Unit 6 : Machine Learning

Prepared by Prof.Kore S.S

(Assistant Professor(CSE) DYPCET,Kolhapur)

Machine Learning

- Machine Learning is defined as a technology that is used to train machines to perform various actions such as predictions, recommendations, estimations, etc., based on historical data or past experience.
- Machine Learning enables computers to behave like human beings by training them with the help of past experience and predicted data.

Techniques in Machine Learning

1. Supervised Learning
2. Unsupervised Learning
3. Reinforcement Learning
4. Semi-supervised Learning

1. Supervised Learning

- Supervised learning, as the name indicates, has the presence of a supervisor as a teacher
- Basically supervised learning is when we teach or train the machine using data that is well-labelled. Which means some data is already tagged with the correct answer. After that, the machine is provided with a new set of examples(data) so that the supervised learning algorithm analyses the training data(set of training examples) and produces a correct outcome from labeled data.

Supervised learning is classified into two categories of algorithms:

- **Classification**: A classification problem is when the output variable is a category, such as “Red” or “blue” , “disease” or “no disease”.
- **Regression**: A regression problem is when the output variable is a real value, such as “dollars” or “weight

Supervised learning deals with or learns with “labeled” data. This implies that some data is already tagged with the correct answer.

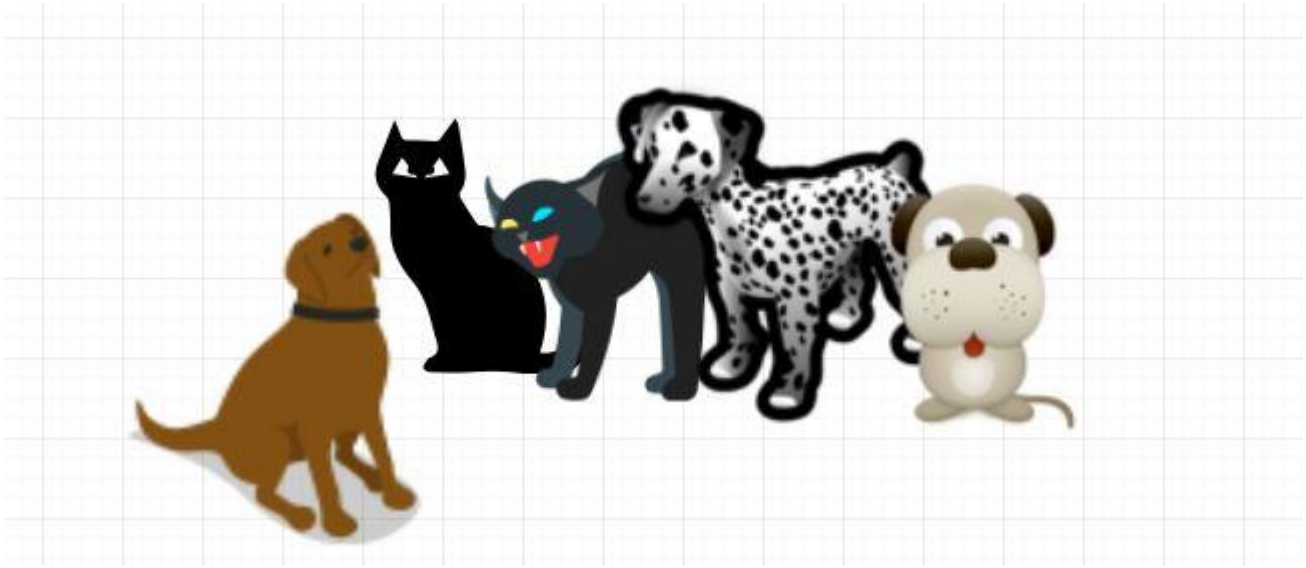
- **Types:-**
- Regression
- Logistic Regression
- Classification
- Naive Bayes Classifiers
- K-NN (k nearest neighbors)
- Decision Trees
- Support Vector Machine

2.Unsupervised learning

- Unsupervised learning is the training of a machine using information that is neither classified nor labeled and allowing the algorithm to act on that information without guidance. Here the task of the machine is to group unsorted information according to similarities, patterns, and differences without any prior training of data.

- Unlike supervised learning, no teacher is provided that means no training will be given to the machine. Therefore the machine is restricted to find the hidden structure in unlabeled data by itself.

- **For instance**, suppose it is given an image having both dogs and cats which it has never seen



- Thus the machine has no idea about the features of dogs and cats so we can’t categorize it as ‘dogs and cats ‘. But it can categorize them according to their similarities, patterns, and differences, i.e., we can easily categorize the above picture into two parts. The first may contain all pics having **dogs** in them and the second part may contain all pics having **cats** in them. Here you didn’t learn anything before, which means no training data or examples.

- Unsupervised learning is classified into two categories of algorithms:
- **Clustering**: A clustering problem is where you want to discover the inherent groupings in the data, such as grouping customers by purchasing behavior.
- **Association**: An association rule learning problem is where you want to discover rules that describe large portions of your data, such as people that buy X also tend to buy Y.

Types of Unsupervised Learning

1. Hierarchical clustering
2. K-means clustering
3. Principal Component Analysis
4. Singular Value Decomposition
5. Independent Component Analysis

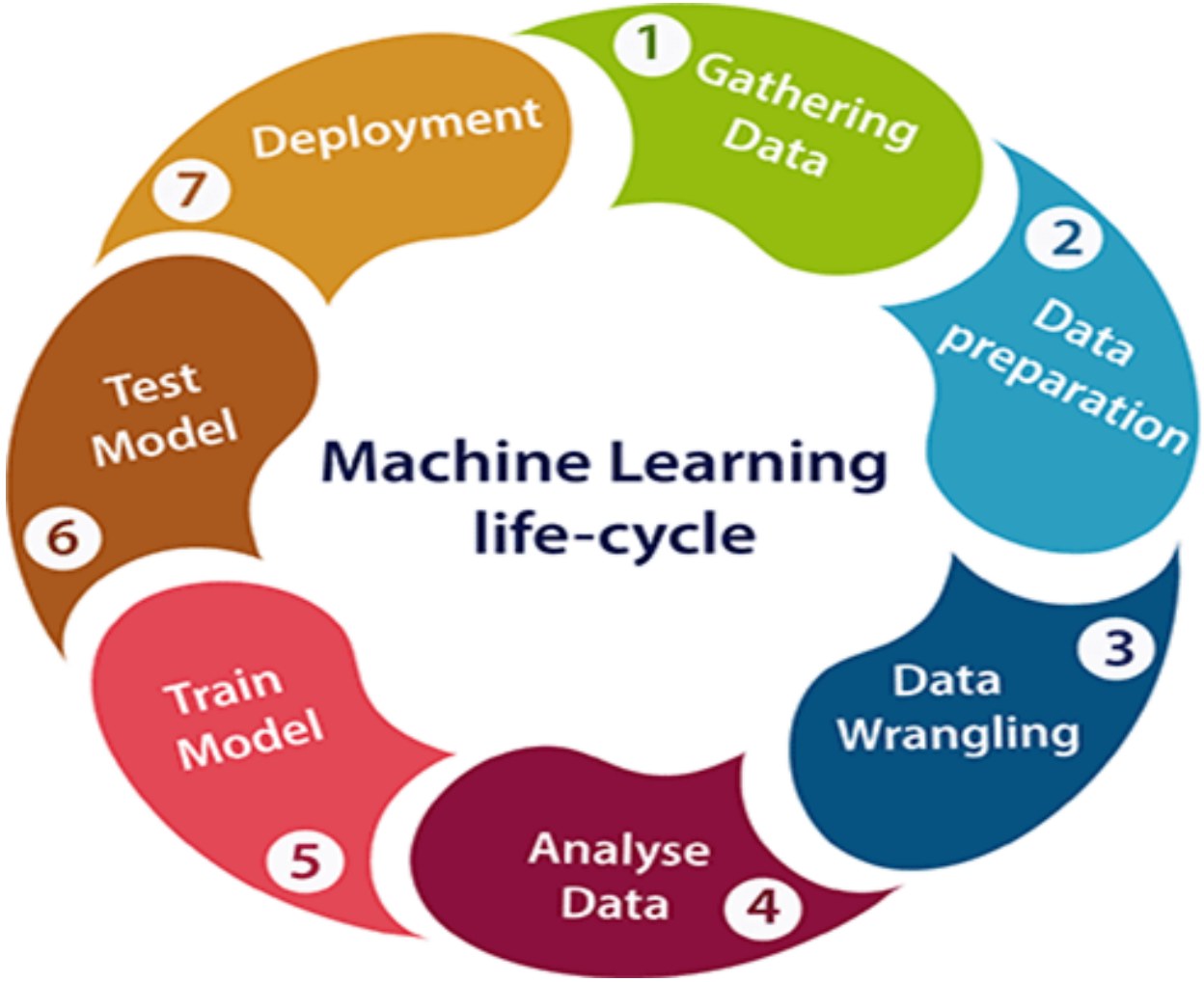
- 3. Reinforcement Learning

Reinforcement Learning is a feedback-based machine learning technique. In such type of learning, agents (computer programs) need to explore the environment, perform actions, and on the basis of their actions, they get rewards as feedback. For each good action, they get a positive reward, and for each bad action, they get a negative reward. The goal of a Reinforcement learning agent is to maximize the positive rewards. Since there is no labeled data, the agent is bound to learn by its experience only.

4. Semi-supervised Learning

Semi-supervised Learning is an intermediate technique of both supervised and unsupervised learning. It performs actions on datasets having few labels as well as unlabeled data. However, it generally contains unlabeled data. Hence, it also reduces the cost of the machine learning model as labels are costly, but for corporate purposes, it may have few labels. Further, it also increases the accuracy and performance of the machine learning model.

Machine learning Life cycle



1. Gathering Data:

- Data Gathering is the first step of the machine learning life cycle. The goal of this step is to identify and obtain all data-related problems.
- In this step, we need to identify the different data sources, as data can be collected from various sources such as **files**, **database**, **internet**, or **mobile devices**.

2. Data preparation

Data preparation is a step where we put our data into a suitable place and prepare it to use in our machine learning training.

- This step can be further divided into two processes:
- **Data exploration:**
  It is used to understand the nature of data that we have to work with. We need to understand the characteristics, format, and quality of data.
  A better understanding of data leads to an effective outcome. In this, we find Correlations, general trends, and outliers.
- **Data pre-processing:**
  Now the next step is preprocessing of data for its analysis.

3. Data Wrangling

- Data wrangling is the process of cleaning and converting raw data into a useable format. It is the process of cleaning the data, selecting the variable to use, and transforming the data in a proper format to make it more suitable for analysis in the next step. It is one of the most important steps of the complete process. Cleaning of data is required to address the quality issues.
- In real-world applications, collected data may have various issues, including:
- **Missing Values**
- **Duplicate data**
- **Invalid data**
- **Noise**

4. Data Analysis

- Now the cleaned and prepared data is passed on to the analysis step. This step involves:
- **Selection of analytical techniques**
- **Building models**
- **Review the result**
- The aim of this step is to build a machine learning model to analyze the data using various analytical techniques and review the outcome. It starts with the determination of the type of the problems, where we select the machine learning techniques such as **Classification**, **Regression**, **Cluster analysis**, **Association**, etc. then build the model using prepared data, and evaluate the model.

5. Train Model

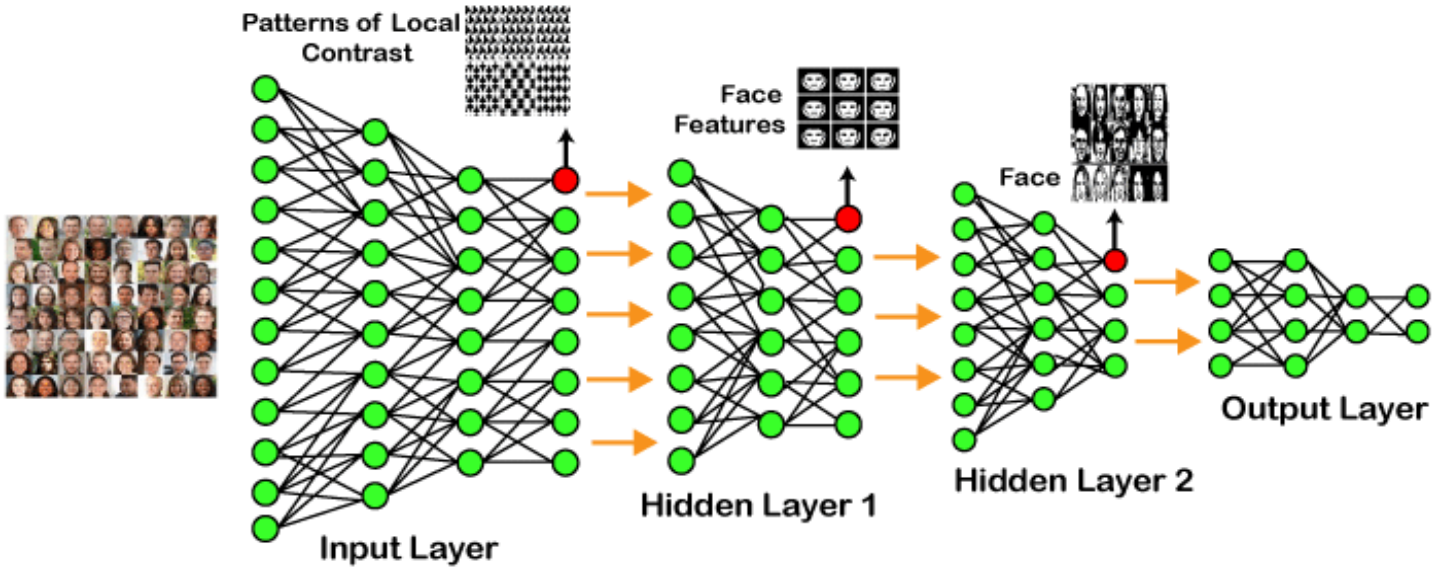
- Now the next step is to train the model, in this step we train our model to improve its performance for better outcome of the problem.
- We use datasets to train the model using various machine learning algorithms. Training a model is required so that it can understand the various patterns, rules, and, features.
- 6. Test Model
- Once our machine learning model has been trained on a given dataset, then we test the model. In this step, we check for the accuracy of our model by providing a test dataset to it.
- 7. Deployment
- The last step of machine learning life cycle is deployment, where we deploy the model in the real-world system.

Topic – Deep Learning

Deep Learning – "Deep learning is a collection of statistical techniques of machine learning for learning feature hierarchies that are actually based on artificial neural networks."

- Deep learning models are capable enough to focus on the accurate features themselves by requiring a little guidance from the programmer and are very helpful in solving out the problem of dimensionality. Deep learning algorithms are used, especially when we have a huge no of inputs and outputs.
- Since deep learning has been evolved by the machine learning, which itself is a subset of artificial intelligence and as the idea behind the artificial intelligence is to mimic the human behavior, so same is "the idea of deep learning to build such algorithm that can mimic the brain".
- Deep learning is implemented with the help of Neural Networks, and the idea behind the motivation of Neural Network is the biological neurons, which is nothing but a brain cell.
- So basically, deep learning is implemented by the help of deep networks, which are nothing but neural networks with multiple hidden layers.

Example of Deep Learning -



Types of Deep Learning Networks –

1. Feed Forward Neural Network
2. Recurrent Neural Network
3. Convolutional Neural Network
4. Restricted Boltzmann Machine
5. Autoencoders

1. Feed Forward Neural Network –

A feed-forward neural network is none other than an Artificial Neural Network, which ensures that the nodes do not form a cycle. In this kind of neural network, all the perceptrons are organized within layers, such that the input layer takes the input, and the output layer generates the output. Since the hidden layers do not link with the outside world, it is named as hidden layers. Each of the perceptrons contained in one single layer is associated with each node in the subsequent layer. It can be concluded that all of the nodes are fully connected. It does not contain any visible or invisible connection between the nodes in the same layer. There are no back-loops in the feed-forward network. To minimize the prediction error, the backpropagation algorithm can be used to update the weight values

Applications:

- Data Compression
- Pattern Recognition
- Computer Vision
- Speech Recognition
- Sonar Target Recognition
- Handwritten Characters Recognition

2. Recurrent Neural Network -

Recurrent neural networks are yet another variation of feed-forward networks. Here each of the neurons present in the hidden layers receives an input with a specific delay in time. The Recurrent neural network mainly accesses the preceding info of existing iterations. For example, to guess the succeeding word in any sentence, one must have knowledge about the words that were previously used. It not only processes the inputs but also shares the length as well as weights crossways time. It does not let the size of the model to increase with the increase in the input size. However, the only problem with this recurrent neural network is that it has slow computational speed as well as it does not contemplate any future input for the current state. It has a problem with reminiscing prior information.

Applications:

- Machine Translation
- Robot Control
- Time Series Prediction
- Speech Recognition
- Speech Synthesis
- Time Series Anomaly Detection
- Rhythm Learning
- Music Composition

3. Convolutional Neural Network -

Convolutional neural networks are a special kind of neural network mainly used for image classification, clustering of images and object recognition. DNNs enable unsupervised construction of hierarchical image representations. To achieve the best accuracy, deep convolutional neural networks are preferred more than any other neural network.

Applications:

- Identify Faces, Street Signs, Tumors.
- Image Recognition.
- Video Analysis.
- NLP.
- Anomaly Detection.
- Drug Discovery.
- Checkers Game.
- Time Series Forecasting

4. Restricted Boltzmann Machine -

RBM's are yet another variant of Boltzmann Machines. Here the neurons present in the input layer and the hidden layer encompasses symmetric connections amid them. However, there is no internal association within the respective layer. But in contrast to RBM, Boltzmann machines do encompass internal connections inside the hidden layer. These restrictions in BMs helps the model to train efficiently

Applications:

- Filtering.
- Feature Learning.
- Classification.
- Risk Detection.
- Business and Economic analysis

5. Autoencoders -

An autoencoder neural network is another kind of unsupervised machine learning algorithm. Here the number of hidden cells is merely small than that of the input cells. But the number of input cells is equivalent to the number of output cells. An autoencoder network is trained to display the output similar to the fed input to force AEs to find common patterns and generalize the data. The autoencoders are mainly used for the smaller representation of the input. It helps in the reconstruction of the original data from compressed data. This algorithm is comparatively simple as it only necessitates the output identical to the input.

- **Encoder:** Convert input data in lower dimensions.
- **Decoder:** Reconstruct the compressed data.

Applications:

- Classification.
- Clustering.
- Feature Compression

Deep learning applications -

- ***Self-Driving Cars***

  In self-driven cars, it is able to capture the images around it by processing a huge amount of data, and then it will decide which actions should be incorporated to take a left or right or should it stop. So, accordingly, it will decide what actions it should take, which will further reduce the accidents that happen every year.

- ***Voice Controlled Assistance***

  When we talk about voice control assistance, then **Siri** is the one thing that comes into our mind. So, you can tell Siri whatever you want it to do it for you, and it will search it for you and display it for you.

- ***Automatic Image Caption Generation***

  Whatever image that you upload, the algorithm will work in such a way that it will generate caption accordingly. If you say blue colored eye, it will display a blue-colored eye with a caption at the bottom of the image.

- ***Automatic Machine Translation***

  With the help of automatic machine translation, we are able to convert one language into another with the help of deep learning.