Probabilistic Artificial Intelligence

Problem Set 1 Solutions

September 28, 2018

## 1. Conditional Independence

Consider the following joint distribution for three random variables, $a, b, c \in \{0, 1\}$.

| $a$ | $b$ | $c$ | $p(a,b,c)$ |
|---|---|---|---|
| 0 | 0 | 0 | 0.192 |
| 0 | 0 | 1 | 0.144 |
| 0 | 1 | 0 | 0.048 |
| 0 | 1 | 1 | 0.216 |
| 1 | 0 | 0 | 0.192 |
| 1 | 0 | 1 | 0.064 |
| 1 | 1 | 0 | 0.048 |
| 1 | 1 | 1 | 0.096 |

Show that $a$ and $b$ are dependent, namely $p(a, b) \neq p(a)p(b)$. But, they are marginally independent given $c$, namely $p(a, b \mid c) = p(a \mid c)p(b \mid c)$. (c.f. Bishop *Pattern Recognition and Machine Learning*, Exercise 8.3)

Solution. For $p(a, b) \neq p(a)p(b)$ it suffices to simply produce example values for which the two expressions differ. Let $a = 0, b = 0$. Then by marginalizing on $c$,

$$p(a = 0, b = 0) = 0.192 + 0.144 = 0.336$$

Where as by marginalizing on $b, c$,

$$p(a = 0) = 0.192 + 0.144 + 0.048 + 0.216 = 0.6$$

And by marginalizing on $a, c$,

$$p(b = 0) = 0.192 + 0.144 + 0.192 + 0.064 = 0.592$$

Thus we can show $p(a = 0) \cdot p(b = 0) = 0.6 \cdot 0.592 = 0.3552 \neq 0.336$.

To show $p(a, b|c) = p(a|c)p(b|c)$, we need to show that it holds for every value of $a, b, c$. Note that $p(c = 0) = 0.192 + 0.048 + 0.192 + 0.048 = 0.48$.

For the case when $a = 0, b = 0, c = 0$,

$$p(a = 0, b = 0|c = 0) = 0.192/0.48 = 0.4$$

| $a$ | $b$ | $p(a,b,c)$ |
|---|---|---|
| 0 | 0 | 0.4 |
| 0 | 1 | 0.1 |
| 1 | 0 | 0.4 |
| 1 | 1 | 0.1 |

Table 1: $p(a,b|c=0)$

| $a$ | $b$ | $p(a,b,c)$ |
|---|---|---|
| 0 | 0 | 0.27692308 |
| 0 | 1 | 0.41538462 |
| 1 | 0 | 0.12307692 |
| 1 | 1 | 0.18461538 |

Table 2: $p(a,b|c=1)$

and

$$p(a=0|c=0)p(b=0|c=0) = (0.192+0.048)/0.48 \cdot (0.192+0.192)/0.48 = 0.4$$

Which are equal.

More generally, we can take the entire table and divide it by $p(c=0)$ and $p(c=1)$ to get the conditional distributions $p(a,b|c=0)$ (Table 1) and $p(a,b|c=1)$ (Table 2) respectively. We can then use these tables to look up the relevant values.

## 2. Bayes Rule

A routine breast cancer mammography screening is performed on a group of people of age fourty. 1% of the participants in the screening actually have breast cancer. 80% of the people in the screening with breast cancer received positive results (has breast cancer) on the mammmography test. 9.6% of people without breast cancer received a positive result on their mammographies. Suppose a person of this age receives a positive result on their mammography. Given the information in this screening, what is the probability that he has breast cancer?

Solution. Our goal is to estimate, $P(C=1|T=1)$ where $C$ means cancer and $T$ means mammography (test). Using Bayes Rule,

$$P(C=1|T=1) = \frac{P(T=1|C=1)P(C=1)}{P(T=1)}$$

where $P(T=1) = P(T=1|C=0)P(C=0) + P(T=1|C=1)P(C=1)$.

Thus using the values provided in the question,

$$P(C=1|T=1) = \frac{0.8 \cdot .01}{.096 \cdot .99 + .8 \cdot .01} \approx .077$$

## 3. Chain rule

Derive the chain rule from the basic rules of probability. (Hint: by the definition of conditional probability $P(A, B) = P(A \mid B)P(B)$). How many factorizations are possible for a distribution on $n$ random variables?

Solution. The chain rule states that

$$P(A_1, \ldots, A_n) = P(A_1)P(A_2|A_1)P(A_3|A_2, A_1)\ldots P(A_n|A_{n-1}, \ldots, A_1)$$

By induction. In the case of $n = 2$ we have simply the definition of conditional probability. Assume true for the $n$-th case. Consider the $n+1$-th case. Again by the definition of conditional probability,

$$P(A_1, \ldots, A_n, A_{n+1}) = P(A_{n+1}|A_1, \ldots, A_n)P(A_1, \ldots, A_n)$$

expand the term $P(A_1, \ldots, A_n)$ using the inductive assumption.

To see that there are $n!$ ways to factor a distribution on $n$ random variables, consider an iterative process where at each step you choose to "factor out" a specific variable. For example, at step 1 you factor out $A_1$ resulting in $P(A_1)P(A_2, \ldots, A_n|A_1)$. Now you have $n-1$ choices left for which variable to factor out. For example, suppose you choose number 42 resulting in $P(A_1)P(A_{42}|A_2, \ldots, A_{41}, A_{43}, \ldots, A_n)P(A_{42}|A_2, \ldots, A_{41}, A_{43}, \ldots, A_n)$. Now you have $n-2$ remaining choices for the next factorization.