Assignment 1: Knowledge Diagnostics

Ash

2018/12/06

1 Rules Of Probability Theory

1. P(A,B) = P(A|B)P(B) : True, because the probability of A given B multiply by the probability of B (right hand side) is just the joint probability of A and B happening together (left hand side).
2. P(A) = P(A|B)P(B) + P(A|not B)P(not B): True, because P(A|B)P(B) = P(A,B) and P(A|not B)P(not B) = P(A, not B), which sum together represents all the possibility of B (as P(not B) = 1 - P(B)), so it must equal the probability of A alone.
3. P(A) = P(A|B)P(B) + P(A|C)P(C) + P(A|D)P(D) : False, because similar with case 2, the right hand side equals P(A,B) + P(A,C) + P(A,D), but B, C, D are not the exhaust list of possible outcomes like B and not B, so the left hand side does not equal the right hand side.
4. P(A|B) = $\frac{P(B|A)P(A)}{P(B)}$: True, this is just Bayes' Rule, and it makes sense as P(B|A)P(A) = P(A,B) = P(A|B)P(B). Intuitively, the probability of A conditioned on B (left hand side) equals the probability of A and B happening together divides by the probability of B happening alone (right hand side).
5. P(A|B)P(B) = P(B|A)P(A) : True, since P(B|A)P(A) = P(A,B) = P(A|B)P(B).

2 Logarithms and probability distributions

1. Normal distribution

$$
\begin{aligned}
ln(pdf) &= ln\left(\frac{1}{\sqrt{2\mu\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}\right) \\
&= ln\frac{1}{\sqrt{2\mu\sigma^2}} + ln\left(e^{-\frac{(x-\mu)^2}{2\sigma^2}}\right) \\
&= ln\frac{1}{\sqrt{2\mu\sigma^2}} - \frac{(x-\mu)^2}{2\sigma^2}
\end{aligned}
$$

2. Exponential distribution

$$ln(pdf) = ln(\lambda e^{-\lambda x}) = ln\lambda - \lambda x$$

3. Gaussian distribution (with scale parameter θ and shape parameter k)

$$\begin{aligned} ln(pdf) &= ln\left(\frac{1}{\Gamma(k)\theta^k}x^{k-1}e^{-\frac{x}{\theta}}\right) \\ &= ln\frac{1}{\Gamma(k)\theta^k} + ln(x^{k-1}) - \frac{x}{\theta} \end{aligned}$$

3 Normal distribution

In this problem the integral from $-\infty$ to ∞ $\left(\int_{-\infty}^{\infty}\right)$ will be used interchangeably with the integral symbol $\left(\int\right)$.

According to the law of unconscious statistician:

$$E[f(x)] = \int_{-\infty}^{\infty} f(x)pdf_x(x)dx = \int (x^3+3)\frac{1}{\sqrt{2\pi\sigma^2}}e^{\frac{(x-\mu)^2}{2\sigma^2}}dx \tag{1}$$

Let

$$t = \frac{x-\mu}{\sigma}$$

so

$$x = \sigma t + \mu$$

and

$$dx = \sigma dt$$

and when $x \to \infty$ we have $t \to \infty$, $x \to -\infty$ we have $t \to -\infty$.

Substitute the variable onto (1) we have:

$$\begin{aligned} E[f(x)] &= \int \left[(\sigma t+\mu)^3 + 3\right]\frac{1}{\sqrt{2\pi\sigma^2}}e^{-t^2/2}\sigma dt \\ &= \frac{1}{\sqrt{2\pi}}\int \left(\sigma^3t^3 + 3\sigma^2\mu t^2 + 3\sigma\mu^2 t + \mu^3 + 3\right)e^{-t^2/2}dt \\ &= \frac{1}{\sqrt{2\pi}}\left[\sigma^3\int t^3e^{-t^2/2}dt + 3\sigma^2\mu\int t^2e^{-t^2/2}dt + 3\sigma\mu^2\int te^{-t^2/2}dt + \left(\mu^3+3\right)\int e^{-t^2/2}dt\right] \end{aligned}$$

We have the Gaussian integral

$$\int e^{-a(z+b)^2}dz = \sqrt{\frac{\pi}{a}}$$

So

$$\left(\mu^3+3\right)\int e^{-t^2/2}dt = \left(\mu^3+3\right)\sqrt{2\pi}$$

For $3\sigma\mu^2 \int te^{-t^2/2}dt$, let

$$u = 1 \Rightarrow du = 0$$
$$v = e^{-t^2/2} \Rightarrow dv = -te^{-t^2/2}dt$$

so integration by parts give us:

$$\begin{aligned}3\sigma\mu^2 \int te^{-t^2/2}dt &= 3\sigma\mu^2\left[-e^{-t^2/2}|_{-\infty}^{\infty} + \int 0e^{-t^2/2}\right] \\ &= 3\sigma\mu^2(0-0) = 0\end{aligned}$$

For $3\sigma^2\mu \int t^2e^{-t^2/2}dt$ let $u = t$ and $v = e^{-t^2/2}$ and doing the same thing as above with integration by parts we have

$$3\sigma^2\mu \int t^2e^{-t^2/2}dt = 3\sigma^2\mu\sqrt{2\pi}$$

For $\sigma^3 \int t^3e^{-t^2/2}dt$ let $u = t^2$ and $v = e^{-t^2/2}$ and doing the same thing as above with integration by parts we have

$$\sigma^3 \int t^3e^{-t^2/2}dt = 0$$

Combining all of the above we have

$$E[f(x)] = \frac{1}{\sqrt{2\pi}}\left[0 + 3\sigma^2\mu\sqrt{2\pi} + 0 + \left(\mu^3+3\right)\sqrt{2\pi}\right] = 3\sigma^2\mu + \mu^3 + 3$$

We have

$$P(f(x) \leq 0.5) = P(x^3 + 3 \leq 0.5) = P(x \leq \sqrt[3]{-2.5})$$

We also have

$$P(x \leq \sqrt[3]{-2.5}) = cdf_x(\sqrt[3]{-2.5}) = \frac{1}{2}\left[1 + erf\left(\frac{\sqrt[3]{-2.5}-\mu}{\sigma\sqrt{2}}\right)\right]$$

So

$$P(f(x) > 0.5) = 1 - P(f(x) \leq 0.5) = 1 - \frac{1}{2}\left[1 + erf\left(\frac{\sqrt[3]{-2.5}-\mu}{\sigma\sqrt{2}}\right)\right]$$

Simulation on Python code:

```python
from scipy.special import erf
import numpy as np

m = 2
s = 4

def prob(mu=m, sigma=s):
    return 1/2 - 1/2*erf((-2.5**(1/3)-mu)/(sigma*np.sqrt(2)))

def drawfx(mu=m, sigma=s):
    x = np.random.normal(mu, sigma, 1)
    return x**3+3

def checkfx(trial):
    count = 0
    for _ in range(trial):
        fx = drawfx()
        if fx > 0.5:
            count +=1
    print(float(count/trial))

print(prob())
checkfx(1000000)
```

The above returns the results (for 1 million trials):

```
0.799350126082228
0.798838
```

The analytical solution matches reasonably well with the simulation result.

4 Marginal and conditional probabilities

Because we are working in the working population of the country the instructions tell us:

P(young|working) = P(young) (given working) = 0.273

P(unemployed|working, young) = P(unemployed|young) (given working) = 0.382

P(unemployed|working, not young) = P(unemployed|not young) (given working) = 0.224

We can calculate:

$$\begin{aligned}
P(unemployed) &= P(unemployed \mid young)P(young) + P(unemployed \mid notyoung)P(notyoung) \\
&= 0.382 \times 0.273 + 0.224 \times (1 - 0.273) = 0.267 \\
P(notunemployed) &= 1 - P(unemployed) = 1 - 0.267 = 0.733 \\
P(young \mid unemployed) &= \frac{P(young, unemployed)}{P(unemployed)} \\
&= \frac{P(unemployed \mid young)P(young}{P(unemployed)} \\
&= \frac{0.382 \times 0.273}{0.267} = 0.391 \\
P(notyoung \mid unemployed) &= \frac{P(notyoung, unemployed)}{P(unemployed)} \\
&= \frac{P(unemployed \mid notyoung)P(notyoung}{P(unemployed)} \\
&= \frac{0.224 \times (1 - 0.273)}{0.267} = 0.610
\end{aligned}$$

5 Inference

From the expert we know that:

$$P(Trained \mid CanRead) = P(Trained \mid R) = 0.6$$
$$P(Trained \mid CannotRead) = P(Trained \mid -R) = 0.1$$

Our initial belief is $P(Olivia = R) = P(O = R) = 0.3$, by which we can infer that $P(Olivia = -R) = P(O = -R) = 0.7$. After receiving information from the expert and with the knowledge that Olivia is receiving training we need to update our belief according to Bayes' Rule:

$$P(O = R \mid Trained) = \frac{P(Trained \mid O = R)P(O = R)}{P(Trained)} \tag{2}$$

$$= \frac{P(Trained \mid O = R)P(O = R)}{P(Trained \mid O = R)P(O = R) + P(Trained \mid O = -R)P(O = -R)} \tag{3}$$

Since Olivia is in our population of interest (the population of kids going to Grade 1) and assume that the expert gives correct statistics we can say that

$$P(Trained \mid O = R) = P(Trained \mid R) = 0.6$$

because the probability of a child being trained given that he/she can read at Grade 1 is 0.6, and we know that Olivia is a child going to Grade 1 so the probability of Olivia being trained given that she can read at Grade 1 is also 0.6. Similarly

$$P(Trained \mid O = -R) = P(Trained \mid -R) = 0.1$$

.

Plugging the probabilities into (3) we have

$$P(O = R \mid Trained) = \frac{P(Trained \mid O = R)P(O = R)}{P(Trained \mid O = R)P(O = R) + P(Trained \mid O = -R)P(O = -R)}$$
$$= \frac{0.6 \times 0.3}{0.6 \times 0.3 + 0.1 \times 0.7} = 0.72$$

So now we are 72% sure that Olivia can read at Grade 1 because we know she is being trained.

The above inference is based on Bayes' Rule of conditional probability, which is a form of updating the prior ($P(O = R)$) to the posterior with the data and likelihood we have. The main variable here is the prior, and they are related through Bayes' Rule:

$$posterior = \frac{likelihood \times prior}{data}$$

Here the prior (probability of Olivia being able to read) is updated via the probability that she is trained given she is able to read (because the main piece of information we received is the fact that her parents are training her) divide by the probability of a child being trained in general, which we can calculate thanks to the data given by the expert.

6 Calculating probabilities

1. There are 16 people in the CS146 class.

2.

a) There are 16! ways to arrange 16 people in the class in a row with order. Assuming we are arranging the people in the increasing order of their birthday, and if I am the oldest then I will be at the end of the arrangement. Among the other 15 there are 15! way to arrange them in order. Therefore the probability of me being at the end of the arrangement (the oldest) is

$$P(I = oldest) = \frac{15!}{16!} = \frac{1}{16} = 6.25\%$$

b) Similar with a) there are 6.25% I'm the youngest.

c) If I am at position ith in the arrangement then there are $16 - i$ people older than me and there are $i - 1$ people younger than me (for example, if I'm at the 15th then there is 16-15=1 person older than me and 15-1=14 people younger than me). There are $15C(i - 1)$ ways to choose $i - 1$ people among the 15 in the group (excluding me) and there $(i - 1)!$ ways to put them in sequential order. For the remaining $16 - i$ people there are $(16 - i)!$ ways to put them in order (we don't need to choose these people since the choice of $i-1$ people define these people). Overall, if I'm at position ith in the arrangement there are:

$$N_i = C_{i-1}^{15}(i-1)!(16-i)!$$

ways to put the rest in sequential order. For example, if I'm at position 15th then there are $N_{15} = C_{14}^{15}(14)!(1)!$ ways to put others in order.

If I'm older than at least half of them then I must be at position from 9th to 16th. The probability is then:

$$P(olderthanhalf) = \frac{\sum_{i=9}^{16} C_{i-1}^{15}(i-1)!(16-i)!}{16!} = 0.5 = 50\%$$

7 Python script

Simulation in Python

```
def one(trial):
    count = 0
    for _ in range(trial):
        nums = np.random.uniform(size=16)
        if nums[0] == nums.max(0):
            count+=1
    print(float(count/trial))

def two(trial):
    count = 0
    for _ in range(trial):
        nums = np.random.uniform(size=16)
        if nums[0] == nums.min(0):
            count+=1
    print(float(count/trial))

def three(trial):
    count = 0
    for _ in range(trial):
        nums = np.random.uniform(size=16)
        nums_sorted = np.sort(nums)
        if nums[0] >= nums_sorted[8]:
            count+=1
    print(float(count/trial))

one(1000000)
two(1000000)
three(1000000)
```

which has the results:

```
    0.062659
    0.062554
    0.499827
```

confirm the answers in section 6.

8 Logarithms

1. A 10:1 odd means the probability of the event happening is $\frac{10}{11}$. The logit is:

$$
\begin{aligned}
logit_2 &= log_2 \frac{\frac{10}{11}}{1 - \frac{10}{11}} = 3.32 \\
logit_{10} &= log_{10} \frac{\frac{10}{11}}{1 - \frac{10}{11}} = 1.00 \\
logit_e &= log_e \frac{\frac{10}{11}}{1 - \frac{10}{11}} = 2.30
\end{aligned}
$$

2. A 100:1 odd means the probability of the event happening is $\frac{100}{101}$. The logit is:

$$
\begin{aligned}
logit_2 &= log_2 \frac{\frac{100}{101}}{1 - \frac{100}{101}} = 6.64 \\
logit_{10} &= log_{10} \frac{\frac{100}{101}}{1 - \frac{100}{101}} = 2.00 \\
logit_e &= log_e \frac{\frac{100}{101}}{1 - \frac{100}{101}} = 4.61
\end{aligned}
$$

3. Because the main reason of using logit is that it linearize the probability manipulations we perform: multiplying probabilities under logit now can be addition, which is more comfortable computation-wise. Therefore, as long as we are consistence with the base we are using we can use pretty much any base we want.

9 Bags and cookies

Let Jar 1 (J1) contains chocolate (C) and vanilla (V) and Jar 2 (J2) contains chocolate and caramel (Ca). According to Bayes' Rule we have:

$$
\begin{aligned}
P(J1 \mid C) &= \frac{P(C \mid J1)P(J1)}{P(C)} \\
&= \frac{P(C \mid J1)P(J1)}{P(C \mid J1)P(J1) + P(C \mid J2)P(J2)} \\
&= \frac{\frac{1}{3} \times \frac{1}{2}}{\frac{1}{3} \times \frac{1}{2} + \frac{1}{4} \times \frac{1}{2}} = \frac{4}{7}
\end{aligned}
$$

since $P(J1) = P(J2) = 1/2$ because there are equal chance of selecting either jar a priori, $P(C \mid J1) = 1/3$ and $P(C \mid J2) = 1/4$ because there are 3 and 4 cookies in jar 1 and jar 2 respectively.

Now the priori changed since you have $P(J1) = 4/7$ probability of this jar being jar 1 and $P(J2) = 1 - 4/7 = 3/7$ probability of the other bag being jar 1. We have:

$$P(C) = P(C \mid J1)P(J1) + P(C \mid J2)P(J2) = \frac{1}{3} \times \frac{3}{7} + \frac{1}{4} \times \frac{4}{7} = \frac{2}{7}$$

$$P(V) = P(V \mid J1)P(J1) + P(V \mid J2)P(J2) = \frac{2}{3} \times \frac{3}{7} + \frac{0}{4} \times \frac{4}{7} = \frac{2}{7}$$

$$P(Ca) = P(Ca \mid J1)P(J1) + P(Ca \mid J2)P(J2) = \frac{0}{3} \times \frac{3}{7} + \frac{3}{4} \times \frac{4}{7} = \frac{3}{7}$$

Sanity check: $P(C) + P(V) + P(Ca) = 2/7 + 2/7 + 3/7 = 1$, which intuitively makes sense.