Probabilistic Artificial Intelligence

Solutions to Problem Set 1

Sep 23, 2019

1. Bayes rule

As a result of a medical screening, one of the tests revealed a serious disease in a person. The test has a high accuracy of 99% (the probability of a positive response in the presence of a disease is 99% and the probability of a negative response in the absence of a disease is also 99%). However, the disease is quite rare and occurs only in one person per 10,000. Calculate the probability of the examined person having the identified disease.

Solution

Let us start by defining some events that we will reason about later. For ease of writing, let us call the person in question X.

$$\begin{aligned} D &= \text{X has the disease,} \\ P &= \text{The test shows a positive response.} \end{aligned}$$

Now we can translate the information in the question in terms of D and P:

$$\begin{aligned} &\Pr\{D\} = 10^{-4} && \text{the disease is rare} \\ &\Pr\{P|D\} = \Pr\{P^c|D^c\} = 0.99 && \text{the test is accurate} \end{aligned}$$

What the question wants is $\Pr\{D|P\}$. Let us compute this probability using Bayes theorem:

$$\Pr\{D|P\} = \Pr\{P|D\} \cdot \frac{\Pr\{D\}}{\Pr\{P\}}$$

From the quantities above, we have everything except $\Pr\{P\}$. No worries, because we can compute it using law of total probability:

$$\begin{aligned} \Pr\{P\} &= \Pr\{P|D\}\Pr\{D\} + \Pr\{P|D^c\}\Pr\{D^c\} \\ &= 0.99 \times 10^{-4} + 0.01 \times (1 - 10^{-4}) \\ &= 0.010098. \end{aligned}$$

Hence, $\Pr\{D|P\} = 0.99 \times 10^{-4}/0.010098 \approx 0.0098 = 0.98\%$.

2. Conditional Probabilities

For each statement below, either prove that it is true, or give a counterexample showing that it is false. Let a, b, c be events in some probability space.

(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 (*provided that* $P(c) > 0$)

Proof. From chain rule one gets

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

and

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

From the question we have $P(a|b,c) = P(b|a,c)$, 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).

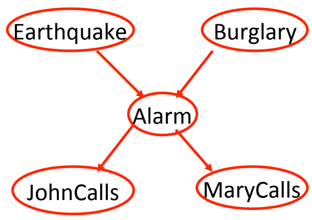

Figure 1: Example from lecture slides: casual parametrization

Counterexample. If $a = \text{JohnCalls}$, $b = \text{Burglary}$, $c = \text{Earthquake}$, then

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

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

$$b \not\perp c \mid \text{Alarm}$$

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

(c). False.

The statement is equivalent to the statement: independence implies conditional independence.

Counterexample. Suppose two independent random variables each taking 0 and 1 with probability 0.5. For example, a and b represent results of tossing a coin twice in a row (a happens if the result of the first experiment is head; b happens if the result of the second experiment is head). Let's say $c = a \cap b$. We want to show, that although the sequential results of tossing a coin are independent, i.e., $P(a|b) = P(a) = 0.5$, they are not conditionally independent: $P(\neg a|\neg c) = 2/3$, but $P(\neg a|\neg b, \neg c) = 1/2$. Therefore the statement $a \perp b \mid c$ is false.

3. Random Walk on Graphs

Let G be a simple connected finite graph. We start at a vertex of G. At every second, we move to one of the neighbors of the current vertex uniformly at random, e.g., if the vertex has 3 neighbors, we move to one of them, each with probability $1/3$. Let u and v be two vertices of G. Starting from u, what is the probability that the walk visits v eventually? Does the same result hold for infinite graphs?

Solution

For brevity, let us define some notation. For vertices x, v of G, define $E_{x,v}$ be the event "starting from x, the random walk visits v eventually"[1] and let $p_{x,v}$ be its probability. Also denote by $N(x)$ the set of neighbors of x in G and $\deg(x)$ be the size of $N(x)$. By $x \to y$ we mean the event "the walk is at x and goes to y in the next step".

One can write

$$\begin{aligned}
p_{x,v} = \Pr\{E_{x,v}\} &= \sum_{y \in N(x)} \Pr\{E_{x,v} \cap (x \to y)\} \\
&= \sum_{y \in N(x)} \Pr\{x \to y\} \cdot \Pr\{E_{x,v} \mid (x \to y)\} \\
&= \sum_{y \in N(x)} \Pr\{x \to y\} \cdot \Pr\{E_{y,v}\} \\
&= \frac{1}{\deg(x)} \sum_{y \in N(x)} p_{y,v},
\end{aligned}$$

where in the first line, we used the law of total probability, in the second we used the definition of conditional probability. Third line follows from the definition of $E_{y,v}$, and in the end we used the fact that we move to each neighbor uniformly at random. In short, we have the following recursion

$$p_{x,v} = \frac{1}{\deg(x)} \sum_{y \in N(x)} p_{y,v} \tag{3}$$

with *boundary condition* $p_{v,v} = 1$. Solving this recursion gives our desired result. Take the minimum probability among $p_{x,v}$ for all vertices x and let the minimizer be $p_{z,v}$. Expanding (3) gives

$$p_{z,v} = \frac{1}{\deg(z)} \sum_{y \in N(z)} p_{y,v} \geq \frac{1}{\deg(z)} \sum_{y \in N(z)} p_{z,v} = p_{z,v}.$$

Hence, for all the neighbors y of z it holds that $p_{y,v} = p_{z,v}$. By applying the same argument to the neighbors of z and neighbors of neighbors of z and so on, we get that (since the graph is finite and connected) $p_{x,v} = p_{z,v}$ for all x. But we had $p_{v,v} = 1$, hence $p_{x,v} = 1$ for all x. That is, *starting from any vertex, the random walk will visit v with probability 1.*

[1] Rigor test for interested reader: Why this is an event? What is the correct probability space?

Remark and further notes If the graph is disconnected, the result obviously does not hold: the walk has no chance to go from one connected component to another one. For infinite connected graphs, the result does not always hold. The following are some examples:[2]

- Let G be the graph with vertices $\mathbb{Z}$ and assume that each integer i is connected to $i \pm 1$. Random walk on this graph visits every integer with probability 1. The same result holds if $G = \mathbb{Z}^2$ and every vertex (i, j) is connected to $(i \pm 1, j \pm 1)$.
- If $G = \mathbb{Z}^3$ the result does *not* hold anymore. Indeed, the probability of coming back to the origin $(0, 0, 0)$ is approximately $\frac{1}{3}$. This saying from Kakutani is famous regarding this fact:

  "A drunk man will find his way home, but a drunk bird may get lost forever…"

4. Multivariate Gaussian Distribution

A vector-valued random variable $x \in \mathbb{R}^n$ is said to have a multivariate normal distribution with mean $\mu \in \mathbb{R}^n$ and covariance matrix $\Sigma \in \mathbf{S}^n_{++}$ if its pdf is:

$$p(x; \mu, \Sigma) = \frac{1}{(2\pi)^{n/2} \det(\Sigma)^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right)$$

In this exercise you will need to show that:

1. Marginal of a joint Gaussian is Gaussian,
2. Conditional of a joint Gaussian is Gaussian.

Specifically, consider $x = \begin{bmatrix} x_A \\ x_B \end{bmatrix}$, $\mu = \begin{bmatrix} \mu_A \\ \mu_B \end{bmatrix}$, and $\Sigma = \begin{bmatrix} \Sigma_{AA} & \Sigma_{AB} \\ \Sigma_{BA} & \Sigma_{BB} \end{bmatrix}$, the joint distribution:

$$p(x) = p(x_A, x_B) = \frac{1}{Z} \exp\left(-\frac{1}{2} \begin{bmatrix} x_A - \mu_A \\ x_B - \mu_B \end{bmatrix}^T \begin{bmatrix} \Sigma_{AA} & \Sigma_{AB} \\ \Sigma_{BA} & \Sigma_{BB} \end{bmatrix}^{-1} \begin{bmatrix} x_A - \mu_A \\ x_B - \mu_B \end{bmatrix}\right).$$

The following notations can ease the computation:

$$V = \begin{bmatrix} V_{AA} & V_{AB} \\ V_{BA} & V_{BB} \end{bmatrix} = \begin{bmatrix} \Sigma_{AA} & \Sigma_{AB} \\ \Sigma_{BA} & \Sigma_{BB} \end{bmatrix}^{-1}, \quad \begin{bmatrix} x_A - \mu_A \\ x_B - \mu_B \end{bmatrix} = \begin{bmatrix} \Delta_A \\ \Delta_B \end{bmatrix}$$

[2] For interested reader…

1. Compute the marginal of joint Gaussians::

$$p(x_A) = \frac{1}{Z}\int_{x_B} p(x_A, x_B) dx_B.$$

2. Compute the conditional of joint Gaussians:

$$p(x_B|x_A) = \frac{p(x_A, x_B)}{p(x_A)}.$$

Solution

1.

$$\begin{aligned} p(x_A) =& \frac{1}{Z}\int_{x_B} \exp\left(-\frac{1}{2}\begin{bmatrix}\Delta_A \\ \Delta_B\end{bmatrix}^T \begin{bmatrix} V_{AA} & V_{AB} \\ V_{BA} & V_{BB}\end{bmatrix} \begin{bmatrix}\Delta_A \\ \Delta_B\end{bmatrix}\right) dx_B \qquad = \\ =& \frac{1}{Z}\exp\left(-\frac{1}{2}\left[\Delta_A^T(V_{AA} - V_{AB}V_{BB}^{-1}V_{BA})\Delta_A\right]\right)\cdot \\ &\cdot \int_{x_B} \exp\left(-\frac{1}{2}\left[(\Delta_B + V_{BB}^{-1}V_{BA}\Delta_A)^T V_{BB}(\Delta_B + V_{BB}^{-1}V_{BA}\Delta_A)\right]\right) dx_B \end{aligned}$$

$$\begin{aligned} p(x_A) =& \frac{1}{Z_A}\exp\left(-\frac{1}{2}\left[\Delta_A^T(V_{AA} - V_{AB}V_{BB}^{-1}V_{BA})\Delta_A\right]\right) = \\ =& \frac{1}{Z_A}\exp\left(-\frac{1}{2}\left[\Delta_A^T\Sigma_{AA}^{-1}\Delta_A\right]\right) \end{aligned}$$

Note: $\frac{1}{2}z^T Az + b^T z + c = \frac{1}{2}(z + A^{-1}b)^T A(z + A^{-1}b) + c - b^T A^{-1}b$

2.

$$\begin{aligned} p(x_B|x_A) =& \frac{p(x_A, x_B)}{p(x_A)} = \\ =& \frac{1}{Z'}\exp\left(-\frac{1}{2}\begin{bmatrix}x_A - \mu_A \\ x_B - \mu_B\end{bmatrix}^T \begin{bmatrix}\Sigma_{AA} & \Sigma_{AB} \\ \Sigma_{BA} & \Sigma_{BB}\end{bmatrix}^{-1} \begin{bmatrix}x_A - \mu_A \\ x_B - \mu_B\end{bmatrix}\right) = \\ =& \frac{1}{Z'}\exp\left(-\frac{1}{2}\left[\Delta_A^T(V_{AA} - V_{AB}V_{BB}^{-1}V_{BA})\Delta_A\right]\right)\cdot \\ &\cdot \exp\left(-\frac{1}{2}\left[(\Delta_B + V_{BB}^{-1}V_{BA}\Delta_A)^T V_{BB}(\Delta_B + V_{BB}^{-1}V_{BA}\Delta_A)\right]\right) = \\ =& \frac{1}{Z''}\exp\left(-\frac{1}{2}\left[(\Delta_B + V_{BB}^{-1}V_{BA}\Delta_A)^T V_{BB}(\Delta_B + V_{BB}^{-1}V_{BA}\Delta_A)\right]\right) \end{aligned}$$

$$x_B|x_A \sim \mathcal{N}(\underbrace{\mu_B - V_{BB}^{-1}V_{BA}(x_A - \mu_A)}_{=\mu_{B|A}}; \underbrace{\Sigma_{BB} - \Sigma_{BA}\Sigma_{AA}^{-1}\Sigma_{AB}}_{=\Sigma_{B|A}=V_{BB}^{-1}})$$

Extra

Based on the obtained results, you can extend the jupyter notebook (.ipynb file) from the first Tutorial to implement and visualise the marginal and the conditional distributions in 2D case.