Artificial Intelligence and Machine Learning

Barbara Caputo

1. *You go to see the doctor about an ingrowing toenail. The doctor selects you at random to have a blood test for swine flu, which for the purposes of this exercise we will say is currently suspected to affect 1 in 10,000 people in Australia. The test is 99% accurate, in the sense that the probability of a false positive is 1%. The probability of a false negative is zero. You test positive. What is the new probability that you have swine flu?*

Let $P(D)$ be the probability you have swine flu.

Let $P(T)$ be the probability of a positive test.

We wish to know $P(D|T)$.

Bayes theorem says

$$P(D|T) = \frac{P(T|D)P(D)}{P(T)}$$

which in this case can be rewritten as

$$P(D|T) = \frac{P(T|D)P(D)}{P(T|D)P(D) + P(T|ND)P(ND)}$$

where $P(ND)$ means the probability of not having swine flu.

We have

$P(D) = 0.0001$ (the a priori probability you have swine flu).

$P(ND) = 0.9999$

$P(T|D) = 1$ (if you have swine flu the test is always positive).

$P(T|ND) = 0.01$ (1% chance of a false positive).

Plugging these numbers in we get

$$P(D|T) = \frac{1 \times 0.0001}{1 \times 0.0001 + 0.01 \times 0.9999} \approx 0.01$$

That is, even though the test was positive your chance of having swine flu is only 1%. (This is essentially the "defense attorney's" argument we discussed in the lectures, though in this case it's not a fallacy because your $P(D)$ is indeed very low.)

Now imagine that you went to a friend's wedding in Mexico recently, and (for the purposes of this exercise) it is know that 1 in 200 people who visited Mexico recently come back with swine flu. Given the same test result as above, what should your revised estimate be for the probability you have the disease?

However, if you went to Mexico recently then your starting $P(D)$ is 0.005. In this case

$$P(D|T) = \frac{1 \times 0.005}{1 \times 0.005 + 0.01 \times 0.995} \approx 0.33$$

and you should be a lot more worried.

2. *Imagine that, while in Mexico, you also took a side trip to Las Vegas, to pay homage to the TV show CSI. Late one night in a bar you meet a guy who claims to know that in the casino at the Tropicana there are two sorts of slot machines: one that pays out 10% of the time, and one that pays out 20% of the time [note these numbers may not be very realistic]. The two types of machines are coloured red and blue. The only problem is, the guy is so drunk he can't quite remember which colour corresponds to which kind of machine. Unfortunately, that night the guy becomes the vic in the next CSI episode, so you are unable to ask him again when he's sober.*

Next day you go to the Tropicana to find out more. You find a red and a blue machine side by side. You toss a coin to decide which machine to try first; based on this you then put the coin into the red machine. It doesn't pay out. How should you update your estimate of the probability that this is the machine you're interested in? What if it had paid out - what would be your new estimate then?

Let $P(R)$ be the probability that the red machine is the one that pays out more often; similarly $P(B) = 1 - P(R)$.

Let $P(J)$ be the probability of payout ("jackpot").

We are interested in $P(R|NJ)$.

Bayes theorem says

$$P(R|NJ) = \frac{P(NJ|R)P(R)}{P(NJ|R)P(R) + P(NJ|B)P(B)}$$

We start with $P(R) = P(B) = 0.5$. This gives us

$$P(R|NJ) = \frac{0.8 \times 0.5}{0.8 \times 0.5 + 0.9 \times 0.5} \approx 0.47$$

If the red machine did pay out then we have

$$P(R|J) = \frac{P(J|R)P(R)}{P(J|R)P(R) + P(J|B)P(B)}$$

Substituting in....

$$P(R|J) = \frac{0.2 \times 0.5}{0.2 \times 0.5 + 0.1 \times 0.5} \approx 0.66$$

That is, a pay-out makes you a lot more confident about which is the good machine than no pay-out. This makes sense since paying out is a rare event, so you would expect it to give you a lot of information.

You may like to take this further by considering how many times you'd have to fail to get a payout from the red machine before being say 90% confident it's not that machine.

I. Emily hits 60% of her free throws in basketball games. She had 25 free throws in last week's game. Use this information to answer the next **two** questions.

1. What is the average number of hits?
   (a) 10
   (b) 15
   (c) 20
   (d) 25

ANSWER: (b)

REASON:

The average (or the expected value) is $n \times p = 25 \times 0.60 = 15$.

2. What is the standard deviation of Emily's hit ?

(a) 6

(b) 3

(c) 3.2

(d) $\sqrt{6}$

ANSWER: (d)

REASON:

The standard deviation is $\sqrt{np(1-p)} = \sqrt{25 \times 0.60 \times (1-0.60)} = \sqrt{6} = 2.45.$

In the previous question, suppose Emily had 7 free throws in yesterday's game.

1. What is the probability that she made at least 5 hits?
   (a) 0.2613
   (b) 0.1306
   (c) 0.0280
   (d) 0.1586
   (e) 0.4199

ANSWER: (e)

REASON:

Denote Y the hits in Emily's 7 free throws. The event that she made at least 5 hits is then

$$(Y \geq 5) = (Y = 5 \text{ or } 6 \text{ or } 7).$$

So,

$$P(Y \geq 5) = P(Y = 5) + P(Y = 6) + P(Y = 7)$$

So,

$$P(Y \geq 5) = P(Y = 5) + P(Y = 6) + P(Y = 7)$$

$$\begin{aligned}
P(Y=5) &= \frac{7!}{5!(7-5)!}(.6)^5(.4)^{7-5} = \frac{7 \times 6 \times 5!}{5! \times 2!}(.6)^5(.4)^2 \\
&= \frac{7 \times 6}{2 \times 1}(.6)^5(.4)^2 \text{ (note that 5! was factored out)} \\
&= 21 \times (.6)^5(.4)^2 = 0.2613 \\
P(Y=6) &= \frac{7!}{6!(7-6)!}(.6)^6(.4)^{7-6} = \frac{7 \times 6!}{6! \times 1!}(.6)^6(.4)^1 \\
&= \frac{7}{1}(.6)^6(.4)^1 \text{ (note that 6! was factored out)} \\
&= 7 \times (.6)^6 \times .4 = 0.1306 \\
P(Y=7) &= \frac{7!}{7!(7-7)!}(.6)^7(.4)^{7-7} = \frac{7!}{7! \times 0!}(.6)^7(.4)^0 \\
&= (.6)^7 \times 1 \text{ (note that 7! was factored out)} \\
&= 0.0280.
\end{aligned}$$

Note that, $0! = 1$ and $(.4)^0 = 1$ above.
It follows that the probability of interest is $0.2613 + 0.1306 + 0.0280 = 0.4199$.

1. X is a normally normally distributed variable with mean $\mu = 30$ and standard devia

4. Find

a) $P(x < 40)$

b) $P(x > 21)$

c) $P(30 < x < 35)$

1. Note: What is meant here by area is the area under the standard normal curve.

a) For $x = 40$, the z-value $z = (40 - 30) / 4 = 2.5$

Hence $P(x < 40) = P(z < 2.5)$ = [area to the left of 2.5] = 0.9938

b) For $x = 21$, $z = (21 - 30) / 4 = -2.25$

Hence $P(x > 21) = P(z > -2.25)$ = [total area] - [area to the left of -2.25]

= 1 - 0.0122 = 0.9878

c) For $x = 30$, $z = (30 - 30) / 4 = 0$ and for $x = 35$, $z = (35 - 30) / 4 = 1.25$

Hence $P(30 < x < 35) = P(0 < z < 1.25)$ = [area to the left of $z = 1.25$] - [area to the

0]

= 0.8944 - 0.5 = 0.3944