
UNIT 9 INTRODUCTION TO MACHINE LEARNING METHODS

Structure

- 9.0 Introduction
- 9.1 Objectives
- 9.2 Introduction to Machine Learning
- 9.3 Techniques of Machine Learning
- 9.4 Reinforcement Learning and Algorithms
- 9.5 Deep Learning and Algorithms
- 9.6 Ensemble Methods
- 9.7 Summary
- 9.8 Solutions/ Answers
- 9.9 Further Readings

9.0 INTRODUCTION

After Artificial Intelligence was introduced, in Computing World. There was a need for a machine that would automatically make things better. This needs to be kept in check, so there should be some rules that apply to all learning processes.

The main goal of machine learning, even at its most basic level, is to be able to analyse and adapt data on its own and make decisions based on calculations and analyses. Machine learning is a way to try to improve computers by imitating how the human brain learns. A computer that doesn't have intelligence is just a fast machine for processing data. The devices that don't have AI or ML are just data processing units that use the information they are given. Machine Learning is what we need to make devices that can make decisions based on data.

To get to this level of intelligence, you need to put algorithms and data into a machine in a way that lets it make decisions.

For example, Real-time GPS data is used by Maps on devices to show the quickest and fastest route. Several algorithms, such as the shortest path (Dijkstra's algorithm) and the travelling salesman (an algorithm that works like water flow), can be used to make the decision (WFA). These are algorithms that have been used and can be made better, but they are useful for learning. Here, we can see that the Machine, which is your computer or mobile device, uses GPS coordinates, traffic data based on density, and predefined map routes to figure out the fastest way to get from Point A to Point B..

This is one of the simplest examples that can help us understand how Machine Learning can help in independent decision making by devices and how it can help in making decision making easier and more accurate.

The accuracy of the data as a whole is a topic of debate since decisions based on the data might be accurate, but it is one of the issues whether or not they are acceptable within the limitations. Consequently, it is necessary to set these boundaries for the entirety of the machine learning algorithms and engines.

Simplest example would be if we instruct an auto driving car to reach a destination at a specified time. IT should also work within the legal boundaries of the land not to break traffic rules to achieve the desired result. The Boundaries and restriction cannot be ignored as they are very important for any Self learning system.

Data/Inputs is a soul of all the business .Data has been a key component for making any decision . Data is the key to successes from prehistoric era. The more you have the data more is the probability of making the right decision. Machine learning is the key to unlock new world where customer data , corporate data , demographic data, or related dimension data relevant to the decision can help you make right and more informed decision to stay ahead of competition.

Both artificial intelligence and statistics research groups contributed to the development of machine learning. Companies like Google, Microsoft, Facebook, and Amazon all use machine learning as part of their decision-making processes.

The most common applications of machine learning nowadays are to interpret and investigate cyber phenomena, to extract and project the future values of those phenomena, and to detect anomalies.

There are a number of open-source solutions for machine learning that can be used with API calls or without programming. Some of the Open-source Machine Learning projects, such as Weka, Orange, and Rapid-Miner. To see how data that has been processed by an algorithm looks, you can put the results into tools like Tableau, Pivotal, or Spotfire and use them to make dashboards and workflow strategies.

Michie et al. (D. Michie, 1994) says that Machine Learning usually refers to automatic computing procedures based on logical or binary operations that learn how to do a task by looking at a series of examples. Machine learning is used in a lot of ways today, but whether or not they are all ready is up for debate. There is a lot of room for improvement when it comes to accuracy, which is a process that never ends and changes every day.

9.1 OBJECTIVES

After going through this unit, you should be able to:

- Understand the basics of Machine learning
- Identify various techniques of Machine Learning
- Understand the concept of Reinforcement Learning
- Understand the concept of Deep Learning
- Understand Ensemble Methods

9.2 INTRODUCTION TO MACHINE LEARNING

Understanding data, describing the characteristics of a data collection, and locating hidden connections and patterns within that data are all necessary steps in the process of developing a model. These steps can be accomplished through the application of statistics, data mining, and machine learning. When it comes to finding solutions to business issues, the methods and tools that are employed by these fields share a lot in common with one another.

The more conventional forms of statistical investigation are the origin of a great deal of the prevalent data mining and machine learning techniques. Data scientists have a background in technology and also have expertise in areas such as statistics, data mining, and machine learning. This allows them to collaborate effectively across all fields.

The process of "data mining" refers to the extraction of information from data that is latent, previously unknown, and has the potential to be beneficial. Building computer algorithms that can automatically search through large databases for recurring structures or patterns is the goal of this project. In the event that robust patterns are discovered, it is likely that they will generalise to enable accurate predictions on future data.

In the renowned book "Data Mining: Practical Machine Learning Tools and Techniques," written by Ian Witten and Eibe Frank, the subject matter is thoroughly covered. The activity known as "data mining" refers to the practise of locating patterns within data. The procedure needs to be fully automatic or, at the very least, semiautomatic. The patterns that are found have to be significant in the sense that they lead to some kind of benefit, most commonly an economic one. Consistently and in substantial amounts, the statistics are there to be found.

Machine learning, on the other hand, is the core of data mining's technical infrastructure. It is used to extract information from the raw data that is stored in databases; this information is then expressed in a form that is understandable and can be applied to a range of situations.

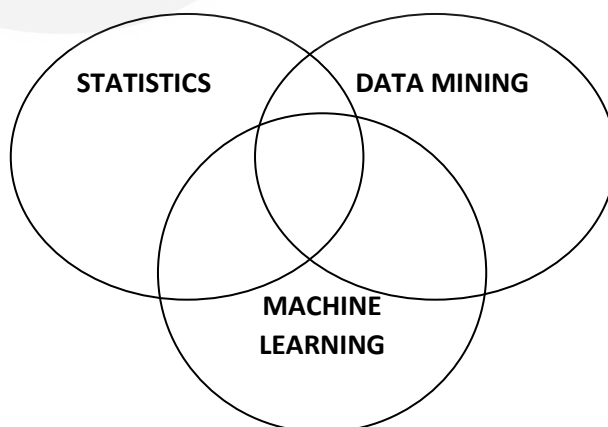


Figure 1 : Machine Learning – Data Mining - Statistics

We learned the differences between Machine Learning and Data Mining from the conversation; nevertheless, because all three, machine learning, data mining, and statistics, are intertwined, we must grasp their relationships. So, how do machine learning and statistics differ? In reality, there is no clear cut

between machine learning and statistics because data analysis techniques are a continuum—and a multidimensional one at that. Some are derived from statistical skills, while others are more strongly linked to the type of machine learning that has emerged from computer science. Both sides have had quite diverse traditions throughout history. If forced to choose one point of emphasis, statistics may have been more concerned with testing hypotheses, whereas machine learning has been more interested with articulating the generalisation process as a search through possible hypotheses. This, however, is an exaggeration: Many machine learning algorithms do not require any searching at all, and statistics is significantly more than just hypothesis testing.

Most learning algorithms use statistical tests to build rules or trees and fix models that are "overfitted," or too dependent on the details of the examples that were used to make them. So, a lot of statistical thinking goes into the techniques we will talk about in this unit. Statistical tests are used to evaluate and validate machine learning models and algorithms.

Machine learning is when a computer learns how to do a task by using algorithms that are logical and can be turned into models that can be used. Artificially intelligent communities are the main reason why Machine Learning is growing. The most important factor contributing to this expansion was that it assisted in the collection of statistical and computational methods that could automatically construct usable models from data. Companies such as Google, Microsoft, Facebook, and Netflix have been putting in consistent effort over the past decade to make this more accurate and mature.

The primary function or application of machine learning algorithms can be summarized as follows:

- (a) To gain an understanding of the cyber phenomenon that produced the data that is being investigated;
- (b) To abstract the understanding of underlying phenomena in the form of a model;
- (c) To predict the future values of a phenomenon by using the model that was just generated; and
- (d) To identify anomalous behavior exhibited by a phenomenon that is being observed.

There are various open-source implementations of machine learning algorithms that can be utilised with either application programming interface (API) calls or non-programmatic applications. These methods can also be used in conjunction with each other. Weka, Orange, and Rapid Miner are a few instances of open-source application programming interfaces. These algorithms' outputs can be fed into visual analytics tools like Tableau4 and Spotfire5, which can then be used to build dashboards and actionable pipelines.

Almost all of the Frameworks have emphasised decision-tree techniques, in which classification is determined by a series of logical processes. Given enough data (which may be a lot!), these are capable of representing even the most complex problems. Other techniques, such as genetic algorithms and inductive logic procedures (ILP), are currently in development and, in theory,

would allow us to deal with a wider range of data, including cases where the number and type of attributes vary, where additional layers of learning are superimposed, and where attributes and classes are organised hierarchically, and so on. Machine Learning seeks to provide classification phrases that are basic enough for humans to understand. They must be able to sufficiently simulate human reasoning in order to provide insight into the decision-making process. Background knowledge, including statistical techniques, can be used in development, but operation is assumed to be without human interference.

The expression “To learn” can be understood as :

- To learn means to acquire knowledge.to gain knowledge or understanding of something through experiencing it or through learning about it (some art or practice)
- To gain experience with something, learn a new skill, or master a talent.
- To memorize (something), to acquire something through the experience, example, or practice of doing so.

Machine Learning is a methodology for automatically improving computer systems through the process of developing using experience and implementing a learning process. There are various techniques for imparting machine learning and we will learn about few of those in the subsequent section.

Check Your Progress - 1

Q1 How machine learning differs from Artificial intelligence ?

.....

.....

.....

Q2 Briefly discuss the major function or use of Machine learning algorithms.

.....

.....

.....

9.3 TECHNIQUES OF MACHINE LEARNING

Machine learning uses various algorithms to improve, describe, and predict outcomes by repeated learning from data. It is possible to make models that are more accurate as the algorithms learn from the training data. A machine learning model is what you get when you use data to train your machine learning algorithm. After it has been trained, a model will give you an output when you give it an input. A predictive model is made, for example, by a predictive algorithm. Then, when you put data into the predictive model, you'll get a prediction based on the data that was used to train the model. At the moment, analytics models can't be made without machine learning.

Machine learning approaches are needed to make prediction models more accurate. Depending on the type and amount of data and the business problem

being solved, there are different ways to approach the problem. In this section, we talk about the machine learning cycle.

The Machine Learning Cycle: Making a machine learning application is similar to making a machine learning algorithm work, which is an iterative process. You can't just train a model once and leave it alone, because data changes, preferences change, and new competitors come along. So, when your model goes into production, you need to keep it updated. Even though you won't need as much training as when you created the model, don't expect it to run on its own.

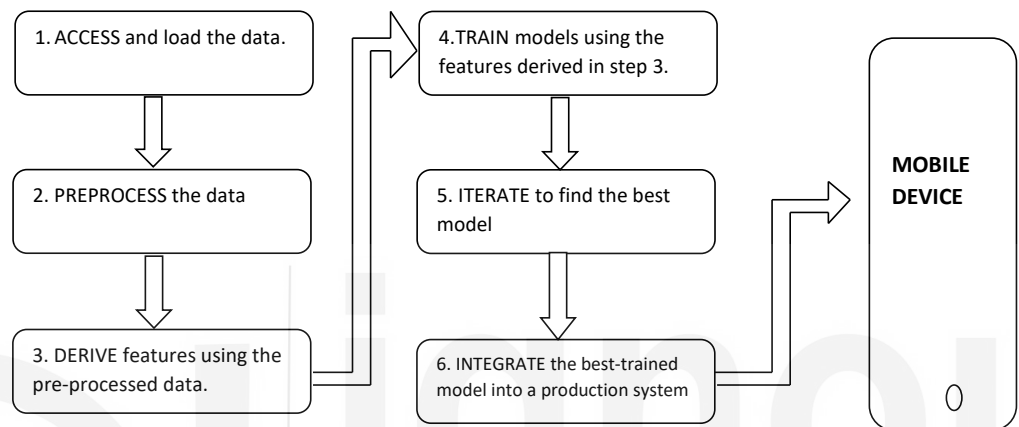


Figure 2 : Machine Learning Cycle at a Glance

To use machine learning techniques effectively, you need to know how they work. You can't just use them without knowing how they work and expect to get good results. Different techniques work for different kinds of problems, but it's not always clear which techniques will work in a given situation. You need to know something about the different kinds of solutions. We talk about a very large number of techniques.

One step in the machine learning cycle is choosing the right machine learning algorithm. So, let's look at how the machine learning cycle works.

The steps in the machine learning cycle are as follows:

1. Data Identification
2. Data Preparation
3. Selection of machine learning algorithm:
4. Training the algorithm to develop a model
5. Evaluating the model
6. Deploying the model
7. Performing Prediction
8. Assess the predictions

When your model has reached the point where it can make accurate predictions, you can restart the process by re-evaluating it using questions such as "Is all of the information important?" Exist any more data sets that could be used to improve the accuracy of the predictions? You may maintain the usefulness of

your applications that are based on machine learning by continually improving the models and assessing new approaches.

When should you use machine learning? Think about using machine learning when you have a hard task or problem that involves a lot of data and many different factors but no formula or equation to solve it. For example, machine learning is a good choice if you need to deal with situations like face and speech recognition, fraud detection by analysing transaction records, automated trading, energy demand forecasting, predicting shopping trends, and many more.

When it comes to machine learning, there's rarely a straight line from the beginning to the end. Instead, you'll find yourself constantly iterating and trying out new ideas and methods.

This unit talks about a step-by-step process for machine learning and points out some important decision points along the way. The most common problem with machine learning is getting your data in order and finding the right model. Here are some of the most important things to worry about with the data:

- **Data comes in all shapes and sizes :** There are many different kinds of data. Datasets from the real world can be messy, with missing values, and may be in different formats. You might just have simple numeric data. But sometimes you have to combine different kinds of data, like sensor signals, text, and images from a camera that are being sent in real time.
- **Preprocessing your data might require specialized knowledge and tools :** You might need specialised tools and knowledge to prepare your data before you use it. For example, you need to know a lot about image processing to choose features to train an object detection algorithm. Preprocessing needs to be done in different ways for different kinds of data.
- **It takes time to find the best model to fit the data :** Finding the best model to fit the data takes time. Finding the right model is like walking a tightrope. Highly flexible models tend to fit the data too well by trying to explain small differences that could just be noise. On the other hand, models that are too simple might assume too much. Model speed, accuracy, and complexity are always at odds with each other.

Does it appear to be a challenge? Try not to let this discourage you. Keep in mind that the process of machine learning relies on trial and error. You merely go on to the next method or algorithm in the event that the first one does not succeed. On the other hand, a well-organized workflow will assist you in getting off to a good start.

Every machine learning workflow begins with three questions:

- What kind of information do you have to work with?
- How do you want to learn something from it?
- How and where will these new ideas come from?

Your answers to these questions help you decide whether to use supervised or unsupervised learning algorithm, before proceeding to other details we will discuss these two types of learning algorithms.

Machine Learning - I Fundamentally Machine learning involves two classes of Learning algorithms:

- a) Supervised learning, which requires training a model with data whose inputs and outputs are already known in order for the model to be able to predict future outputs, such as whether or not an email is authentic or spam or whether or not a tumor is cancerous. Classification models classify given data into categories. Imaging for medical purposes, speech recognition, and rating credit are a few examples of typical applications.
- b) Unsupervised learning analyses data to uncover previously unknown patterns or structures. It is used to infer conclusions from sets of data that contain inputs but no tagged answers. The most prevalent method of learning without being observed is clustering. Exploratory data analysis is used to uncover hidden patterns or groups in data. Clustering can be used for gene sequence analysis, market research, and object recognition.

Note: In semi-supervised learning, algorithms are trained on small sets of labelled data before being applied to unlabeled data, like in unsupervised learning. When there is a dearth of quality data, this method is frequently used.

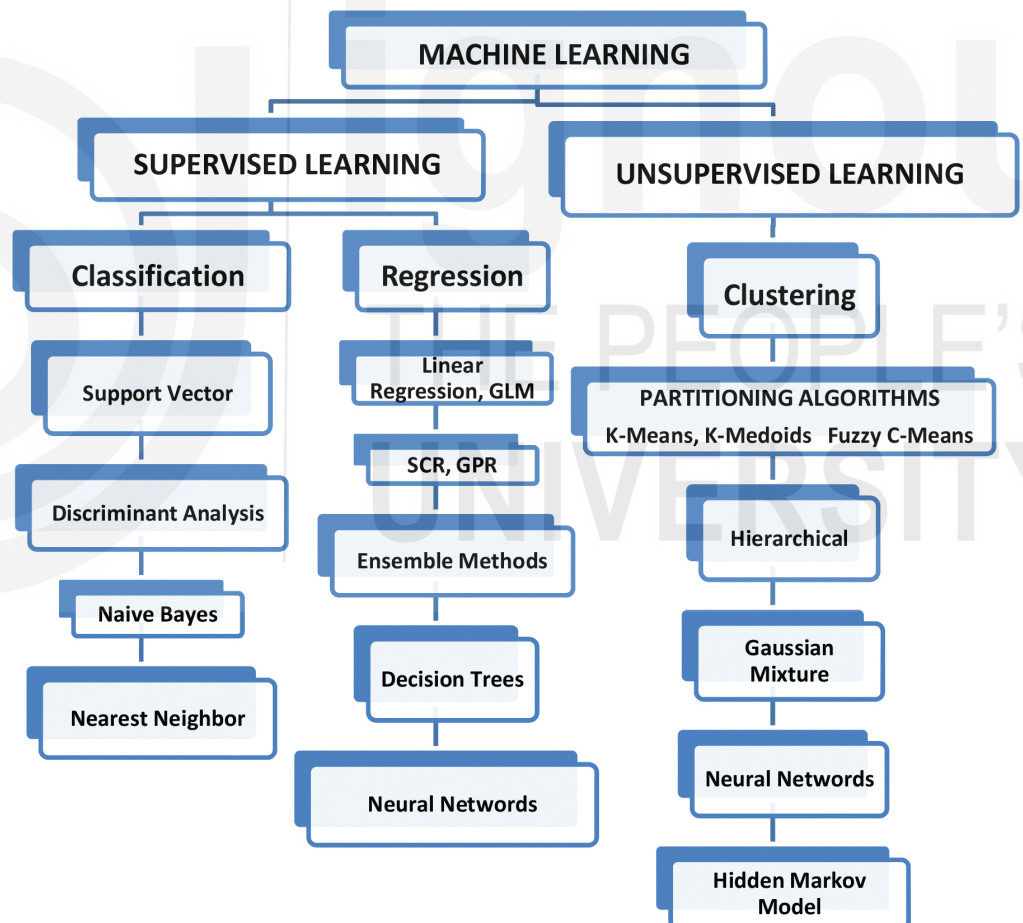


Figure – 3 Machine Learning Algorithms

"How Do You Choose Which Algorithm to Use?" is a crucial question. There are numerous supervised and unsupervised machine learning algorithms, each with its own learning strategy. This can make picking the appropriate one difficult. There is no alternative solution or strategy that will work for everyone. It takes some trial and error to find the proper algorithm. Even the most seasoned data scientists can't predict whether or not an algorithm would work without putting

it to the test. However, the size and type of data you're working with, the insights you want to gain from the data, and how those insights will be used all go into the algorithm you choose.

- *If you need to train a model to produce a forecast, such as the future value of a continuous variable like temperature or a stock price, or a classification, such as determining what kind of automobile is on a webcam footage, go with supervised learning.*
- *If you want to look at your data and train a model to identify an appropriate way to represent it internally, for as by grouping it, use unsupervised learning.*

The purpose of supervised machine learning is to create a model capable of making predictions based on data even when there is ambiguity. A supervised learning technique trains a model to generate good predictions about the response to new data using a known set of input data and previous responses to the data (output).

Using Supervised Learning to Predict Heart Attacks as an Example: Assume doctors want to determine if someone will suffer a heart attack in the coming year. They have information on former patients' age, weight, height, and blood pressure. They know if any of the previous patients had heart attacks within a year. The challenge is to create a model using existing data that can predict if a new person will have a heart attack in the coming year.

Supervised Learning Techniques: Every supervised learning method may be broken down into one of two categories: classification or regression. Methods like as classification and regression, which are employed in supervised learning, are put to use in the development of models that are able to forecast the future.

- **Classification techniques :** Classification methods make predictions about discrete outcomes, such as whether an e-mail is genuine or spam or if a tumour is cancerous or not. Classification models classify incoming data into categories. Imaging for medical purposes, speech recognition, and rating credit are a few examples of typical applications.
- **Regression methods :** Predictions can be made with regression algorithms about things like shifts in temperature or alterations in the quantity of power consumed. The most typical applications are stock price predicting, handwriting recognition, electricity load forecasting, acoustic signal processing, and other similar tasks.

Note:

- Is it possible to tag or categorise your data? Use classification techniques if your data can be divided into distinct groups or classes.
- Working with a collection of data? Use regression techniques if your answer is a real number, such as the temperature or the time until a piece of equipment fails.
- Binary vs. Multiclass Classification: Before you start working on a classification problem, figure out whether it's a binary or multiclass problem.

A single training or test item (instance) can only be classified into two classes in a binary classification task, such as determining whether an email is real or spam. If you wish to train a model to categorise a picture as a dog, cat, or other animal, for example, a multiclass classification problem might be separated into more than two categories. Remember that a multiclass classification problem is more difficult to solve since it necessitates a more sophisticated model. Certain techniques (such as logistic regression) are specifically intended for binary classification situations. These methods are more efficient than multiclass algorithms during training.

Now it's time to talk about the role of algorithms in machine learning. Algorithms are a very important part of how machine learning works, so it's important to talk about both of them. Discussion about algorithms and machine learning go hand in hand. They're the most important part of learning. In the world of computers, algorithms have been used for a long time to help us solve hard problems. They are a set of computer instructions for working with, changing, and interacting with data. An algorithm can be as simple as adding a column or as complicated as figuring out how to recognize anyone's face in a picture.

For an algorithm to work, it must be written as a programme that a computer can understand. Machine learning algorithms are usually written in either Java, Python, or R. Each of these languages has machine learning libraries that support a wide range of machine learning algorithms.

Active user communities for these languages share code and talk about ideas, problems, and ways to solve business problems. Machine learning algorithms are different from other algorithms. Most of the time, a programmer starts by typing in the algorithm. Machine learning turns the process around. With machine learning, the data itself creates the model. When you add more data to an algorithm, it gets harder to understand. As the machine learning algorithm gets more and more information, it can make more accurate algorithms.

It's a mix of science and art, to choose the right kind of machine learning algorithm. If you ask two data scientists to solve the same business problem, they might do it in different ways. But data scientists can figure out which machine learning algorithms work best if they know the different kinds. So, the most important step after getting the data in the right format is to choose the right machine learning algorithm.

As a result of our earlier discussion, we understood that choosing a right algorithm for machine learning is a process of trial and error. There is also a contradiction between certain aspects of the algorithms, such as:

- the amount of time spent in training;
- the amount of memory needed;
- the accuracy with which predictions are made on new data; and
- the level of transparency or interpretability (how easily you can understand the reasons an algorithm makes its predictions)

Let's take a closer look at the most commonly used machine learning algorithms.

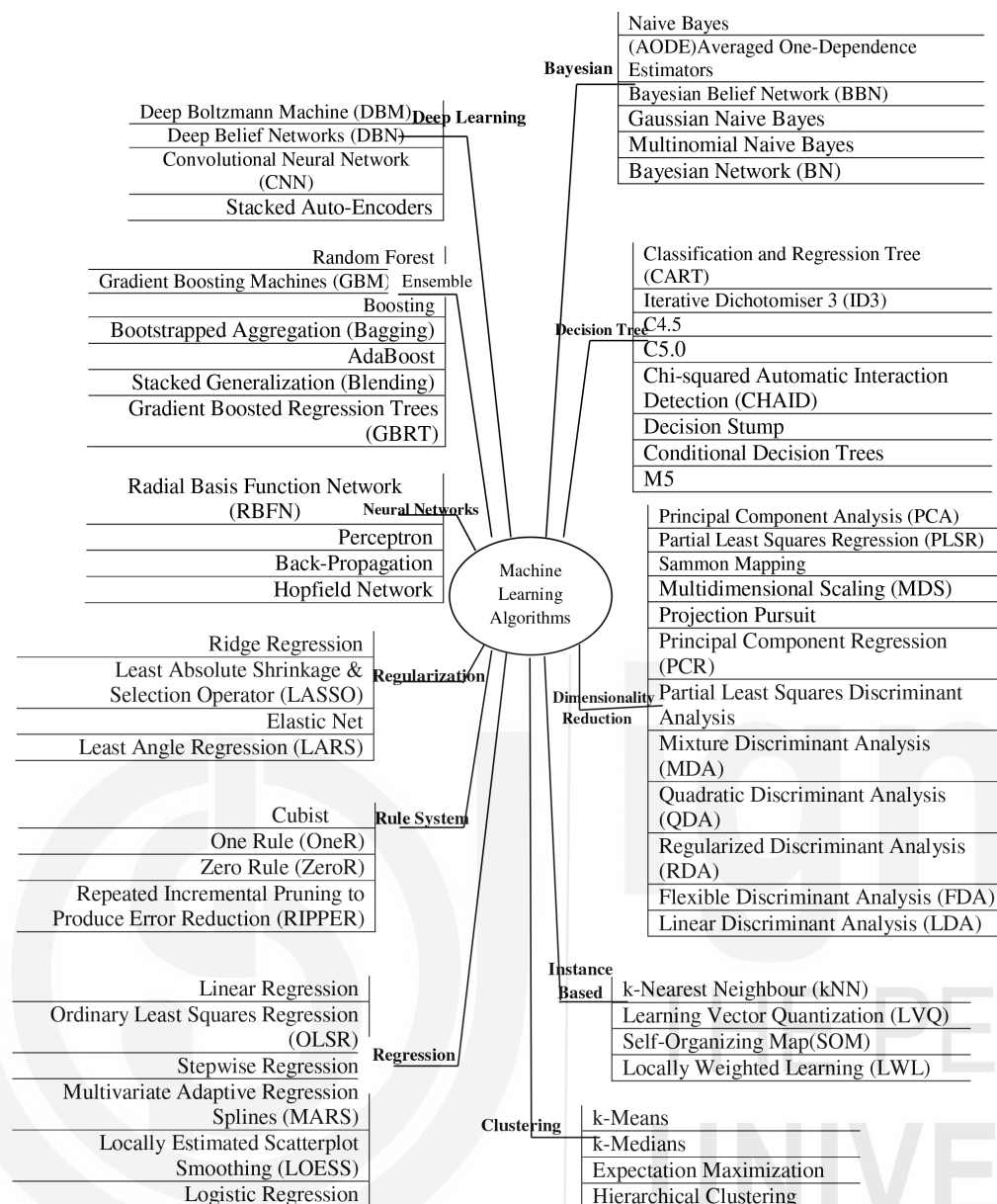


Figure – 4 : Types of machine learning algorithms

A brief discussion of the main types of machine learning algorithms, is given below :

- **Bayesian:** Regardless of what the data shows, data scientists can use Bayesian algorithms to save their ideas about how models should look. Given how much attention is devoted to how the data shapes the model, you might ask why anyone would be interested in Bayesian algorithms. When you don't have much data to work with, Bayesian techniques come in handy.

If you already knew something about a part of the model and could code that part directly, a Bayesian algorithm might make sense. Consider a medical imaging system that looks for signs of lung disease. These estimates can be incorporated into the model if a study published in a journal calculates the likelihood of various lung diseases based on a person's lifestyle.

- **Clustering:** Clustering is an easy-to-understand approach. Objects with comparable properties are combined (in a cluster). A cluster's contents are

more similar than those of other clusters. Because the data are not labelled, clustering is a sort of unsupervised learning. Based on the parameters, the algorithm determines what each item is made of and assigns it to the appropriate group.

- **Decision tree :** Decision tree algorithms show what will happen when a choice is made by using a structure with branches. Decision trees can be used to show all the possible outcomes of a choice. A decision tree shows all the possible outcomes at each branch. The likelihood of the outcome is shown as a percentage for each node.

Sometimes, online sales use decision trees. You might want to figure out who is most likely to use a 50% off coupon before sending it to them. Customers can be split into four groups:

- a) Customers who are likely to use the code if they get a personal message.
- b) Customers who will buy no matter what.
- c) Customers who will never buy.
- d) Customers who are likely to be upset if someone tries to reach out to them.

If you send out a campaign, it's obvious that you don't want to send items to three of the groups since they will either ignore them or respond negatively. You'll get the best return on investment (ROI) if you go after the convenience.

A decision tree will assist you in identifying these four client categories and organizing prospects and customers according to who will respond best to the marketing campaign.

- **Dimensionality reduction :** Dimensionality reduction allows systems to eliminate redundant data. These approaches remove data that is redundant, outliers, or otherwise useless. Sensor data and other IoT use cases can benefit from dimensionality reduction. The status of a sensor in an IoT system can be communicated using thousands of data points. Storing and analyzing "on" data is wasteful and wastes storage. Furthermore, minimizing redundant data increases machine learning system performance. Finally, data visualization is also benefited from dimensionality reduction.
- **Instance based :** Instance-based algorithms are used to classify new data points based on training data. Because there's no training phase, these algorithms are called "lazy learners." Instead, instance-based algorithms compare new data to training data and classify it based on how similar it is. Data sets with random changes, irrelevant data, or missing values are not good for instance-based learning.

They can be quite good at finding patterns.

For example, instance learning is used in spatial and chemical structure analysis. There are many instance-based algorithms used in biology, pharmacology, chemistry, and engineering.

- **Neural networks and deep learning :** A neural network is an artificial intelligence system that attempts to solve problems in the same way that the human brain does. This is accomplished by the utilisation of many layers of interconnected units that acquire knowledge from data and infer linkages. In a neural network, the layers can be connected to one another in various ways. When referring to the process of learning that takes place within a neural network with multiple hidden layers, the term "deep learning" is frequently used. Models built with neural networks are able to adapt to new information and gain knowledge from it. Neural networks are frequently utilised in situations in which the data in question is not tagged or is not organised in a particular fashion. The field of computer vision is quickly becoming one of the most important applications for neural networks. Today, one can find applications for deep learning in a diverse range of contexts.

The process of deep learning is utilised to assist self-driving autos in figuring out what is going on in their surroundings. Deep learning algorithms analyse the unstructured data that is being collected by the cameras as they capture pictures of the environment around them. This allows the system to make judgments in what is essentially real time. The apps that radiologists use to better analyse medical images also include deep learning as an integral part of their design.

- **Linear regression :** Regression algorithms are important in machine learning and are often used for statistical analysis. Regression algorithms help analysts figure out how data points are related.

Regression algorithms can measure how strongly two variables in a set of data are linked to each other. Regression analysis can also be used to predict the values of data in the future based on their past values. But it's important to remember that regression analysis is based on the idea that correlation means cause. Regression analysis can lead to wrong conclusions if you don't understand the context of the data.

- **Regularization to avoid over-fitting :** The process of regularisation involves modifying models in such a way that they no longer fit too well into the data . Any model that is used for machine learning can benefit from using regularisation. For example, you can regularise a decision tree model. Models that are excessively intricate and have a tendency to be overfit can become easier to grasp with the help of regularisation. A model that has been overfit to the available data will produce accurate predictions when additional data sets are added to it. When a model is developed for a particular set of data, but that model is unable to produce accurate predictions when applied to a more general set of data, this is an example of overfitting.
- **Rule-based machine learning :** Rule-based machine learning algorithms describe data with the help of rules about relationships. A rule-based system is different from a machine learning system, which builds a model that can be used on all the data. Rule-based systems are easy to understand in general: if X data is put in, do Y. A rule-based approach to machine learning, on the other hand, can get very complicated as systems get more

complicated. For example, a system might have 100 rules that are already set. As the system gets more and more data and learns how to use it, it is likely that hundreds of rules will be broken. When making a rule-based approach, it's important to make sure it doesn't get so complicated that it stops being clear.

Think about how hard it would be to make an algorithm based on rules to apply the GST codes.

Check your progress - 2

Q3 Discuss the various phases of Machine Learning.

.....

.....

.....

.....

Q4 When should we use machine learning ?

.....

.....

.....

.....

Q5 Compare the concept of Classification, Regression and Clustering? List the algorithms in respective categories.

.....

.....

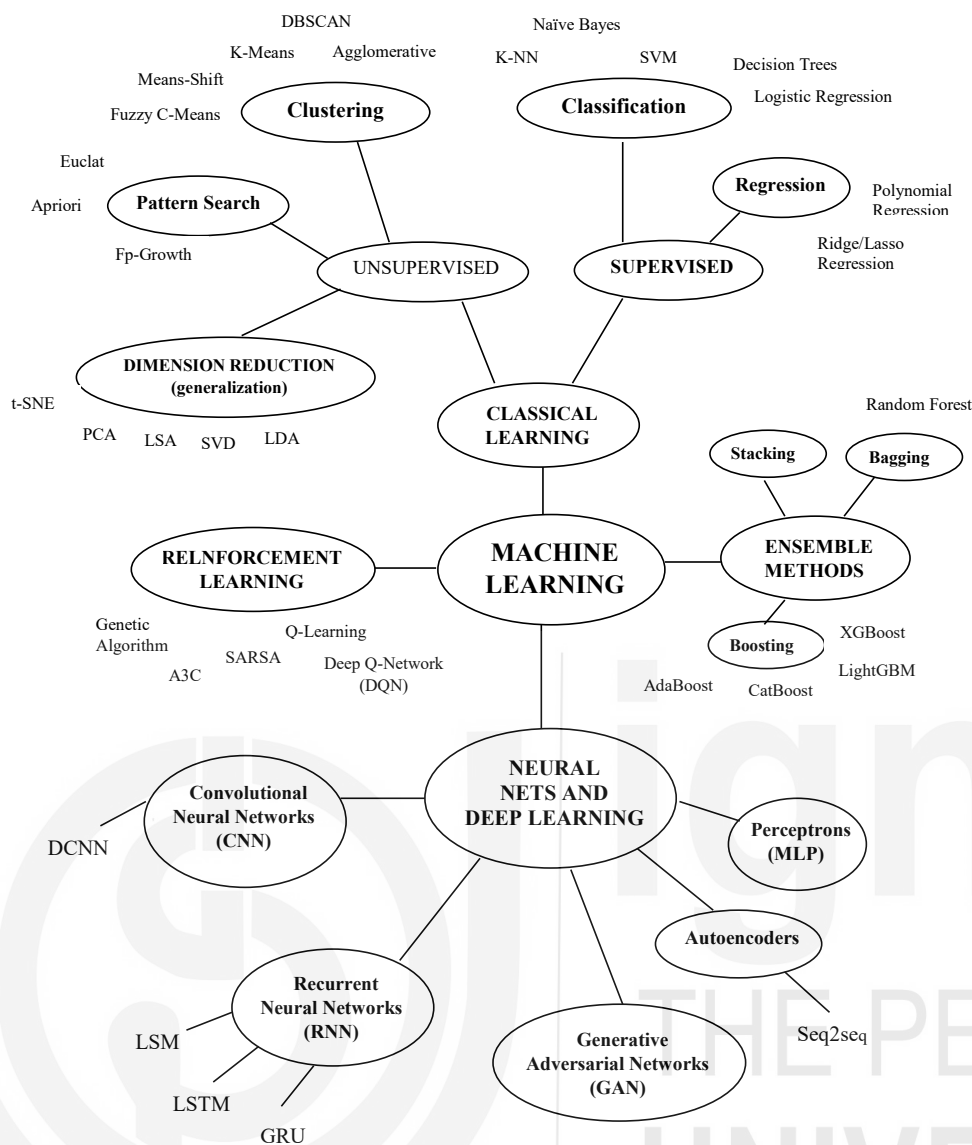
.....

.....

9.4 REINFORCEMENT LEARNING AND ALGORITHMS

According to what we observed in the previous section, learning can be broken down into three main categories: supervised, unsupervised, and semi-supervised. However, in addition to these two categories, there are also other types of learning, such as reinforcement learning (RL), deep learning (DL), adaptive learning, and so on.

The graph shown below, depicts the various branches and sub-branches of Machine learning, including the various algorithms involved in each sub-branch. Let's understand them in brief, as the entire coverage of the said Machine Learning techniques is out of the scope of this unit. We will begin our discussion with Reinforcement learning.



In Reinforcement Learning (RL), algorithms get a set of instructions and rules and then figure out how to handle a task by trying things out and seeing what works and what doesn't. As a way to help the AI find the best way to solve a problem, decisions are either rewarded or punished.

Machine learning models are taught through reinforcement learning to make a series of decisions. It is set up so that an Agent talks to an Environment.

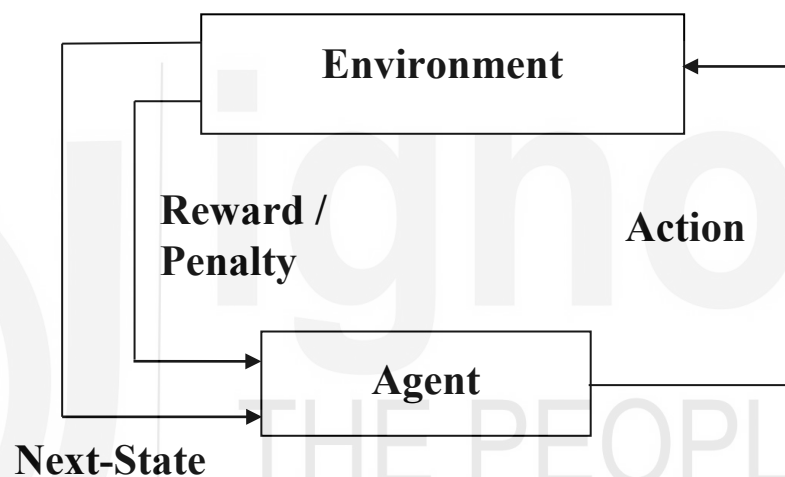
Reinforcement Learning (RL) is a type of Machine Learning in which the agent gets a delayed reward in the next time step to evaluate how well it did in the previous time step. It was mostly used in games, like Atari and Mario, where it could do as well as or better than a person. Since Neural Networks have been added to the algorithm, it has been able to do more complicated tasks.

In reinforcement learning, an AI system is put in a situation that is like a game (i.e. a simulation). The AI system tries until it finds a solution to the problem. Slowly but surely, the agent learns how to reach a goal in an uncertain, potentially complicated environment, but we can't expect the agent to slip upon the perfect solution by accident. This is where the interactions come into play, the Agent is provided with the State of the Environment which becomes the input/basis for

the Agent to take Action. An Action first gives the Agent a Reward. (Note that rewards can be both positive and negative depending on the fitness function for the problem.) Based on this reward, the Policy (ML model) inside the Agent adapts and learns. Second, it affects the Environment and changes its State, which means the input for the next cycle changes.

This cycle will keep going until the best Agent is created. This cycle tries to imitate the way that organisms learn over the course of their lives. Most of the time, the Environment is reset after a certain number of cycles or if something goes wrong. Note that you can run more than one Agent at the same time to get to the solution faster, but each Agent runs on its own, independently.

Reinforcement Learning (RL) refers to a kind of Machine Learning method in which the agent receives a delayed reward in the next time step to evaluate its previous action. It was mostly used in games. Typically, a RL setup is composed of two components, an agent and an environment.



The following are the meanings of the different parts of reinforcement learning:

1. **AGENT** : The agent is the person who learns and makes decisions.
2. **ENVIRONMENT** : The agent's environment is where it learns and decides what to do.
3. **ACTION** : A group of things that the agent can do.
4. **STATE** : How the agent is doing in its environment.
5. **REWARD** : The environment gives the agent a reward for each action they choose. Usually a scalar value.
6. **POLICY** : Policy is the agent's way of deciding what to do (its control strategy), which is a mapping from situations to actions.
7. **VALUE FUNCTION** : A way to map states to real numbers, where the value of a state is the long-term reward that can be earned by starting in that state and following a certain policy.
8. **FUNCTION APPROXIMATOR** : is a term for the problem of figuring out what a function is by looking at training examples. Decision trees, neural

networks, and nearest-neighbor methods are all examples of standard approximators.

9. **MODEL** : The agent's view of the environment, which maps state-action pairs to probability distributions over states. Note that not every agent that learns from its environment uses a model of its environment.

In spite of the fact that there is a large number of RL algorithms, it does not appear that there is a comparison that is exhaustive of each of them. It is quite challenging to determine which algorithms should be used for which type of activity. This section will attempt to provide an introduction to several well-known algorithms.

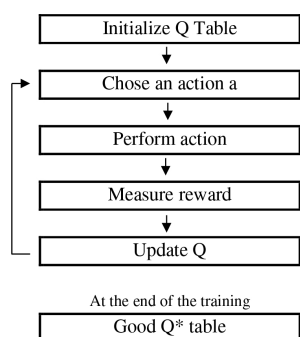
The algorithms for reinforcement learning may be broken down into two broad categories: model-free and model-based. In this section, we will analyse the key differences between these two types of reinforcement learning algorithms.

Model-Free Vs Model Based RL: The model is used to perform a simulation of the dynamic processes that take place in the environment. In other words, the model learns the transition probability $T(s_1|(s_0, a))$ from the present state s_0 and action a to the next state s_1 , and it does so by pairing the two states together. If the agent is able to successfully learn the transition probability, then the agent will be aware of how probable it is to reach a particular state given the present state and activity. On the other hand, as the state space and the action space grow, model-based algorithms become less practical.

On the other hand, model-free algorithms acquire new information through an iterative process of trial and error. As a consequence of this, it does not need any additional space in order to store every possible combination of states and actions.

Within the realm of Model-Free RL, policy optimization serves as a subclass, and it is comprised of two distinct sorts of policies. i.e. On-Policy Vs Off-Policy: The value is learned by an on-policy agent based on its current action "a" which is derived from the current policy, but the value is learned by an off-policy agent's counterpart based on the action "a*" which is received from another policy. This policy is referred to as the greedy policy in Q-learning.

The Q-learning or value-iteration methods are the next subcategory that is included in Model-Free RL. Q-learning is responsible for the acquisition of the action-value function. How advantageous would it be to perform a certain action at a certain state? In its most basic form, the action "a" receives a scalar value that is determined by the state "s". The algorithm is shown in the following chart, which does a good job of conveying its details.



Lets extend our discussion to some more Reinforcement Learning Algorithms i.e. DQN and SARSA

Deep Q Neural Network (DQN): It is Q-learning with Neural Networks . The motivation behind is simply related to big state space environments where defining a Q-table would be a very complex, challenging and time-consuming task. Instead of a Q-table Neural Networks approximate Q-values for each action based on the state.

State–action–reward–state–action (SARSA): SARSA algorithm is a slight variation of the popular Q-Learning algorithm. For a learning agent in any Reinforcement Learning algorithm it's policy can be of two types:-

- **On Policy:** In this, the learning agent learns the value function according to the current action derived from the policy currently being used.
- **Off Policy:** In this, the learning agent learns the value function according to the action derived from another policy.

The Q-Learning technique is considered an Off Policy technique that employs the greedy learning strategy in order to acquire knowledge of the Q-value. On the other hand, the SARSA approach is an On Policy and it makes advantage of the action that is being performed by the current policy in order to learn the Q-value.

Text mining, facial recognition, city planning, and targeted marketing are some of the applications which are actually the implementation of unsupervised learning algorithms. In a similar manner, the classification methods that fall under the supervised learning umbrella have applications in the areas of fraud detection, spam detection, diagnostics, picture classification, and score prediction. Similarly , reinforcement learning has a wide range of applications in a variety of fields, including the gaming industry, manufacturing, inventory management, and the financial sector, among many others..

Check Your Progress – 3

Q6 What is Reinforcement Learning ? List the components involved in it.

.....

.....

.....

Q7 Briefly discuss the various algorithms of Reinforcement Learning.

.....

.....

.....

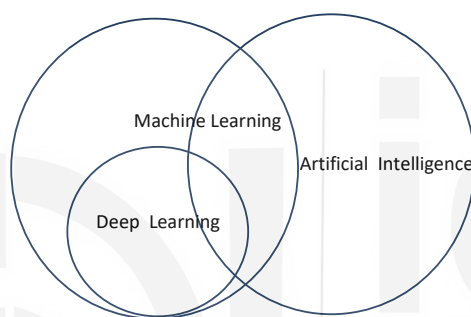
9.5 DEEP LEARNING AND ALGORITHMS

Deep learning is a type of machine learning that uses artificial neural networks and representation learning. It is also called deep structured learning or differential programming. Deep learning is a way for machines to learn through

deep neural networks. It is used a lot to solve practical problems in fields like computer vision (image), natural language processing (text), and automated speech recognition (audio). Machine learning is often thought of as a tool with several algorithms. However, deep learning is actually just a subset of approaches that mostly use neural networks, which are a type of algorithm loosely based on the human brain.

A deep learning model learns to solve classification tasks directly from images, text, or sound. A neural network architecture is commonly used to implement deep learning. The number of layers in a network defines the depth of the network; the more layers, the deeper the network. Traditional neural networks have two or three layers, whereas deep neural networks include hundreds.

Deep learning is especially well-suited to identification applications such as face recognition, text translation, voice recognition, and advanced driver assistance systems, including, lane classification and traffic sign recognition.

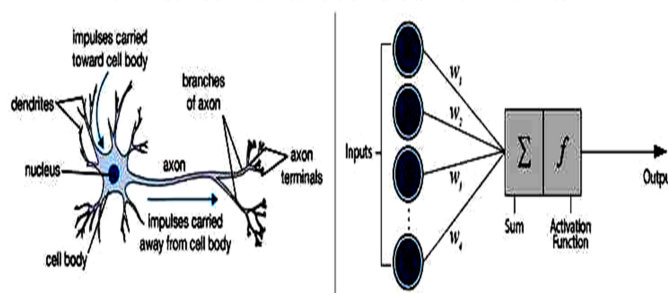


Relation Between Machine learning , Deep Learning and Artificial Intelligence

As seen in the diagram above, machine learning (ML), deep learning (DL), and artificial intelligence (AI) are all related. Deep Learning is a collection of algorithms inspired by the human brain's workings in processing the data and creating patterns for use in decision making, which are expanding and improving or refining the idea of a single model architecture termed Artificial Neural Network (ANN). Later in this course, we shall go deeper into neural networks. However for now, a quick overview of neural networks is provided below, followed by a discussion of the various Deep Learning algorithms, such as CNN, RNN, Auto Encoders, GAN, and others..

Neural Networks: Just like the human brain, Neural Networks consist of Neurons. Each Neuron takes in signals as input, multiplies them by weights, adds them together, and then applies a non-linear function. These neurons are arranged in layers and stacked close to each other.

Biological Neuron versus Artificial Neural Network



Neural Networks have proven to be effective function approximators. We can presume that every behaviour and system can be represented mathematically at some point (sometimes an incredible complex one). If we can find that function, we will know everything there is to know about the system. However, locating the function can be difficult. As a result, we must use Neural Networks to estimate it.

A deep neural network is one that incorporates several nonlinear processing layers, makes use of simple pieces that work in parallel, and takes its cues from the biological nervous systems of living things. There is an input layer, numerous hidden layers, and an output layer that make up this structure. Each hidden layer takes as its input the information that was output by the layer that came before it and is connected to the other layers via nodes, also known as neurons.

To understand the basic deep neural networks we need to have brief understanding of various algorithms, the same are given below:

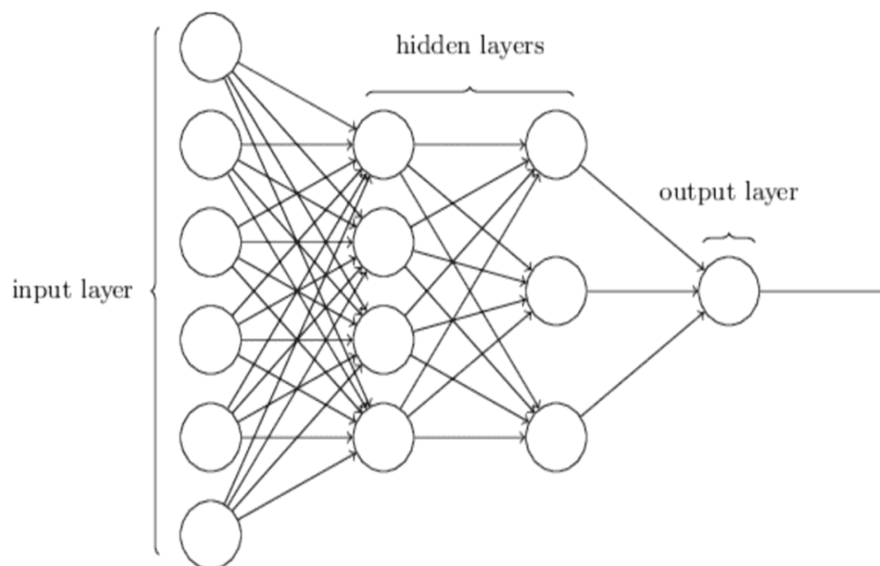
Back Propagation Neural Networks: Propagation in reverse Back propagation is an iterative process that allows neural networks to learn the desired function by utilising large quantities of data and learning from their previous mistakes. We feed the network data, and in return, it provides us with an output. We start by comparing the output to what we want using a loss function. Then, based on the gap that we find, we iteratively adjust the weights of the various variables. This non-linear optimization procedure is termed stochastic gradient descent, and it is used to make the necessary modification to the weights.

After some time, the network will improve its ability to provide the output to a very high standard. As a result, the training is complete. As a result, we are able to come close to approximating our function. In addition, if we give the network an input for which we do not know the corresponding output, it will provide us with an answer based on the approximated function.

To further understand, let's look at an example. Let's say we need to recognise pictures that contain a tree. Photos are input into the network, and the system produces results. We might evaluate the results in light of our current situation and make adjustments to the network accordingly.

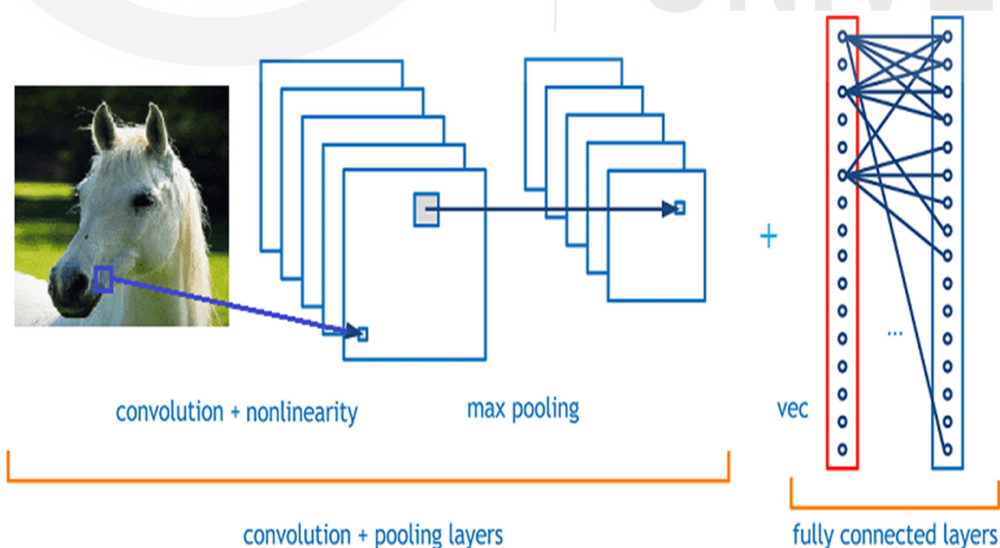
As more photographs are passed via the network, the number of errors that occur decreases. We can now feed it an unknown image, and it will tell us whether or not it has a tree. or not, that is astounding either way.

Feed-forward neural networks (FNN) : Typically, feed-forward neural networks, also known as FNN, are completely connected, this implies that each neuron in one layer is connected to each neuron in the layer next to it. A "Multilayer Perceptron" is the name given to the structure shown below and that is the topic of discussion here. A multilayer perceptron, has the ability to learn associations between the data that are not linear, in contrast to a single-layer perceptron, which can only learn patterns that can be separated in a linear manner. FNN are exceptionally well on tasks like classification and regression. Contrary to other machine learning algorithms, they don't converge so easily. The more data they have, the higher their accuracy.



Convolutional Neural Networks (CNN) The term "convolution" refers to the function that is utilised by convolutional neural networks (CNN). The idea that underlies them is that rather than linking each neuron with all of the ones that come after it, we just connect it with a select few of those that come after it (the receptive field). They strive to regularise feed-forward networks in order to avoid overfitting, which is when the model is unable to generalise its findings since it can only learn from the data it has already seen. Because of this, they are particularly skilled at determining how the data are related to one another spatially. As a result, computer vision is their primary application, which includes image classification, video identification, medical image analysis, and self-driving automobiles. These are the types of tasks where they achieve near-superhuman results.

Due to their adaptability, they are also ideal for merging with other types of models, such as Recurrent Networks and Auto-encoders. The recognition of sign languages is one such example.



Face Recognition Based on Convolutional Neural Network

Recurrent Neural Networks (RNN) are utilised in time series forecasting because they are ideal for time-related data. They employ some type of feedback,

in which the output is fed back into the input. You can think of it as a loop that passes data back to the network from the output to the input. As a result, they are able to recall previous data and use it to make predictions.

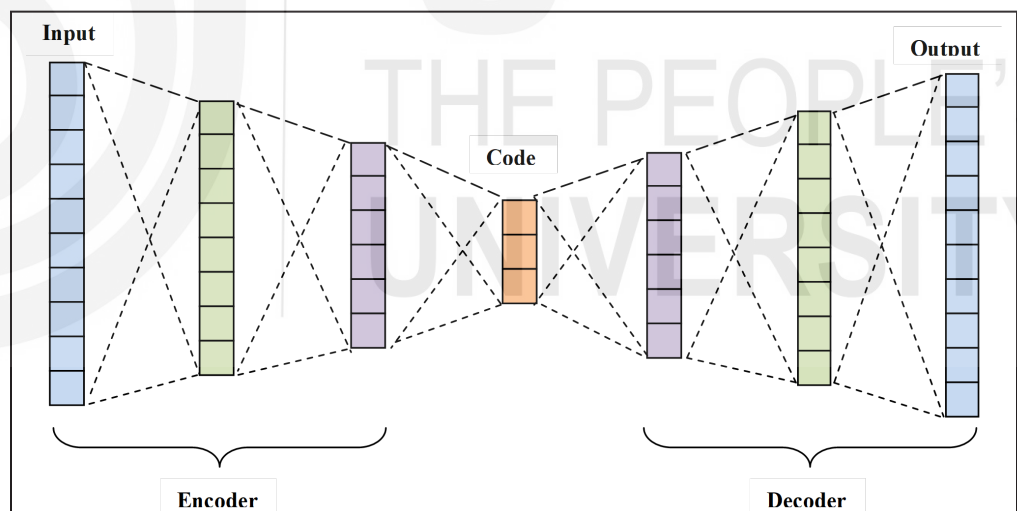
Researchers have transformed the original neuron into more complicated structures such as GRU units and LSTM Units to improve performance. Language translation, speech production, and text to speech synthesis have all employed LSTM units extensively in natural language processing.

Recursive Neural Networks : Another type of recurrent network is the recursive neural network, which is set up in a tree-like manner. As a result, they can simulate the hierarchical structures training dataset's.

They're frequently utilised in NLP applications like audio-to-text transcription and sentiment analysis because they're related to binary trees, contexts, and natural-language-based parsers. They are, however, typically much slower than Recurrent Networks.

Auto-Encoders (Auto Encoder Neural Networks) are a type of unsupervised technique that is used to reduce dimensionality and compress data. Their technique is to try and make the output equal to the input. They are attempting to recreate the data.

An encoder and a decoder are included in Auto-Encoders. The encoder receives the input and encodes it in a lower-dimensional latent space. Whereas, the decoder is used to decode that vector back to the original input.



Restricted Boltzmann Machines (RBM) are stochastic neural networks that can learn a probability distribution over their inputs and so have generative capabilities. They differ from other networks in that they only have input and hidden layers (no outputs).

They take the input and create a representation of it in the forward phase of the training. They rebuild the original input from the representation in the backward pass. (This is similar to autoencoders, but in a single network.)

Several RBMs are piled on top of each other to form a Deep Belief Network. They have the same appearance as Fully Connected layers, but they are trained differently.

Generative Adversarial Networks (GANs): Ian Goodfellow introduced Generative Adversarial Networks (GANs) in 2016, and they are built on a basic but elegant idea: You need to create data, such as photos. What exactly do you do?

You must construct two models. You teach the first one to make up fake data (generator) and the second one to tell the difference between actual and fake data (discriminator). And you turned them against one another.

The generator develops better and better at image production, as its ultimate purpose is to mislead the discriminator. As its purpose is to avoid being tricked, the discriminator improves its ability to identify fake from real images. As a result, we now have extremely realistic fake data from the discriminator.

Video games, astronomical imagery, interior design, and fashion are all examples of Generative Adversarial Networks at action. Essentially, you can utilise GANs if you have photos in your fields. Do you recall the movie Deep Fakes? That was all created by GANs.

Transformers are also very new, and they are mostly employed in language applications because recurrent networks are becoming obsolete. They are based on the concept of "attention," which instructs the network to focus on a certain data piece.

Instead of complicating LSTM units, you may use Attention mechanisms to assign varying weights to different regions of the input based on their importance. The attention mechanism is simply another weighted layer whose sole purpose is to change the weights such that some parts of the inputs are given greater weight than others.

In actuality, transformers are made up of stacked encoders (encoder layer), stacked decoders (decoder layer), including several attention layers (self-attentions and encoder-decoder attentions)

Graph Neural Networks: Deep Learning does not operate well with unstructured data in general. And there are many circumstances in which unstructured data is organised as a graph in the actual world. Consider social networks, chemical molecules, knowledge graphs, and location information.

Graph Neural Networks are used to model graph data. This implies they locate and convert the connections between nodes in a network into integers. As if it were an embedding. As a result, they can be utilized in any other machine learning model to perform tasks such as grouping, classifying, and so on.

Check Your Progress – 4

Q8 What is Deep Learning ?How Deep learning relates to AI & ML.

.....

.....

.....

.....

.....

.....

.....

.....

9.6 ENSEMBLE METHODS

Ensemble learning is a general meta approach to machine learning that combines predictions from different models to improve predictive performance.

Although you can create an apparently infinite number of ensembles for any predictive modelling problem, the subject of ensemble learning is dominated by three methods. Bagging, stacking, and boosting. They are the three primary classes of ensemble learning methods, and it's essential to understand each one thoroughly.

- **Bagging Ensemble learning** is the process of fitting multiple decision trees to various samples of the same dataset and averaging the results.
- **Stacking Ensemble learning** is fitting multiple types of models to the same data and then using another model to learn how to combine the predictions in the best way possible.
- **Boosting Ensemble Learning** entails successively adding ensemble members that correct prior model predictions and produce a weighted average of the predictions.

Now let's discuss each of the learning method in some detail

(I) Bagging Ensemble learning Bagging ensemble learning involves fitting numerous decision trees to various samples of the same dataset, and then averaging the results of those tree fittings to provide a final prediction.

In most cases, this is accomplished by making use of a single machine learning method, which is nearly invariably an unpruned decision tree, and by training each model on a separate sample from the same training dataset. After then, straightforward statistical approaches such as voting or averaging are utilised in order to aggregate the predictions that were generated by each individual participant in the ensemble.

The manner in which each individual data sample is prepared to train members of the ensemble constitutes the most essential component of the technique. Every model receives its own unique, customised portion of the dataset to use for testing. Rows (examples) are selected at random from the dataset, and once selected, they are replaced.

When a row is selected, it is added back to the dataset that it was learned from, so that it can be selected once more from the same training dataset. This indicates that within a specific training dataset, a row of data may be selected 0 times, 1 times, or multiple times.

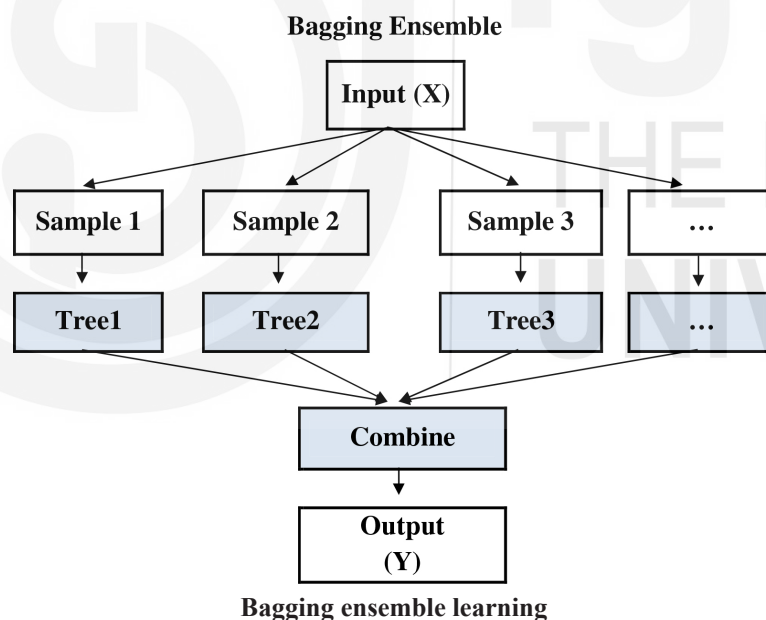
This type of sample is known as a bootstrap sample. In the field of statistics, this approach is a way for estimating the statistical value of a limited data sample. It is typically applied to somewhat limited data sets. You can get a better overall estimate of the desired quantity if you make a number of distinct bootstrap samples, estimate a statistical quantity, and then determine the average of the estimates. This is in comparison to the situation in which you would just estimate the quantity based on the dataset.

In the same way, several training datasets can be compiled, put to use in the process of estimating a predictive model, and then put to use in order to produce predictions. The majority of the time, it is preferable to take the average of the predictions made by all of the models rather than to fit a single model directly to the dataset used for training.

The following is a concise summary of the most important aspects of bagging:

- Take samples of the training dataset using bootstrapping.
- Unpruned decision trees fit on each sample.
- Voting or taking the average of all the predictions.

In a nutshell, bagging has an effect because it modifies the training data that is used to fit each individual member of the ensemble. This results in skillful but unique models.



It is a comprehensive strategy that is simple to expand upon. For instance, additional alterations can be made to the dataset that was used for training, the method that was used to fit the training data can be modified, and the manner in which predictions are constructed can be altered.

Many popular ensemble algorithms are based on this approach, including:

- Bagged Decision Trees (canonical bagging)
- Random Forest
- Extra Trees

(II) Stacking Ensemble learning: Stacked Generalization, sometimes known as "stacking" due to its abbreviated form, is an ensemble strategy that searches for a diverse group of members by varying the types of models that are fitted to the training data and utilising a model to aggregate predictions. It requires fitting of various kinds of models, applied to the same data, and then using another model to find out how to integrate the predictions in the best way possible. This process is known as model fitting.

There is a specific vocabulary for stacking. The individual models that comprise an ensemble are referred to as level-0 models, whereas the model that integrates all of the predictions is referred to as a level-1 model.

Although there are often only two levels of models applied, you are free to apply as many levels as you see fit. For instance, instead of a single level-1 model, we might have three or five level-1 models and a single level-2 model that integrates the forecasts of level-1 models to generate a prediction. This would allow us to make more accurate predictions.

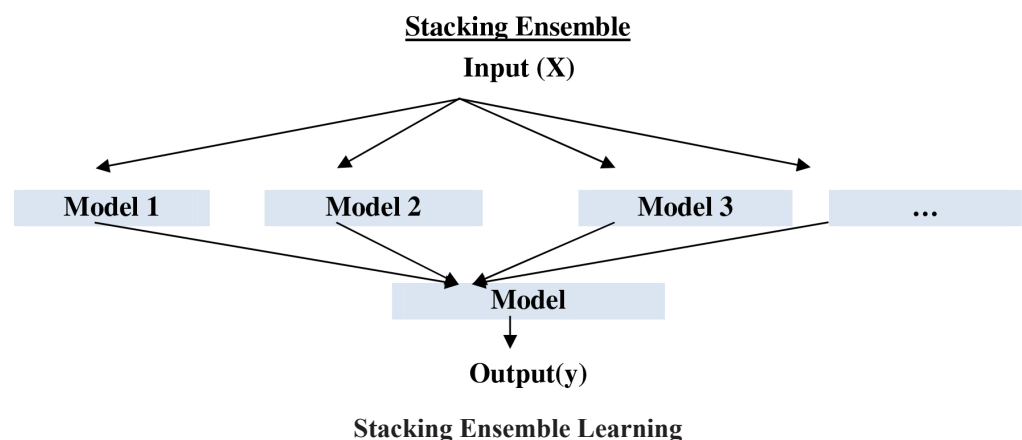
It is possible to integrate the predictions using any machine learning model, but the majority of users prefer linear models, such as linear regression for regression and logistic regression for binary classification. Because of this, it is more likely that the more difficult components of the model will be included in the lower-level ensemble member models, and that straightforward models will be used to learn how to apply the various predictions.

The key elements of stacking are summarized below:

- Unchanged training dataset.
- Separate machine learning algorithms for respective ensemble member.
- Machine learning model to learn how to combine predictions in the best way.

The diversity of the ensemble is a direct result of the many diverse machine learning models that serve as the ensemble's members.

As a consequence of this, it is recommended to make use of a variety of models that can be learnt or constructed in a wide variety of methods. Because of this, it is ensured that they will make separate assumptions, and as a consequence, it is less probable that their errors in prediction would be linked to one another.



Many popular ensemble algorithms are based on this approach, including:

- Stacked Models (canonical stacking)
- Blending
- Super Ensemble

(III) Boosting Ensemble learning: Boosting is an ensemble strategy that aims to alter the training data so that it focuses on examples that earlier models that fit the training data got wrong. Boosting tries to do this by focusing on examples that prior models got wrong. In order for it to function, members are added to the ensemble one at a time, and when each new member is added, the predictions that were produced by the model that came before it are refined. A weighted average of the forecasts is what we get as a result.

The fact that boosting ensembles may correct errors in forecasts is the single most important advantage of using them. The models are calibrated and introduced to the ensemble one at a time, which means that the second model attempts to fix what the first model indicated, and so on and so forth.

The majority of the time, this is accomplished using weak learners, which are relatively straightforward decision trees that only make a single or a few decisions at a time. The forecasts of the weak learners are merged by simple voting or by average, but the importance of each learner's input is weighted according to how well they performed or how much they know. The objective is to create a "strong-learner" out of a number of "weak-learners," each of which was designed to accomplish a particular task.

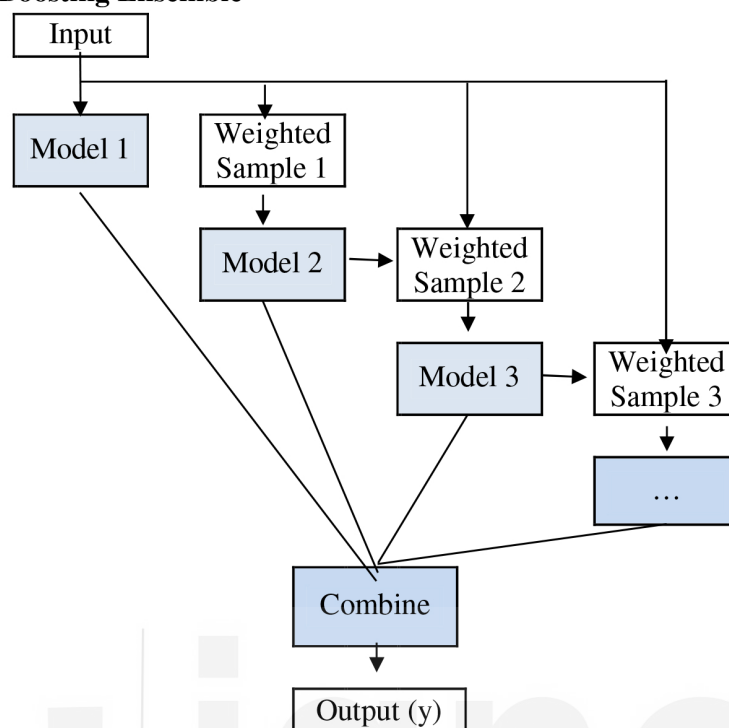
Majority of the time the training dataset is left unchanged ; instead, the learning algorithm is adjusted to pay more or less attention to certain examples (rows of data) depending on how well they were predicted by ensemble members who were added earlier. For instance, a weight could be assigned to each row of data in order to demonstrate the level of focus that a learning algorithm must maintain on the model while it is doing so.

The key elements of boosting are summarized below:

- Give more weight to examples that are hard to guess when training.
- Add members of the ensemble one at a time to correct the predictions of earlier models.
- Use a weighted average of models to combine their predictions.

The idea of turning a group of weak learners into a group of strong learners was first thought of in theory, and many algorithms were tried but didn't work very well. Until the Adaptive Boosting (AdaBoost) algorithm was made, it wasn't clear that boosting was a good way to put together a group of methods.

Since AdaBoost, many boosting methods have been made, and some, like stochastic gradient boosting, may be among the best ways to use tabular (structured) data for classification and regression.

Boosting Ensemble**Boosting Ensemble Learning**

To summarize, many popular ensemble algorithms are based on this approach, including:

- AdaBoost (canonical boosting)
- Gradient Boosting Machines
- Stochastic Gradient Boosting (XGBoost and similar)

This completes our tour of the standard ensemble learning techniques.

Check Your Progress – 5

Q10 What is Ensemble Learning ?

.....

.....

.....

Q11 Briefly discuss the various Ensemble Methods.

.....

.....

.....

9.7 SUMMARY

In this unit we discussed about the basic concepts of machine learning and also about the various Machine learning algorithms. The unit also covers the understanding of reinforcement learning and its related algorithms. There after

we discussed the concept of Deep Learning and various techniques involved in Deep Learning. The unit finally discussed about the Ensemble Learning and its related methods. The unit

9.8 SOLUTIONS/ANSWERS

Check Your Progress - 1

Q1 How machine learning differs from Artificial intelligence ?

Solution: Refer to Section 9.2

Q2 Briefly discuss the major function or use of Machine learning algorithms

Solution: Refer to Section 9.2

Check your progress - 2

Q3 Discuss the various phases of Machine Learning.

Solution: Refer to Section 9.3

Q4 When should we use machine learning ?

Solution: Refer to Section 9.3

Q5 Compare the concept of Classification, Regression and Clustering? List the algorithms in respective categories. 9.3

Solution: Refer to Section 9.3

Check Your Progress – 3

Q6 What is Reinforcement Learning ? List the various components involved in Reinforcement Learning.

Solution: Refer to Section 9.4

Q7 Briefly discuss the various algorithms of Reinforcement Learning.

Solution: Refer to Section 9.4

Check Your Progress – 4

Q8 What is Deep Learning ?How Deep learning relates to AI & ML.

Solution: Refer to Section 9.5

Q9 Briefly discuss the various algorithms of Reinforcement Learning.

Solution: Refer to Section 9.5

Check Your Progress – 5

Q10 What is Ensemble Learning ? 9.6

Solution: Refer to Section 9.6

Q11 Briefly discuss the various Ensemble Methods. 9.6

Solution: Refer to Section 9.6

9.9 FURTHER READINGS

- Prof. Ela Kumar, “Artificial Intelligence” Edition: First, Publisher: Dreamtech Press, (2020) ISBN: 9789389795134
- Machine learning an algorithm perspective, Stephen Marsland, 2nd Edition, CRC Press,, 2015.
- Machine Learning, Tom Mitchell, 1st Edition, McGraw- Hill, 1997.
- Machine Learning: The Art and Science of Algorithms that Make Sense of Data, Peter.

