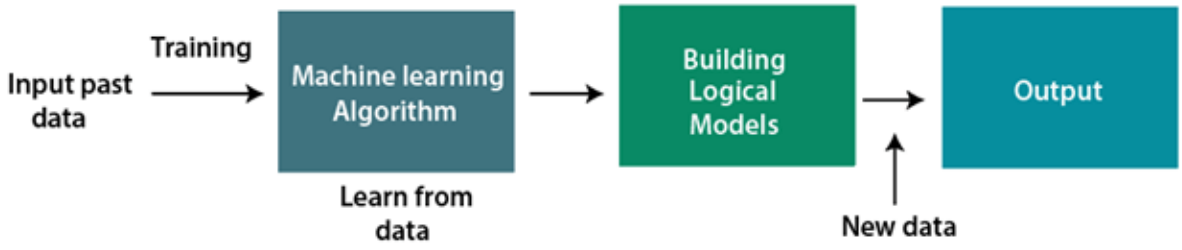
What is Machine Learning

- Machine Learning is said as a subset of artificial intelligence that is mainly concerned with the development of algorithms which allow a computer to learn from the data and past experiences on their own.
- The term machine learning was first introduced by Arthur Samuel in 1959
- Machine learning enables a machine to automatically learn from data, improve performance from experiences, and predict things without being explicitly programmed.

How does Machine Learning work

- A Machine Learning system learns from historical data, builds the prediction models, and whenever it receives new data, predicts the output for it.
- The accuracy of predicted output depends upon the amount of data, as the huge amount of data helps to build a better model which predicts the output more accurately.
- Suppose we have a complex problem, where we need to perform some predictions, so instead of writing a code for it, we just need to feed the data to generic algorithms, and with the help of these algorithms, machine builds the logic as per the data and predict the output.
- Machine learning has changed our way of thinking about the problem.
- The below block diagram explains the working of Machine Learning algorithm:



Features of Machine Learning:

- Machine learning uses data to detect various patterns in a given dataset.
- It can learn from past data and improve automatically.
- It is a data-driven technology.
- Machine learning is much similar to data mining as it also deals with the huge amount of the data.

Importance of Machine Learning:

- Machine learning is used in self-driving cars, cyber fraud detection, face recognition, and friend suggestion by Facebook, etc.
- Various top companies such as Netflix and Amazon have built machine learning models that are using a vast amount of data to analyze the user interest and recommend product accordingly.
- Rapid increment in the production of data
- Solving complex problems, which are difficult for a human
- Decision making in various sector including finance
- Finding hidden patterns and extracting useful information from data.

TYPES OF MACHINE LEARNING

1. **Supervised learning**
    - In supervised learning, the machine is taught by example.
    - The operator provides the machine learning algorithm with a known dataset that includes desired inputs and outputs, and the algorithm must find a method to determine how to arrive at those inputs and outputs.
    - While the operator knows the correct answers to the problem, the algorithm identifies patterns in data, learns from observations and makes predictions.
    - The algorithm makes predictions and is corrected by the operator – and this process continues until the algorithm achieves a high level of accuracy/performance.

There are two main types of supervised learning: classification and regression.

- **In classification,** the goal is to predict a discrete class label for an input. For example, given an image of a handwritten digit, the task is to predict the digit (0-9) that the image represents.
- **In regression**, the goal is to predict a continuous value for an input. For example, given data about a house, the task is to predict its price.

Example: Suppose we have an image of different types of fruits. The task of our supervised learning model is to identify the fruits and classify them accordingly. So, to identify the image in supervised learning, we will give the input data as well as output for that, which means we will train the model by the shape, size, color, and taste of each fruit. Once the training is completed, we will test the model by giving the new set of fruit. The model will identify the fruit and predict the output using a suitable algorithm.

2. **Unsupervised learning**

Unsupervised learning is a method of machine learning where an algorithm is trained on unlabeled data. The goal of unsupervised learning is to discover the underlying structure or pattern in the data, rather than to predict a specific output. Unlike supervised learning, there is no explicit feedback provided to the algorithm about the correct output for a given input.

There are two main types of unsupervised learning: clustering and dimensionality reduction.

- **In clustering**, the goal is to group similar data points together into clusters. For example, a clustering algorithm might group customers with similar purchasing habits into the same cluster.
- **In dimensionality reduction**, the goal is to reduce the number of features in the data while preserving as much information as possible. This can be useful for reducing noise in the data, visualizing high-dimensional data, or making the data more amenable to other forms of analysis.

The process of unsupervised learning typically involves the following steps:

1. Collecting and preparing the data
2. Selecting an appropriate algorithm and training it on the data
3. Evaluating the performance of the algorithm by visualizing the results and interpreting the discovered patterns or clusters
4. Using the trained algorithm to transform the data in some way

Unsupervised Learning algorithm are like PCA, K-means, Apriori algorithm, etc which are widely used in various domains as well.

The main differences between Supervised and Unsupervised learning are given below:

Supervised Learning	Unsupervised Learning
Supervised learning algorithms are trained using labeled data.	Unsupervised learning algorithms are trained using unlabeled data.
Supervised learning model takes direct feedback to check if it is predicting correct output or not.	Unsupervised learning model does not take any feedback.
Supervised learning model predicts the output.	Unsupervised learning model finds the hidden patterns in data.
In supervised learning, input data is provided to the model along with the output.	In unsupervised learning, only input data is provided to the model.
The goal of supervised learning is to train the model so that it can predict the output when it is given new data.	The goal of unsupervised learning is to find the hidden patterns and useful insights from the unknown dataset.
Supervised learning needs supervision to train the model.	Unsupervised learning does not need any supervision to train the model.
Supervised learning can be categorized in **Classification** and **Regression** problems.	Unsupervised Learning can be classified in **Clustering** and **Associations** problems.
Supervised learning can be used for those cases where we know the input as well as corresponding outputs.	Unsupervised learning can be used for those cases where we have only input data and no corresponding output data.
Supervised learning model produces an accurate result.	Unsupervised learning model may give less accurate result as compared to supervised learning.

Supervised learning is not close to true Artificial intelligence as in this, we first train the model for each data, and then only it can predict the correct output.	Unsupervised learning is closer to the true Artificial Intelligence as it learns similarly as a child learns daily routine things by his experiences.
It includes various algorithms such as Linear Regression, Logistic Regression, Support Vector Machine, Multi-class Classification, Decision tree, Bayesian Logic, etc.	It includes various algorithms such as Clustering, KNN, and Apriori algorithm.

3. **Rote learning**
    - Rote learning is possible on the basis of memorization.
    - This technique mainly focuses on memorization by avoiding the inner complexities.
    - So, it becomes possible for the learner to recall the stored knowledge.

Learning by memorization: which avoids understanding the inner complexities the subject that is being learned.

Learning something from Repeating: saying the same thing and trying to remember how to say it; it does not help to understand, it helps to remember, like we learn a poem, song, etc.

For example: When a learner learns a poem or song by reciting or repeating it, without knowing the actual meaning of the poem or song.

4. **Induction learning**
    - Induction learning is carried out on the basis of supervised learning.
    - In this learning process, a general rule is induced by the system from a set of observed instances.
    - However, class definitions can be constructed with the help of a classification method.

Some practical examples of induction are:

Credit risk assessment.

- The x is the property of the customer.
- The f(x) is credit approved or not.

Disease diagnosis.

- The x is the characteristics of a given patient.
- The f(x) is the patient's disease.

Face recognition.

- The x are bitmaps of the faces we want to recognize.
- The f(x) is a name assigned to that face.

Inductive Learning may be helpful in the following four situations:

- **Problems in which no human expertise is available**. People cannot write a program to solve a problem if they do not know the answer. These are areas ripe for exploration.

- **Humans can complete the task, but no one knows how to do it**. There are situations in which humans can do things that computers cannot or do not do well. Riding a bike or driving a car are two examples.
- **Problems where the desired function is frequently changing**. Humans could describe it and write a program to solve it, but the problem changes too frequently. It is not economical. The stock market is one example.
- **Problems where each user requires a unique function**. Writing a custom program for each user is not cost-effective. Consider Netflix or Amazon recommendations for movies or books.

5. **Explanation-Based Learning (EBL)**

Explanation-based learning (EBL) is a form of machine learning that uses previously acquired knowledge to improve the performance of future tasks. The core idea behind EBL is that, by understanding the reasons behind the success or failure of a previous task, the system can more effectively adapt to new situations.

The process of EBL typically consists of two main phases:

Learning Phase: In this phase, the system is presented with a set of examples and uses them to learn a set of rules or a decision tree.

Generalization Phase: In this phase, the system uses the knowledge acquired in the learning phase to solve new problems. In order to solve the new problem, the system generates explanations for its previous decisions and uses this information to make new decisions.

EBL Architecture:

EBL model during training

During training, the model generalizes the training example in such a way that all scenarios lead to the Goal Concept, not just in specific cases. (As shown in Fig 1)

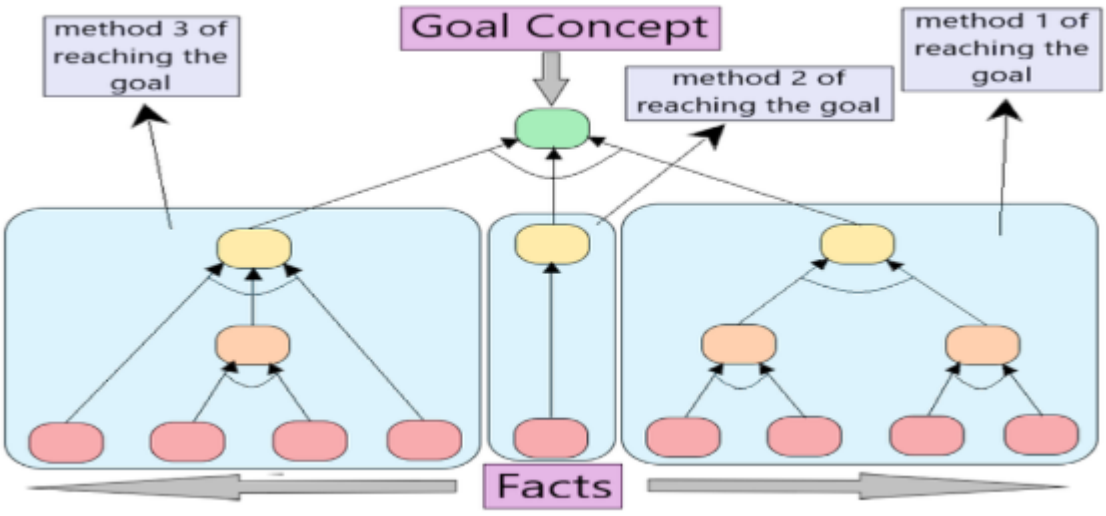


Fig 1 : Training EBL Model

EBL model after training

Post training, EBL model tends to directly reach the hypothesis space involving the goal concept. (As shown in Fig 2)

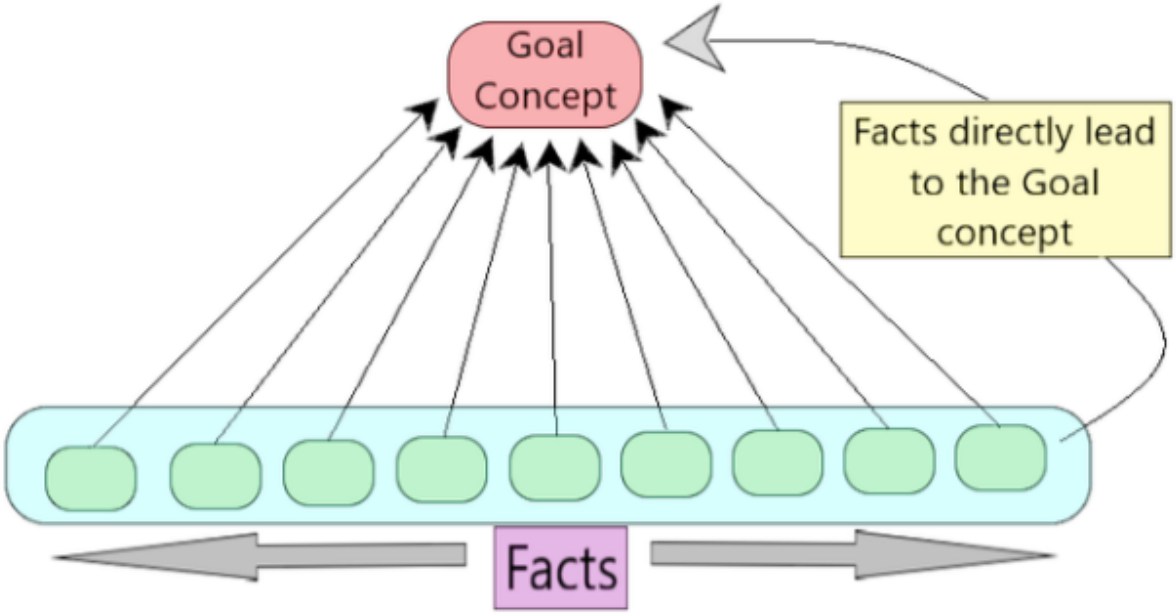


Fig 2 : Trained EBL Model

6. **Genetic algorithm (GA) / learning by evolution**

A genetic algorithm (GA) is a search heuristic that is inspired by the process of natural selection. GAs are used to find approximate solutions to optimization and search problems in AI and computer science.

The basic idea behind a genetic algorithm is to use a population of candidate solutions, called individuals or chromosomes, that evolve over time towards better solutions. The evolution is guided by a fitness function that measures the quality of an individual in terms of how well it solves the problem at hand.

The process of a genetic algorithm typically consists of the following steps:

- **Initialization:** A population of individuals is generated randomly.
- **Evaluation:** Each individual in the population is evaluated using the fitness function.
- **Selection:** Individuals with high fitness values are more likely to be selected to produce offspring.
- **Crossover:** The selected individuals are combined to create new individuals through a process called crossover or recombination.
- **Mutation:** Small random changes are introduced to the new individuals through a process called mutation.
- **Replacement:** The new individuals replace the old ones in the population and the process continues.
- The genetic algorithm terminates when a stopping criterion is met, for example when a certain number of generations has passed, or when a solution with a fitness above a certain threshold is found.
- **Termination:** After replacement has been done, a stopping criterion is used to provide the basis for termination. The algorithm will terminate after the

threshold fitness solution has been attained. It will identify this solution as the best solution in the population.

Genetic Algorithm is widely used for optimization in various domains such as manufacturing, engineering, and finance. Also, it is commonly used to optimize the parameters of other machine learning algorithms, such as neural networks, decision trees and support vector machines.

After replacement has been done, a stopping criterion is used to provide the basis for termination. The algorithm will terminate after the threshold fitness solution has been attained. It will identify this solution as the best solution in the population.

7. **Learn by analogy**

Learning by analogy is a form of machine learning in which a system learns new knowledge by comparing it to previously acquired knowledge. The system finds similarities between the new knowledge and the previously acquired knowledge, and then uses the previously acquired knowledge to understand the new knowledge.

The process of learning by analogy typically consists of the following steps:

Representation: The system represents the new knowledge and the previously acquired knowledge in a way that allows for direct comparison.

Comparison: The system compares the new knowledge to the previously acquired knowledge, looking for similarities and differences.

Mapping: The system maps the new knowledge to the previously acquired knowledge, creating a link between the two.

Generalization: The system generalizes from the previously acquired knowledge to the new knowledge, allowing it to make predictions about the new knowledge.

The main advantage of learning by analogy is that it allows the system to use previously acquired knowledge to understand new knowledge, thus avoiding the need to learn everything from scratch.

Learning by analogy has been applied to various AI-related tasks, such as natural language understanding, problem solving, and concept learning. It's also used in areas like computer vision and robotics, allowing systems to understand and adapt to new situations by comparing them to similar situations they've encountered before. However, it's not without its challenges and criticism, one of the main ones being that it's hard to define and identify the similarity, and can suffer from the problem of overgeneralization.