**Marginalization**

In the previous segment, you understood what it means for a model to be generative or discriminative. Prof. Raghavan also mentioned that generative models can perform prediction even when some of the feature values are not available. In the following lecture, you will understand how this happens with the help of a joint probability distribution.

Replay

Mute

Current Time 7:12

/

Duration 7:12

Loaded: 100.00%

Playback Rate

Auto

Playback Rate

1.5x

Fullscreen

Let's try to understand the concepts using another example. Consider a situation where you are trying to predict whether a person has lung cancer depending upon the following factors with their levels shown below:

1. Lung cancer

|  |  |
| --- | --- |
| lung cancer  = no | L(0) |
| lung cancer = yes | L(1) |

1. Smoker

|  |  |
| --- | --- |
| Smoker = no | S(0) |
| Smoker  = yes | S(1) |

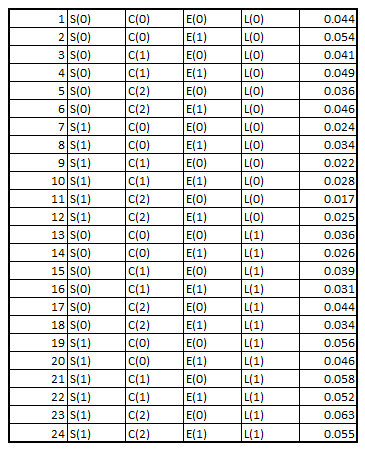
1. Smoking Circle (Number of friends who smoke along with him irrespective of whether he smokes or not)

|  |  |
| --- | --- |
| Smoking Circle = 1 | C(0) |
| Smoking Circle = 3 | C(1) |
| Smoking Circle = 5+ | C(2) |

1. Exercise

|  |  |
| --- | --- |
| Does not exercise regularly | E(0) |
| Exercises Regularly | E(1) |

You can see that different variables/ features have a different number of levels. A joint probability distribution here will be defined as a probability distribution over all the different combinations of the levels of the features as shown below:



**Joint Probability Distribution**

The last column in the above table represents the probability of occurrence of each of the combination of events.

Before we move forward, let us first revise what the conditional probability expression means.

p(A|B) = p(A, B) / p(B)

where,

* p(A|B) refers to the probability of occurrence of event A conditioned that event B has occurred
* p(A, B) is the probability of occurrence of both the events A and B

When the events are independent, p(A, B) becomes p(A) x p(B).

Note that the symbol '⊥' in A ⊥ B means that A is independent of B.

You can refer to the following .csv file to answer the questions that follow.

**[JPD CSV](https://cdn.upgrad.com/UpGrad/temp/b9c29aed-36b4-41da-bad4-a2baaaefd4d0/JPD+Platform.xlsx" \o "JPD Platform.xlsx" \t "_blank)**

[file\_download](https://cdn.upgrad.com/UpGrad/temp/b9c29aed-36b4-41da-bad4-a2baaaefd4d0/JPD+Platform.xlsx" \o "JPD Platform.xlsx" \t "_blank)**[Download](https://cdn.upgrad.com/UpGrad/temp/b9c29aed-36b4-41da-bad4-a2baaaefd4d0/JPD+Platform.xlsx" \o "JPD Platform.xlsx" \t "_blank)**

keyboard\_arrow\_leftkeyboard\_arrow\_rightQuestions:3/3

**Conditional Probability**

How would you break up the conditional probability p(S(1) | L(1), E(0))?

Top of Form



p(S(1),L(1),E(0))p(L(1),E(0))

**Feedback :**

*Carefully, look at the expression mentioned by the professor.*

**Correct!**



p(L(1),E(0))p(S(1),L(1),E(0))

**Feedback :**

*Carefully, look at the expression mentioned by the professor.*

**Incorrect!**

Bottom of Form

**close**Your answer is **Incorrect.**

Continue

Attempt 1 of 1

In order to calculate P(L(1)), you summed up rows 13 - 24 which according to the law of total probability is essentially

∑E∑C∑Sp(S,C,E,L(1))

In the next lecture, you'll learn more about the computational feasibility of the Marginalization process.

Replay

Mute

Current Time 3:13

/

Duration 3:13

Loaded: 100.00%

Playback Rate

Auto

Playback Rate

1.5x

Fullscreen

Marginalization is a computationally expensive process to make an inference/prediction. You have seen the number of summations one needs to perform to make an inference over partial data. This problem shall be solved with the help of graphs which you shall see in the upcoming segments.

keyboard\_arrow\_leftkeyboard\_arrow\_rightQuiz Summary

* **Question 1**

keyboard\_arrow\_right**Correct**

* **Question 2**

keyboard\_arrow\_right**Correct**

* **Question 3**

keyboard\_arrow\_right**Incorrect**

* **Question 4**

keyboard\_arrow\_right**Incorrect**

A terminology one should be aware of is that calculating the conditional probability p(S(1) | L(1), E(0)) is often called the **inference task**. This will become more clear in the upcoming modules.

Having understood the steps involved in the marginalisation process, let's go through the following lecture where Prof. summarises the whole process.

Play Video

Probabilistic Graphical Models belong to the generative class of models which can model all the relationship between the variables and answer any question asked even when partial data is provided. The process of marginalization used to answer such questions is computationally very inefficient and hence we need to reduce the size of the joint probability distribution table to make it quicker. You'll see how we do this in the next couple of segments.