Chapter 2 Solutions for Machine Learning a Probabilistic Perspective

January 2, 2017

1 README

All solutions authored by Matt Johnson < rjohnson0186@gmail.com >. I have no independent verification that these solutions are correct. Use at your own risk.

2 Exercise 2.1 (a)

Event space is $\{(B_1, B_2), (B_1, G_2), (G_1, G_2), (G_1, B_2)\}$. But I know that neighbor has at least one (could be more) boy. Therefore, event space is now $\{(B_1, B_2), (B_1, G_2), (G_1, B_2)\}$. So proability that one of the children is a girl is 2/3.

3 Exercise 2.1 (b)

Either I observed child #1 is a boy or child #2 is a boy. Suppose it was child #1 that I observed, I condition on that event to obtain: {(B_1, B_2), (B_1, G_2)}. Thus, probability child #2 is a girl is 1/2. Suppose it was child #2 that I observed, then {(B_1, B_2), (G_1, B_2)}. So, probability child #1 is a girl is 1/2. Either way, probability is 1/2.

4 Exercise 2.2 (a)

Let G the binary random variable person is guilty. Let B be the binary random variable person has the rare blood type. Presecutor asserts $P(G = 1|B = 1) = 0.99$. But, $P(G = 1|B = 1) = \frac{P(B=1|G=1)P(G=1)}{P(B=1)}$ Assuming no error in blood type testing, and also, that the guilty person does indeed, have the blood type, then $P(B = 1|G = 1) = 1.0$. Thus, $P(G = 1|B = 1) = \frac{1.0 \cdot P(G=1)}{0.01}$. If I let this equation equal 0.99, then this implies $P(G = 1) = 0.00989 = 0.989\%$, or, that about 1% of the population is guilty. Assuming the population of the town is 800,000, and that only a person from the town can be guilty, then about 8,000 would be guilty (assuming no other evidence can be used, which is probably not the case–see solution to Exercise 2.2 (b)). But surely, the prosecutor would argue that this can't be the case.

5 Exercise 2.2 (b)

In part (a), we reasoned about 8,000 could be consider guilty (restricting our population to the town). But the defender cannot discount this evidence. First, the evidence does certainly increase the posterior probability the defendant is guilty, due to matching blood. Also, surely there exist other indicators than blood, that also must be conditioned on. If these other indicators are accounted for, this may further reduce the suspected 8,000 people to a lower number.

6 Exercise 2.3

First, I prove a lemma, that $E[X+Y] = E[X] + E[Y]$. The expectation of $X+Y$ is $E[X+Y] = \sum_x \sum_y p(x,y)(x+y) = \sum_x \sum_y p(x,y)x + \sum_x \sum_y p(x,y)y = \sum_x p(x) + \sum_y p(y)y = E[X] + E[Y] = \mu_X + \mu_Y$. This is known as "linearity of expecation".

Now, the variance of X + Y is $\text{Var}[X+Y] = \sum_{x,y} p(x,y)(x+y-E[X+Y])^2 = \sum_{x,y} p(x,y)((x-\mu_X)+(y-\mu_Y))^2$. Expanding this, we get: $\sum_x \sum_y p(x,y)(x-\mu_X)^2 + \sum_x \sum_y p(x,y)(y-\mu_Y)^2 + 2\sum_x \sum_y p(x,y)(x-\mu_X)(y-\mu_Y) = \text{Var}[X] + \text{Var}[Y] + 2\text{Cov}[X,Y]$ After marginalization of Y and X for Cov[X] and Cov[Y], respectively.

7 Exercise 2.4

This problem can be represented graphically as D –> T where D and T are binary random variables. D is for "PatientHasDisease" and T is "TestIsPostive". We wish to calculate $P(D=1|T=1)$. By Bayes' rule: $P(D=1|T=1) = \frac{p(T=1|D=1)P(D=1)}{P(T=1|D=1)P(D=1)+P(T=1|D=0)P(D=0)} = \frac{0.99 \cdot 0.0001}{0.99 \cdot 0.0001 + P(T=1|D=0) \cdot 0.9999}$. Now, what is $P(T=1|D=0)$? We are also told that $P(T=0|D=0) = 0.99$. Therefore, $P(T=1|D=0) = 1.0 - p(T=0|D=0) = 0.01$. So answer is $\frac{0.99 \cdot 0.0001}{0.99 \cdot 0.0001 + 0.01 \cdot 0.9999} \approx 0.0098 = 0.98\% \approx 1\%$. Compare this to prior of 0.01%. Thus, patient is $0.98/0.001 = 980$ times more likely since test result is known. But…, posterior probability is still only 1%

8 Exercise 2.5

Let Z, C, and H all be random variables with three states each, that represent the three doors:

- Z models the probability of where the prize actually resides
- C models the door chosen by the contestant
- H models the door opened by the host

We desire to find $P(Z|C,H)$. We solve this problem graphically, as C –> H <– Z. This representation allows us to easily specify the priors and CPDs required. The joint distribution is $P(Z,C,H) = P(C)P(Z)P(H|Z,C)$. Now, $P(Z|C,H) = \frac{P(Z,C,H)}{P(C,H)} = \frac{P(C)P(Z)P(H|Z,C)}{\sum_Z P(C)P(Z)P(H|Z,C)} = \frac{P(Z)P(H|Z,C)}{\sum_Z P(Z)P(H|Z,C)}$ And assuming a uniform distribution for $P(Z)$, this drops out as well. Thus, $P(Z|C,H) = \frac{P(H|Z,C)}{\sum_Z P(H|Z,C)}$. Let's suppose contestant picks door 1, now, either host will open door 2 or door 3. Let's work out math for scenario that host opens door 2. $P(Z=1|C=1,H=2) = \frac{P(H=2|Z=1,C=1)}{P(H=2|Z=1,C=1)+P(H=2|Z=2,C=1)+P(H=2|Z=3,C=1)} = \frac{0.5}{0.5+0.0+1.0} = \frac{1}{3}$. OK, now $P(Z=2|C=1,H=2) = \frac{P(H=2|Z=2,C=1)}{P(H=2|Z=1,C=1)+P(H=2|Z=2,C=1)+P(H=2|Z=3,C=1)} = \frac{0.0}{0.5+0.0+1.0} = 0.0$ A bunch of complicated math for an intuitive result, the prize can't be behind door 2! Because that's the door the host opened. But this means that $P(Z = 3|C = 1, H = 2) = \frac{2}{3}$. Therefore, we should switch to the door we did not pick.

9 Exercise 2.6 (a)

$P(H|E_1, E_2) = \frac{P(E_1,E_2|H)P(H)}{P(E_1,E_2)} = \frac{P(E_1,E_2|H)P(H)}{\sum_H P(E_1,E_2|H)P(H)}$

(ii) is oversufficient (because of $P(e1, e2)$), but still sufficient.

10 Exercise 2.6 (b)

Now, with conditional independence assumption, we get $\frac{P(E_1|H)P(E_1|H)P(H)}{P(E_1,E_2)} = \frac{P(E_1,E_2|H)P(H)}{\sum_H P(E_1|H)P(E_1|H)P(H)}$

(i) is oversufficient (due to $P(e_1, e_2)$ term), but still sufficient. (iii) is sufficient.

11 Exercise 2.7

$P(X_1, X_2, X_3) = P(X_2, X_1)P(X_3|X_1, X_2) = P(X_1)P(X_2)P(X_3|X_1, X_2) = P(X_1)P(X_2)P(X_3|X_1)$, but we can't write $P(X_3|X_1)$ as $P(X_3)P(X_1)$ because we can only assume pairwise independence assumption. Thus, pairwise does not imply mutual independence.

12 Exercise 2.8

Suppose $p(x, y|z) = g(x, z)h(y, z)$. Integrating both sides with respect to y, we get $\int_y p(x, y|z)dy = g(x, z) \int_y h(y, z)dy = p(x|z)$. Now, with x: $\int_x p(x, y|z)dx = h(y, z) \int_x g(x, z)dx = p(y|z)$. Multiplying both equations: $p(y|z)p(x|z) = g(x, z)h(y, z) \int_y h(y, z)dy \int_x g(x, z)dx = g(x, z)h(y, z) \int_x \int_y h(y, z)g(x, z)dxdy = g(x, z)h(y, z) \int_x \int_y p(x, y|z)dxdy = g(x, z)h(y, z) = p(x, y|z) = p(y|z)p(x|z)$

13 Exercise 2.9 (a)

TODO

14 Exercise 2.9 (b)

TODO

15 Exercise 2.10

Suppose I have a p.d.f like $f(x)$, then, the probabilty of observing some event is roughly $f(x)dx$. But now, suppose I transform the random variable x into another random variable, say, $u(x)$. Well, the probability, to remain invariant, must be $g(u(x))|\frac{dx}{du}|du$. Because after I transform, I must adjust the volume of space I am investigating. This ratio gives me the factor it changed by. For the problem at hand, let $u(x) = \frac{1}{x}$. Thus, $g(u(x))|\frac{dx}{du}| = \frac{b^a}{\Gamma(a)}\frac{1}{x}^{a-1}e^{\frac{-b}{x}} \cdot \frac{1}{x^2} = \frac{b^a}{\Gamma(a)}x^{-(a+1)}e^{\frac{-b}{x}}$

16 Exercise 2.11

TODO

17 Exercise 2.12

$I(X,Y) = \sum_x \sum_y p(x,y)\log\frac{p(x,y)}{p(x)p(y)} = \sum_x \sum_y p(x,y)\log p(x,y) - \sum_x \sum_y p(x,y)\log p(x) - \sum_x \sum_y p(x,y)\log p(y) = H(X) + H(Y) - H(X,Y)$. But, $H(X,Y) = H(X) + H(Y|X) = H(Y) + H(X|Y)$. Either choice yields the proof. Of course, I'd also have to show that the relationship of the joint and conditional entropy holds.

18 Exercise 2.13

TODO

19 Exercise 2.14

TODO

20 Exercise 2.15

TODO

21 Exercise 2.16

TODO