**Bayes Theorem**

Bayes theorem allows computation of conditional probabilities. It means that we wish to calculate how probable an event is given that another event has already happened.

Bayes theorem is as follows. Let A and B be to independent events. Then:

: The probability of event A given that event B has occurred. This is the **conditional probability**. It’s also called the **posterior probability**.

The probability that event A and B have happen together. This is the **joint probability** of A and B.

The probability that event A happens. This is called the **prior probability**.

The probability that event B happens. That is a combination of **all possible outcomes** that result in event B happening.

Notice though, that:

And since ,

Thus,

The ratio is called the likelihood ratio. Thus the **posterior probability** is equals to the **prior probability** multiplied by the **likelihood ratio**.

And also

Is not rare that event A is a hypothesis and event B is an evidence. So we try to determine the probability that hypothesis A is true given evidence B.

**Example 1.** Imagine that we use a test to detect cancer. Let us consider the following data:

* 0.5% of people have cancer (99.5% do not).
* 80% of tests detect cancer when it is there, that is give true positives (20% miss it). This is called test sensitivity.
* 10% of mammograms detect cancer when it’s not there. Therefore 90% correctly return true negatives, this is called specificity.

Now assume a woman receives a positive test for breast cancer. What is the probability that she actually has breast cancer?

The following table summarizes:

|  |  |  |
| --- | --- | --- |
|  | w=P[cancer]=0.005 | P[Not cancer]=0.995 |
| Test Positive | y=0.8 | 0.1 |
| Test Negative | 0.2 | z=0.9 |
| Total | 1 | 1 |

|  |  |  |
| --- | --- | --- |
|  | P[A]=w | P[Not A]=1-w |
| B | P[True Positive]=P[B|A]=y | P[False Positive]=P[B|Not A]=1-z |
| Not B | P[False Negative]=P[Not B|A]=1-y | P[True Negative]=P[Not B|Not A]=z |

y is called the **sensitivity** or true positive rate and it represents the probability of detection. It measures the proportion of positives that are correctly identified.

z is called the **specificity** or true negative rate. It represents the proportion of negatives that are correctly identified.

With that, we may construct the following table:

|  |  |  |  |
| --- | --- | --- | --- |
|  | A | Not A | Total |
| B | a=P[B|A]P[A]=w·y | b=P[B|Not A]P[Not A]=(1-w)·(1-z) | a+b |
| Not B | c=P[Not B|A]P[A]=w·(1-y) | d=P[Not B|Not A]P[Not A]=(1-w)·z | c+d |
| Total | a+c | b+d | a+b+c+d |

|  |  |  |
| --- | --- | --- |
|  | w=P[cancer]=0.005 | P[Not cancer]=0.995 |
| Test Positive | a=(0.005)(0.8)=0.004 | b=(0.995)(0.1)=0.0995 |
| Test Negative | c=(0.005)(0.2)=0.001 | d=(0.995)(0.9)=0.8955 |
| Total | 0.005 | 0.995 |

Since ,

and all possibilities for B are

We find,

Returning to our cancer test where w=0.005, y=0.8 and z=0.9 we find that

Second view, Tree. Assume 1,000 samples.¿:

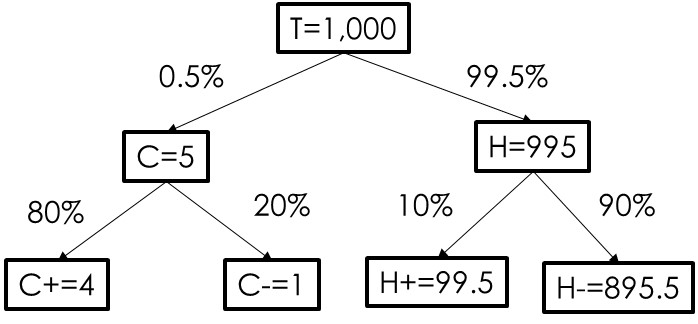


Table view:

|  |  |  |  |
| --- | --- | --- | --- |
|  | C=Cancer | H=Not Cancer |  |
| +=Positive | C+=4 | H+=99.5 | 103.5 |
| -=Not Positive | c-=1 | H-=895.5 | 896.5 |
|  | 5 | 995 | 1000 |

Thus

**Example 2**: Assume you are police chief of a community. This is the data about criminals in your district in jail population:

|  |  |  |  |
| --- | --- | --- | --- |
|  | Jail Population | |  |
|  | Jailed | Not Jailed | Total |
| Criminal | 95 | 10 | 105 |
| Not criminal | 5 | 300 | 305 |
| Total | 100 | 310 | 410 |

, recall

, precision

**Example 3**: Bayesian Spam Filter. Given that a message contains certain word (from a collection of spam words) we wish to find the probability that the message is spam.

Event A: The message is spam

Event B: The message contains a spam word

We have the following data:

1. 5% of messages are spam
2. If a message contains a spam word it will be spam 75% of the time. This is the sensitivity.
3. If a message does not contain a spam word it will not be spam 90% of the time. This is the specificity.

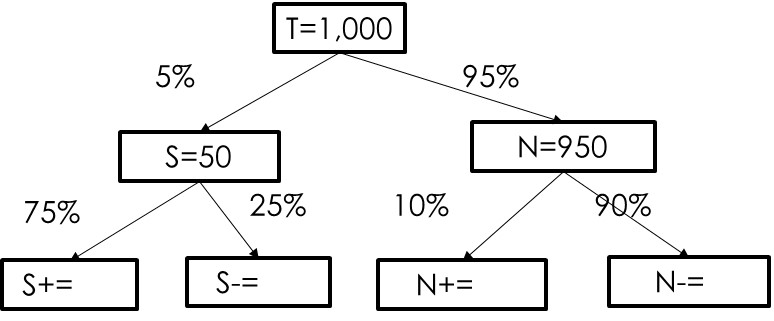


Table view:

|  |  |  |  |
| --- | --- | --- | --- |
|  | S=Spam | N=Not Spam |  |
| +=Spam Word | S+= | N+= | 132.5 |
| -=No Spam Word | S-= | N-= | 867.5 |
|  | 50 | 950 | 1000 |

Thus

|  |  |  |
| --- | --- | --- |
|  | P[Spam]=w=0.05 | P[Not Spam]=1-w=0.95 |
| Word | P[True Positive]=  P[Word|Spam]=y=0.75 | P[False Positive]  =P[Word|Not Spam]=0.1 |
| Not Word | P[False Negative]  =P[Not Word|Spam]=0.25 | P[True Negative]  =P[Not Word|Not Spam]=0.9 |

|  |  |  |
| --- | --- | --- |
|  | P[Spam]=w=0.05 | P[Not Spam]=1-w=0.95 |
| Word | a=w·y=0.05(0.75)=0.0375 | b=(1-w)·(1-z)=0.95(0.1)=0.095 |
| Not Word | c=w·(1-y)=0.05(0.25)=0.0125 | d=(1-w)·z=0.95(0.9)=0.855 |

Thus

This filter only catches 28.3% of spam.

If the Bayesian filter were to keep learning and increase the sensitivity to 90% and the specificity to 95% we would have:

The filter now detects 48.6% of spam mail.