Story of Raj & Simran with Naïve Bayes Concept

Introduction to Conditional Probability and Bayes theorem

1. **Events – Union, Intersection & Disjoint events**

1.1 EVENTS

An event is simply the outcome of a random experiment. Getting a head when we toss a coin is an event. Getting a 6 when we roll a fair die is an event. We associate probabilities to these events by defining the event and the sample space.

The **sample space** is nothing but the collection of all possible outcomes of an experiment. This means that if we perform a task again and again, all the possible results of the task are listed in the sample space.

For example: A sample space for a single throw of a die will be {1,2,3,4,5,6}. One of these is bound to occur if we throw a die. The sample space exhausts all the possibilities that can happen when that experiment is performed.

An event can also be a combination of different events.

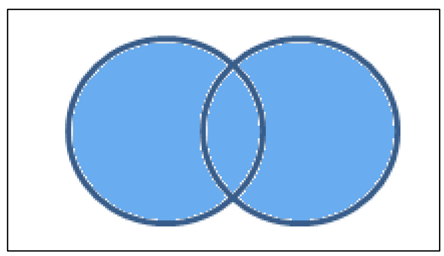
1.2 Union of Events

We can define an event (C) of getting a 4 or 6 when we roll a fair die. Here event C is a **union** of two events:

Event A = Getting a 4

Event B = Getting a 6

P (C) = P (A U B)

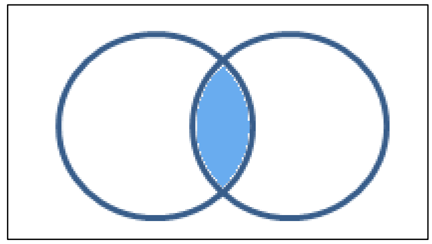


In simple words, we can say that we should consider the probability of (A U B) when we are interested in combined probability of two (or more) events.

1.3. Intersection of Events

Let's look at another example.

Let C be the event of getting a multiple of 2 and 3 when you throw a fair die.



Event A = Getting a multiple of 2 when you throw a fair die

Event B = Getting a multiple of 3 when you throw a fair die

Event C = Getting a multiple of 2 and 3

Event C is an intersection of event A & B.

Probabilities are then defined as follows.

P (C) = P (A$\cap$ B)

We can now say that the shaded region is the probability of both events A and B occurring together.

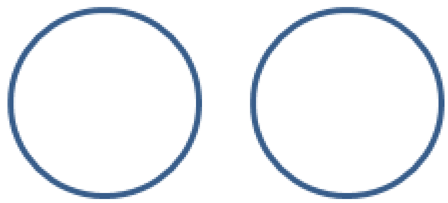
1.4 Disjoint Events

What if, we come across a case when any two events cannot occur at the same time.

For example: Let's say we have a fair die and we have only one throw.

Event A = Getting a multiple of 3

Event B = Getting a multiple of 5



We want both event A & B should occur together.

Let's find the sub space for Event A & B.

Event A = {3,6}

Event B = {5}

Sample Space= {1,2,3,4,5,6}

As we can see, there is no case for which event A & B can occur together. Such events are called disjoint event. To represent this using a Venn diagram.

2. Independent, Dependent & Exclusive Events

Suppose we have two events – event A and event B.

If the occurrence of event A doesn't affect the occurrence of event B, these events are called independent events.

Let's see some examples of independent events.

- Getting heads after tossing a coin AND getting a 5 on a throw of a fair die.
- Choosing a marble from a jar AND getting heads after tossing a coin.
- Choosing a 3 card from a deck of cards, replacing it, AND then choosing an ace as the second card.
- Rolling a 4 on a fair die, AND then rolling a 1 on a second roll of the die.

In each of these cases the probability of outcome of the second event is not affected at all by the outcome of the first event.

Probability of independent events

In this case the probability of P (A and B) = P (A) * P (B)

Let's take an example here. Suppose we win the game if we pick a red marble from a jar containing **4 red and 3 black marbles** and we get heads on the toss of a coin. What is the probability of winning?

Let's define event A, as getting red marble from the jar

Event B is getting heads on the toss of a coin.

We need to find the probability of both getting a red marble and a head in a coin toss.

$$P (A) = 4/7$$

$$P (B) = 1/2$$

We know that there is no effect of the color of the marble on the outcome of the coin toss.

P (A and B) = P (A) * P (B)

P (A and B) = (4/7) * (1/2) = (2/7)

Probability of dependent events

Next, can we think of examples of dependent events?

In the above example, let's define event A as getting a Red marble from the jar. We then keep the marble out and then take another marble from the jar.

Will the probabilities in the second case still be the same as that in the first case?

Let's see. So, for the first time there are 4/7 chances of getting a red marble. Let's assume we got a red marble on the first attempt. Now, for second chance, to get a red marble we have 3/6 chances.

If we didn't get a red marble on the first attempt but a white marble instead. Then, there were 4/6 chances to get the red marble second time. **Therefore, the probability in the second case was dependent on what happened the first time.**

Mutually exclusive and Exhaustive events

Mutually exclusive events are those events where two events cannot happen together.

The easiest example to understand this is the toss of a coin. Getting a head and a tail are mutually exclusive because we can either get heads or tails but never both at the same in a single coin toss.

A set of events is collectively exhaustive when the set should contain all the possible outcomes of the experiment. One of the events from the list must occur for sure when the experiment is performed.

For example, in a throw of a die, {1,2,3,4,5,6} is an exhaustive collection because, it encompasses the entire range of the possible outcomes.

Consider the outcomes "even" (2,4 or 6) and "not-6" (1,2,3,4, or 5) in a throw of a fair die. They are collectively exhaustive but not mutually exclusive.

3. Conditional Probability

Conditional probabilities arise naturally in the investigation of experiments where an outcome of a trial may affect the outcomes of the subsequent trials.

We try to calculate the probability of the second event (event B) given that the first event (event A) has already **happened**. If the probability of the event changes when we take the first event into consideration, we can safely say that the probability of event B is dependent of the occurrence of event A.

Let's think of cases where this happens:

- Drawing a second ace from a deck given we got the first ace
- Finding the probability of having a disease given you were tested positive
- Finding the probability of liking Harry Potter given we know the person likes fiction

Here we can define, 2 events:

- Event A is the probability of the event we're trying to calculate.
- Event B is the condition that we know or the event that has happened.

We can write the conditional probability as P(A/B), the probability of the occurrence of event A given that B has already happened.

$$P\left(\frac{A}{B}\right) = \frac{\text{P(A and B)}}{P(B)} = \frac{\text{Probability of the occurence of both A and B}}{\text{Probability of B}}$$

Let's play a simple game of cards for you to understand this. Suppose you draw two cards from a deck and you win if you get a jack followed by an ace (without replacement). What is the probability of winning, given we know that you got a jack in the first turn?

Let event A be getting a jack in the first turn

Let event B be getting an ace in the second turn.

We need to find P(B/A)

P(A) = 4/52

P(B) = 4/51 {no replacement}

P(A and B) = 4/52*4/51= 0.006

$$P\left(\frac{B}{A}\right) = \frac{\text{P(A and B)}}{P(A)} = \frac{0.006}{0.077} = 0.078$$

Here we are determining the probabilities when we know some conditions instead of calculating random probabilities. Here we knew that he got a jack in the first turn.

4. Bayes Theorem

The Bayes theorem describes the probability of an event based on the prior knowledge of the conditions that might be related to the event. If we know the conditional probability P(A/B), we can use the Bayes rule to find out the reverse probabilities P(B/A).

$$P\left(\frac{A}{B}\right) = \frac{P\ (A \cap B)}{P\ (B)}$$

$$P\left(\frac{B}{A}\right) = \frac{P\ (A \cap B)}{P\ (A)}$$

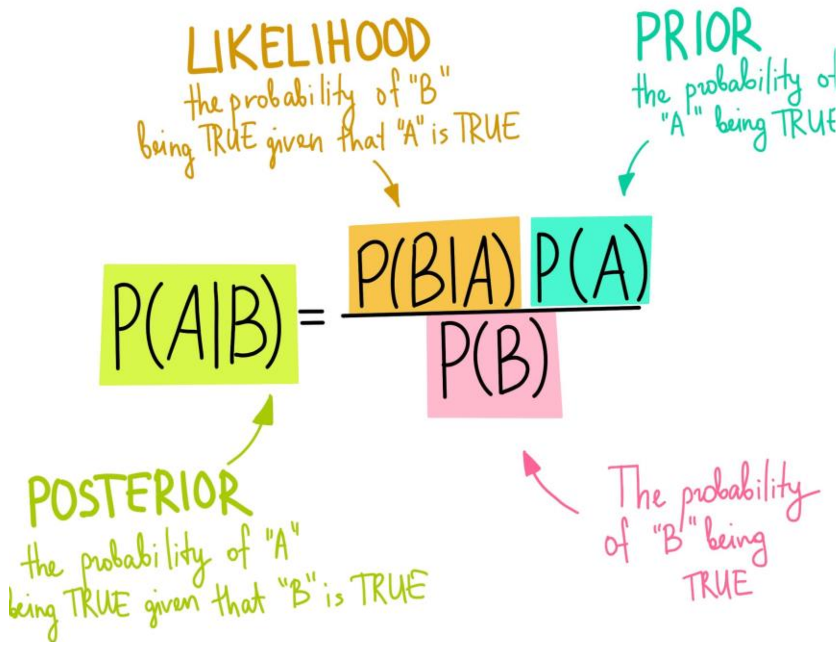
$$P\ (A \cap B) = P\left(\frac{A}{B}\right) * P\ (B) = P\left(\frac{B}{A}\right) * P\ (A)$$

$$P\left(\frac{B}{A}\right) = P\left(\frac{A}{B}\right) * \frac{P\ (B)}{P\ (A)}$$

The above statement is the general representation of the **Bayes rule**.

$$\boldsymbol{P(^{A}/_{B}\ ) = \frac{P\left(\frac{B}{A}\right) * P(A)}{P(B)}}$$

$$\boldsymbol{Posterior = \frac{Likelihood * Prior}{Evidence}}$$



Where,

Posterior Probability: Which represents the degree to which believe a given model accurately describes the situation given the available data and all of our prior information.

Likelihood: Which describes how well the model predicts the data

Prior Probability: Which describes the degree to which we believe the model accurately describe reality based on all of our prior information.

If multiple events A_i form an exhaustive set with another event B and K is the class of A. We can write the equation as

$$P\left(\frac{A=k}{B_1 B_2 \ldots .Bn}\right) = \frac{P\left(\frac{B_1}{A=k}\right) * P\left(\frac{B_2}{A=k}\right) * \ldots \ldots \ldots ..P\left(\frac{B_n}{A=k}\right) * P(A=k)}{P(B_1) * P(B_2) * \ldots ..* P(B_n)}$$

5 Naïve Bayes classifiers

Naive Bayes classifiers are a family of simple probabilistic classifiers based on applying **Bayes' theorem** with strong (**naive**) **independence** assumptions between the features. It is a popular method for text categorization *i.e.* classifying documents as belonging to one category or the other (such as spam or legitimate, sports or politics, etc.) using word frequencies as the features. With appropriate preprocessing, it is comparable to more advanced methods.

The method uses conditional probabilities and is based on frequency tables. Essentially,

$$prob(outcome|data) = \frac{prob(data|outcome) * prob(outcome)}{prob(data)}$$

$$P\left(\frac{B}{A}\right) = P\left(\frac{A}{B}\right) * \frac{P(B)}{P(A)}$$

Three popular Naive Bayes algorithms:

1. Bernoulli Naive Bayes
2. Multinomial Naive Bayes
3. Gaussian Naive Bayes

Example to understand

Raj and Simran are travelling in a train together but Raj is not able to figure out whether Simran is a male or a female.

However, Raj is good at mathematics and he knows how to applies Bayes theorem for conditional probabilities.

Raj notices that Simran has longer hair and tries to calculate how many males and females have longer hair in the entire population.

Understanding the situation

- 180 passengers in the train
- 90 are male and 90 are female.
- 85 females out of 90 have long hair and 7 out of 90 males have long hair.

	Males	Females	Total
Longer Hair	7	85	92
Shorter Hair	83	5	88
Total	90	90	180

Probability of Simran being a female

P(Male) = 90/180 = 0.5
P(Female) = 90/180 = 0.5
P (Long Hair) = 92/180
P(Short Hair) = 88/180
P(Long/Male) = 7/92
P(Long/Female) = 85/92

```
P(Female/Long) = P(Long/Female) x P(Female)/ P(Long)
               =    (85/92)     x  (90/180) /(92/180)
               =    0.9038
               = 90.38 %

P(Male/Long) = P(Long/Male) x P(Male)/ P(Long)
             =    (7/92)    x (90/180)/(92/180)
             =  0.0744
             =7.44 %
P(Female/Long) is higher than P(Male/Long) so now Raj has
decided Simran is female.
```

Example 2:Golf Data set

Outlook	Temp	Humidity	Windy	Play Golf
Rainy	Hot	High	False	No
Rainy	Hot	High	True	No
Overcast	Hot	High	False	Yes
Sunny	Mild	High	False	Yes
Sunny	Cool	Normal	False	Yes
Sunny	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Rainy	Mild	High	False	No
Rainy	Cool	Normal	False	Yes
Sunny	Mild	Normal	False	Yes
Rainy	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Sunny	Mild	High	True	No

Frequency Table

		Play Golf	
		Yes	No
Outlook	Sunny	3	2
	Overcast	4	0
	Rainy	2	3

		Play Golf	
		Yes	No
Humidity	High	3	4
	Normal	6	1

		Play Golf	
		Yes	No
Temp.	Hot	2	2
	Mild	4	2
	Cool	3	1

		Play Golf	
		Yes	No
Windy	False	6	2
	True	3	3

Likelihood Table

		Play Golf	
		Yes	No
Outlook	Sunny	3/9	2/5
	Overcast	4/9	0/5
	Rainy	2/9	3/5

		Play Golf	
		Yes	No
Humidity	High	3/9	4/5
	Normal	6/9	1/5

		Play Golf	
		Yes	No
Temp.	Hot	2/9	2/5
	Mild	4/9	2/5
	Cool	3/9	1/5

		Play Golf	
		Yes	No
Windy	False	6/9	2/5
	True	3/9	3/5

P(B)=	P(Outlook=Rainy)*	P(Temp=cool)*	P(Humidity=High)*	P(Windy=True)
	0.357142857	0.285714286	0.5	0.428571429
	0.021865889			

P(yes/B)=	P(Rainy/yes)*	P(cool/yes)*	P(High/yes)*	P(True/yes)*	P(yes)	/P(B)
	0.222222222	0.333333333	0.333333333	0.333333333	0.642857	0.021866
	0.241975309					

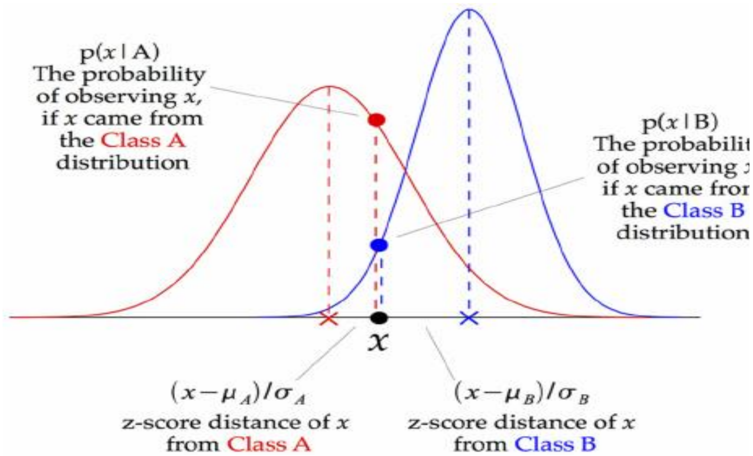
P(No/B)=	P(Rainy/No)*	P(cool/No)*	P(High/No)*	P(True/No)*	P(No)	/P(B)
	0.6	0.2	0.8	0.6	0.357143	0.021866
	0.9408					

P(No/B) > P(yes/B) so value of play golf=No

Cool	High	TRUE	NO

Gaussian naive Bayes

When the predictors take up a continuous value and are not discrete, we assume that these values are sampled from a **normal distribution**.



$$P(x_i|y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} exp\left(-\frac{(x_i-\mu_y)^2}{2\sigma_y^2}\right)$$

Example 3: Person classification

Create Data

Dataset is containing data on eight individuals. We will use the dataset to construct a classifier that takes in the height, weight, and foot size of an individual and outputs a prediction for their gender.

Person	height (feet)	weight (lbs)	foot size (inches)
male	6	180	12
male	5.92	190	11
male	5.58	170	12
male	5.92	165	10
female	5	100	6
female	5.5	150	8
female	5.42	130	7
female	5.75	150	9

?	6	130	8

Step 1: Calculate mean and variance of all variables group by target values.

Person	mean (height)	variance (height)	mean (weight)	variance (weight)	mean (foot size)	variance (foot size)
male	5.855	0.035033333	176.25	122.9166667	11.25	0.91666667
female	5.4175	0.097225	132.5	558.3333333	7.5	1.66666667

Step 2: Derive Bayes formula

As per Bayes rule: -

$$Posterior = \frac{Likelihood * Prior}{Evidence}$$

X={male,female}
Prior=P(X)
Likelihood=P(height/X)*P(Weight/X)*P(footsize/X)
Evidence=P(height/male) *P(Weight/male) *P(footsize/male) *P(male)+P(height/female) *P(Weight/female) *P(footsize/female) *P(female)

Posterior(X)=P(height/X)*P(Weight/X)*P(footsize/X)*P(X)/Evidence

Step 3: Calculate probability density of each of the terms of the likelihood

$$P(x_i|y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} exp\left(-\frac{(x_i-\mu_y)^2}{2\sigma_y^2}\right)$$

X={male,female}

Person	P(Height=6/x)	P(weight=130/x)	P(Foot sixe=8/x)
male	1.578883183	5.98674E-06	0.001311221
female	0.223458727	0.016789298	0.2866907

Step 4: Calculate Posterior(X)

Posterior(male)=P(height=6/ male)*P(Weight=130/ male)*P(footsize=8/ male)*P(male)/Evidence
=1.578883183*5.98674E-06*0.001311221*0.5/ **Evidence**
=6.19707E-09/ **Evidence**

Posterior(female)=P(height=6/ female)*P(Weight=130/ female)*P(footsize=8/ male)*P(female)/Evidence
=0.223458727 * 0.016789298 * 0.2866907*0.5/ **Evidence**
= 0.000537791

Because the numerator of the posterior for female is greater than male, then we predict that the person is female.

female	6	130	8