Chapter 3.8 Learning Using Bayes Rule

Jim Albert and Monika Hu

Chapter 3 Conditional Probability

Introduction

- Probabilities are conditional in that one's opinion about an event is dependent on our current state of knowledge.
- As we gain new information, our probabilities can change.
- Bayes' rule provides a mechanism for changing our probabilities when we obtain new data.

Blood test example

- You are given a blood test for a rare disease. The proportion of people who currently have this disease is 0.1.
- The blood test comes back with two results: positive, which is some indication that you may have the disease, or negative.
- It is possible that the test will give the wrong result. If you have the disease, it will give a negative reading with probability 0.2. Likewise, it will give a false positive result with probability 0.2.
- Suppose that you have a blood test and the result is positive. Should you be concerned that you have the disease?

Two alternatives

- There are two possible alternatives: you have the disease, or you don't have the disease.
- Before you have a blood test, you assign probabilities to "have disease" and "don't have disease" that reflect the plausibility of these two models.
- You assign the event "have disease" a probability of 0.1 By the complement property, this implies that the event "don't have disease" has a probability of 1- 0.1 = 0.9.

Data

- The new information that one obtains to learn about the different models is called data.
- Here the data is the result of the blood test. Here the two possible data results are a positive result ($+$) or a negative result ($-$).
- If one "has the disease," the probability of a $+$ observation is 0.8 and the probability of a $-$ observation is 0.2. One writes

  $$P(+ \mid \mathrm{disease}) = 0.8, \;\; P(- \mid \mathrm{disease}) = 0.2.$$

  Likewise, if you don't have the disease

  $$P(+ \mid \mathrm{no\,disease}) = 0.2, \;\; P(- \mid \mathrm{no\,disease}) = 0.8.$$

Probability of interest

- Suppose you take the blood test and the result is positive $(+)$ – what is the chance you really have the disease?

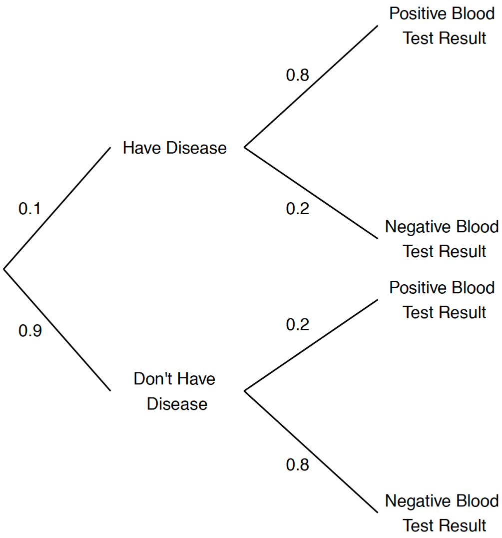
$$P(\text{disease} \mid +).$$

- Here the focus is on the so-called inverse probability – the probability of having the disease given a positive blood test result.
- We describe the computation of this inverse probability using two methods.

Method 1: Using a tree diagram

- A person either has or does not have the disease, and given the person's disease state, he or she either gets a positive or negative test result.
- One represents the outcomes by a tree diagram where the first set of branches corresponds to the disease states and the second set of branches corresponds to the blood test results.

The tree diagram



Computation

- By the definition of conditional probability,

$$P(\text{disease} \mid +) = \frac{P(\text{disease} \cap +)}{P(+)}.$$

- One finds the numerator $P(\text{disease} \cap +)$ by use of the multiplication rule:

$$\begin{aligned} P(\text{disease} \cap +) &= P(\text{disease})P(+ \mid \text{disease}) \\ &= 0.1 \times 0.8 = 0.08. \end{aligned}$$

- In the tree diagram, one is multiplying probabilities along the disease/$+$ branch to find this probability.

Computation

- To find the denominator $P(+)$, note that there are two ways of getting a positive blood test result.
- These two outcomes are the disease/+ and no disease/+ branches of the tree. One finds the probability of each outcome, and then sum the outcome probabilities:

$$\begin{aligned} P(+) &= P(\mathrm{disease} \cap +) + P(\mathrm{no\,disease} \cap +) \\ &= P(\mathrm{disease})P(+ \mid \mathrm{disease}) + P(\mathrm{no\,disease})P(+ \mid \mathrm{no\,dise} \\ &= 0.1 \times 0.8 + 0.9 \times 0.2 \\ &= 0.26. \end{aligned}$$

- So the probability of having the disease, given a positive blood test result is

$$P(\mathrm{disease} \mid +) = \frac{P(\mathrm{disease} \cap +)}{P(+)} = \frac{0.08}{0.26} = 0.31.$$

Method 2: Using a Bayes' box

- Suppose there are 1000 people in the community – one places "1000' in the lower right corner of the table.

		$+$	$-$	TOTAL
Disease	Have disease			
status	Don't have disease			
TOTAL				1000

Method 2: Using a Bayes' box

- One knows that the chance of getting the disease is 10% – so one expects 10% of the $1000 = 100$ people to have the disease and the remaining 900 people to be disease-free.

		$+$	$-$	TOTAL
Disease	Have disease			100
status	Don't have disease			900
TOTAL				1000

Method 2: Using a Bayes' box

- One knows the test will err with probability 0.2. So if 100 people have the disease, one expects 20% of 100 = 20 to have a negative test result and 80 will have a positive result
- Likewise, if 900 people are disease-free, then 20% of 900 = 180 will have an incorrect positive result and the remaining 720 will have a negative result

		$+$	$-$	TOTAL
Disease	Have disease	80	20	100
status	Don't have disease	180	720	900
TOTAL				1000

Method 2: Using a Bayes' box

- Now one is ready to compute the probability of interest $P(\text{disease} \mid +)$ from the table of counts.
- Restrict attention to the $+$ column of the table – there are 260 people had a positive test result. Of these 260, 80 actually had the disease, so

$$P(\text{disease} \mid +) = \frac{80}{260} = 0.31.$$

Comments

- Initially you had a small probability of 0.10 of having the disease
- The new probability of having the disease (0.31) is larger than the initial probability since a positive blood test was observed
- Note that since the new probability is under 0.5, it is still unlikely you have the disease