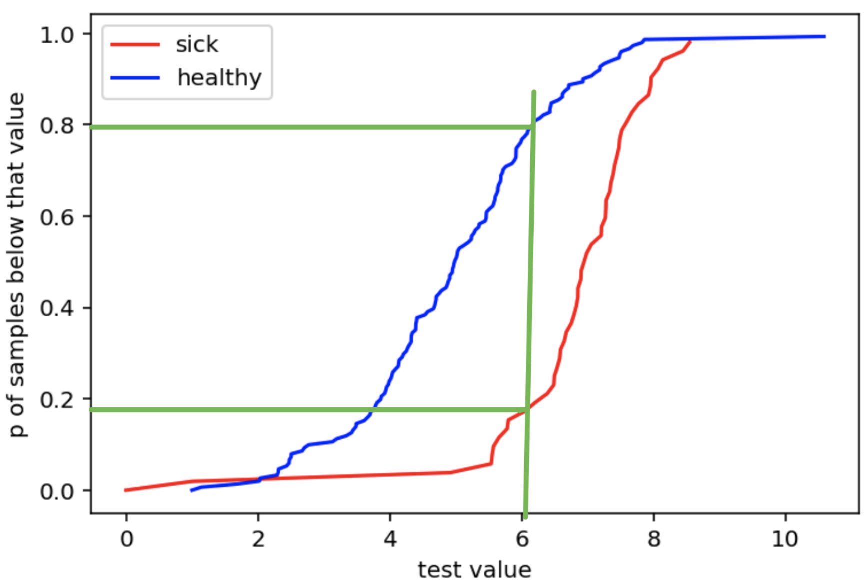
Questions (and answers) before DM101 Lesson 4

I got my hands on characteristics of one of the fast Covid tests. It measures the concentration of McPN15 in saliva: you are considered sick if the concentration is above 6, and healthy if it's below.

The curves below, though, tell the whole story. They were obtained by testing people who were known to be healthy and people who are known to be sick (as established beyond doubt by some other tests). The x-axis shows the concentration of McPN15, and the values on y-axis correspond to the proportion of people (in each group) for whom the concentration was below that value. Now consider the currently used threshold of 6. The graph tells us that 80 % of healthy people have the concentration of McPN15 smaller or equal to 6. In turn, only 18 % of sick people have McPN15 below or equal to 6.



We will now examine the characteristics of this test. In some of the following questions, you may need an additional piece of information: generally, 10 % of people who take the test are actually sick. (This is not the same as saying that the test is positive in 10 % of cases!) Also, remember that Bayes is your friend.

1. What is the probability for a sick person to be correctly diagnosed as sick? And what proportion of sick people is missed?

2. What is the probability that a healthy person is recognized as healthy? What proportion of healthy people of unnecessarily put in isolation?

3. What is the probability that the test gives the correct result? (Don't forget: only 10 % of people are actually sick!) That is, the probability that we encounter a person who is sick and classified as sick or a person who is healthy and classified as healthy?

4. If you are classified as sick: what is the probability that you are actually sick? (Compare this question with question 1. Be aware of the difference!)

5. If you are classified as healthy: what is the probability that you indeed are?

The goal of above questions was to help you think about the mathematics that is going on here, and help you thinking about the following, deeper and more important questions:

6. What happens with those numbers if we increase the classification threshold from 6 to 7? Or decrease it to 5?

7. Can you think of some semi-objective, principled criteria for choosing the threshold? Try looking at it from utilitarianistic perspective.

Answers

The whole task is just an exercise in conditional probabilities and Bayesian rule.

We will use + and − to denote positive and negative test results, and *sick* and *healthy* will refer to person's actual status.

The curve gives us two probabilities.

"*The graph tells us that 80 % of healthy people have the concentration of McPN15 smaller or equal to 6*": this translates to

$$p(-|healthy) = 0.8.$$

"*In turn, only 18 % of sick people have McPN15 below or equal to 6.*"

$$p(-|sick) = 0.18.$$

1. What is the probability for a sick person to be correctly diagnosed as sick? And what proportion of sick people is missed?

$$p(+|sick) = 1 - p(-|sick) = 0.72$$

$$p(-|sick) = 0.18$$

2. What is the probability that a healthy person is recognized as healthy? What proportion of healthy people of unnecessarily put in isolation?

$$p(-|healthy) = 0.8$$

$$p(+|healthy) = 1 - p(-|healthy) = 0.2$$

3. What is the probability that the test gives the correct result? (Don't forget: only 10 % of people are actually sick!) That is, the probability that we encounter a person who is sick and classified as sick or a person who is healthy and classified as healthy?

$$\begin{aligned} & p(+, sick) + p(-, healthy) \\ = \; & p(+|sick)\; p(sick) + p(-|healthy)\; p(healthy) \\ = \; & 0.82 \times 0.1 + 0.8 \times 0.9 \\ = \; & 0.082 + 0.72 \\ = \; & 0.802 \end{aligned}$$

   Intuitively, for every tested person, there is a 10 % chance that he's sick, and for such our accuracy is 82 %; and there is a 90 % he's healthy, and for those our accuracy is 80 %.

   You can also say that the total accuracy is the weighted average of the accuracies.

4. If you are classified as sick: what is the probability that you are actually sick? (Compare this question with question 1. Be aware of the difference!)

   Question 1 asks about $p(+|sick)$, while this is about $p(sick|+)$. We can compute this using the Bayesian rule:

$$p(sick|+) = \frac{p(+|sick)\; p(sick)}{p(+)}$$

   The problem is that we don't know the probability that the test result is positive. But we can compute this, too, because we know the probability of getting the positive result for someone who is sick (which we multiply by the probability of being sick) and the probability of positive result for healthy people (which we multiply by the probability of being healthy). Mathematically:

$$\begin{aligned} p(+) &= p(+, sick) + p(+, healthy) \\ &= p(+|sick)\; p(sick) + p(+|healthy)\; p(healthy) \\ &= 0.82 \times 0.1 + 0.2 \times 0.9 = \\ &= 0.082 + 0.18 \\ &= 0.262 \end{aligned}$$

   Thus

$$p(sick|+) = \frac{p(+|sick)\; p(sick)}{p(+)} = \frac{0.82 \times 0.1}{0.262} = \frac{0.082}{0.262} = 0.313$$

5. If you are classified as healthy: what is the probability that you indeed are?

$$p(-) = 1 - p(+) = 0.738$$

$$p(healthy|-) = \frac{p(-|healthy)\, p(healthy)}{p(-)} = \frac{0.8 \times 0.9}{0.738} = 0.976$$

6. What happens with those numbers if we increase the classification threshold from 6 to 7? Or decrease it to 5?

   If we decrease the threshold we will increase the number of positive results *in both groups, among the healthy and the sick*. The test will be less likely to overlook a sick person, but at the same time, there will be a higher chance that somebody who tests positive is actually healthy.

   Increasing the threshold works in the opposite direction.

   The table below shows a the above two metrics and accuracy. We could also show some other numbers, but these are those that we perhaps most often think about.

threshold	$p(+\|sick)$	$p(healthy\|+)$	accuracy
5	0.94	0.83	0.544
6	0.72	0.69	0.802
7	0.45	0.67	0.855
8	0.05	0.00	0.905

7. Can you think of some semi-objective, principled criteria for choosing the threshold? Try looking at it from utilitarianistic perspective.

   A common answer to this would be: we would like the fewest possible number of mistakes. This is the same as if to say that we'd like to make the highest possible number of correct predictions. Among the above, the best threshold is 7, with 90.5 % accuracy. This is, however, useless: this threshold will detect only 5 % of sick people. Its high accuracy stems from not making mistakes among the healthy, which represent the large majority of those tested.

   A threshold of 5 or 6 may be favourable and sensible. Yet: even the accuracy with threshold 6 is only 80 %. If we simply classify everybody as healthy, we reach the accuracy of 90 %.

   This, though, only shows that accuracy is a very poor measure for assessing the usefulness of models.

   A more objective criteria is to set the cost of both types of mistakes (or costs or benefits of some other criteria) and find the threshold that optimizes it. This is why I dropped the hint about utilitarianism.