Bayes' Theorem

- Bayes' theorem is also known as Bayes' Rule or Bayes' law, which is used to determine the probability of a hypothesis with prior knowledge.
- It depends on the conditional probability.

The formula for Bayes' theorem is given as:

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

Where,

P(A|B) is Posterior probability: Probability of hypothesis A on the observed event B.

P(B|A) is Likelihood probability: Probability of the evidence given that the probability of a hypothesis is true.

P(A) is Prior Probability: Probability of hypothesis before observing the evidence.

P(B) is Marginal Probability: Probability of Evidence.

- Machine Learning is one of the most emerging technology of Artificial Intelligence.
- We are living in the 21th century which is completely driven by new technologies and gadgets in which some are yet to be used and few are on its full potential.
- Similarly, Machine Learning is also a technology that is still in its developing phase.
- There are lots of concepts that make machine learning a better technology such as supervised learning, unsupervised learning, reinforcement learning, perceptron models, Neural networks, etc.

- In this article "Bayes Theorem in Machine Learning", we will discuss another most important concept of Machine Learning theorem i.e., Bayes Theorem.
- But before starting this topic you should have essential understanding of this theorem such as what exactly is Bayes theorem, why it is used in Machine Learning, examples of Bayes theorem in Machine Learning and much more.
- So, let's start the brief introduction of Bayes theorem.

Introduction to Bayes Theorem in Machine Learning

- Bayes theorem is given by an English statistician, philosopher, and Presbyterian minister named Mr. Thomas Bayes in 17^{th} century.
- Bayes provides their thoughts in decision theory which is extensively used in important mathematics concepts as Probability.
- Bayes theorem is also widely used in Machine Learning where we need to predict classes precisely and accurately.
- An important concept of Bayes theorem named Bayesian method is used to calculate conditional probability in Machine Learning application that includes classification tasks.
- Further, a simplified version of Bayes theorem (Naïve Bayes classification) is also used to reduce computation time and average cost of the projects.
- Bayes theorem is also known with some other name such as Bayes rule or Bayes Law.
- Bayes theorem helps to determine the probability of an event with random knowledge.
- It is used to calculate the probability of occurring one event while other one already occurred.
- It is a best method to relate the condition probability and marginal probability.
- Bayes Theorem is used to estimate the precision of values and provides a method for calculating the conditional probability.
- However, it is hypocritically a simple calculation but it is used to easily calculate the conditional probability of events where intuition often fails.
- Some of the data scientist assumes that Bayes theorem is most widely used in financial industries but it is not like that.
- Other than financial, Bayes theorem is also extensively applied in health and medical, research and survey industry, aeronautical sector, etc.

What is Bayes Theorem?

- Bayes theorem is one of the most popular machine learning concepts that helps to calculate the probability of occurring one event with uncertain knowledge while other one has already occurred.

Bayes' theorem can be derived using product rule and conditional probability of event X with known event Y:

- According to the product rule we can express as the probability of event X with known event Y as follows;

1. P(X ? Y)= P(X|Y) P(Y)   {equation 1}

    - Further, the probability of event Y with known event X:

1. P(X ? Y)= P(Y|X) P(X)   {equation 2}

Mathematically, Bayes theorem can be expressed by combining both equations on right hand side. We will get:

$$P(X|Y) = \frac{P(Y|X).P(X)}{P(Y)}$$

Here, both events X and Y are independent events which means probability of outcome of both events does not depends one another.

The above equation is called as Bayes Rule or Bayes Theorem.

- P(X|Y) is called as posterior, which we need to calculate. It is defined as updated probability after considering the evidence.
- P(Y|X) is called the likelihood. It is the probability of evidence when hypothesis is true.

- P(X) is called the prior probability, probability of hypothesis before considering the evidence
- P(Y) is called marginal probability. It is defined as the probability of evidence under any consideration.

Hence, Bayes Theorem can be written as:

posterior = likelihood * prior / evidence

Prerequisites for Bayes Theorem

- While studying the Bayes theorem, we need to understand few important concepts.

These are as follows:

1. Experiment

- An experiment is defined as the planned operation carried out under controlled condition such as tossing a coin, drawing a card and rolling a dice, etc.

2. Sample Space

- During an experiment what we get as a result is called as possible outcomes and the set of all possible outcome of an event is known as sample space.

  For example, if we are rolling a dice, sample space will be:

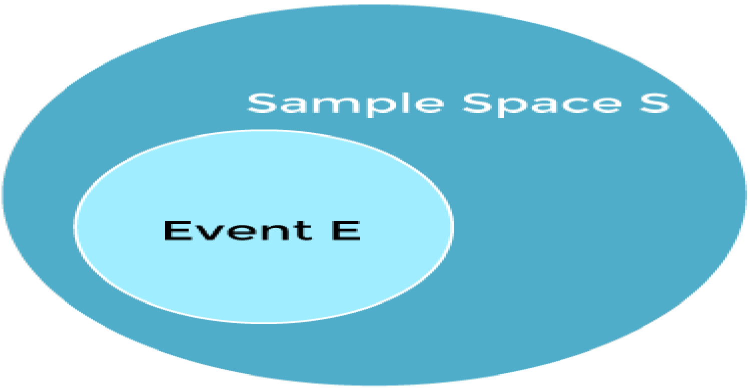
  S1 = {1, 2, 3, 4, 5, 6}

  Similarly, if our experiment is related to toss a coin and recording its outcomes, then sample space will be:

  S2 = {Head, Tail}

3. Event

- Event is defined as subset of sample space in an experiment.
- Further, it is also called as set of outcomes.



Assume in our experiment of rolling a dice, there are two event A and B such that;

A = Event when an even number is obtained = {2, 4, 6}

B = Event when a number is greater than 4 = {5, 6}

- Probability of the event A "P(A)"= Number of favourable outcomes / Total number of possible outcomes

  P(E) = 3/6 =1/2 =0.5
- Similarly, Probability of the event B "P(B)"= Number of favourable outcomes / Total number of possible outcomes

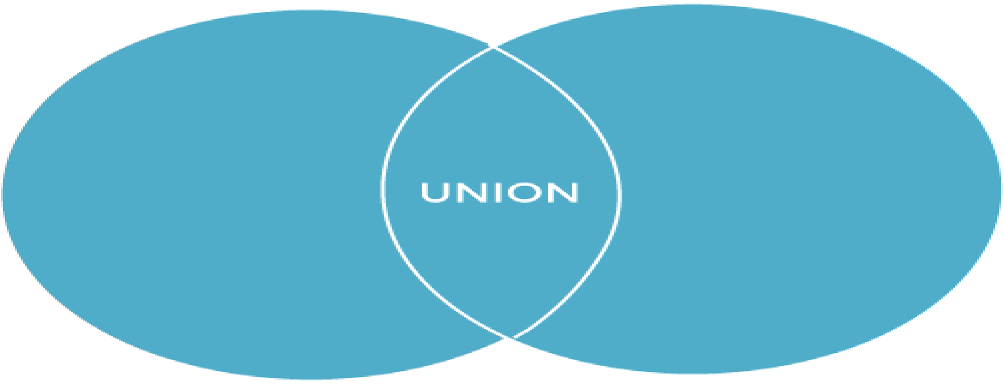
  =2/6

  =1/3

  =0.333

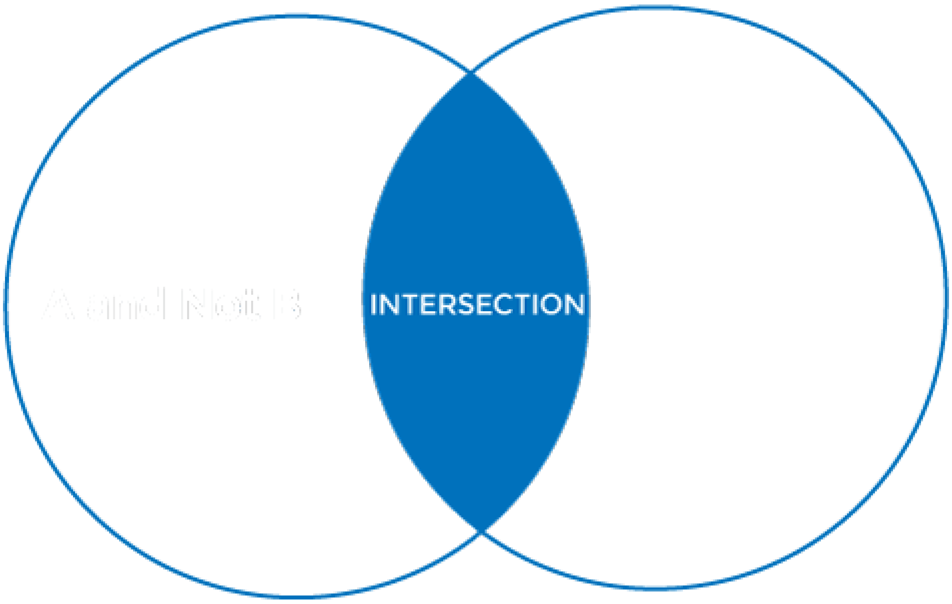
Union of event A and B:

A∪B = {2, 4, 5, 6}

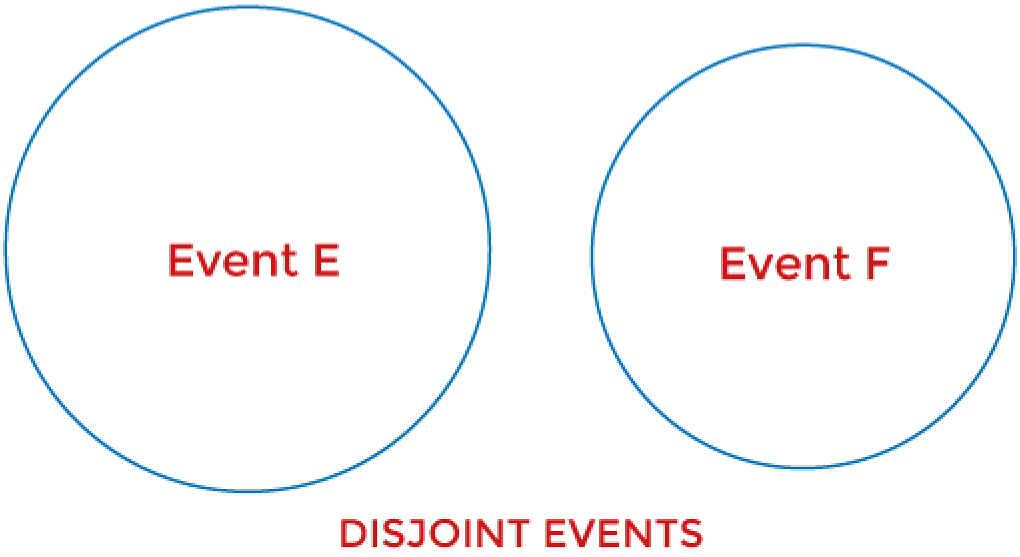


Intersection of event A and B:

A∩B= {6}



- Disjoint Event: If the intersection of the event A and B is an empty set or null then such eventsare known as disjoint event or mutually exclusive events also.



4. Random Variable:

- It is a real value function which helps mapping between sample space and a real line of an experiment.
- A random variable is taken on some random values and each value having some probability.
- However, it is neither random nor a variable but it behaves as a function which can either be discrete, continuous or combination of both.

5. Exhaustive Event:

- As per the name suggests, a set of events where at least one event occurs at a time, called exhaustive event of an experiment.
- Thus, two events A and B are said to be exhaustive if either A or B definitely occur at a time and both are mutually exclusive for e.g., while tossing a coin, either it will be a Head or may be a Tail.

6. Independent Event:

- Two events are said to be independent when occurrence of one event does not affect the occurrence of another event.

- In simple words we can say that the probability of outcome of both events does not depends one another.

Mathematically, two events A and B are said to be independent if:

$P(A \cap B) = P(AB) = P(A)*P(B)$

7. Conditional Probability:

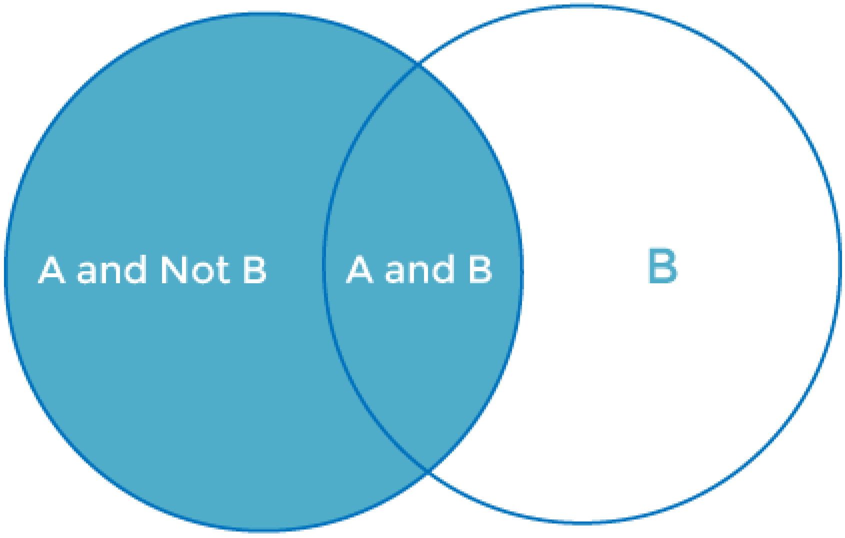
- Conditional probability is defined as the probability of an event A, given that another event B has already occurred (i.e. A conditional B). This is represented by P(A|B) and we can define it as:

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

8. Marginal Probability:

- Marginal probability is defined as the probability of an event A occurring independent of any other event B.
- Further, it is considered as the probability of evidence under any consideration.

$$P(A) = P(A|B)*P(B) + P(A|{\sim}B)*P({\sim}B)$$



Here ~B represents the event that B does not occur.