P and $\mathbf{P}$ (slides from Lecture/Week 2).

2. You take a test T to tell whether you have a disease D. The test comes back positive. You know that test is 95% accurate (the probability of testing positive when you do have the disease is 0.95, and the probability of testing negative when you don't have the disease is also 0.95). You also know that the disease is rare, only 1 person is 10,000 gets the disease.

   What is the probability that you have the disease?

   How would this change if the disease was more common, say affecting 1 person in 100?

3. Consider two tests, A and B, for a virus.

   Test A is 95% effective at recognizing the virus when it is present (that is 95% of the time that the virus is present, the test detects it), but has a 10% false positive rate (that is, 10% of the time it indicates the virus is present when it is not).

   Test B is 90% effective at recognizing the virus, but has a 5% false positive rate.

   The two test use different, independent, methods of indentifying the virus.

   1% of all people have the virus.

   Joe tests positive for the virus using test A. Bob tests positive using test B. Who is more likely to have the virus?

4. Look at the meningitis example on the nodes from Lecture/Week 2. Carry out the calculation without using the value for $P(s)$.

   This uses the trick of considering the denominator to be a normalization constant α.

   You'll need a value of $P(s|\neg m)$ to do this — use $P(s|\neg m) = 0.1$.

5. Now the optional computational part of the tutorial.

   On KEATS you can find the file `underwear.py`. This is the "wrong underwear" example from the notes coded using a Python package called `pomegranate`:

   `https://pomegranate.readthedocs.io/en/latest/index.html`

   `pomegranate` provides tools for probabilistic inference. You can install it using:

   ```
   conda install pomegranate
   ```

   if you have Anaconda installed[1]. Otherwise, see the webpage.

   With `pomegranate` installed, you can run the underwear example with:

   ```
   python underwear.py
   ```

   from the command line. As set up, this gives you the result for the first observation of the wrong underwear. (You can check this against the value in the slides.) To run the calculation for the second observation of the wrong underwear, you need to change the value of `prior_cheat`, the prior probability of cheating.

   What do you change it to? For the second observation, the prior is the value you just obtained for the first observation. Do this, and verify that you get the same result as in the slides.

   Now try building your own probability model for the Meningitis example.

[1] If you don't have Anaconda installed, you can install that from `https://www.anaconda.com/download`. If you are planning to do much with Python, Anaconda is pretty handy to have since it makes it easy to install new packages.

6CCS3AIN, Tutorial 02 Answers (Version 1.0)

1. (a)

$$\begin{aligned}
P(cavity) &= \sum_{\omega \models cavity} P(\omega) \\
&= P(cavity \land catch \land toothache) + P(cavity \land \neg catch \land toothache) \\
&\quad + P(cavity \land catch \land \neg toothache) + P(cavity \land \neg catch \land \neg toothache) \\
&= 0.108 + 0.012 + 0.072 + 0.008 \\
&= 0.2
\end{aligned}$$

(b)

$$\begin{aligned}
\mathbf{P}(Toothache) &= \langle P(toothache), P(\neg toothache) \rangle \\
P(toothache) &= P(cavity \land catch \land toothache) + P(cavity \land \neg catch \land toothache) \\
&\quad + P(\neg cavity \land catch \land toothache) + P(\neg cavity \land \neg catch \land toothache) \\
&= 0.108 + 0.012 + 0.016 + 0.064 \\
&= 0.2 \\
\mathbf{P}(Toothache) &= \begin{pmatrix} 0.2 \\ 0.8 \end{pmatrix}
\end{aligned}$$

The second probability is computed either in the same way as the first, or by knowing that the two sum to 1.

(c)

$$\begin{aligned}
P(toothache|cavity) &= \frac{P(toothache \land cavity}{P(cavity)} \\
P(toothache \land cavity) &= P(cavity \land catch \land toothache) + P(cavity \land \neg catch \land toothache) \\
&= 0.108 + 0.012 \\
&= 0.12
\end{aligned}$$

and so:

$$P(toothache|cavity) = 0.6$$

Similarly (or by subtraction):

$$P(\neg toothache|cavity) = 0.4$$

and:

$$\mathbf{P}(Toothache|cavity) = \begin{pmatrix} 0.6 \\ 0.4 \end{pmatrix}$$

(d)

$$\begin{aligned}
P(catch \lor cavity) &= P(cavity \land catch \land toothache) + P(cavity \land catch \land \neg toothache) \\
&\quad + P(\neg cavity \land catch \land toothache) + P(\neg cavity \land catch \land \neg toothache) \\
&\quad + P(cavity \land \neg catch \land toothache) + P(cavity \land \neg catch \land \neg toothache) \\
&= 0.108 + 0.072 + 0.016 + 0.144 + 0.012 + 0.008 \\
&= 0.36
\end{aligned}$$

(e)

$$\mathbf{P}(Cavity|toothache \lor catch) = \frac{\mathbf{P}(Cavity \land (toothache \lor catch))}{P(toothache \lor catch)}$$

In a similar way to the previous question, we can compute:

$$P(toothache \lor catch) = 0.416$$

(in this case we sum up all the values on slide 32 except those in the last column where we have $\neg cavity$ and $\neg toothache$). Then:

$$\begin{aligned}
P(cavity \wedge (toothache \vee catch)) &= 0.108 + 0.012 + 0.072 \\
&= 0.192 \\
P(cavity|toothache \vee catch) &= 0.462 \\
\mathbf{P}(Cavity|toothache \vee catch) &= \begin{pmatrix} 0.462 \\ 0.538 \end{pmatrix}
\end{aligned}$$

2. The problem tells us that:

$$\begin{aligned}
P(t|d) &= 0.95 \\
P(\neg t|\neg d) &= 0.95 \\
P(d) &= 0.0001
\end{aligned}$$

And:

$$\begin{aligned}
P(t|\neg d) &= 1 - P(\neg t|\neg d) \\
&= 0.05
\end{aligned}$$

Now,

$$\begin{aligned}
P(d|t) &= \frac{P(d \wedge t)}{P(t)} \\
&= \frac{P(t|d)P(d)}{P(t|d)P(d) + P(t|\neg d)P(\neg d)} \\
&= \frac{0.95 \cdot 0.0001}{0.95 \cdot 0.0001 + 0.05 \cdot 0.9999} \\
&= 0.00189
\end{aligned}$$

So the probability of having the disease once you have the positive test is 0.00189, which is not large, despite the accuracy of the test. The reason is that the disease is very unlikely, meaning that the prior probability is low.

Repeating the calculation with $P(d) = 0.01$ gives a probability of having the disease of 0.16, which is obviously much larger (though still much less than 0.95, which many people will say, without doing the calculation, is the chance of having the disease if the test comes back positive.)

3. We have:

$$\begin{aligned}
P(a|v) &= 0.95 & P(b|v) &= 0.9 \\
P(a|\neg v) &= 0.1 & P(b|\neg v) &= 0.05
\end{aligned}$$

Now, for Joe we want:

$$\begin{aligned}
P(v|a) &= \frac{P(a|v)P(v)}{P(a|v)P(v) + P(a|\neg v)P(\neg v)} \\
&= \frac{0.95 \cdot 0.01}{0.95 \cdot 0.01 + 0.99 \cdot 0.1} \\
&= 0.0876
\end{aligned}$$

and, as before, because the virus is unlikely even after an accurate test, there is not that great a chance of having the virus.

Carrying out a similar calculation for Bob, we find that:

$$P(v|b) = 0.154$$

so even though the second test is less accurate, the lower false positive rate means that Bob is about twice as likely to have the virus as Joe.

4. We have, from the slides for Lecture/Week 2:

$$P(m) = 0.0001$$
$$P(s|m) = 0.8$$

We also know that $P(s|\neg m) = 0.1$. Then, using $\odot$ to denote the Hadamard product (not important that you know what it is, it's just important that you multiply row-wise; you can also see it as a set of 2 equations)

$$\begin{aligned}
\mathbf{P}(M|s) &= \alpha \mathbf{P}(s|M) \odot \mathbf{P}(M) \\
\mathbf{P}(M|s) &= \alpha \begin{pmatrix} P(s|m) \\ P(s|\neg m) \end{pmatrix} \odot \begin{pmatrix} P(m) \\ P(\neg m) \end{pmatrix} \\
&= \alpha \begin{pmatrix} 0.8 \\ 0.1 \end{pmatrix} \odot \begin{pmatrix} 0.0001 \\ 0.9999 \end{pmatrix} \\
&= \alpha \begin{pmatrix} 0.8 \cdot 0.0001 \\ 0.1 \cdot 0.9999 \end{pmatrix} \\
&= \alpha \begin{pmatrix} 0.00008 \\ 0.09999 \end{pmatrix} \\
&\approx \begin{pmatrix} 0.0008 \\ 0.9992 \end{pmatrix}
\end{aligned}$$

5. If you did this bit of the tutorial, you should be able to check that your code works by checking the answers against the calculations above.