Baye's Theorem

Baye's Theorem, named after 18th century British mathematician Thomas Bayes, is a mathematical formula for determining conditional probability. Baye's theorem allows us to update predicted probabilities of an event by incorporating new information.

$$\underbrace{P(A|B)}_{\text{Posterior probability}} = \frac{\underbrace{P(B|A)}_{\text{Likelihood}} \cdot \overbrace{P(A)}^{\text{Prior Probability}}}{\underbrace{P(B)}_{\text{Marginal Likelihood}}}$$

$$\boxed{\begin{aligned} &\circ\ P(A|B) = \frac{P(A \cap B)}{P(B)} \\ &\circ\ P(B|A) = \frac{P(B \cap A)}{P(A)} \end{aligned}}$$

(Conditional Probability)

- $P(A|B) \cdot P(B) = P(A \cap B)$
- $P(B|A) \cdot P(A) = P(B \cap A)$

Now, because $P(A \cap B)$ and $P(A \cap B)$ are equal, we can write

$$P(A \cap B) = P(A|B) \cdot P(B) = P(B|A) \cdot P(A)$$

Now, we can also write

$$P(A|B) \cdot \underline{P(B)} = P(B|A) \cdot P(A)$$

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)} \quad \text{(Baye's Theorem)}$$

Here, A is hypothesis and B is evidence / data.

How to read?

We have to find the probability of hyphothesis (A), given that we have observed some evidence / we have been given some data. On the basis of that given data, we have to find the probability of hyphothesis.

Likelihood – Probability of the evidence given that hyphothesis is true (P(B|A))

Prior Probability – Probability of the hyphothesis before considering the evidence.

Marginal Likelihood – Pure probability of evidence / data.

Example :-

Fruit	Yellow	Sweet	Long	Total
Orange	350	450	0	650
Banana	400	300	350	400
Others	50	100	50	150
Total	800	850	400	1200

Using the Naive Baye's Classifier, find the fruit which is yellow, sweet as well as long.

Solution :-

$$\underset{\text{find}}{P(A} \,|\, \underset{\text{Given}}{B)} = \frac{P(B|A) \cdot P(A)}{P(B)}$$

$$P(\text{Yellow} \,|\, \text{Orange}) = \frac{P(\text{Orange} \,|\, \text{Yellow}) \cdot P(\text{Yellow})}{P(\text{Orange})}$$

$$P(\text{Yellow} \mid \text{Orange}) = \frac{\frac{350}{800} \times \frac{800}{1200}}{\frac{650}{1200}} = 0.5$$

$$P(\text{Sweet} \mid \text{Orange}) = \frac{P(\text{Orange} \mid \text{Sweet}) \cdot P(\text{Sweet})}{P(\text{Orange})}$$

$$P(\text{Sweet} \mid \text{Orange}) = \frac{\frac{450}{850} \times \frac{850}{1200}}{\frac{650}{1200}} = 0.69$$

$$P(\text{Long} \mid \text{Orange}) = \frac{P(\text{Orange} \mid \text{Sweet}) \cdot P(\text{Sweet})}{P(\text{Orange})}$$

$$P(\text{Long} \mid \text{Orange}) = \frac{\frac{0}{400} \times \frac{400}{1200}}{\frac{650}{1200}} = 0$$

We have calculated all the possible values for the fruit orange. Now we will multiply all the values to get the probability of fruit being Orange, which is yellow, sweet and long.

Because there are 0 oranges which are long, we will ultimately get 0 probability.

$$P(\text{Fruit} \mid \text{Orange}) = 0.53 \times 0.69 \times 0 = 0$$

Just like above calculations, we can find probability for other fruits as well.

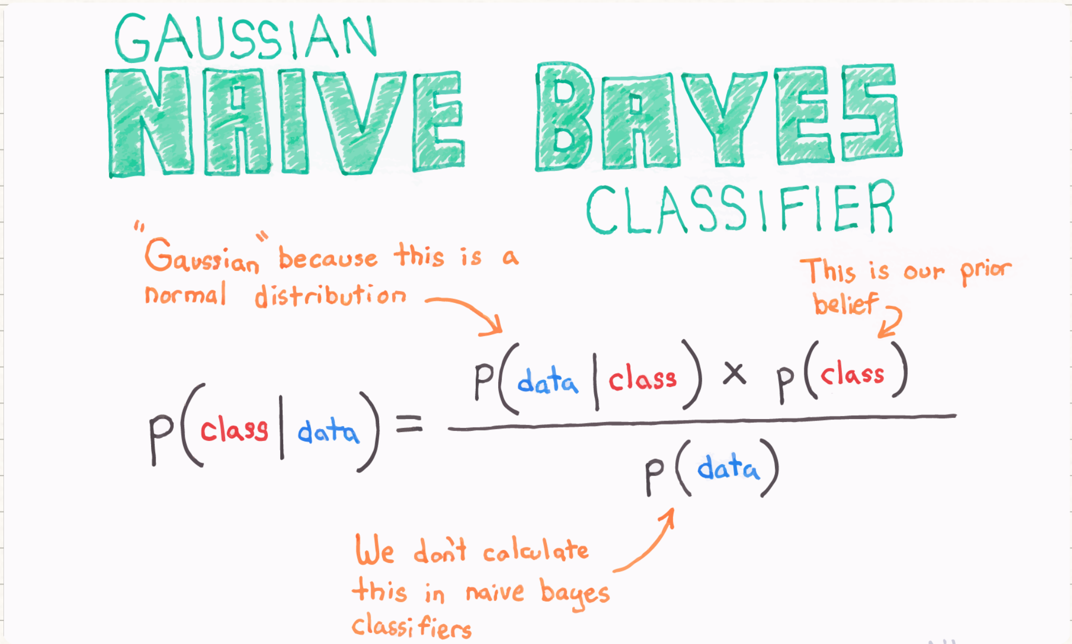
$$P(\text{Fruit} \mid \text{Banana}) = 1 \times 0.75 \times 0.87 = 0.65$$

$$P(\text{Fruit} \mid \text{Others}) = 0.33 \times 0.66 \times 0.33 = 0.072$$

So from the above result, we can clearly see that Banana wins.

Q. Why Naive Baye's Theorem is called Naive?

A. Because it assumes that each input variable is independent. This is a strong assumption and unrealistic for real data, however, this technique is very efficient on a large range of complex problems.



Bayes' Theorem

Bayes can do magic!

Ever wondered how computers learn about people?

Example:

An internet search for "movie automatic shoe laces" brings up "Back to the future"

Has the search engine watched the movie? No, but it knows from lots of other searches what people are **probably** looking for.

And it calculates that probability using Bayes' Theorem.

Bayes' Theorem is a way of finding a probability when we know certain other probabilities.

The formula is:

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

Which tells us:	how often A happens *given that B happens*, written **P(A|B)**,
When we know:	how often B happens *given that A happens*, written **P(B|A)**
	and how likely A is on its own, written **P(A)**
	and how likely B is on its own, written **P(B)**

Let us say P(Fire) means how often there is fire, and P(Smoke) means how often we see smoke, then:

P(Fire|Smoke) means how often there is fire when we can see smoke
P(Smoke|Fire) means how often we can see smoke when there is fire

So the formula kind of tells us "forwards" P(Fire|Smoke) when we know "backwards" P(Smoke|Fire)

Example:

dangerous fires are rare (1%)

but smoke is fairly common (10%) due to barbecues,

and 90% of dangerous fires make smoke

We can then discover the **probability of dangerous Fire when there is Smoke**:

$$P(\text{Fire}|\text{Smoke}) = \frac{P(\text{Fire})\,P(\text{Smoke}|\text{Fire})}{P(\text{Smoke})}$$

$$= \frac{1\% \times 90\%}{10\%}$$

$$= 9\%$$

So it is still worth checking out any smoke to be sure.

Example: Picnic Day

You are planning a picnic today, but the morning is cloudy

Oh no! 50% of all rainy days start off cloudy!

But cloudy mornings are common (about 40% of days start cloudy)

And this is usually a dry month (only 3 of 30 days tend to be rainy, or 10%)

What is the chance of rain during the day?

We will use Rain to mean rain during the day, and Cloud to mean cloudy morning.

The chance of Rain given Cloud is written P(Rain|Cloud)

So let's put that in the formula:

$$P(\text{Rain}|\text{Cloud}) = \frac{P(\text{Rain})\,P(\text{Cloud}|\text{Rain})}{P(\text{Cloud})}$$

P(Rain) is Probability of Rain = 10%

P(Cloud|Rain) is Probability of Cloud, given that Rain happens = 50%

P(Cloud) is Probability of Cloud = 40%

$$P(\text{Rain}|\text{Cloud}) = \frac{0.1 \times 0.5}{0.4} = .125$$

Or a 12.5% chance of rain. Not too bad, let's have a picnic!

Just 4 Numbers

Imagine 100 people at a party, and you tally how many wear pink or not, and if a man or not, and get these numbers:

	Pink	*notPink*
Man	5	35
notMan	20	40

Bayes' Theorem is based off just those 4 numbers!

Let us do some totals:

	Pink	*notPink*	
Man	5	35	*40*
notMan	20	40	*60*
	25	*75*	*100*

And calculate some probabilities:

the probability of being a man is $P(\text{Man}) = \frac{40}{100} = 0.4$

the probability of wearing pink is $P(\text{Pink}) = \frac{25}{100} = 0.25$

the probability that a man wears pink is $P(\text{Pink}|\text{Man}) = \frac{5}{40} = 0.125$

the probability that a person wearing pink is a man **P(Man|Pink) = ...**

And then the puppy arrives! Such a cute puppy.

But all your data is **ripped up**! Only 3 values survive:

P(Man) = 0.4,

P(Pink) = 0.25 and

P(Pink|Man) = 0.125

Can you discover **P(Man|Pink)** ?

Imagine a pink-wearing guest leaves money behind ... was it a man? We can answer this question using Bayes' Theorem:

$$P(\text{Man}|\text{Pink}) = \frac{P(\text{Man})\ P(\text{Pink}|\text{Man})}{P(\text{Pink})}$$

$$P(\text{Man}|\text{Pink}) = \frac{0.4 \times 0.125}{0.25} = 0.2$$

Note: if we still had the raw data we could calculate directly $\frac{5}{25} = 0.2$

Being General

Why does it work?

Let us replace the numbers with letters:

	B	notB	
A	s	t	s+t
notA	u	v	u+v
	s+u	t+v	s+t+u+v

Now let us look at **probabilities**. So we take some ratios:

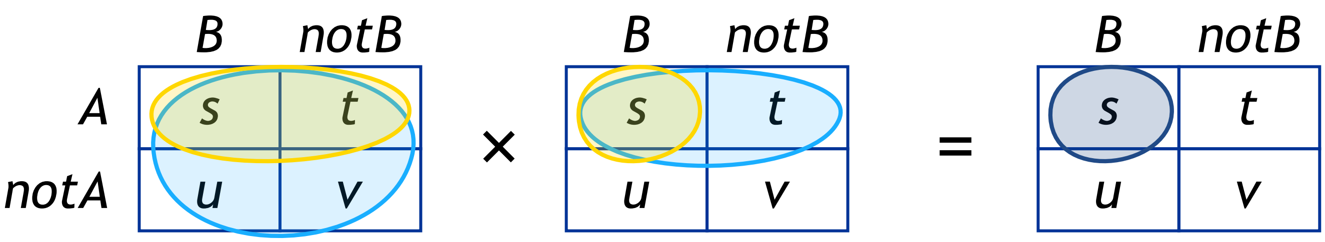
the overall probability of "A" is $P(A) = \frac{s+t}{s+t+u+v}$

the probability of "B given A" is $P(B|A) = \frac{s}{s+t}$

And then multiply them together like this:

$$P(A) \times P(B|A) = P(A)\,P(B|A)$$

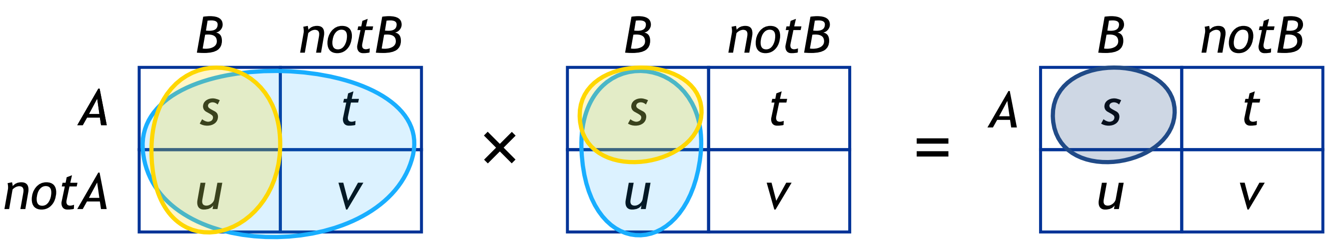
$$\frac{s+t}{s+t+u+v} \times \frac{s}{s+t} = \frac{s}{s+t+u+v}$$



Now let us do that again but use **P(B)** and **P(A|B)**:

$$P(B) \times P(A|B) = P(B)\,P(A|B)$$

$$\frac{s+u}{s+t+u+v} \times \frac{s}{s+u} = \frac{s}{s+t+u+v}$$



Both ways get the **same result** of $\frac{s}{s+t+u+v}$

So we can see that:

$$P(B)\ P(A|B) = P(A)\ P(B|A)$$

Nice and symmetrical isn't it?

It actually *has* to be symmetrical as we can swap rows and columns and get the same top-left corner.

And it is also **Bayes Formula** ... just divide both sides by P(B):

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

Remembering

First think "AB AB AB" then remember to group it like: "AB = A BA / B"

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

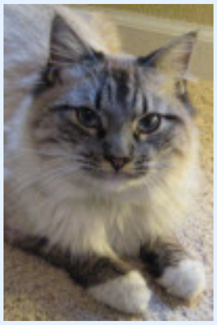
Cat Allergy?

One of the famous uses for Bayes Theorem is False Positives and False Negatives.

For those we have two possible cases for "A", such as **Pass**/**Fail** (or Yes/No etc)

Example: Allergy or Not?

Hunter says she is itchy. There is a test for Allergy to Cats, but this test is not always right:



For people that **really do** have the allergy, the test says "Yes" **80%** of the time

For people that **do not** have the allergy, the test says "Yes" **10%** of the time ("false positive")

If 1% of the population have the allergy, and **Hunter's test says "Yes"**, what are the chances that Hunter really has the allergy?

We want to know the chance of having the allergy when test says "Yes", written **P(Allergy|Yes)**

Let's get our formula:

$$P(\text{Allergy}|\text{Yes}) = \frac{P(\text{Allergy})\ P(\text{Yes}|\text{Allergy})}{P(\text{Yes})}$$

P(Allergy) is Probability of Allergy = 1%

P(Yes|Allergy) is Probability of test saying "Yes" for people with allergy = 80%

P(Yes) is Probability of test saying "Yes" (to anyone) = ??%

Oh no! We **don't know** what the **general** chance of the test saying "Yes" is ...

... but we can calculate it by adding up those **with**, and those **without** the allergy:

1% have the allergy, and the test says "Yes" to 80% of them

99% do **not** have the allergy and the test says "Yes" to 10% of them

Let's add that up:

$$P(\text{Yes}) = 1\% \times 80\% + 99\% \times 10\% = 10.7\%$$

Which means that about 10.7% of the population will get a "Yes" result.

So now we can complete our formula:

$$P(\text{Allergy}|\text{Yes}) = \frac{1\% \times 80\%}{10.7\%} = 7.48\%$$

$$P(\text{Allergy}|\text{Yes}) = \text{about } \mathbf{7\%}$$

This is the same result we got on False Positives and False Negatives.

In fact we can write a special version of the Bayes' formula just for things like this:

$$P(A|B) = \frac{P(A)P(B|A)}{P(A)P(B|A) + P(\text{not } A)P(B|\text{not } A)}$$

"A" With Three (or more) Cases

We just saw "A" with two cases (A and not A), which we took care of in the bottom line.

When "A" has 3 or more cases we include them all in the bottom line:

$$P(A1|B) = \frac{P(A1)P(B|A1)}{P(A1)P(B|A1) + P(A2)P(B|A2) + P(A3)P(B|A3) + \text{...etc}}$$

Example: The Art Competition has entries from three painters: Pam, Pia and Pablo

Pam put in 15 paintings, 4% of her works have won First Prize.

Pia put in 5 paintings, 6% of her works have won First Prize.

Pablo put in 10 paintings, 3% of his works have won First Prize.

What is the chance that Pam will win First Prize?

$$P(\text{Pam}|\text{First}) = \frac{P(\text{Pam})P(\text{First}|\text{Pam})}{P(\text{Pam})P(\text{First}|\text{Pam}) + P(\text{Pia})P(\text{First}|\text{Pia}) + P(\text{Pablo})P(\text{First}|\text{Pablo})}$$

Put in the values:

$$P(\text{Pam}|\text{First}) = \frac{(15/30) \times 4\%}{(15/30) \times 4\% + (5/30) \times 6\% + (10/30) \times 3\%}$$

Multiply all by 30 (makes calculation easier):

$$P(\text{Pam}|\text{First}) = \frac{15 \times 4\%}{15 \times 4\% + 5 \times 6\% + 10 \times 3\%}$$

$$= \frac{0.6}{0.6 + 0.3 + 0.3}$$

$$= 50\%$$

A good chance!

Pam isn't the most successful artist, but she did put in lots of entries.

Now, back to Search Engines.

Search Engines take this idea and scale it up a lot (plus some other tricks).

It makes them look like they can read your mind!

It can also be used for mail filters, music recommendation services and more.

Question 1 Question 2 Question 3 Question 4 Question 5
Question 6 Question 7 Question 8 Question 9 Question 10