TP6 - Medical diagnosis

RO11 - Apprentissage pour la robotique

Gallego, Natalia
AST
ENSTA Paris
natalia.gallego@ensta-paris.fr

I. Introduction

In this practical work, the analysis and diagnosis of respiratory diseases was addressed by using a support system based on probabilities and medical tests. The main objective was to model the diagnosis of diseases such as tuberculosis, lung cancer and bronchitis, using a data set that includes information about the patients' medical history.

The analysis began with the evaluation of the probability of the patient suffering from each disease, given his previous information, and was then updated as test results were obtained.

The work focused on the use of conditional probabilities to update the diagnosis as more data was obtained, applying the concept of Bayesian inference at each step of the process. In addition, the relevance of each medical test (stethoscope and x-ray) in clinical decision making and the efficiency of their use in the diagnosis of respiratory diseases were discussed.

II. Problem information

The data given by the problem are presented below. The interpretation of these will be used throughout the work.

Symbol	Description
A	Recently visited Asia
S	Smoker status
T	Has Tuberculosis
C	Has Cancer
B	Has Bronchitis
St	Positive stethoscope findings
X	Positive X-ray findings

Expression	Probability
$P(\mathbf{A})$	0.1
$P(\mathbf{S})$	0.3
$P(\mathbf{T} \mid \mathbf{A})$	0.1
$P(\mathbf{T} \mid \neg\mathbf{A})$	0.01
$P(\mathbf{C} \mid \mathbf{S})$	0.2
$P(\mathbf{C} \mid \neg\mathbf{S})$	0.02
$P(\mathbf{B} \mid \mathbf{S})$	0.6
$P(\mathbf{B} \mid \neg\mathbf{S})$	0.8
$P(\mathbf{St} \mid \mathbf{B})$	0.6
$P(\mathbf{St} \mid \mathbf{C})$	0.6
$P(\neg\mathbf{St} \mid \neg\mathbf{B} \cup \neg\mathbf{C})$	0.99
$P(\mathbf{X} \mid \mathbf{T})$	0.7
$P(\mathbf{X} \mid \mathbf{C})$	0.7
$\mathbf{P}(\neg\mathbf{X} \mid \neg\mathbf{T} \cup \neg\mathbf{C})$	0.98

III. Question 1 : Model with a Bayesian network

A Bayesian network is a statistical model that represents the conditional dependencies of random variables using an acyclic directed graph. In this graph, variables are represented as nodes and the dependencies between them are represented as arrows, which indicate the causal relationship between the variables.

Each node has an associated conditional probability distribution that describes the probability of that event occurring given its parents in the graph.

Taking into account the above, and the information provided by the problem, the following Bayesian network is created.

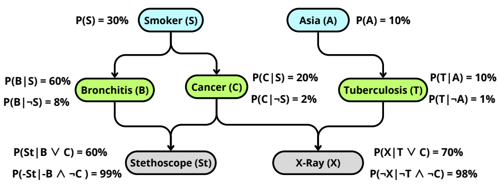

Fig. 1. CIFAR-10 Dataset

IV. Question 2 : If the patient ¬S and ¬A, can we infer what disease?

If a patient is a non-smoker ($\neg\mathbf{S}$) and has not recently visited Asia ($\neg\mathbf{A}$), we can use the given probabilities to infer the probability of each disease.

Considering the diagram in Figure 1, we see that Bronchitis ($\mathbf{B}$) and Cancer ($\mathbf{C}$) depend only on whether the patient is a smoker ($\mathbf{S}$). And Tuberculosis ($\mathbf{T}$) depends only on whether the patient recently traveled to Asia ($\mathbf{A}$).

With that in mind, we can rewrite the probabilities in the following way and use the information given by the problem to find the probability of each:

$$P(\mathbf{B} \mid \neg\mathbf{S} \wedge \neg\mathbf{A}) = P(\mathbf{B} \mid \neg\mathbf{S}) = 0.8 \quad (1)$$

$$P(\mathbf{C} \mid \neg\mathbf{S} \wedge \neg\mathbf{A}) = P(\mathbf{C} \mid \neg\mathbf{S}) = 0.02 \quad (2)$$

$$P(\mathbf{T} \mid \neg\mathbf{S} \wedge \neg\mathbf{A}) = P(\mathbf{T} \mid \neg\mathbf{A}) = 0.01 \quad (3)$$

Given the high probability of bronchitis (80%) in non-smokers, bronchitis ($\mathbf{B}$) is the most likely diagnosis based on prior probabilities alone.

V. Question 3 : According to Question 2, the doctor decides to auscultate the patient's lungs with a stethoscope?

Based on the inference from point 2, the doctor would suspect that bronchitis ($\mathbf{B}$) is the most likely disease, with a probability of 0.80, since the patient is a nonsmoker ($\neg\mathbf{S}$) and has not visited Asia recently ($\neg\mathbf{A}$). Given this suspicion, the doctor decides to listen to the patient's lungs with a stethoscope ($\mathbf{St}$) because with X-rays ($\mathbf{X}$) it is not possible to verify this disease, and also the stethoscope test ($\mathbf{St}$) is relatively effective at detecting bronchitis ($\mathbf{B}$), with a 60% probability of detecting either bronchitis ($\mathbf{B}$) or lung cancer ($\mathbf{C}$).

Furthermore, since the probabilities of tuberculosis and lung cancer are low for this patient (0.01 and 0.02, respectively), the doctor prioritizes a test that focuses specifically on the most likely condition, bronchitis ($\mathbf{B}$), before moving on to more specific tests (such as an X-ray ($\mathbf{X}$)) that would be more appropriate if cancer or tuberculosis were suspected.

A. The stethoscope test is negative. What is the new inferred diagnosis?

We will take into account that the conditions of not being a smoker ($\neg\mathbf{S}$) and not having been to Asia recently ($\neg\mathbf{A}$) are followed for the rest of the problem. This allows us to rewrite the probabilities as:

$$P(\mathbf{B} \mid \neg\mathbf{S}) = P(\mathbf{B}) = 0.8$$
$$P(\mathbf{C} \mid \neg\mathbf{S}) = P(\mathbf{C}) = 0.02$$
$$P(\mathbf{T} \mid \neg\mathbf{A}) = P(\mathbf{T}) = 0.01$$

Using Bayes' theorem we obtain the following expressions:

$$P(\mathbf{B} \mid \neg\mathbf{St}) = \frac{P(\neg\mathbf{St} \mid \mathbf{B}) \cdot P(\mathbf{B})}{P(\neg\mathbf{St})} \quad (4)$$

$$P(\mathbf{C} \mid \neg\mathbf{St}) = \frac{P(\neg\mathbf{St} \mid \mathbf{C}) \cdot P(\mathbf{C})}{P(\neg\mathbf{St})} \quad (5)$$

Calculate $P(\neg\mathbf{St})$ from the following definition:

$$\begin{aligned} P(\neg\mathbf{St}) = & P(\neg\mathbf{St} \mid \mathbf{B}) \cdot P(\mathbf{B}) \\ & + P(\neg\mathbf{St} \mid \mathbf{C}) \cdot P(\mathbf{C}) \\ & + P(\neg\mathbf{St} \mid \neg\mathbf{B} \cup \neg\mathbf{C}) \cdot P(\neg\mathbf{B} \cup \neg\mathbf{C}) \end{aligned} \tag{6}$$

$$\begin{aligned} P(\neg\mathbf{B} \cup \neg\mathbf{C}) &= 1 - [P(\mathbf{B}) + P(\mathbf{C}) - P(\mathbf{B} \cap \mathbf{C})] \\ &= 1 - (0.8 + 0.02 - (0.8)(0.02)) \\ &= 0.196 \end{aligned} \tag{7}$$

Using 6 and 7 we get the result of 8.

$$\begin{aligned} P(\neg\mathbf{St}) &= (0.4)(0.8) + (0.4)(0.02) \\ &+ (0.99)(0.0196) \\ &= 0.52204 \end{aligned} \tag{8}$$

Finally, evaluating in (4) and (5)

$$P(\mathbf{B} \mid \neg\mathbf{St}) = \frac{(1-0.6)(0.8)}{0.52204} = 0.613 \tag{9}$$

$$P(\mathbf{C} \mid \neg\mathbf{St}) = \frac{(1-0.6)(0.02)}{0.52204} = 0.0153 \tag{10}$$

After the negative stethoscope result, the most likely diagnosis remains bronchitis ($\mathbf{B}$) with a probability of 61.3%. The likelihood of cancer is very low at around 1.53%.

VI. Question 4 : The doctor orders an X-Ray ($\mathbf{X}$) that is positive. What is the new inferred diagnosis?

Seeking to know the new probabilities of each of the diseases using the information provided by X-rays, we again use Bayes' theorem.

We calculate the updated probability for tuberculosis ($\mathbf{T}$) and Cancer ($\mathbf{C}$):

$$P(\mathbf{T} \mid \mathbf{X}) = \frac{P(\mathbf{X} \mid \mathbf{T})P(\mathbf{T})}{P(\mathbf{X})} \tag{11}$$

$$P(\mathbf{C} \mid \mathbf{X}) = \frac{P(\mathbf{X} \mid \mathbf{C})P(\mathbf{C})}{P(\mathbf{X})} \tag{12}$$

We calculate the probability of having a positive result using X-rays.

$$\begin{aligned} P(\mathbf{X}) = & P(\mathbf{X} \mid \mathbf{T}) \cdot P(\mathbf{T}) \\ & + P(\mathbf{X} \mid \mathbf{C}) \cdot P(\mathbf{C}) \\ & + P(\mathbf{X} \mid \neg\mathbf{T} \cup \neg\mathbf{C}) \cdot P(\neg\mathbf{T} \cup \neg\mathbf{C}) \end{aligned} \tag{13}$$

We calculate the value of a positive result when the patient has neither cancer ($\mathbf{C}$) nor tuberculosis ($\mathbf{T}$).

$$\begin{aligned} P(\mathbf{X} \mid \neg\mathbf{T} \cup \neg\mathbf{C}) &= 1 - P(\neg\mathbf{X} \mid \neg\mathbf{T} \cup \neg\mathbf{C}) \\ &= 1 - 0.98 = 0.02 \end{aligned} \tag{14}$$

It is also necessary to calculate the probability that the patient has neither cancer ($\mathbf{C}$) nor tuberculosis ($\mathbf{T}$):

$$\begin{aligned} P(\neg\mathbf{T} \cup \neg\mathbf{C}) &= 1 - [P(\mathbf{T}) + P(\mathbf{C}) - P(\mathbf{T} \cap \mathbf{C})] \\ &= 1 - (0.01 + 0.02 - (0.01)(0.02)) \\ &= 0.9702 \end{aligned} \tag{15}$$

Replacing the values found in 14 and 15 within the expression 13 we obtain the following result:

$$\begin{aligned} P(\mathbf{X}) &= (0.3)(0.01) + (0.3)(0.02) \\ &+ (0.02)(0.9702) \\ &= 0.20304 \end{aligned} \tag{16}$$

Replacing the value of the expression 16 in equations 11 and 12 we obtain the following:

$$P(\mathbf{T} \mid \mathbf{X}) = \frac{(0.7)(0.02)}{0.20304} = 0.00345 \tag{17}$$

$$P(\mathbf{C} \mid \mathbf{X}) = \frac{(0.7)(0.01)}{0.20304} = 0.0069 \tag{18}$$

Since both the probabilities of having tuberculosis ($\mathbf{T}$) and cancer ($\mathbf{C}$) are very low, the most probable diagnosis is that the patient have some other condition (such as bronchitis ($\mathbf{B}$)).

The new inferred diagnosis remains the same as that obtained from question two, which indicates that the patient has bronchitis ($\mathbf{B}$).

VII. Question 5 : Was the X-Ray needed?

No, X-ray ($\mathbf{X}$) was not strictly necessary for diagnosis in this case.

Based on the patient's characteristics (non-smoker ($\neg\mathbf{S}$) and not having visited Asia recently ($\neg\mathbf{A}$)), the most likely diagnosis was bronchitis ($\mathbf{B}$), which has a higher probability in this scenario compared to tuberculosis ($\mathbf{T}$) or cancer ($\mathbf{C}$). The negative stethoscope test ($\mathbf{St}$) result also indicated the low possibility of cancer.

Even without the X-ray ($\mathbf{X}$), based on the patient's history, the odds of having tuberculosis ($\mathbf{T}$) (1%) or cancer ($\mathbf{C}$) (2%) were already very low. Bronchitis ($\mathbf{B}$), on the other hand, was more likely given the patient's non-smoking status and lack of recent travel to Asia. After the X-ray ($\mathbf{X}$), the odds of having tuberculosis ($\mathbf{T}$) (0.34%) or cancer ($\mathbf{C}$) (0.69%) remained extremely low.

The X-ray ($\mathbf{X}$) was not necessary because the patient's characteristics already pointed towards bronchitis ($\mathbf{B}$) with a high degree of confidence. The X-ray ($\mathbf{X}$) did not significantly change the outcome and could therefore be considered an unnecessary additional step in the diagnostic process for this particular case.