Probability :- probability is simply how likely something is to happen.

Probability Means Possibility

$$\text{Prob of event to happen } P(E) = \frac{\text{No. of favourable outcom}}{\text{total No of outcomes}}$$

Conditional probability is defined as the likelihood of an event or outcome occurring, based on the occurance of a Previous event or outcome.

Conditional probability is calculated by Multiplying the probability of the preceding event by the updated probability of the succeeding, or conditional event.

$$\text{formula :- } P(B/A) = \frac{P(A \text{ and } B)}{P(A)}$$

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

Example of Conditional Probability :- P(A/B) is the Probability of event A occuring, given that B occurs;

Ex :- Given that you drew a card, what's is the Prob. that it's a four

$$P\left(\frac{\text{four}}{\text{Red}}\right) = 2/26 = 1/13$$

So out of the 26 Red cards (Given a Red card), There are Two fours so $2/26 = 1/13$ As

Mutually exclusive or disjoint event :- In logic and Probability theory; Two events are mutually exclusive or disjoint if they cannot both occur at the same time. A clear example is the set of outcomes of a single coin toss;

which can result is either heads or tails, but not both.

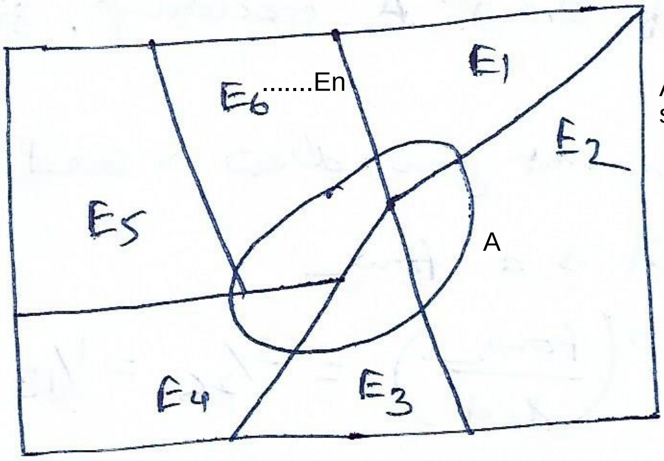
Theorem of total Probability / Rule of elimination :→

suppose the events $E_1, E_2, --- E_n$ are exhaustive and mutually exclusive; in this case, if $P(Ei > 0)$ for each i then

$$P(A) = \sum_{i=1}^{n} P(Ei) P(A/Ei)$$

$$= \sum_{i=1}^{n} P(Ei \cap A)$$

the conditional Probability being defined since none of $E_1, E_2 -- E_n$ has zero Probability; $Bi \neq 0$



A is an arbitrary event which is the subset of E1, E2, E3----En.

Baye's Theorem:— It is given by British Mathematician Thomas Bayes in 1973.

If $E_1, E_2, E_3 --- E_n$ are mutually disjoint events with probability

$$P(E_i) \neq 0 \quad (i = 1, 2, 3 --- n)$$

Then for any arbitrary event A, which is a subset of $\bigcup_{i=1}^{n} E_i$; such that

$$P(A) > 0, \text{ we have}$$

$$P\left(\frac{E_i}{A}\right) = \frac{P(E_i) \cdot P(A/E_i)}{\sum_{i=1}^{n} P(E_i) \cdot P(A/E_i)}$$

$$= \frac{P(E_i)\, P(A/E_i)}{P(A)}$$

(i) $P(E_1), P(E_2) --- P(E_n)$ are called Prior Probabilities because they exists before we gitten any information from the experiment.

(ii) The Probability $P(A|E_i)$ where $i = 1, 2, 3 --- n$ are called likelyhood, because they indicate how likely the event A under candidate is to occur given each and every prior Probability.

(iii) $P(E_i/A)$: it is called {E_i given A} are called Posterior probabilities, bcs they are determination after the results of experiment are known.

Q. Consider Guwahati (G) and Delhi (D) whose temp can be classified as high (H), Medium (M) and low (L). Let $P(H_G)$ denote the probability that Guwahati has high temperature, similarly, $P(M_G)$ and $P(L_G)$ denotes the probability of Guwahati having Medium and low temperature respectively. Similarly we use $P(H_D)$, $P(M_D)$ and $P(L_D)$ for Delhi.

The following table gives the conditional prob. for Delhi's temperature given Guwahati temperature

	H_D	M_D	L_D
H_G	0.4	0.48	0.12
M_G	0.1	0.65	0.25
L_G	0.01	0.50	0.49

Consider the first row in table above. The first entry denotes that if Guwahati has high temp (H_G) then the prob of Delhi also having a high temp (H_D) is 0.40; i.e $P(H_D/H_G) = 0.40$

Similarly, the next two entries are $P(M_D/H_G) = 0.48$ and $P(L_D/H_G) = 0.12$; similarly for other row

If it is known as $P(H_G) = 0.2$; $P(M_G) = 0.5$ and $P(L_G) = 0.3$; Then the probability that Guwahati has high temperature given that Delhi has high temperature is---

Sol^n:-

$$P\left(\frac{H_G}{H_D}\right) = \frac{P(H_G \cap H_D)}{P(H_D)} \Rightarrow \frac{P(H_D/H_G) \cdot P(H_G)}{\sum_{i=1}^{3} P\left(\frac{H_D}{H_G}\right) \cdot P(H_G)}$$

where $P(H_D) = P\left(\frac{H_D}{H_G}\right)P(H_G) + P\left(\frac{H_D}{M_G}\right)P(M_G) + P\left(\frac{H_D}{L_G}\right)P(L_G)$

$$P\left(\frac{H_G}{H_D}\right) = \frac{0.4 * 0.2}{0.4*0.2 + 0.1*0.5 + 0.01*0.3} = \frac{0.08}{0.08 + 0.05 + 0.003}$$

$$\Rightarrow \frac{0.08}{0.133} \Rightarrow 0.60 \text{ Ans}$$

Ex:- in a certain Assembly plant, Three machines B_1, B_2 and B_3 make 30%, 45%, and 25%, respectively of The Products, it is known from Past experience that 2%, 3% and 2% of The Products made by each Machine, respectively are defective; Now suppose that a finished Product is randomly selected; what is The Probability that it is defective.

Sol^n:- Consider The following Events

A : Product is defective ~~made~~

B_1, B_2, B_3 are the products made by machine B_1, B_2, B_3, respectively; Applying the rule of elimination we can write

$$P(A) = P(B_1) \cdot P(A/B_1) + P(B_2) \cdot P(A/B_2) + P(B_3) \cdot P(A/B_3)$$

$$P(B_1) = \frac{30}{100} = 0.3$$

$$P(B_2) = \frac{45}{100} = 0.45$$

$$P(B_3) = \frac{25}{100} = 0.25$$

$$P(A/B_1) = \frac{P(A \cap B_1)}{P(B_1)} = 2/100 = 0.02$$

$$P(A/B_2) = 3/100 = \frac{P(A \cap B_2)}{P(B_2)} = 0.03$$

$$P(A/B_3) = \left(\frac{P(A \cap B_3)}{P(B_3)}\right) = 2/100 = 0.02$$

$P(A) = 0.3 \times 0.02 + 0.45 \times 0.03 + 0.25 \times 0.02$

$\Rightarrow 0.006 + 0.0135 + 0.005$

$\Rightarrow 0.0245$

* With reference to this example, if a product was chosen randomly and found to be defective, what is the probability that it was made by machine B_3.

Soln:- using Bayes Rule to write

$$P(B_3/A) = \frac{P(B_3) \cdot P(A/B_3)}{P(B_1)\,P(A/B_1) + P(B_2)\,P(A/B_2) + P(B_3) \cdot P(A/B_3)}$$

$$\Rightarrow \frac{0.25 \times 0.02}{0.0245}$$

$$\Rightarrow \frac{0.005}{0.0245} \Rightarrow 10/49 \quad \text{Ans}$$

Ex:- A factory Produces a certain type of outputs by Three types of machine. The respective daily Production figure are

Machine I: 3,000 Units;

Machine II: 2500 Units

Machine III: 4500 Units

Past experience shows that 1 Percent of The output Produced by machine I is defective. The corresponding fraction of defectives for The other Two machines are 1.2 Percent and 2 Percent respectively. An Item is drawn at random from The day's Production run and is found to be defective. What is The Probability that it comes from The output of

(i) Machine (I) (ii) Machine (II) (iii) Machine (III)

Solution:- Let E_1, E_2 and E_3 denote the events that the output is Produced by machine I, II and III respectively and let A denote the event that the output is defective. Then we have

$$P(E_1) = \frac{3000}{10,000} = 0.30$$

$$P(E_2) = \frac{2500}{10,000} = 0.25$$

$$P(E_3) = \frac{4500}{10,000} = 0.45$$

$P(A/E_1) = 1\% = 0.01;$

$P(A/E_2) = 1.2\% = 0.012$

$P(A/E_3) = \cancel{\frac{4500}{10,000}} = \cancel{0.45}\ 2\% = 0.02$

the Probability that an item selected at random from day's Production is defective is given by:

$$P(A) = \sum_{i=1}^{3} P(E_i \cap A) = \sum_{i=1}^{3} P(E_i) \cdot P(A/E_i)$$

$$= 0.30 \times 0.01 + 0.25 * 0.012 + 0.45 * 0.02$$

$$= 0.015$$

By Baye's rule, The required Probabilities are given by:

(i) $P\left(\frac{E_1}{A}\right) = \frac{P(E_1) \cdot P(A/E_1)}{P(A)} = \frac{0.003}{0.015} = \frac{1}{5}$

(ii) $P(E_2/A) = \frac{P(E_2) \cdot P(A/E_2)}{P(A)} = \frac{0.003}{0.015} = 1/5$

(iii) $P(E_3/A) = \frac{P(E_3) \cdot P(A/E_3)}{P(A)} = \frac{0.009}{0.015} = 3/5$

The Probabilies in (i), (ii) and (iii) are known as Posterior Probabilies of events E_1, E_2 and E_3 Respectively.

Bayes classification Methods :→

Bayesian classifiers are statistical classifiers. They can predict class membership probabilities such as the probability that a given tuple belongs to a particular class.

Bayesian classification is based on Bayes' Theorem, described next. Studies comparing classification algorithms have found a simple Bayesian classifier known as the naive Bayesian classifier to be comparable in performance with decision Tree and selected neural network classifiers. Bayesian classifiers have also exhibited high accuracy and speed when applied to a large databases.

Naive Bayesian classifiers assume that the effect of an ~~algorithm~~ attribute value on a given class is independent of the values of the other attributes. This assumption is called class conditional independence. It is made to simplify the computations involved, and, in this ~~sense~~ sense, is considered "Naive".

Bayes' Theorem :—

Bayes' Theorem is named after Thomas Bayes; ~~a man~~ let X be a data tuple. In Bayesian terms, X is considered "evidence." As usual, it is ~~considered~~ described by measurements made on a set of n attributes. Let H be some hypothesis such as that the data tuple X belongs to a specified class C. For classification problems, we want to determine $P(H/X)$, The probability that the hypothesis H holds given the "evidence" or observed tuple X. In other words, we are lookip for The Prob. that tuple X belongs to class C, given That we know the attribute description of X.

$P(H/X)$ is the Posterior Probability, ~~of H~~ conditioned on X. For Example, suppose our world of data tuples is confined to customers described by the attribute age and Income, respectively, and that X is a 35-year-old customer with an income of 40,000\$. suppose that H is the hypothesis that our customer will buy a computer. Then $P(H/X)$ reflects The probability that customer X will buy a computer given that we know The customer's age and income.

In contrast, $P(H)$ is the prior probability, ~~of~~ of H. for our example, this is the Probability that any given customer will buy a computer, regardless of age, income, or any other Information, for That matter. The Posterior Prob, $P(H/X)$, is based on more Information (eg customer information) Than the Prior Probability, $P(H)$, which is Independent of X.

Similarly, $P(X/H)$ is the Posterior Probability of X conditioned on H. That is, it is the Probability that a customer, X, is 35 years old and earns 40000\$, given that we know the customer will buy a Computer.

$P(X)$ is The prior Probability of X. using our example, it is the Probability that a Person from our set of customers is 35 years old and earns \$40,000.

"How are these probabilities estimated?

$P(H)$, $P(X/H)$, and $P(X)$ may be estimated from the given data, as we shall see next. Bayes' theorem is useful in that it provides a way of calculating the posterior probability, $P(H/X)$, from $P(H)$, $P(X/H)$, and $P(X)$.

Bayes' theorem is

$$P(H/X) = \frac{P(X/H)\,P(H)}{P(X)} \Rightarrow \text{Posterior} = \frac{\text{likelihood} \times \text{Prior}}{\text{Evidence}}$$

Now that we will proceed for Example and Definition:→

Naive Bayesian classification:→ The Naive Bayesian classifier, or simple Bayesian classifier, works as follows:

(1) Let D be a training set of tuples and their associated class labels. As usual, each tuple is represented by an n-dimensional attribute vector, $X = (x_1, x_2, \ldots x_n)$, depicting n measurements made on the tuple from n attributes, respectively, $A_1, A_2, \ldots A_n$.

(2) Suppose that there are m classes, $C_1, C_2, \ldots C_m$. Given a tuple X, the classifier will predict that X belongs to the class having the highest posterior probability, conditioned on X. That is, the naive Baysian classifier predicts that tuple X belongs to the class C_i if and only if

$$P(C_i/X) > P(C_j/X) \quad \text{for } 1 \le j \le m,\ j \ne i$$

Thus, we maximize $P(C_i/X)$. The class C_i for which $P(C_i/X)$ is maximized is called the maximum posteriori hypothesis. By Bayes' theorem

$$P(C_i/X) = \frac{P(X/C_i)P(C_i)}{P(X)}$$

(3) As $P(X)$ is constant for all classes, only $P(X/C_i)P(C_i)$ needs to be maximized. If the class prior probabilities are not known, then it is commonly assumed that the classes are equally likely, that is $P(C_1)=P(C_2)=$ --- $P(C_m)$, and we would therefore maximize $P(X/C_i)$. Otherwise, we maximize $P(X/C_i)P(C_i)$. Note that class prior probabilities may be estimated by $P(C_i) = |C_{i,D}| / |D|$, where $|C_{i,D}|$ is the number of training tuples of class C_i in D.

(4) Given data sets with many attributes, it would be extremely computationally expensive to compute $P(X/C_i)$. To reduce computation in evaluating $P(X/C_i)$, the naive assumption of class-conditional independence is made. This presumes that the attributes' values are conditionally independent of one another, given the class label of the tuple (i.e that there are no dependence relationships among the attributes). Thus

$$P(X/C_i) = \prod_{k=1}^{n} P(x_k/C_i)$$

$$= P(x_1/C_i) \times P(x_2/C_i) \times \cdots \times P(x_n/C_i)$$

we can easily estimate the Probability of $P(x_1/C_i), P(x_2/C_i), --- P(x_n/C_i)$ from the traning tuples. Recall that here x_k refers to the value of attribute A_k for tuple X. For each attribute, we look at whether the ~~value~~ attribute is categorical or continuous - valued.

For Instance, to compute $P(X/C_i)$, we consider the following: If A_k is categorical, then $P(x_k/C_i)$ is the number of tuples of class C_i in D having x_k for A_k, divided by $|C_{i,D}|$, the number of tuples of class C_i in D.

(5) To Predict The class label of X, $P(X/C_i)P(C_i)$ is evaluated for each class C_i. The classifier Predicts that the class label of tuple X is the class C_i if and only if

$$P(X/C_i)P(C_i) > P(X/C_j)P(C_j) \quad \text{for } 1 \le j \le m, j \ne i$$

In other words, the Predicted class label is the class C_i for which $P(X/C_i)P(C_i)$ is the maximum.

"How to Predict a class label using naive Bayesian classification"! →

Ex:- we wish to predict the class label of a tuple using naive Bayesian classification, given the same training data as in Example 8.3 for decision tree induction

Class-labeled Training tuples from The All electronics Customer Database

RID	age	Income	Student	credit-rating	class: Buys-Computer
1	youth	high	no	fair	no
2	youth	high	no	excellent	no
3	middle-aged	high	no	fair	yes
4	Senior	medium	no	fair	yes
5	Senior	low	yes	fair	yes
6	Senior	low	yes	excellent	no
7	middle-aged	low	yes	excellent	yes
8	youth	medium	no	fair	no
9	youth	low	yes	fair	yes
10	senior	medium	yes	fair	yes
11	youth	medium	yes	excellent	yes
12	middle-aged	medium	no	excellent	yes
13	middle-aged	high	yes	fair	yes
14	senior	Medium	no	excellent	no

The data tuples are described by the attribute age, Income, student, and credit-rating. The class label attribute, buys-computer has two distinct values (namely, {yes, no}). Let C_1 correspond to the class buys-computer = yes and C_2 correspond to buys-computer = no. The tuple we wish to classify is

X = (age = youth, income = Medium, student = yes, credit-rating = fair) we need to maximize $P(X/C_i)$, for $i=1,2$, $P(C_i)$, The Prior Probability of each class, can be computed based on The Training tuples:

$P(\text{buys\_computer} = \text{yes}) = 9/14 = 0.643$
$P(\text{buys\_computer} = \text{no}) = 5/14 = 0.357$

To compute $P(X/C_i)$, for $i=1,2$, we compute the following conditional probabilities:

$P(\text{age} = \text{youth} \mid \text{buys computer} = \text{yes}) = 2/9 = 0.222$
$P(\text{age} = \text{youth} \mid \text{buys\_computer} = \text{no}) = 3/5 = 0.6$
$P(\text{income} = \text{Medium} \mid \text{buys computer} = \text{yes}) = 4/9 = 0.444$
$P(\text{Income} = \text{Medium} \mid \text{buys\_computer} = \text{no}) = 2/5 = 0.4$
$P(\text{Student} = \text{yes} \mid \text{buys\_computer} = \text{yes}) = 6/9 = 0.667$
$P(\text{Student} = \text{yes} \mid \text{buys\_computer} = \text{no}) = 1/5 = 0.200$
$P(\text{credit\_rating} = \text{fair} \mid \text{buys\_computer} = \text{yes}) = 6/9 = 0.667$
$P(\text{credit\_rating} = \text{fair} \mid \text{buys\_computer} = \text{no}) = 2/5 = 0.400$

using These Probabilities, we obtain

$P(X \mid \text{buys\_computer} = \text{yes}) = P(\text{age} = \text{youth} \mid \text{buys\_computer} = \text{yes})$
$\times P(\text{Income} = \text{Medium} \mid \text{buy\_computer} = \text{yes}) \times P(\text{Student} = \text{yes} \mid$
$\text{buy\_computer} = \text{yes}) \times P(\text{Credit\_Rating} = \text{fair} \mid \text{buy\_computer} = \text{yes})$

$= 0.222 \times 0.444 \times 0.667 \times 0.667 = 0.044$

Similarly,

$P(X \mid \text{buy\_computer} = \text{no}) = 0.600 \times 0.400 \times 0.2 \times 0.4 = 0.019$

To find the class, C_i, that maximizes $P(X/C_i)P(C_i)$, we compute

$$P(X \mid buys\text{-}computer = yes)\ P(buys\text{-}computer = yes)$$

$$= 0.044 \times 0.643 = 0.028$$

$$P(X \mid buys\text{-}computer = no)\ P(buys\text{-}computer = no)$$

$$= 0.015 \times 0.357$$

$$= 0.007$$

Therefore, The naive Bayesian classifier predicts buys-computer = yes for tuple X.

What if I encounter probability value of zero? We need to compute $P(X/C_i)$ for each class $(i=1,2,\ldots m)$ to find the class C_i for which $P(X/C_i)P(C_i)$ is the maximum.

Data for height classification :-

Name	Gender	height	output 1	output 2
Kristina	F	1.6 M	Short	Medium
JIM	M	2 M	tall	Medium
maggie	F	1.9 M	Medium	tall
Martha	F	1.88m	Medium	tall
stephanie	F	1.7 M	Short	Medium
Bob	M	1.85m	M	M
Kathy	F	1.6 m	S	M
Dave	M	1.7 m	S	M
Worth	M	2.2 m	t	T
Steven	M	2.1 m	tall	T
Debbie	F	1.8 m	M	M
todd	M	1.95m	m	M
Kim	F	1.9 M	M	T
Amy	F	1.8 m	m	M
Wynette	F	1.75 M	M	M

Ex:- Using the output 1 classification results for table. There are four tuple classified as Short, eight as Medium, and three as tall. TO facilitate classification. we devide the height attribute into six ranges

(0, 1.6), (1.6, 1.7), (1.7, 1.8), (1.8, 1.9), (1.9, 2.0), (2.0, ∞)

table shows the count and subsequent probabilities associated with the attributes, with these training data, we estimates the prior probabilities.

$$P(\text{short}) = \frac{4}{15} = 0.267$$

$$P(\text{medium}) = 8/15 = 0.533$$

and $P(\text{tall}) = 3/15 = 0.2$

we use these values to classify a new tuple.

for Example:- suppose we wish to classify

$t = (\text{adam}, M, 1.95m)$. By using these values and the associated probabilities of gender and height, we obtain the following estimates:

Conditional Prob $$P(A|B) = \frac{n(A \cap B)}{n(B)}$$

so $$P(t/\text{short}) = \frac{P(t \cap \text{short})}{P(\text{Short})} = \frac{\frac{1}{4} \times 0}{1/4} = 0$$

$$P(t/\text{medium}) = \frac{2/8 \times 1/8}{2/8} \to 0.125 = \frac{P(t \cap \text{med})}{t(\text{med})}$$

$$P(t/\text{tall}) = \frac{3/3 \times 1/3}{3/3} = 0.333 = \frac{P(t \cap \text{tall})}{t(\text{tall})}$$

Combining these, we get

Likelihood of being short $= 0 \times 0.267 = 0$

Likelihood of being Medium $= 0.125 \times 0.533 = 0.066$

Likelihood of being tall $= 0.33 \times 0.2 = 0.066$

we estimates $P(t)$ by summing up these individual likelihood values since t will be either short or Medium or tall

$$P(t) = 0 + 0.0166 + 0.066$$

$$\Rightarrow 0.0826$$

finally, we obtain the actual prob. of each event

$$P\left(\frac{short}{t}\right) = \frac{P(short \cap t)}{P(t)} = \frac{P(short) . P(t/short)}{P(t)}$$

where $P(t) = P(short) . P(t/short) + P(med) . P(t/med) + P(tall) . P(t/tall)$

$$P\left(\frac{short}{t}\right) = \frac{0.267 \times 0}{0.0826} = 0$$

$$P\left(\frac{medium}{t}\right) = \frac{0.533 \times 0.125}{0.0826} \Rightarrow 0.806$$

$$P(tall/t) = \frac{0.2 * 0.333}{0.0826} = 0.807$$

Therefore based on these prob. we classify the new tuple as tall because it has the highest probability table:- prob. associated with attributes

Attribute	Value	Count			Probabilities		
		Short	Medium	tall	Short	Medium	tall
Gender	M	1	2	3	1/4	2/8	3/3
	F	3	6	0	3/4	6/8	0/3
Height	(0,1.6)	2	0	0	2/4	0	0
	(1.6,1.7)	2	0	0	2/4	0	0
	(1.7,1.8)	0	3	0	0	3/8	0
	(1.8,1.9)	0	4	0	0	4/8	0
	(1.9,2)	0	1	1	0	1/8	1/3
	(2,∞)	0	0	2	0	0	2/3

Ex:- Classify the tuple X= {Color ='Red', type ='SUV', origin = 'Domestic'} using NB classification. Training data is given in the following table where class label is {Stolen}

Color	Type	origin	Stolen
Red	Sports	Domestic	yes
Red	Sports	Domestic	No
Red	Sports	Domestic	yes
yellow	Sports	Domestic	No
yellow	Sports	Imported	yes
yellow	SUV	Imported	No
yellow	SUV	Imported	yes
yellow	SUV	Domestic	No
Red	SUV	Imported	No
Red	Sports	Imported	Yes.