Machine Learning (21CSC305P)

Unit-1 Notes

What is Machine Learning?

Machine learning (ML) is a discipline of artificial intelligence (AI) that provides machines with the ability to automatically learn from data and past experiences while identifying patterns to make predictions with minimal human intervention.

Machine learning methods enable computers to operate autonomously without explicit programming. ML applications are fed with new data, and they can independently learn, grow, develop and adapt. Machine learning derives insightful information from large volumes of data by leveraging algorithms to identify patterns and learn in an iterative process. ML algorithms use computation methods to learn directly from data instead of relying on any predetermined equation that may serve as a model.

The performance of ML algorithms adaptively improves with an increase in the number of available samples during the 'learning' processes. For example, deep learning is a sub-domain of machine learning that trains computers to imitate natural human traits like learning from examples. It offers better performance parameters than conventional ML algorithms.

Today, with the rise of big data, IoT and ubiquitous computing, machine learning has become essential for solving problems across numerous areas, such as:

- Computational finance (credit scoring, algorithmic trading)
- Computer vision (facial recognition, motion tracking, object detection)
- Computational biology (DNA sequencing, brain tumour detection, drug discovery)
- Automotive, aerospace and manufacturing (predictive maintenance)
- Natural language processing (voice recognition)

How does machine learning work?

There are some steps you would follow when creating a machine learning model.

Choose and prepare a training data set

Training data is information that is representative of the data the machine learning application will ingest to tune model parameters. Training data is sometimes labeled, meaning it has been tagged to call out classifications or expected values the machine learning mode is required to predict. Other training data may be unlabeled so the model will have to extract features and assign clusters autonomously. For labeled, data should be divided into a training subset and a testing subset. The former is used to train the model and the latter to evaluate the effectiveness of the model and find ways to improve it.

Select an algorithm to apply to the training data set

The type of machine learning algorithm you choose will primarily depend on a few aspects:

- Whether the use case is prediction of a value or classification which uses labeled training data or the use case is clustering or dimensionality reduction which uses unlabeled training data
- How much data is in the training set
- The nature of the problem the model seeks to solve

For prediction or classification use cases, you would usually use regression algorithms such as ordinary least square regression or logistic regression. With unlabeled data, you are likely to rely on clustering algorithms such as k-means or nearest neighbor. Some algorithms like neural networks can be configured to work with both clustering and prediction use cases.

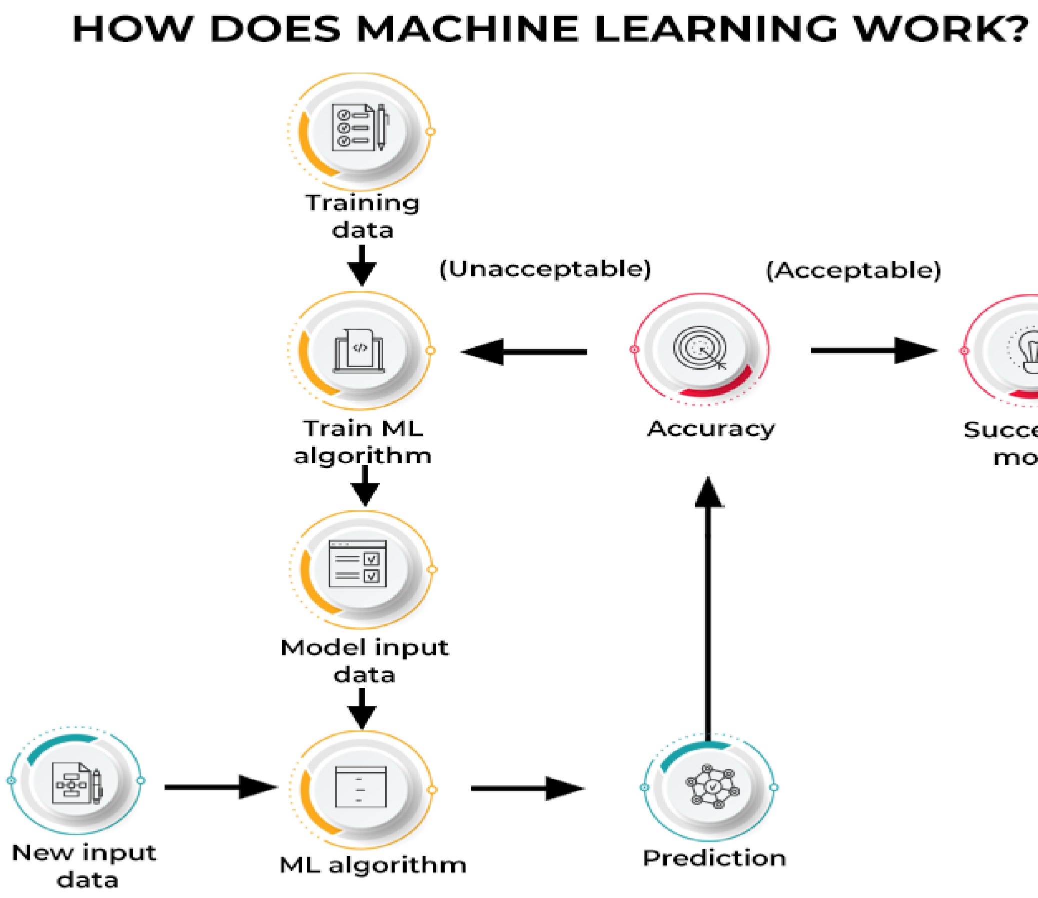
Train the algorithm to build the model

Training the algorithm is the process of tuning model variables and parameters to more accurately predict the appropriate results. Training the machine learning algorithm is usually iterative and uses a variety of optimization methods depending upon the chosen model. These optimization methods do not require human intervention which is part of the power of machine learning. The machine learns from the data you give it with little to no specific direction from the user.

Use and improve the model

The last step is to feed new data to the model as a means of improving its effectiveness and accuracy over time. Where the new information will come from depends on the nature of the problem to be solved. For instance, a machine learning model for self-driving cars will ingest real-world information on road conditions, objects and traffic laws.

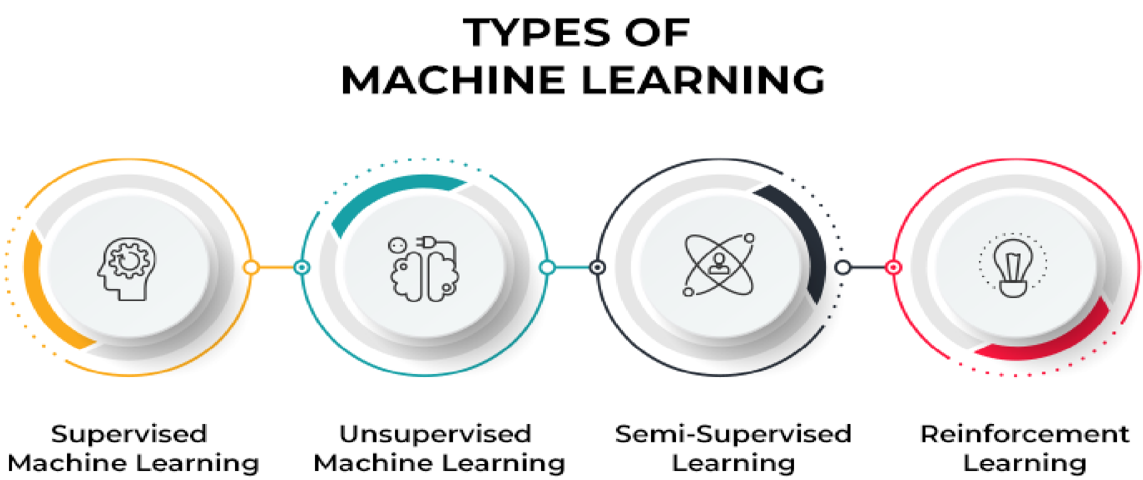
Machine learning algorithms are molded on a training dataset to create a model. As new input data is introduced to the trained ML algorithm, it uses the developed model to make a prediction.



Further, the prediction is checked for accuracy. Based on its accuracy, the ML algorithm is either deployed or trained repeatedly with an augmented training dataset until the desired accuracy is achieved.

Types of Machine Learning

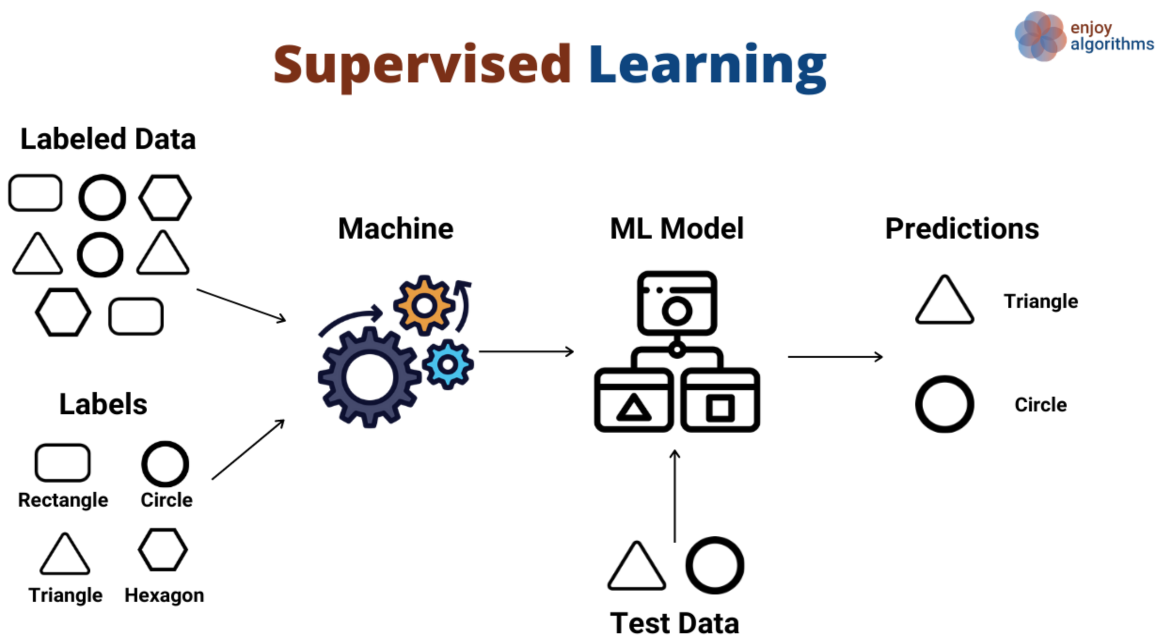
Machine learning algorithms can be trained in many ways, with each method having its pros and cons. Based on these methods and ways of learning, machine learning is broadly categorized into four main types:



1. Supervised machine learning

This type of ML involves supervision, where machines are trained on labeled datasets and enabled to predict outputs based on the provided training. The labeled dataset specifies that some input and output parameters are already mapped. Hence, the machine is trained with the input and corresponding output. A device is made to predict the outcome using the test dataset in subsequent phases.

For example, consider an input dataset of parrot and crow images. Initially, the machine is trained to understand the pictures, including the parrot and crow's color, eyes, shape, and size. Post-training, an input picture of a parrot is provided, and the machine is expected to identify the object and predict the output. The trained machine checks for the various features of the object, such as colour, eyes, shape etc., in the input picture, to make a final prediction. This is the process of object identification in supervised machine learning.



The primary objective of the supervised learning technique is to map the input variable (a) with the output variable (b). Supervised machine learning is further classified into two broad categories:

Classification: These refer to algorithms that address classification problems where the output variable is categorical; for example, yes or no, true or false, male or female, etc. Real-world applications of this category are evident in spam detection and email filtering. Some known classification algorithms include the Random Forest Algorithm, Decision Tree Algorithm, Logistic Regression Algorithm and Support Vector Machine Algorithm.

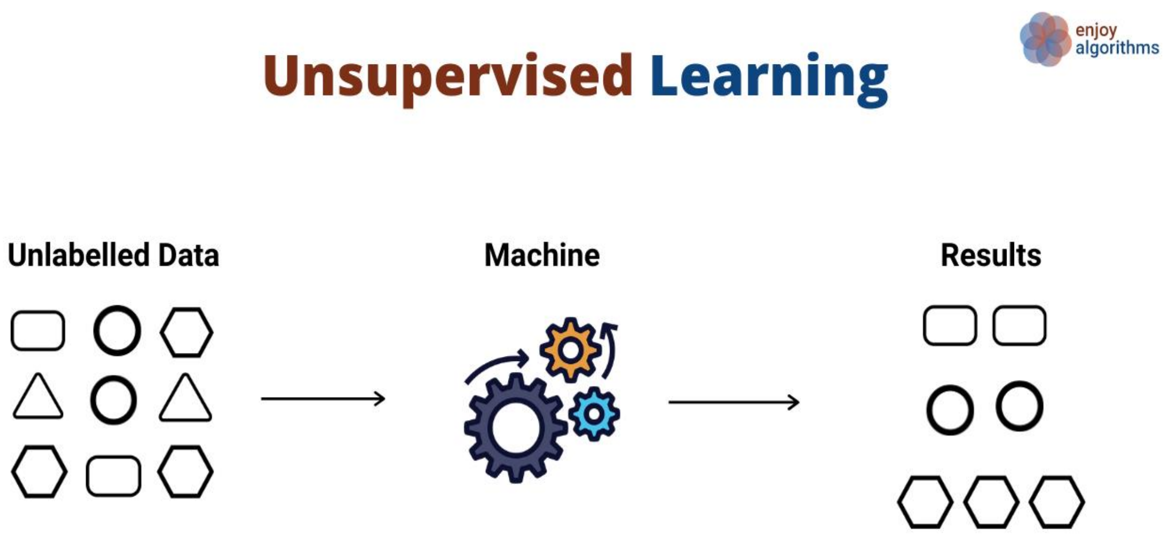
Regression: Regression algorithms handle regression problems where input and output variables have a linear relationship. These are known to predict continuous output variables.

Examples include weather prediction, market trend analysis, etc. Popular regression algorithms include the Simple Linear Regression Algorithm, Multivariate Regression Algorithm, Decision Tree Algorithm, and Lasso Regression.

2. Unsupervised machine learning

Unsupervised learning refers to a learning technique that's devoid of supervision. Here, the machine is trained using an unlabeled dataset and is enabled to predict the output without any supervision. An unsupervised learning algorithm aims to group the unsorted dataset based on the input's similarities, differences, and patterns.

For example, consider an input dataset of images of a fruit-filled container. Here, the images are not known to the machine learning model. When we input the dataset into the ML model, the task of the model is to identify the pattern of objects, such as color, shape, or differences seen in the input images and categorize them. Upon categorization, the machine then predicts the output as it gets tested with a test dataset.



Unsupervised machine learning is further classified into two types:

Clustering: The clustering technique refers to grouping objects into clusters based on parameters such as similarities or differences between objects. For example, grouping customers by the products they purchase. Some known clustering algorithms include the K-Means Clustering Algorithm, Mean-Shift Algorithm, DBSCAN Algorithm, Principal Component Analysis, and Independent Component Analysis.

Association: Association learning refers to identifying typical relations between the variables of a large dataset. It determines the dependency of various data items and maps associated variables. Typical applications include web usage mining and market data analysis. Popular algorithms obeying association rules include the Apriori Algorithm, Eclat Algorithm, and FP-Growth Algorithm.

3. Semi-supervised learning

Semi-supervised learning comprises characteristics of both supervised and unsupervised machine learning. It uses the combination of labeled and unlabeled datasets to train its algorithms. Using both types of datasets, semi-supervised learning overcomes the drawbacks of the options mentioned above.

Consider an example of a college student. A student learning a concept under a teacher's supervision in college is termed supervised learning. In unsupervised learning, a student self-learns the same concept at home without a teacher's guidance. Meanwhile, a student revising the concept after learning under the direction of a teacher in college is a semi-supervised form of learning.

4. Reinforcement learning

Reinforcement learning is a feedback-based process. Here, the AI component automatically takes stock of its surroundings by the hit & trial method, takes action, learns from experiences, and improves performance. The component is rewarded for each good action and penalized for every wrong move. Thus, the reinforcement learning component aims to maximize the rewards by performing good actions.

Unlike supervised learning, reinforcement learning lacks labeled data, and the agents learn via experiences only. Consider video games. Here, the game specifies the environment, and each move of the reinforcement agent defines its state. The agent is entitled to receive feedback via punishment and rewards, thereby affecting the overall game score. The ultimate goal of the agent is to achieve a high score. Reinforcement learning is applied across different fields such as game theory, information theory, and multi-agent systems. Reinforcement learning is further divided into two types of methods or algorithms:

Positive reinforcement learning: This refers to adding a reinforcing stimulus after a specific behavior of the agent, which makes it more likely that the behavior may occur again in the future, e.g., adding a reward after a behavior.

Negative reinforcement learning: Negative reinforcement learning refers to strengthening a specific behavior that avoids a negative outcome.

What is the difference between supervised and unsupervised machine learning?

Aspect	Supervised learning	Unsupervised learning
Process	Input and output variables are provided to train model.	Only input data is provided to train model. No output data is used.
Input data	Uses labeled data.	Uses unlabeled data.
Algorithms supported	Supports regression algorithms, instance-based algorithms, classification algorithms, neural networks and decision trees.	Supports clustering algorithms, association algorithms and neural networks.
Complexity	Simpler.	More complex.
Subjectivity	Objective.	Subjective.
Number of classes	Number of classes is known.	Number of classes is unknown.
Primary drawback	Classifying massive data with supervised learning is difficult.	Choosing number of clusters can be subjective.
Primary goal	Train the model to predict output when presented with new inputs.	Find useful insights and hidden patterns.

Machine learning versus deep learning versus neural networks

Since deep learning and machine learning tend to be used interchangeably, it's worth noting the nuances between the two. Machine learning, deep learning, and neural networks are all sub-fields of artificial intelligence. However, neural networks are actually a sub-field of machine learning, and deep learning is a sub-field of neural networks.

The way in which deep learning and machine learning differ is in how each algorithm learns. "Deep" machine learning can use labeled datasets, also known as supervised learning, to inform its algorithm, but it doesn't necessarily require a labeled dataset. The deep learning process can ingest unstructured data in its raw form (e.g., text or images), and it can automatically determine the set of features which distinguish different categories of data from one another. This eliminates some of the human intervention required and enables the use of large amounts of data. Classical, or "non-deep," machine learning is more dependent on human intervention to learn. Human experts determine the set of features to understand the differences between data inputs, usually requiring more structured data to learn.

Neural networks, or artificial neural networks (ANNs), are comprised of node layers, containing an input layer, one or more hidden layers, and an output layer. Each node, or artificial neuron, connects to another and has an associated weight and threshold. If the output of any individual node is above the specified threshold value, that node is activated, sending data to the next layer of the network. Otherwise, no data is passed along to the next layer of the network by that node. The “deep” in deep learning is just referring to the number of layers in a neural network. A neural network that consists of more than three layers—which would be inclusive of the input and the output—can be considered a deep learning algorithm or a deep neural network. A neural network that only has three layers is just a basic neural network. Deep learning and neural networks are credited with accelerating progress in areas such as computer vision, natural language processing, and speech recognition.

Top 5 Machine Learning Applications

Industry verticals handling large amounts of data have realized the significance and value of machine learning technology. As machine learning derives insights from data in real-time, organizations using it can work efficiently and gain an edge over their competitors. Every industry vertical in this fast-paced digital world, benefits immensely from machine learning tech. Here, we look at the top five ML application sectors.

1. Healthcare industry

Machine learning is being increasingly adopted in the healthcare industry, credit to wearable devices and sensors such as wearable fitness trackers, smart health watches, etc. All such devices monitor users’ health data to assess their health in real-time. Moreover, the technology is helping medical practitioners in analyzing trends or flagging events that may help in improved patient diagnoses and treatment. ML algorithms even allow medical experts to predict the lifespan of a patient suffering from a fatal disease with increasing accuracy.

Additionally, machine learning is contributing significantly to two areas:

Drug discovery: Manufacturing or discovering a new drug is expensive and involves a lengthy process. Machine learning helps speed up the steps involved in such a multi-step process. For example, Pfizer uses IBM’s Watson to analyze massive volumes of disparate data for drug discovery.

Personalized treatment: Drug manufacturers face the stiff challenge of validating the effectiveness of a specific drug on a large mass of the population. This is because the drug works only on a small group in clinical trials and possibly causes side effects on some subjects.

To address these issues, companies like Genentech have collaborated with GNS Healthcare to leverage machine learning and simulation AI platforms, innovating biomedical treatments to address these issues. ML technology looks for patients' response markers by analyzing individual genes, which provides targeted therapies to patients.

2. Finance sector

Today, several financial organizations and banks use machine learning technology to tackle fraudulent activities and draw essential insights from vast volumes of data. ML-derived insights aid in identifying investment opportunities that allow investors to decide when to trade. Moreover, data mining methods help cyber-surveillance systems zero in on warning signs of fraudulent activities, subsequently neutralizing them. Several financial institutes have already partnered with tech companies to leverage the benefits of machine learning.

For example,

Citibank has partnered with fraud detection company Feedzai to handle online and in-person banking frauds. PayPal uses several machine learning tools to differentiate between legitimate and fraudulent transactions between buyers and sellers.

3. Retail sector

Retail websites extensively use machine learning to recommend items based on users' purchase history. Retailers use ML techniques to capture data, analyze it, and deliver personalized shopping experiences to their customers. They also implement ML for marketing campaigns, customer insights, customer merchandise planning, and price optimization. According to a September 2021 report by Grand View Research, Inc., the global recommendation engine market is expected to reach a valuation of $17.30 billion by 2028. Common day-to-day examples of recommendation systems include:

When you browse items on Amazon, the product recommendations that you see on the homepage result from machine learning algorithms. Amazon uses artificial neural networks (ANN) to offer intelligent, personalized recommendations relevant to customers based on their recent purchase history, comments, bookmarks, and other online activities. Netflix and YouTube rely heavily on recommendation systems to suggest shows and videos to their users based on their viewing history.

Moreover, retail sites are also powered with virtual assistants or conversational chatbots that leverage ML, natural language processing (NLP), and natural language understanding (NLU) to automate customer shopping experiences.

4. Travel industry

Machine learning is playing a pivotal role in expanding the scope of the travel industry. Rides offered by Uber, Ola, and even self-driving cars have a robust machine learning backend. Consider Uber's machine learning algorithm that handles the dynamic pricing of their rides. Uber uses a machine learning model called 'Geosurge' to manage dynamic pricing parameters. It uses real-time predictive modeling on traffic patterns, supply, and demand. If you are getting late for a meeting and need to book an Uber in a crowded area, the dynamic pricing model kicks in, and you can get an Uber ride immediately but would need to pay twice the regular fare. Moreover, the travel industry uses machine learning to analyze user reviews. User comments are classified through sentiment analysis based on positive or negative scores. This is used for campaign monitoring, brand monitoring, compliance monitoring, etc., by companies in the travel industry.

5. Social media

With machine learning, billions of users can efficiently engage on social media networks. Machine learning is pivotal in driving social media platforms from personalizing news feeds to delivering user-specific ads. For example, Facebook's auto-tagging feature employs image recognition to identify your friend's face and tag them automatically. The social network uses ANN to recognize familiar faces in users' contact lists and facilitates automated tagging. Similarly, LinkedIn knows when you should apply for your next role, whom you need to connect with, and how your skills rank compared to peers. All these features are enabled by machine learning.

What can machine learning do: Machine learning in the real world

Whereas machine learning functionality has been around for decades, it is the more recent ability to apply and automatically compute complex mathematical calculations involving big data that has given it unprecedented sophistication. The realm of machine learning application today is vast ranging from enterprise AIOps to online retail. Some real world examples of machine learning capabilities today include the following:

- Cyber Security using behavioral analytics to determine suspicious or anomalous events that may indicate insider threats, APTs or zero-day attacks.
- Self-driving car projects, such as Waymo (a subsidiary of Alphabet Inc.) and Tesla's Autopilot which is a step below actual self-driving cars.
- Digital assistants like Siri, Alexa and Google Assistant that search the web for information in response to our voice commands.

- User-tailored recommendations that are driven by machine learning algorithms on websites and apps like Netflix, Amazon and YouTube.
- Fraud detection and cyber resilience solutions that aggregate data from multiple systems, unearth clients exhibiting high-risk behavior and identify patterns of suspicious activity. These solutions can use supervised and unsupervised machine learning to classify transactions for financial organizations as fraudulent or legitimate. This is why a consumer can get texts from their credit card company verifying if an unusual purchase using the consumer's financial credentials is legitimate. Machine learning has gotten so advanced in the area of fraud that many credit card companies advertise no-fault to consumers if fraudulent transactions are not caught by the financial organization's algorithms.
- Image recognition has had significant advancements and can be reliably used for facial recognition, reading handwriting on deposited checks, traffic monitoring and counting the number of people in a room.
- Spam filters that detect and block unwanted mail from inboxes.
- Utilities that analyze sensor data to find ways of improving efficiency and cutting costs.
- Wearable medical devices that capture in real time valuable data for use in assessing patient health continuously.
- Taxi apps evaluating traffic conditions in real time and recommending the most efficient route.
- Sentiment analysis determines the tone of a line of text. Good applications of sentiment analysis are Twitter, customer reviews, and survey respondents:
    - Twitter: one way to evaluate brands is to detect the tone of tweets directed toward a person or company. Companies such as Crimson Hexagon and Nuvi provide this real time.
    - Customer reviews: You can detect the tone of customer reviews to evaluate how your company is doing. This is especially useful if there is no rating system paired with free text customer reviews.
    - Surveys: Using sentiment analysis on free text survey responses can give you at a glance evaluation of how your survey respondents feel. Qualtrics has this implemented with their surveys.
- Market segmentation analysis uses unsupervised machine learning to cluster customers according to buying habits to determine different types or personas of customers. This allows you to better know your most valuable or underserved customers.
- It is easy to press ctrl+F to search a document for exact words and phrases, but if you do not know the exact wording you are looking for it can be difficult to search documents. Machine learning can use techniques such as fuzzy methods and topic modelling can make this process much easier by allowing you to search documents without knowing the exact phrasing you are looking for.

Probability theory

Probability theory provides a consistent framework for the quantification and manipulation of uncertainty. In what ways do we have to deal with uncertainty?

- Uncertainty on my measurements, because there's noise
- Uncertainty related to the finite size of datasets

Probability measures the likelihood of an event's occurrence. In situations where the outcome of an event is uncertain, we discuss the probability of specific outcomes to understand their chances of happening. The study of events influenced by probability falls under the domain of statistics. When we talk about probability theory, we can take on:

- A **frequentist interpretation**, in which case probability is defined as the fraction of times an event occurs in an experiment. We're going to observe random variables and we just count how many times a particular event happens — then this defines a probability which you use to make predictions for the future
- A **bayesian approach**, in which case probability takes on a meaning as a quantification of plausibility or the strength of a belief — in a way this is a more modeling-based approach and it's a bit more generic

When we talk about probabilities, we're dealing with **random variables**. These are stochastic variables sampled from a set of possible outcomes, which means that every time I make an observation of such an x, it takes on one of the values in this set of possible values x. This variable can be discrete, it can be continuous and it always comes with a probability distribution which assigns probabilities to a particular event x happening.

Probability is important to us in machine learning because the idea of generalization is that the past is predictive of the future. That is, we really believe that we can look at a bunch of training data and make interesting and useful predictions about data we have never seen before. So, why is generalization possible? It's possible because we're willing to commit ourselves to structure in the data that we expect to see again and again. That is to say, in order to generalize we must make assumptions about the structure of the world. There are many different ways to specify these kinds of assumptions, but one of the most powerful frameworks is the calculus of probabilities. It's helpful to think about all the different ways that noise and uncertainty can come into play in a problem like this.

- We might have noisy data
- The environment might be stochastic

- Uncertainty about what the optimal parameters should be, due to limited training data
- Your model can't be the perfect representation of what you're studying
- Limited amount of computation to make predictions

Probability simply talks about how likely is the event to occur, and its value always lies between 0 and 1 (inclusive of 0 and 1). For example: consider that you have two bags, named A and B, each containing 10 red balls and 10 black balls. If you randomly pick up the ball from any bag (without looking in the bag), you surely don't know which ball you're going to pick up. So here is the need of probability where we find how likely you're going to pick up either a black or a red ball. Note that we'll be denoting probability as P from now on. P(X) means the probability for an event X to occur.

Basic concepts of probability theory

Sample space: The sample space is the collection of all potential outcomes of an experiment. For example, the sample space of flipping a coin is {heads, tails}.

Event: An event is a collection of outcomes within the sample space. For example, the event of flipping a head is {heads}**.**

Probability: The probability of an event is a number between 0 and 1 that represents the likelihood of the event occurring. A chance of 0 means that the event is impossible, and a probability of 1 means that the event is specific.

Random Variable- A random variable is a variable where chance determines its value. It is a variable that assigns a numerical value to each outcome in a sample space of a random experiment. The value of a random variable is determined by the outcome of a random process or experiment.

Discrete Random Variables- A discrete random variable has distinct values that are countable and finite or countably infinite. This data type often occurs when you are counting the number of event occurrences. These values are often represented by integers or whole numbers, other than this they can also be represented by other discrete values.

For example, the number of heads obtained after flipping a coin three times is a discrete random variable. The possible values of this variable are 0, 1, 2, or 3. Other examples are

The number of cars that pass through a given intersection in an hour.

The number of defective items in a shipment of goods.

The number of people in a household.

The number of accidents that occur at a given intersection in a week.

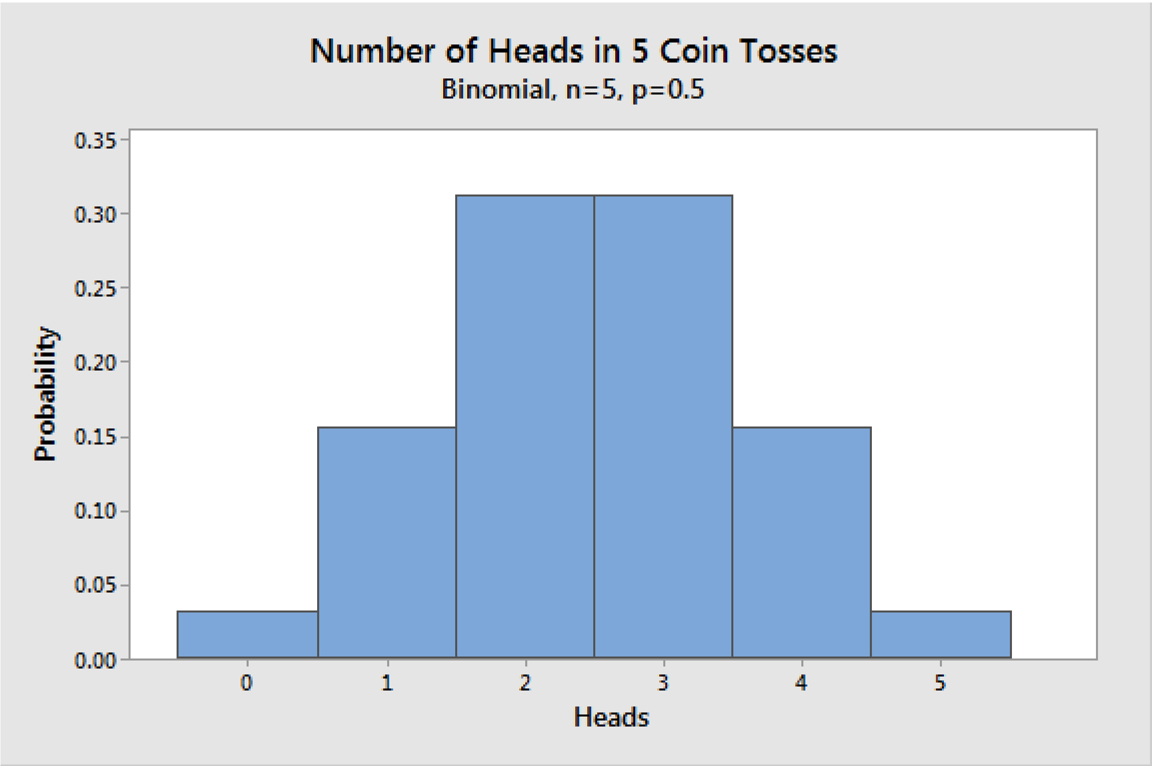
Analysts denote the variable as X and its possible values as $x_1, x_2, \ldots, x_n$.

The probability of X having a value of x for its i^{th} observation equals p_i: $P(X = x_i) = p_i$.

Using this notation, discrete random variables must satisfy these conditions:

- All possible discrete values must have probabilities between zero and one: $0 < p_i \leq 1$.
- The total probability for all possible k values must equal 1: $p_1 + p_2 + p_3 + \ldots + p_k = 1$.

Discrete Example- The number of heads that appear during a series of five coin tosses is a discrete random variable that follows the binomial distribution. We can use that distribution to determine the likelihood of obtaining 0 to 5 heads. The graph below displays the probability for each possible outcome.

- 

Continuous Random Variable- Continuous variable is a type of variable that can take on any value within a given range. Unlike discrete variables, which consist of distinct, separate values, continuous variables can represent an infinite number of possible values, including fractional and decimal values. Continuous variables often represent measurements or quantities.

Example of continuous variables are:

- Height: Height is a continuous variable because it can take on any value within a range (e.g., 150.5 cm, 162.3 cm, 175.9 cm).

- Weight: Weight is continuous because it can be measured with precision and can take on any value within a range (e.g., 55.3 kg, 68.7 kg, 72.1 kg).

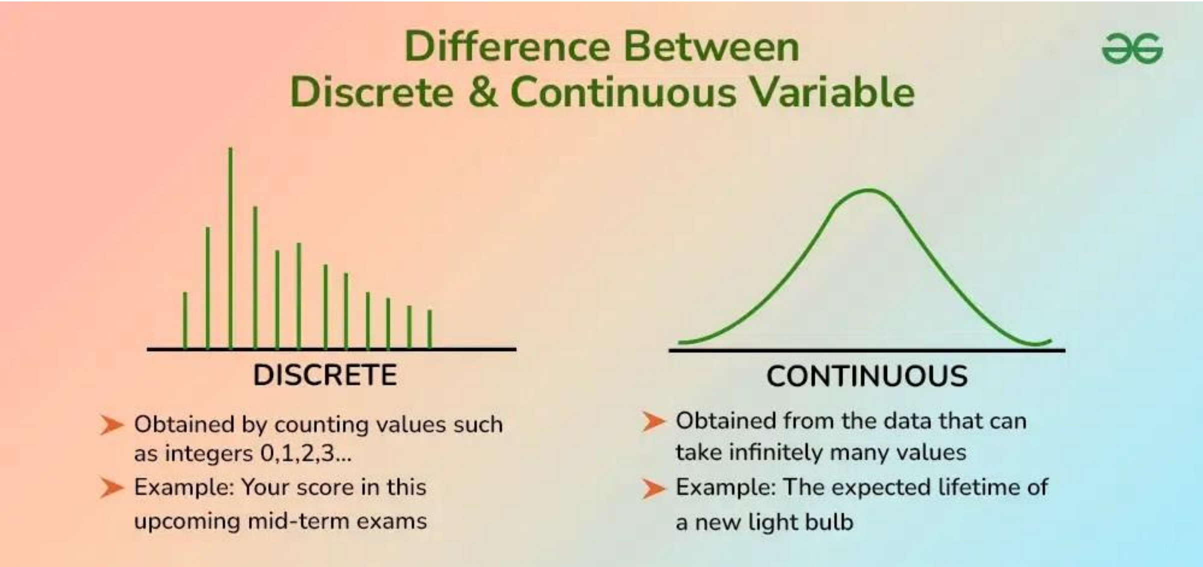
- Time: Time can be measured with precision, and it can take on any value (e.g., 10:30:15.5 AM, 10:45:30.75 AM).
- Analysts denote a continuous random variable as X and its possible values as x, just like the discrete version. However, unlike discrete random variables, the chances of X taking on a specific value for continuous data is zero. In other words: $P(X = x) = 0$, where x is any specific value.
- Instead, probabilities greater than zero only exist for ranges of values, such as $P(a \leq X \leq b)$, where a and b are the lower and upper bounds of the range.

A probability density function (PDF) describes the probability distribution of a continuous random variable. These functions use a curve displaying probability densities, which are ranges of one unit.

Continuous random variables must satisfy the following:

- Probabilities for all ranges of X are greater than or equal to zero: $P(a \leq X \leq b) \geq 0$.
- The total area under the curve equals one: $P(-\infty \leq X \leq +\infty) = 1$.

Difference Between Discrete and Continuous Variable



The difference between continuous and discrete variables is described below:

Aspect	Discrete Variables	Continuous Variable
Nature of Values	They can take only specific or discrete values.	They can take any value within a specific range.
Measurement Scale	Discrete variables are typically measured on a nominal or ordinal scale.	Continuous variables are typically measured on an interval or ratio scale.
Representation	Discrete variables are often represented by bar graphs or histograms.	Continuous variables are often represented by line graphs or smooth curves.
Examples	Examples include the number of students in a class or the outcomes of rolling a die.	Examples include measurements such as length, time, or temperature.
Probability Distributions	Discrete variables have probability mass functions (PMF)	Continuous variables have probability density functions (PDF).
Applications	They are employed in various mathematical contexts and applications where quantities are counted.	They are often employed in various branches of mathematics, including calculus, differential equations, and real analysis, as well as in applied fields such as physics, engineering and statistics.
Mean	The mean of a discrete random variable is $E[X] = \sum x\, P(X = x)$, where $P(X = x)$ is the probability mass function.	Mean of a continuous random variable is $E[X] = \int_{-\infty}^{\infty} x f(x)\,dx$

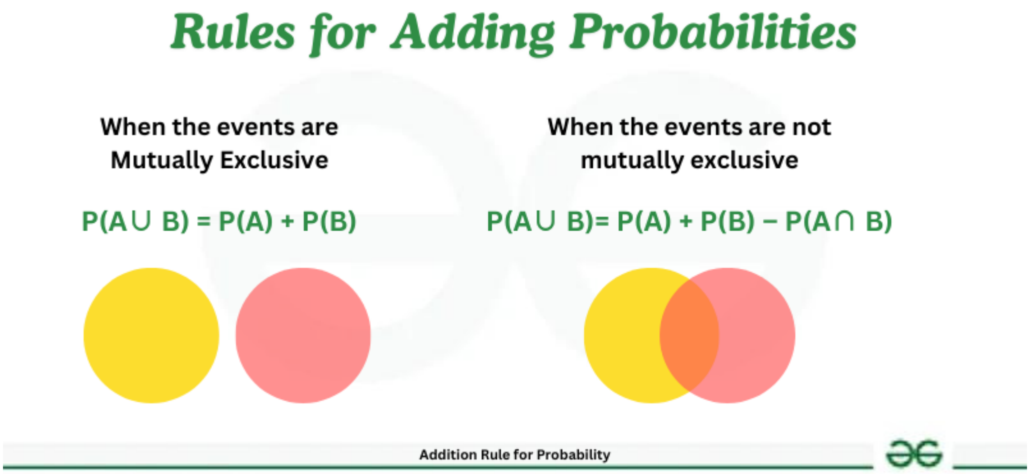
Fundamental Rules

General Addition Rule/Union Rule- It deals with the probability of the union of two events. If A and B are two events, then the probability of either event A or event B occurring is given by: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

This rule applies when events A and B are not mutually exclusive.

where, $P(A \cap B)$ represents the probability of both events A and B occurring simultaneously

The General Addition Rule for Probability is given by P(A or B) = P(A) + P(B) – P(A and B) where A and B are the two events. For mutually exclusive events, P(A and B) = 0. So P(A or B) = P(A) + P(B) for mutually exclusive events.



Mutually exclusive events are events that cannot occur simultaneously. For example, when rolling a six-sided die, the outcomes of getting a 2 and getting a 3 are mutually exclusive because it is not possible to roll both a 2 and a 3 on the same die. In general, if two events A and B are mutually exclusive, then the probability of both events occurring ($P(A \cap B)$) is equal to 0.

Let's say we have a well-shuffled deck. We draw two cards, find the probability of getting either King or a Queen.

Solution: Let's say drawing a king represents an event A while drawing a queen represents an event B. We are asked for the probability for getting either King or Queen. We will use law of adding probabilities here,

Probability (King or Queen) = Probability (King) + Probability (Queen)

We know that there are 4 Kings and 4 Queens in the deck.

P(King) = 4/52=1/13

P(Queen) = 4/52=1/13

Thus, Probability (King or Queen) = 1/13+1/13=2/13

Example 2: In a class of 90 students, 50 took Math, 25 took Physics, 30 took both Math and Physics. Find the number of students who have taken either math or Physics.

Solution: Since the events of choosing math and physics are **non mutually** exclusive, hence

P(Math U Physics)= P(Math)+ P (Physics) -P (Math ∩ Physics)

= 50 +25-30

=45

Complementary Rule

- Rule: The probability of an event not occurring (the complement of event A) is 1 minus the probability of the event occurring.
- Mathematical Expression: P(Ac)=1−P(A)
- Explanation: Since an event and its complement together cover the entire sample space, their probabilities must sum to 1.

Example: Suppose we draw a card from a standard deck of 52 playing cards. Let A be the event of drawing a heart. The probability of not drawing a heart (drawing a card that is not a heart) is given by:

P(A) = 13/52 = 1/4

P(A') = 1 – P(A)

P(A') = 1 – 1/4 = 3/4

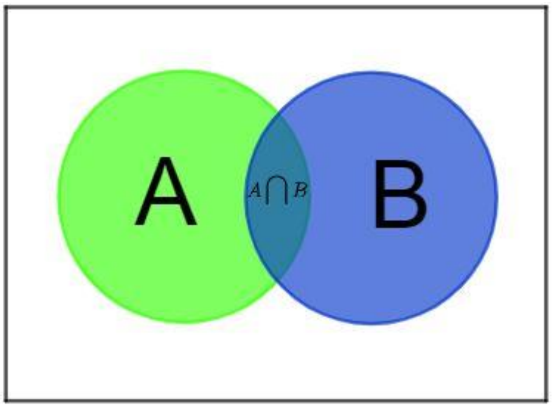
Multiplication Rule/Joint Probability

Multiplication rule of probability applies when we want to find the probability of the intersection of two independent events. If A and B are independent events, then the probability of both events A and B occurring is given by:

P(A∩B) = P(A) × P(B)

This rule holds true when the occurrence of one event does not affect the probability of the other event.

For two independent events, outcomes that do not rely on the occurrence of another event, the joint probability formula is given by **P(A∩B) = P(A) × P(B)**. That is, the probability of both A *and* B occurring is equal to the probability of A times the probability of B.



For the same deck of cards, the probability of drawing both a 7 and a diamond is P(7∩ diamond)= P(7). P(diamond) = 4/52.13/52= 1/52/ (There are four 7's in the deck, thirteen diamonds, but only one 7 of diamonds).

Example: Consider rolling a fair six-sided die. Let A be the event of rolling an even number, and B be the event of rolling a number less than 4. Since these events are independent, the probability of rolling an even number and a number less than 4 is given by:

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

$$P(A) = 3/6 = 1/2$$

$$P(B) = 3/6 = 1/2$$

$$P(A \cap B) = 1/2 \times 1/2 = 1/4$$

Conditional Probability

When event A is already known to have occurred and probability of event B is desired, then P(B, given A)=P(A and B)P(A, given B). It can be vica versa in case of event B.

$$\mathbf{P(B|A)=P(A \cap B)P(A)}$$

This rule quantifies how the probability of one event changes in light of the occurrence of another event.

Example: Let's continue with the example of drawing a card from a standard deck of 52 playing cards. Suppose B is the event of drawing a face card (jack, queen, or king), and A is the event of drawing a heart. The probability of drawing a heart given that the card drawn is a face card is given by:

$$P(A) = 13/52 = 1/4$$

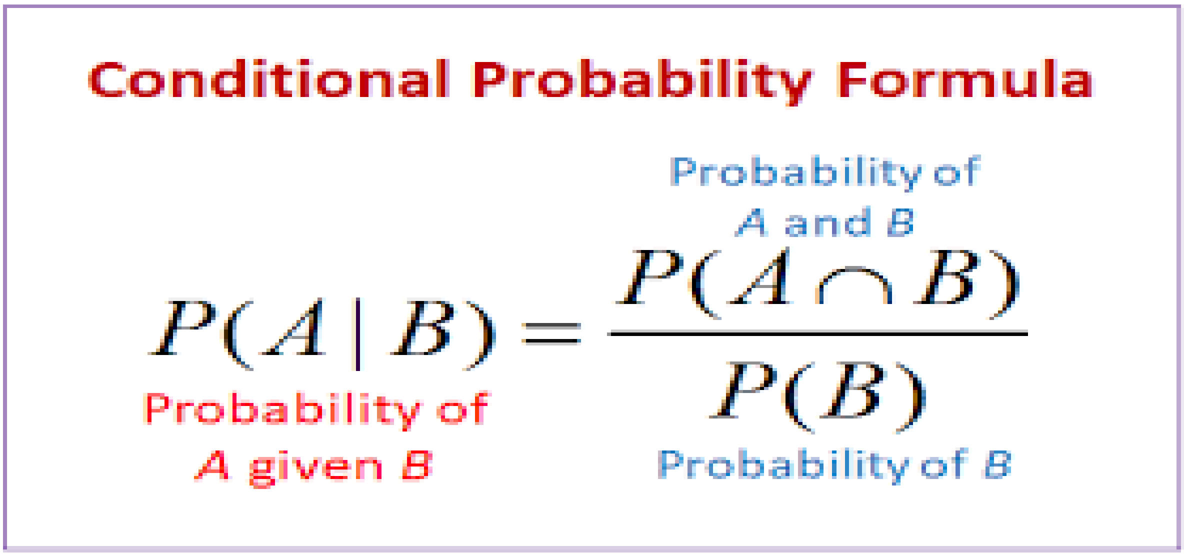
$$P(B) = 12/52$$

$$P(A \cap B) = 3/52$$

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

$$P(A|B) = (3/52)/(12/52) = 1/4$$

This means that given the card drawn is a face card, there is a 1/4 chance that it is also a heart.



Introduction to Bayes Theorem

Bayes' Theorem is a fundamental result in probability theory and statistics that describes how to update the probability of a hypothesis based on new evidence. It provides a way to revise existing predictions or theories given new or additional data.

Bayes theorem (also known as the Bayes Rule or Bayes Law) is used to determine the conditional probability of event A when event B has already occurred.

The general statement of Bayes' theorem is "**The conditional probability of an event A, given the occurrence of another event B, is equal to the product of the event of B, given A and the probability of A divided by the probability of event B.**" i.e.

$$\mathbf{P(A|B) = P(B|A)P(A) / P(B)}$$

where,

- **P(A)** and **P(B)** are the probabilities of events A and B
- **P(A|B)** is the probability of event A when event B happens
- **P(B|A)** is the probability of event B when A happens

Example: Imagine an email filter that classifies emails as spam or not spam. The filter uses certain words to determine whether an email is spam. Consider the word "free":

- The probability that an email is spam (P(Spam) is 0.2 (20% of emails are spam).
- The probability that the word "free" appears in a spam email P(Free|Spam)) is 0.7.

- The probability that the word "free" appears in a non-spam email P(Free|Not Spam)) is 0.1.

We want to find the probability that an email is spam given that it contains the word "free" P(Spam|Free)).

Solution:

1. **Calculate P(Free)**:

$P(Free)=P(Free|Spam)\cdot P(Spam)+P(Free|Not\ Spam)\cdot P(Not\ Spam)$
$=0.7\cdot 0.2+0.1\cdot 0.8=0.14+0.08=0.22$

2. **Apply Bayes' Theorem**:

$P(Spam|Free)=P(Free|Spam)\cdot P(Spam)/P(Free)=0.7\cdot 0.2/0.22\approx 0.636$.

Quantiles

Quantiles offers valuable insights into data distribution and helping in various aspects of analysis. This article describes quantiles, looks at **how to calculate** them, and talks about **how important they are for machine learning applications**. We also discuss the problems with quantiles and how box plots may be used to represent them. For anybody dealing with data in the field of machine learning, having a firm understanding of quantiles is crucial.

Quantiles divide the dataset into equal parts based on rank or percentile. They represent the values at certain points in a dataset sorted in increasing order. General quantiles include the median (50th percentile), quartiles (25th, 50th, and 75th percentiles), and percentiles (values ranging from 0 to 100).

In machine learning and data science, quantiles play an important role in understanding the data, detecting outliers and evaluating model performance.

Types of Quantiles

Quartiles: Quartiles divide a dataset into four equal parts, representing the 25th, 50th (median), and 75th percentiles.

Quintiles: Quintiles divide a dataset into five equal parts, each representing 20% of the data.

Deciles: Deciles divide a dataset into ten equal parts, with each decile representing 10% of the data.

Percentiles: Percentiles divide a dataset into 100 equal parts, with each percentile representing 1% of the data.

Steps to Calculate Quantiles

The steps for calculating quantiles involve:

Sorting the Data: Arrange the dataset in increasing order.

Determine the Position: Calculate the position of the desired quantile based on the given formula: **"Position=(quantile×(n+1))/100"**, where n is the total number of observations.

Interpolation (if needed): Interpolate between two adjacent values to find the quantile if the position is not an integer.

Example with Mathematical Imputation:

Let's consider a dataset: [5, 10, 15, 20, 25, 30, 35, 40, 45, 50].

Median (Q2): There are 10 observations, so the median position is (2×(10+1))/2=5.5. Since, 5.5 is not an integer, we interpolate between the 5th and 6th observations: Median=(25+30)/2=27.5.

First Quartile (Q1): (25×(10+1))/4=13.75. Interpolating between the 13th and 14th observations: Q1=(15+20)/2=17.5.

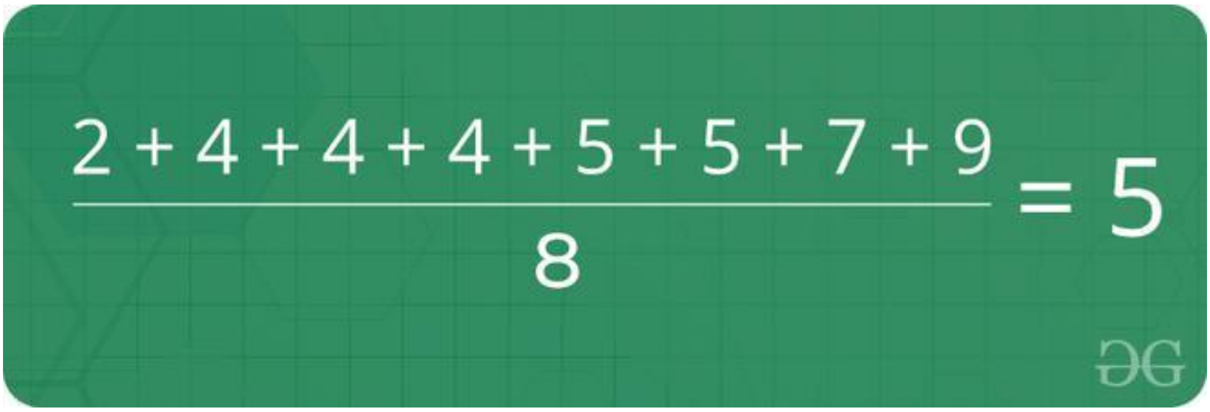
Third Quartile (Q3):(75×(10+1))/4=41.25. Interpolating between the 41st and 42nd observations: Q3=(40+45)/2=42.5.

Mean

Mean is the average of the given numbers which is calculated by **dividing** the **sum of given numbers** by the **total count of numbers**.

Example

Find the mean of the given numbers 2, 4, 4, 4, 5, 5, 7, and 9?



Variance is a measurement value used to find how the data is spread concerning the mean or the average value of the data set. It is used to find how the distribution data is spread out concerning the mean or the average value. The symbol used to define the variance is σ^2. It is the square of the Standard Deviation.

The are two types of variance used in statistics,

- **Sample Variance**
- **Population Variance**

The population variance is used to determine how each data point in a particular population fluctuates or is spread out, while the sample variance is used to find the average of the squared deviations from the mean.

Variance measures the dispersion of a dataset, indicating how much the values differ from the mean. It is the average of the squared differences from the mean.

Population Variance

Population variance is used to find the spread of the given population. The population is defined as a group of people and all the people in that group are part of the population. It tells us about how the population of a group varies with respect to the mean population.

All the members of a group are known as the population. When we want to find how each data point in a given population varies or is spread out then we use the population variance. It is used to give the squared distance of each data point from the population mean.

Sample Variance

If the population data is very large it becomes difficult to calculate the population variance of the data set. In that case, we take a sample of data from the given data set and find the variance of that data set which is called sample variance. While calculating the sample mean we make sure to calculate the sample mean, i.e. the mean of the sample data set not the population mean. We can define the sample variance as the mean of the square of the difference between the sample data point and the sample mean.

Variance Formula

The variance for a data set is denoted by the symbol σ^2. For population data, its formula is equal to the sum of squared differences of data entries from the mean divided by the number of entries. While for sample data, we divide the numerator value by the difference between the number of entries and unity.

Sample Variance Formula

If the data set is a sample the formula of variance is given by,

$$\sigma^2 = \sum (x_i - \bar{x})^2/(n - 1)$$

where,

- **$\bar{x}$ is the mean of sample data set**
- **n is the total number of observations**

Population Variance Formula

$$\sigma^2 = \sum (x_i - \bar{x})^2/n$$

where,

- **$\bar{x}$ is the mean of population data set**
- **n is the total number of observations**

Example: Find the population variance of the data {4,6,8,10}

Solution: Mean = (4+6+8+10)/4 = 7

4	$(4-7)^2$	9
6	$(6-7)^2$	1
8	$(8-7)^2$	1
10	$(10-7)^2$	9

Variance = (9+1+1+9)/4 = 20/4 = 5

Probability Densities

The Probability Density Function(PDF) defines the probability function representing the density of a continuous random variable lying between a specific range of values. In other words, the probability density function produces the likelihood of values of the continuous random variable. Sometimes it is also called a probability distribution function or just a probability function.

Probability Density Function Formula

Let Y be a continuous random variable and F(y) be the cumulative distribution function (CDF) of Y. Then, the probability density function (PDF) f(y) of Y is obtained by differentiating the CDF of Y.

f(y) = ddy[F(y)]dyd[F(y)] **= F'(y)**

If we want to calculate the probability for X lying between the interval a and b, then we can use the following formula:

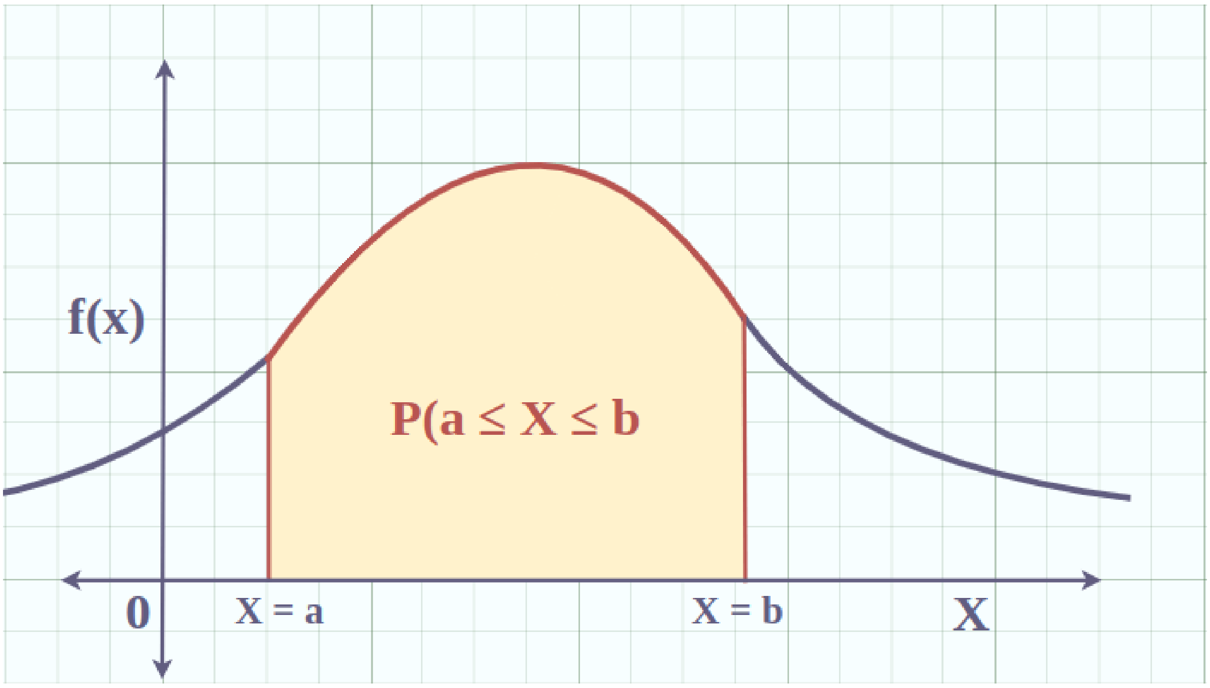
P (a ≤ X ≤ b) = F(b) – F(a) = ∫abf(x)dxa∫bf(x)dx

A Probability Density Function (PDF) is a function that describes the likelihood of a continuous random variable taking on a particular value. Unlike discrete random variables, where probabilities are assigned to specific outcomes, continuous random variables can take on any value within a range. Probability Density Function (PDF) tells us

- **Relative Likelihood**
- **Distribution Shape**
- **Expected Value and Variance, etc.**

Graph for Probability Density Function

If X is continuous random variable and f(x) be the probability density function. The probability for the random variable is given by area under the pdf curve. The graph of PDF looks like bell curve, with the probability of X given by area below the curve. The following graph gives the probability for X lying between interval a and b.



Expectation and Covariance

Expected Value: Random variables are the functions that assign a probability to some outcomes in the sample space. They are very useful in the analysis of real-life random experiments which become complex. These variables take some outcomes from a sample space as input and assign some real numbers to it. The expectation is an important part of random variable analysis. It gives the average output of the random variable.

For a random variable X, the expectation gives an idea of the average value attained by X when the experiment is repeated many times. Since this value is mapped with an outcome in the sample space. Expected value can be used to determine which of the outcomes is most likely to happen when the experiment is repeated many times.

For random variable X which assumes values $x_1, x_2, x_3, \ldots x_n$ with probability $P(x_1)$, $P(x_2)$, $P(x_3)$, … $P(x_n)$

Expectation of X is defined as,

$$E(x) = \sum P(xi)xi\sum P(xi)xi$$

Covariance is a statistical measure that **indicates the direction of the linear relationship between two variables**. It assesses how much two variables change together from their mean values.

Types of Covariance:

- **Positive Covariance**: When one variable increases, the other variable tends to increase as well, and vice versa.

- **Negative Covariance**: When one variable increases, the other variable tends to decrease.

- **Zero Covariance**: There is no linear relationship between the two variables; they move independently of each other

Covariance is calculated by taking the average of the product of the deviations of each variable from their respective means. It is useful for understanding the direction of the relationship but not its strength, as its magnitude depends on the units of the variables.

It is an essential tool for understanding how variables change together and are widely used in various fields, including finance, economics, and science.

Covariance:

1. It is the relationship between a pair of random variables where change in one variable causes change in another variable.

2. It can take any value between – infinity to +infinity, where the negative value represents the negative relationship whereas a positive value represents the positive relationship.

3. It is used for the linear relationship between variables.

4. It gives the direction of relationship between variables.

Covariance Formula

For Population:

$$Covri(x, y) = \frac{\sum_{i=1}^{n}(x_i - x')\,(y_i - y')}{n}$$

For Sample:

$$Covari(x, y) = \frac{\sum_{i=1}^{n}(x_i - x')\ (y_i - y')}{n-1}$$

Here, x’ and y’ = mean of given sample set n = total no of sample xi and yi = individual sample of set