Module 10

Reasoning with Uncertainty - Probabilistic reasoning

Lesson 27

Probabilistic Inference

10.4 Probabilistic Inference Rules

Two rules in probability theory are important for inferencing, namely, the product rule and the Bayes' rule.

Product rule:

$$\begin{aligned} P(A,B|C) &= P(A|B,C)P(B|C) \\ &= P(B|A,C)P(A|C) \end{aligned}$$

Bayes' rule:

$$P(A|B,C) = \frac{P(B|A,C)P(A|C)}{P(B|C)}$$

Used in Bayesian statistics :

$$P(Model|Data) = \frac{P(Model)P(Data|Model)}{P(Data)}$$

Here is a simple example, of application of Bayes' rule.

Suppose you have been tested positive for a disease; what is the probability that you actually have the disease?

It depends on the accuracy and sensitivity of the test, and on the background (*prior*) probability of the disease.

Let P(Test=+ve | Disease=true) = 0.95, so the false negative rate, P(Test=-ve | Disease=true), is 5%.

Let P(Test=+ve | Disease=false) = 0.05, so the false positive rate is also 5%. Suppose the disease is rare: P(Disease=true) = 0.01 (1%).

Let D denote Disease and "T=+ve" denote the positive Tes.

Then,

$$P(D=true|T=+ve) = \frac{P(T=+ve|D=true) * P(D=true)}{P(T=+ve|D=true) * P(D=true) + P(T=+ve|D=false) * P(D=false)}$$

$$= \frac{0.95 * 0.01}{0.95*0.01 + 0.05*0.99} = 0.161$$

So the probability of having the disease given that you tested positive is just 16%. This seems too low, but here is an intuitive argument to support it. Of 100 people, we expect only 1 to have the disease, but we expect about 5% of those (5 people) to test positive. So of the 6 people who test positive, we only expect 1 of them to actually have the disease; and indeed 1/6 is approximately 0.16.

In other words, the reason the number is so small is that you believed that this is a rare disease; the test has made it 16 times more likely you have the disease, but it is still unlikely in absolute terms. If you want to be "objective", you can set the prior to uniform (i.e. effectively ignore the prior), and then get

$$P(D=true|T=+ve) = \frac{P(T=+ve|D=true) * P(D=true)}{P(T=+ve)}$$

$$= \frac{0.95 * 0.5}{0.95*0.5 + 0.05*0.5} = \frac{0.475}{0.5} = 0.95$$

This, of course, is just the true positive rate of the test. However, this conclusion relies on your belief that, if you did not conduct the test, half the people in the world have the disease, which does not seem reasonable.

A better approach is to use a plausible prior (eg P(D=true)=0.01), but then conduct multiple independent tests; if they all show up positive, then the posterior will increase. For example, if we conduct two (conditionally independent) tests T1, T2 with the same reliability, and they are both positive, we get

$$P(D=true|T1=+ve,T2=+ve) = \frac{P(T1=+ve|D=true) * P(T2=+ve|D=true) * P(D=true)}{P(T1=+ve, T2=+ve)}$$

$$= \frac{0.95 * 0.95 * 0.01}{0.95*0.95*0.01 + 0.05*0.05*0.99} = \frac{0.009}{0.0115} = 0.7826$$

The assumption that the pieces of evidence are conditionally independent is called the **naive Bayes** assumption. This model has been successfully used for mainly application including classifying email as spam (D=true) or not (D=false) given the presence of various key words (Ti=+ve if word i is in the text, else Ti=-ve). It is clear that the words are not independent, even conditioned on spam/not-spam, but the model works surprisingly well nonetheless.

In many problems, complete independence of variables do not exist. Though many of them are conditionally independent.

X and Y are conditionally independent given Z iff

$$P(X, Y|Z) = P(X|Z)P(Y|Z)$$

In full: X and Y are conditionally independent given Z iff for any instantiation x, y, z of X, Y,Z we have

$$\begin{aligned} & P(X = x, Y = y|Z = z) \\ = \; & P(X = x|Z = z)P(Y = y|Z = z) \end{aligned}$$

An example of conditional independence:

Consider the following three Boolean random variables:

LeaveBy8, GetTrain, OnTime

Suppose we can assume that:

P(*OnTime* | *GetTrain*, *LeaveBy8*) = P(*OnTime* | *GetTrain*)

but NOT   P(*OnTime* | *LeaveBy8*) = P(*OnTime*)

Then, *OnTime* is dependent on *LeaveBy8*, but *independent* of *LeaveBy8 given GetTrain*.

We can represent P(*OnTime* | *GetTrain*, *LeaveBy8*) = P(*OnTime* | *GetTrain*)

graphically by: *LeaveBy8* -> *GetTrain* -> *OnTime*