Probabilistic Foundations of Artificial Intelligence
Solutions to Problem Set 1

Oct 07, 2016

1. Conditional Probabilities

For each statement below, either prove it is true, or give a counterexample showing it is false. In the following, we assume that all events have non-zero probability.

(a) If $P(a|b,c) = P(b|a,c)$, then $P(a|c) = P(b|c)$

(b) If $P(a|b,c) = P(a)$, then $P(b|c) = P(b)$

(c) If $P(a|b) = P(a)$, then $P(a|b,c) = P(a|c)$

Solution

(a). True.

From Bayes' rule, we get

$$P(a,b,c) = P(a|b,c)P(b|c)P(c) \tag{1}$$

and

$$P(a,b,c) = P(b|a,c)P(a|c)P(c) \tag{2}$$

From the question we have $P(a|b,c) = P(b|a,c)$, and therefore we can rewrite (1) as $P(a,b,c) = P(b|a,c)P(b|c)P(c)$. Combining with (2) we get $P(a|c) = P(b|c)$.

(b). False.

The statement is equivalent to: $a \perp (b,c) \Rightarrow b \perp c$, which is false. See Figure 1 for a counterexample (description below).

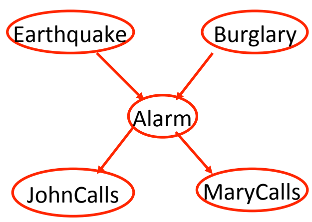


Figure 1: Example from lecture slides: casual parametrization

Counterexample. If $a = JohnCalls$, $b = Burglary$, $c = Earthquake$, then

$$a \perp (b, c) \mid Alarm.$$

However, the event $Burglary$ is dependent with $Earthquake$ if $Alarm$ is observed:

$$b \not\perp c \mid Alarm$$

Therefore, we have identified an example where the statement is false.

(c). False.

Counterexample. Suppose $a \perp b$, each taking 0 and 1 with probability 0.5, and $c = ab$. $P(a = 0|b = 0) = 1/2$. But, when $c = 0$, $P(a = 0) = 2/3$ and $P(a = 0) > P(a = 0|b = 0)$. Therefore the statement $a \perp b \mid c$ is false.

2. Finding the fake coin

Suppose you are given a bag containing n unbiased coins. You are also told that $n - 1$ of these coins are normal, that is, they have a head on one side and a tail in the other. The remaining one is fake and has heads on both sides.

(a) Suppose you pick a coin from the bag uniformly at random, you flip it, and get a head. Given this result, what is the probability that the coin you picked is the fake one? (Note that we ask for a conditional probability.)

(b) Suppose you continue flipping the same coin for a total of k times and you get k heads. Given this result, what is the probability that you picked the fake coin?

(c) Now, suppose you devise the following method to determine if the coin is fake or not. You flip it k times, after which you conclude that it is the fake one if all k flips have resulted in heads, else you conclude that it is normal. What is the probability that using this method you arrive at a wrong conclusion? (Note that this time we ask for an unconditional probability.)

Solution

(a) Define a random variable C corresponding to the coin that takes values N (normal) or F (fake) and a random variable E corresponding to the outcome of the flip that takes values H (heads) or T (tails). From the problem description, we can write down the following probabilities:

$$\begin{aligned}
&P(C = N) = \frac{n-1}{n} \\
&P(C = F) = \frac{1}{n} \\
&P(E = H \mid C = N) = 0.5 \\
&P(E = H \mid C = F) = 1
\end{aligned}$$

To compute the probability $P(C = F \mid E = H)$, we can use Bayes' rule:

$$
\begin{aligned}
P(C = F \mid E = H) &= \frac{P(E = H \mid C = F)P(C = F)}{P(E = H)} \\
&= \frac{P(E = H \mid C = F)P(C = F)}{P(E = H \mid C = N)P(C = N) + P(E = H \mid C = F)P(C = F)} \\
&= \frac{1 * \frac{1}{n}}{0.5 * \frac{n-1}{n} + 1 * \frac{1}{n}} \\
&= \frac{2}{n+1}
\end{aligned}
$$

(b) Since the outcomes of coin flips are conditionally independent given the coin, if we denote by E_k the random variable that corresponds to the outcome of k flips with value H^k when k heads occur, then we have that

$$
\begin{aligned}
P(E_k = H^k \mid C = N) &= 2^{-k} \\
P(E_k = H^k \mid C = F) &= 1.
\end{aligned}
$$

Using Bayes' rule similarly to the previous question we have

$$
\begin{aligned}
P(C = F \mid E_k = H^k) &= \frac{P(E_k = H^k \mid C = F)P(C = F)}{P(E_k = H^k)} \\
&= \frac{P(E_k = H^k \mid C = F)P(C = F)}{P(E_k = H^k \mid C = N)P(C = N) + P(E_k = H^k \mid C = F)P(C = F)} \\
&= \frac{1 * \frac{1}{n}}{2^{-k} * \frac{n-1}{n} + 1 * \frac{1}{n}} \\
&= \frac{1}{1 + (n-1)2^{-k}}
\end{aligned}
$$

Figure 2 shows how the probability of having picked the fake coin among $n = 50$ total coins increases as the number of observed heads k gets larger.

(c) A wrong conclusion is reached when we pick a normal coin and claim it is fake (i.e., we observe k heads) or we pick the fake coin and claim it is normal (i.e., observe tails at least once). Note that the latter event is not possible (happens with probability 0) by definition of the fake coin. Therefore, the probability of a wrong conclusion is

$$
\begin{aligned}
P(C = N, E_k = H^k) &= P(E_k = H^k \mid C = N)P(C = N) \\
&= \frac{n-1}{n} 2^{-k},
\end{aligned}
$$

which is depicted in Figure 3 for $n = 50$.

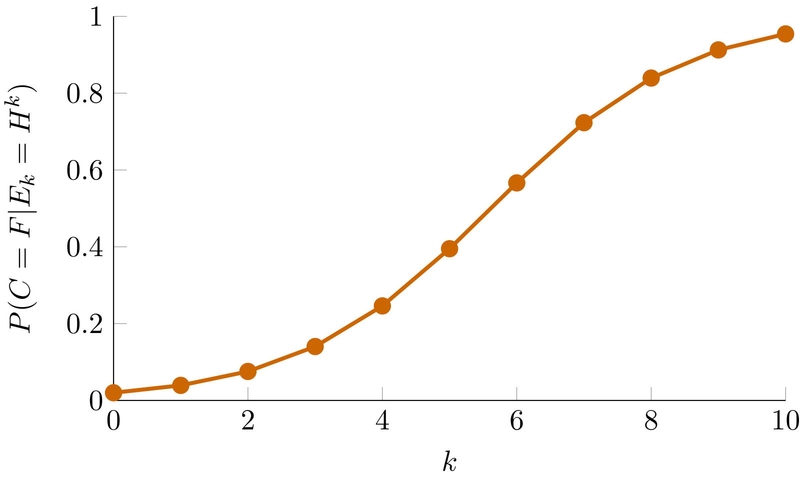


Figure 2: Probability of having picked the fake coin given k observed heads ($n = 50$).

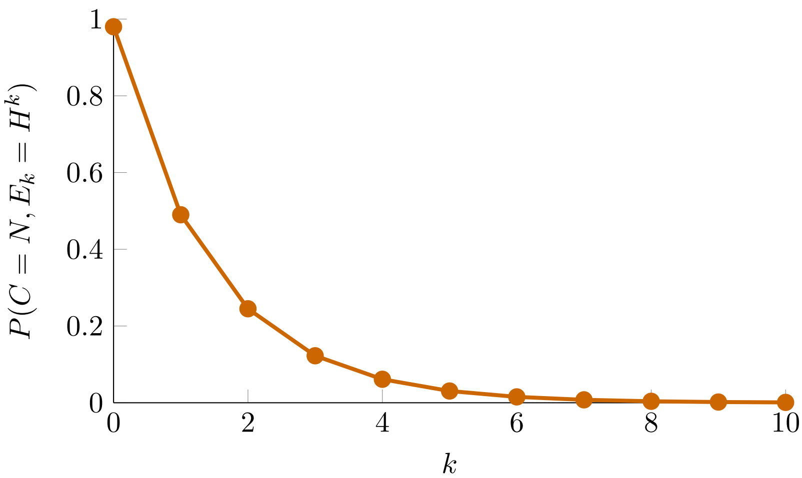


Figure 3: Probability of reaching a wrong conclusion using k flips ($n = 50$).

3. Naive Bayes

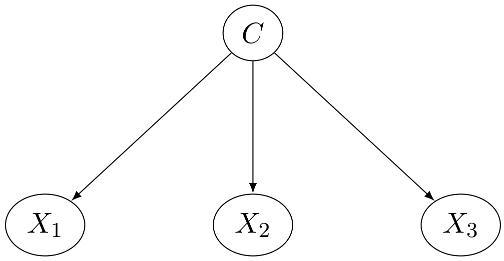
Suppose you have a bag of three biased coins a, b, and c, with probabilities of coming up heads of 0.2, 0.6, and 0.8 respectively. You draw a coin uniformly at random from the bag and flip it three times to generate the sequence of outcomes X_1, X_2, X_3.

(a) Draw the Bayesian Network corresponding to this setup and specify the necessary Conditional Probability Tables (CPTs).

(b) Calculate which coin was most likely to have been drawn from the bag, if two of the observed outcomes were heads and the other was a tail.

Solution

(a) Figure 4 shows the Bayesian network of the coin flipping setup, where C is a random variable that correponds to the selection of the coin. Note that X_1, X_2, X_3 have identical conditional probability tables, since they correspond to independently flipping the same coin three times.



(a) Network structure.

$C = a$	$C = b$	$C = c$
1/3	1/3	1/3

(b) $P(C)$

	$X_i = H$	$X_i = T$
$C = a$	0.2	0.8
$C = b$	0.6	0.4
$C = c$	0.8	0.2

(c) $P(X_i \mid C),\ i = 1, 2, 3$

Figure 4: Structure and conditional probability tables for the coin flipping Bayesian network.

(b) Let A denote the event where two of the coin flips come out as heads and one as tails. First, note that, since the coin flips are conditionally independent and identically distributed given C, all events that contain two heads and one tail have the same conditional probability, that is,

$$\begin{aligned} & P(X_1 = H, X_2 = H, X_3 = T \mid C) \\ &= P(X_1 = H, X_2 = T, X_3 = H \mid C) \\ &= P(X_1 = T, X_2 = H, X_3 = H \mid C) \\ &= [P(X_i = H \mid C)]^2\, P(X_i = T \mid C), \end{aligned}$$

and, consequently,

$$P(A \mid C) = 3\, [P(X_i = H \mid C)]^2 P(X_i = T \mid C). \tag{3}$$

Now, we can use Bayes' rule to compute the probability of the selected coin being z (one of a, b, or c), given the observed event A:

$$P(C = z \mid A) = \frac{P(A \mid C = z) P(C = z)}{P(A)}$$

Since $P(C=a)=P(C=b)=P(C=c)$, it follows from the previous equation that

$$\begin{aligned}
\operatorname*{argmax}_{z\in\{a,b,c\}} P(C=z \mid A) &= \operatorname*{argmax}_{z\in\{a,b,c\}} P(A \mid C=z) \\
&\overset{(3)}{=} \operatorname*{argmax}_{z\in\{a,b,c\}} \left\{[P(X_i=H \mid C=z)]^2 P(X_i=T \mid C=z)\right\} \\
&= \operatorname{argmax}\left\{\underbrace{0.2^2 * 0.8}_{z=a},\ \underbrace{0.6^2 * 0.4}_{z=b},\ \underbrace{0.8^2 * 0.2}_{z=c}\right\} \\
&= \operatorname{argmax}\left\{\underbrace{0.032}_{z=a},\ \underbrace{0.144}_{z=b},\ \underbrace{0.128}_{z=c}\right\} \\
&= b.
\end{aligned}$$

Thus, we conclude that, given the observations of two heads and one tail, the most likely coin to have been flipped is coin b.

Note that, to obtain the actual posterior distribution of C, we would have to normalize the above values inside the argmax:

$$\begin{aligned}
P(C=a \mid A) &= \frac{0.032}{0.032+0.144+0.128} = 0.105 \\
P(C=b \mid A) &= \frac{0.144}{0.032+0.144+0.128} = 0.474 \\
P(C=c \mid A) &= \frac{0.128}{0.032+0.144+0.128} = 0.421.
\end{aligned}$$

However, this computation is not required if we are only interested in finding the most likely coin.