BAYES' THEOREM

BAYES THEOREM

- one of the most famous equations in the world of statistics and probability

$$\boldsymbol{P(A|B) = \frac{P(B|A)P(A)}{P(B)}}$$

A FRAMEWORK FOR UPDATING OUR BELIEFS

- What’s the point of probability ? **=> decision making under uncertainty**
- When you decide on an action, **you are betting** that completing the action will leave you better off than had you not done it.
- **But bets are inherently uncertain**, and our knowledge about the world is never totally exact, so how do you decide whether to go ahead with it or not ?
- **Bayes’ Theorem gives us a quantitative framework for updating our beliefs as the facts around us change.**

THE INTUITION WITH AN EXAMPLE
(HTTPS://MEDIUM.COM/OPEX-ANALYTICS/BAYES-THEOREM-101-6A9A1EA5D4A6)

- It’s 9AM on Monday morning, and you receive an email from your boss. You notice that it seems a little different from her usual notes: the message contains several grammatical errors, and ends by asking you to provide your social security number. Though you first assumed it was a legitimate email, the grammar mistakes and suspicious request convince you to send it right to the spam folder.

- When making that quick decision to ignore the email from your “boss,” you unconsciously estimated several different probabilities. First, you judged the likelihood of a work email’s legitimacy to be fairly high. But then you assessed the probability that such a weird email could come from your boss to be low. You also have some general sense that phishing emails tend to be weird in a few specific ways, and you know that phishing scams are common enough that this particular email could plausibly be harmful.

- With all this information swirling around in your head, you decide that the email is most likely spam. That's pretty much all conditional probability is: determining the probability of an event *given* the other information that you know.

TAKING A CLOSER LOOK AT THE FORMULA

$$\boldsymbol{P(A|B) = \frac{P(B|A)P(A)}{P(B)}}$$

- **P(A|B)** – is the probability of A given that B has already happened.
- **P(B|A)** – is the probability of B given that A has already happened. It looks circular and arbitrary for now…
- **P(A)** – is the unconditional probability of A occurring.
- **P(B)** – is the unconditional probability of B occurring.

TAKING A CLOSER LOOK AT THE FORMULA

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

P(A|B) is an example of a conditional probability – one that measures probability over only certain states of the world (states where B has occurred).

P(A) is an example of an unconditional probability and is measured over all states of the world.

VIDEO EXPLANATION

- https://www.youtube.com/watch?v=HZGCoVF3YvM&t=542s

EXAMPLE

- Suppose that you are a recently graduated student. You have yet to hear back from some of the companies you interviewed with and are getting nervous. So you decide to calculate the probability that a specific company will make you an offer given that it's been 3 days and they still have not called you yet.

- Let's rewrite the formula in terms of our example. Here, outcome A is "receiving an offer" and outcome B is "no phone call for 3 days". So we can write our formula as:

P(A|B) = P(B|A) * P(A)/P(B)

P(Offer|NoCall) = P(NoCall|Offer) * P(Offer) / P(NoCall)

EXAMPLE

- **P(Offer|NoCall),** the probability of receiving an offer given no phone call for 3 days

  => hard to estimate

- Let’s estimate the rest (arbitrarily):
  - **P(NoCall|Offer)** = 40% => no phone call for 3 days given that you have an offer.
  - **P(Offer)** = 20% => landing a job offer in general.
  - **P(NoCall)** = 90% => not getting a call back from the company for 3 days (pass on you, or they might still be interviewing other candidates)
- Plug in our estimates:

P(Offer|NoCall) = 40% * 20%/90% = 8.9%

THE INTUITION BEHIND THE FORMULA

- Remember Bayes' Theorem is a framework for updating our beliefs…
- **So where do our beliefs come in?**
- **The prior =>** P(A) ⇔P(Offer)
  - this is our prior belief about how likely it is to receive an offer.
  - Or the likelihood that you will receive an offer at the exact moment that you exit the interview room.

$$P(A|B) = P(B|A) * P(A)/P(B)$$

THE INTUITION BEHIND THE FORMULA

- Now, new information has come in
  - 3 days have gone by and the company has yet to call you.
  - So we use the other parts of the equation to adjust our prior for the new event that has occurred.

$$P(Hypothesis \mid Evidence) = \underbrace{P(Hypothesis)}_{\text{Prior}} * \frac{\overbrace{P(Evidence \mid Hypothesis)}^{\text{Scaler}}}{\underbrace{P(Evidence)}_{\text{Normalizer}}}$$

THE INTUITION BEHIND THE FORMULA

- We use **P(Evidence|Hypothesis)** to flip the problem around by asking, "What is the probability of observing this evidence in a world where our hypothesis is true?"
- So in our example, we want to know how likely it is to go 3 days without a phone call in a world where the company has definitely decided to make us an offer.
- **P(Evidence|Hypothesis) => the scaler**
- **When we multiply it against the prior, the scaler scales the prior up or down depending on whether the evidence helps or hurts our hypothesis** – in our case, the scaler reduces the prior because more days going by without a phone call would be an increasingly bad sign…
- **The scaler is the mechanism that Bayes' Theorem utilizes to adjust our prior beliefs.**

THE INTUITION BEHIND THE FORMULA

$$\underbrace{\boxed{P(Hypothesis \mid Evidence) * P(Evidence)}}_{\text{Joint Probability}} = \underbrace{P(Hypothesis)}_{\text{Prior}} * \underbrace{P(Evidence \mid Hypothesis)}_{\text{Scaler}}$$

- **If we leave P(Evidence) on the left side of our equation, we get a joint probability, which considers all states of the world. But we we only want the states where the evidence has occurred => P(Hypothesis|Evidence)**

THE INTUITION BEHIND THE FORMULA

- **Dividing the product of prior and scaler by P(Evidence) means we reduce the possible states to the ones where the evidence has occurred.**

TO SUMMARIZE

- P(Offer) =>prior
- Fresh out of the interview, we start with a prior – there is a 20% chance that we will get the job we just interviewed for.

TO SUMMARIZE

- As more days pass, we use the scaler to scale down our prior.
- For example, after 3 days have gone by, we estimate that in a world where we get the job, there is just a 40% chance that the company would have waited this long to call you. Multiplying scaler and prior we get 20% * 40% = 8%.
- P(NoCall|Offer) = scaler

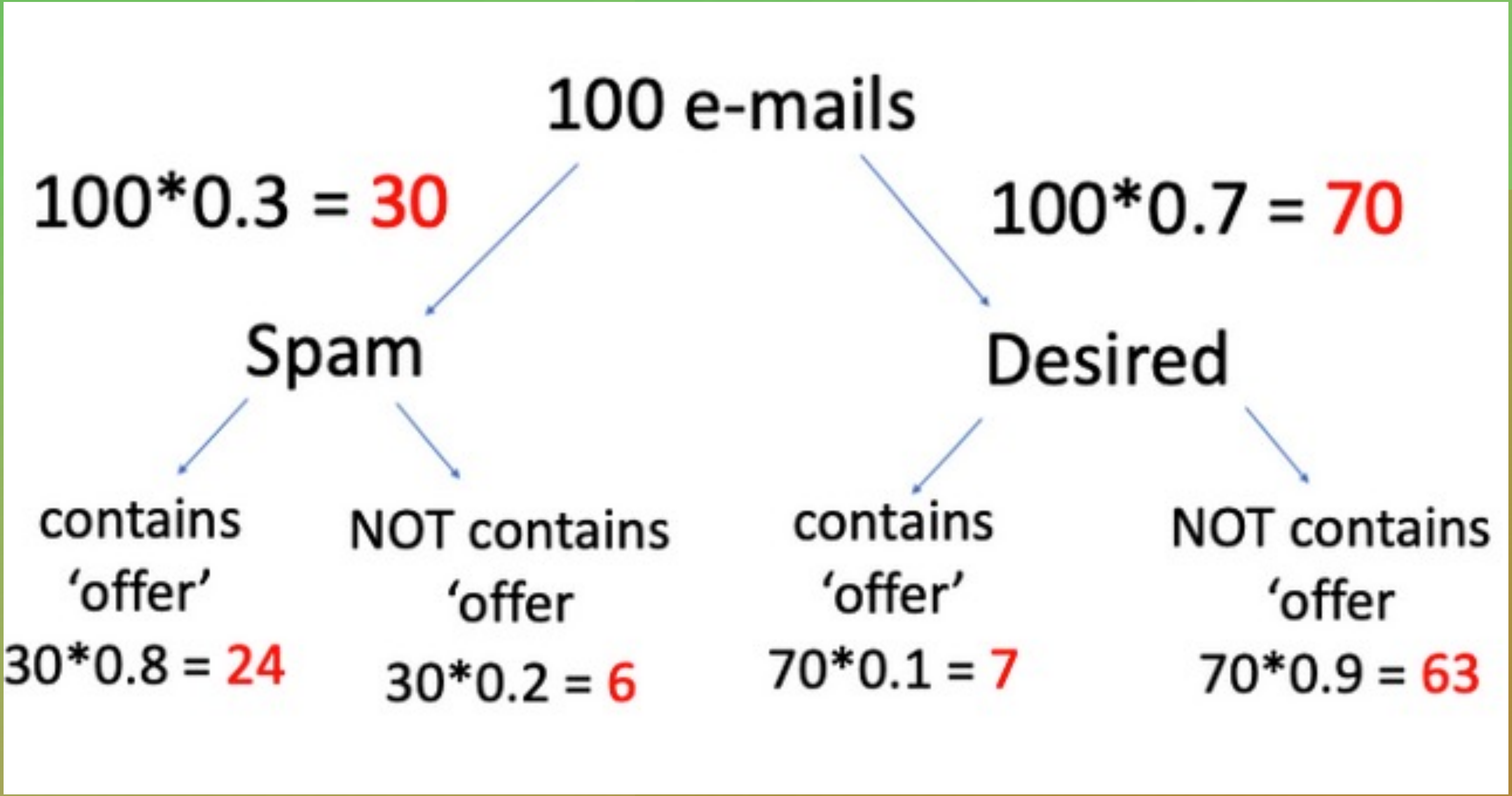
TO SUMMARIZE

- P(NoCall) => normalizer
- Finally, we recognize that the 8% is calculated over all states of the world.
- But we only care about states of the world where we have not received a phone call from the company for 3 days post interview.
- In order to capture only those states, we estimate the unconditional probability of not receiving a call for 3 days to be 90% – this is our normalizer. We divide our previously calculated 8% by the normalizer, 8% / 90% = 8.9%, to get our final answer.
- **So in the states of the world where we have not heard back from the company for 3 days, there is an 8.9% chance that we will receive an offer.**

PRACTICE PROBLEM1

- Using Bayes for NLP to predict spams based on the content of an email.
- Assume that the word *offer* occurs in 80% of the spam messages
- Also assume *offer* occurs in 10% of desired e-mails (hams)
- If 30% of the received e-mails are considered to be spams, and I receive a new message which contains *offer*, what is the probability that this new email is a spam?

- Draw a tree diagram to help you and consider the case where you have a sample of 100 emails : you can first find the solution by counting, and then try and find it using the theorem.


100 e-mails
100*0.3 = 30
100*0.7 = 70
Spam
Desired
contains 'offer'
30*0.8 = 24
NOT contains 'offer
30*0.2 = 6
contains 'offer'
70*0.1 = 7
NOT contains 'offer
70*0.9 = 63

PRACTICE PROBLEM1

- P(contains *offer*|spam) = 0.8 (given in the question)
- P(spam) = 0.3 (given in the question)
- P(contains *offer*) = 0.3*0.8 + 0.7*0.1 = 0.31

$$P(spam|contains\ offer) = \frac{P(contains\ offer|spam) * P(spam)}{P(contains\ offer)}$$

PRACTICE PROBLEM1

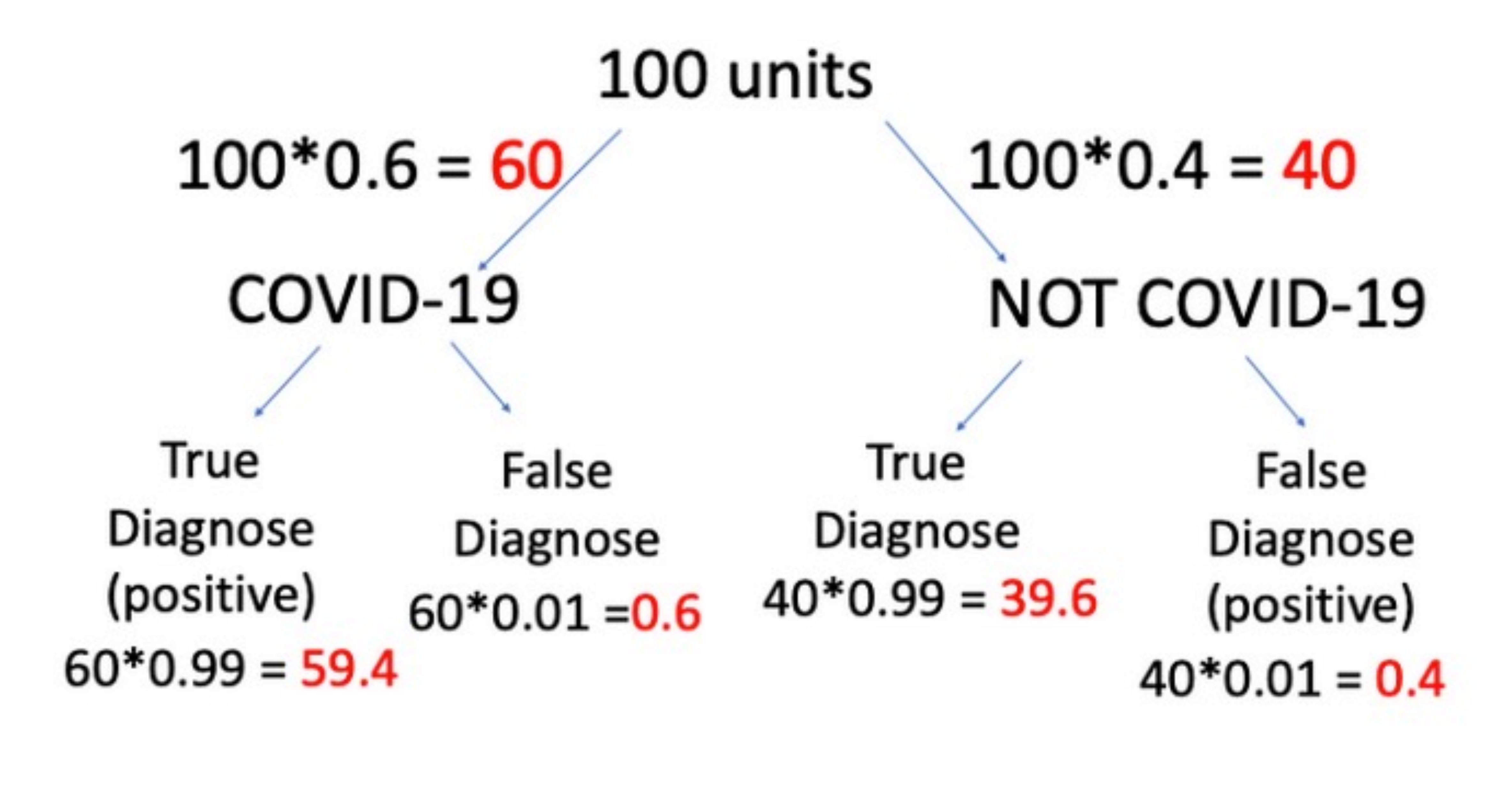
- Both results should be the same :

$$P(spam|contains\ offer) = \frac{0.8 * 0.3}{0.31} = 0.774$$

PRACTICE PROBLEM 2

- Covid-19 tests are common nowadays, but some test results can be wrong…
- Let’s assume:
  - a diagnostic test has 99% accuracy
  - and 60% of all people have Covid-19.
- If a patient tests positive, what is the probability that they actually have the disease?

- Same as previously: take a sample of 100 patients first and find the probability using counts and then use the theorem.


100 units
100*0.6 = 60
100*0.4 = 40
COVID-19
NOT COVID-19
True Diagnose (positive)
60*0.99 = 59.4
False Diagnose
60*0.01 =0.6
True Diagnose
40*0.99 = 39.6
False Diagnose (positive)
40*0.01 = 0.4

$$P(covid19|positive) = \frac{P(positive|covid19) * P(covid19)}{P(positive)}$$

- P(positive|covid19) = 0.99
- P(covid19) = 0.6
- P(positive) = 0.6*0.99+0.4*0.01=0.598

$$P(covid19|positive) = \frac{0.99 * 0.6}{0.598} = 0.993$$

PRACTICE PROBLEM 3 (MONTY HALL IS BACK)

- You're on a gameshow called "**Let's Make a Deal**". There are 3 closed doors in front of you.
- Behind each door is a prize. One door has a **car**, one door has **breath mints**, and one door has a **bar of soap**. You'll get the prize behind the door you pick, but you don't know which prize is behind which door. Obviously you want the car!
- Imagine you pick **door A**.
- After picking **door A**, the host of the show, Monty Hall, now opens **door B,** revealing a bar of soap. He then asks you if you'd like to change your guess. Should you?
- By working through Bayes Theorem, we can calculate the actual odds of winning the car if we stick with **door A**, or switch to **door C**.

PRACTICE PROBLEM 3

- The posteriors we want to compute :

1. $P(prize=A|opened=B)$ vs. 2. $P(prize=C|opened=B)$

PRACTICE PROBLEM 3

- **Priors**
  - The probability of any door being correct before we pick a door is 1/3. Prizes are randomly arranged behind doors and we have no other information. So the **prior**, P(A), of any door being correct is **1/3**.

1. $P(prize = A)$, the prior probability that door A contains the car = 1/3
2. $P(prize = C)$, the prior probability that door C contains the car = 1/3

PRACTICE PROBLEM 3

- **Likelihood**
  - If the car is behind door A, then Monty can open door B or C. So the probability of opening either is 50%.

1. $P(opens = B|prize = A) = \frac{1}{2}$, the likelihood Monty opened door B if door A is correct

  - If the car is in fact behind door C then Monty can only open door B. He cannot open A, the door we picked. He also cannot open door C because it has the car behind it.

2. $P(opens = B|prize = C) = 1$, the likelihood Monty opened door B if door C is correct

PRACTICE PROBLEM 3

- **Numerator: P(A) x P(B|A)**
- $P(prize = A) \times P(opens = B|prize = A) \ = \ 1/3 \times 1/2 \ = \ 1/6$
- $P(prize = C) \times P(opens = B|prize = C) \ = \ 1/3 \times 1 \ = \ 1/3$

PRACTICE PROBLEM 3

- **Normalize**

- This is the marginal probability P(opens=B) which is the total probability, removing dependence from any event:
  - In this case:

$$\sum P(opens = B|prize = A)P(prize = A), P(opens = B|prize = C)P(prize = C)$$

$$P(opens = B) = \frac{1}{6} + \frac{1}{3} = \frac{1}{2}$$

- Putting everything together:

1. $P(prize = A|opens = B) = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{1}{3}$

2. $P(prize = C|opens = B) = \frac{\frac{1}{3}}{\frac{1}{2}} = \frac{2}{3}$

=> the prize is more likely to be hidden behind door C, so we should switch !