

Introduction to Machine Learning

Agenda

- Introduction to Machine learning
- Supervised
- Unsupervised
- Semi-Supervised Learning

Introduction to Machine Learning

- In the real world, we are surrounded by humans who can learn everything from their experiences with their learning capability, and we have computers or machines which work on our instructions. But can a machine also learn from experiences or past data like a human does? So here comes the role of **Machine Learning**.
- Machine Learning is said as a subset of artificial intelligence
- The term machine learning was first introduced by Arthur Samuel in 1959.
- **Definition of Machine learning:** It enables a machine to automatically learn from data, improve performance from experiences, and predict things without being explicitly programmed.



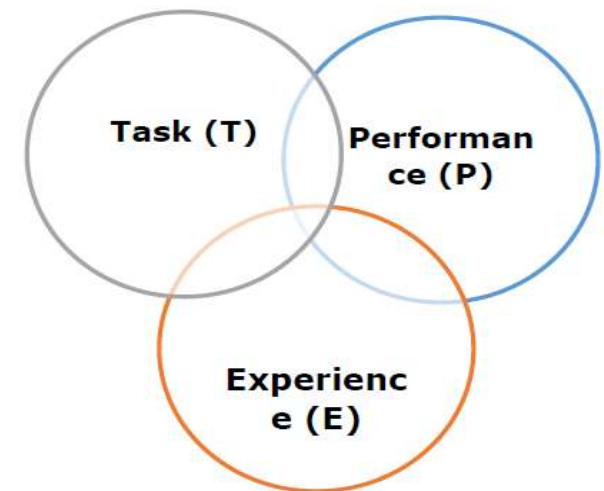
Another Definition: we must need to understand the following formal definition of ML given by professor Mitchell –

“A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E.”

The above definition is basically focusing on three parameters, also the main components of any learning algorithm, namely Task(T), Performance(P) and experience (E). In this context, we can simplify this definition as –

ML is a field of AI consisting of learning algorithms that –

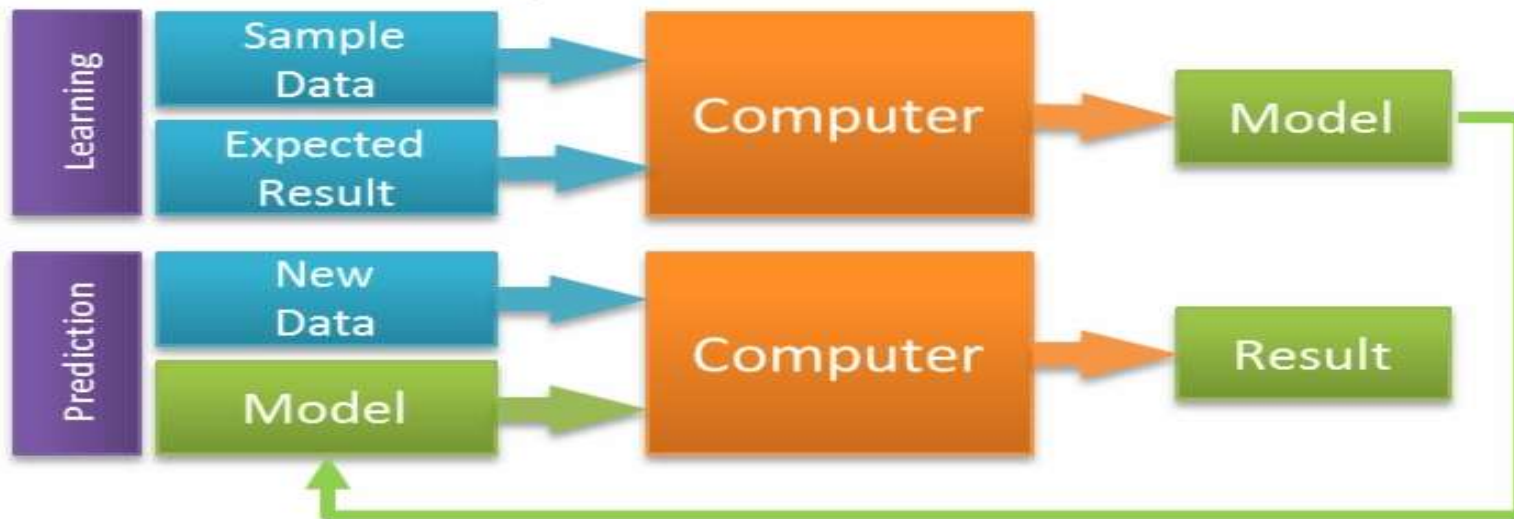
- * Improve their performance (P)
 - * At executing some task (T)
 - * Over time with experience (E)
- Based on the above, the following diagram represents a Machine Learning Model



Traditional modeling:



Machine Learning:



8 steps for approaching a Machine Learning application

- **Step1: - Defining the problem**

- **Step2: - Data Collection**

Preparing your data

- **Step3: - Data Preparation**

Clean that which may require it (remove duplicates, correct errors, deal with missing values, normalization, data type conversions, etc.). Split into training and evaluation sets

- **Step4: - Choose a Model**

Different algorithms are for different tasks; choose the right one

- **Step5: - Train the Model**

The goal of training is to answer a question or make a prediction correctly as often as possible

- **Step6: - Evaluate the Model**

Test the model against previously unseen data

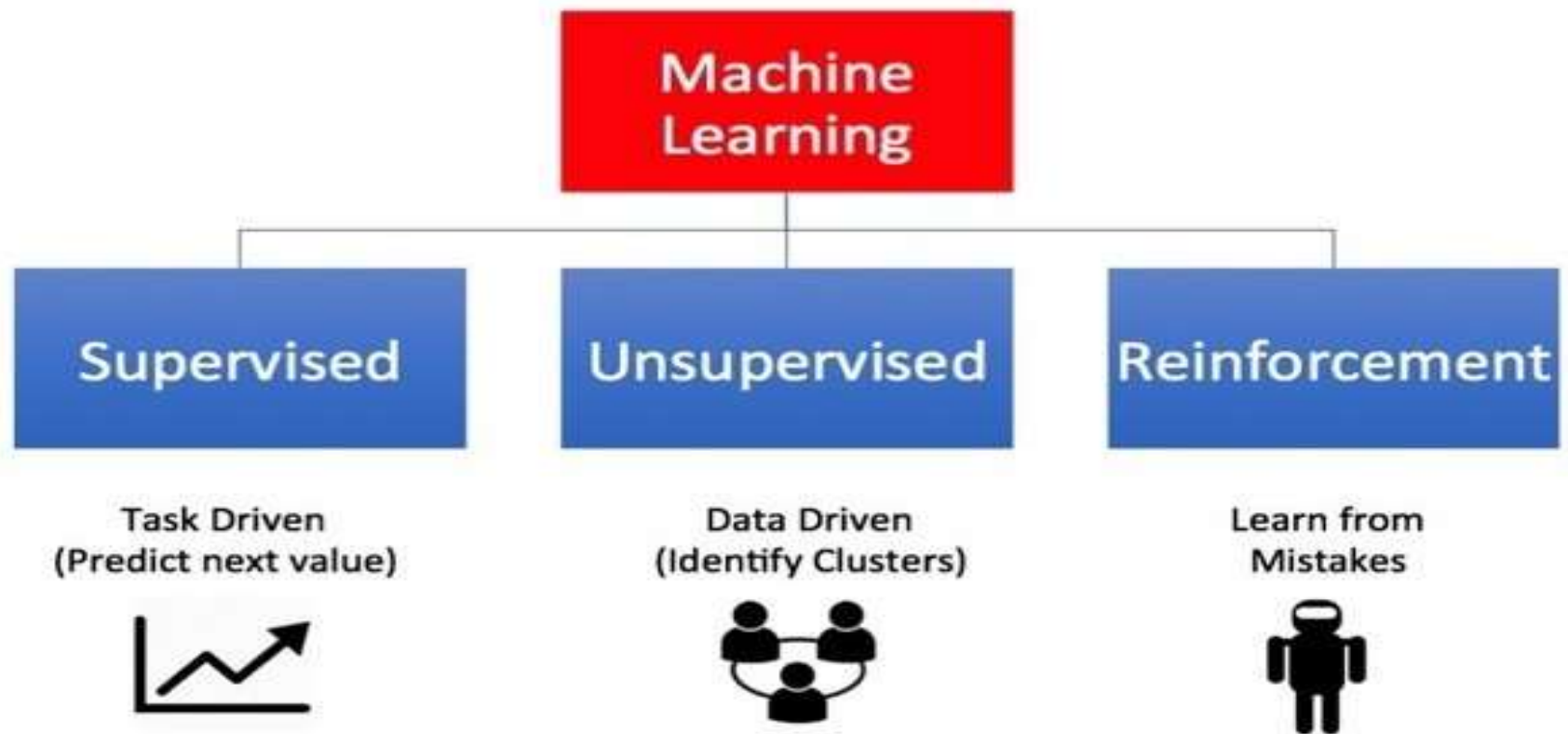
- **Step7: - Parameter Tuning**

Tune model parameters for improved performance

- **Step8: - Make Predictions**

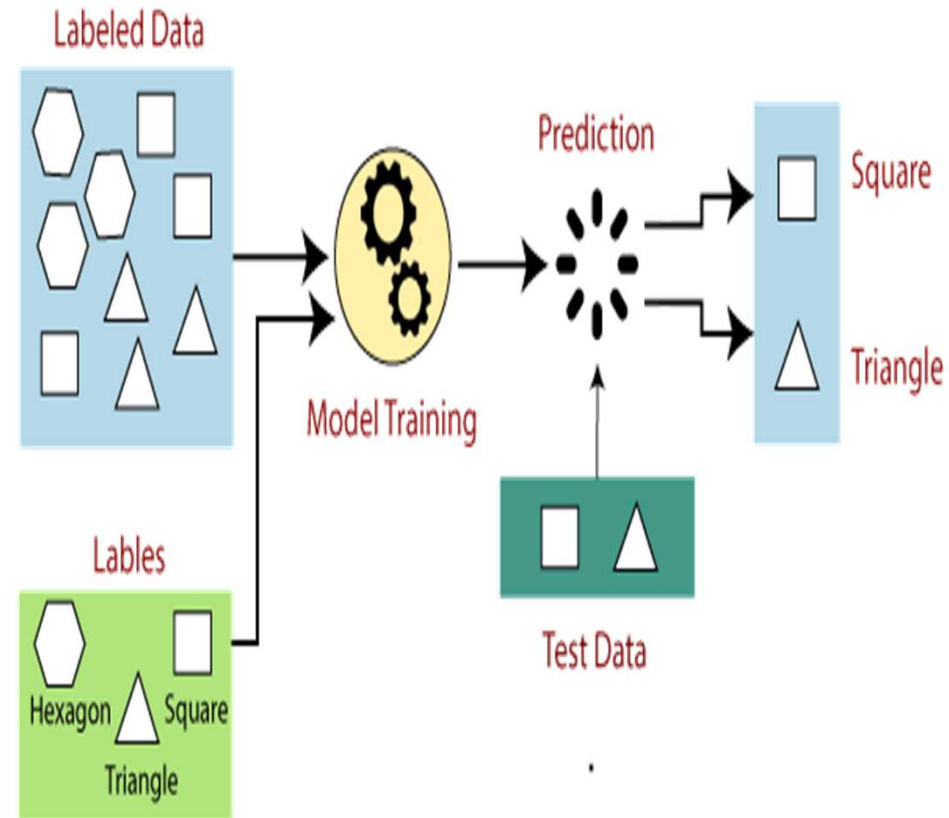
Using further (test set) data which have, until this point, are used to test the model

Types of Machine Learning



Supervised

- In supervised learning, models are trained using **labeled** dataset, where the model learns about each type of data.
- Once the training process is completed, the model is tested on the basis of test data (a subset of the training set), and then it predicts the output.
- In the real-world, supervised learning can be used for **Risk Assessment, Image classification, Fraud Detection, spam filtering**, etc.



- **Classification**

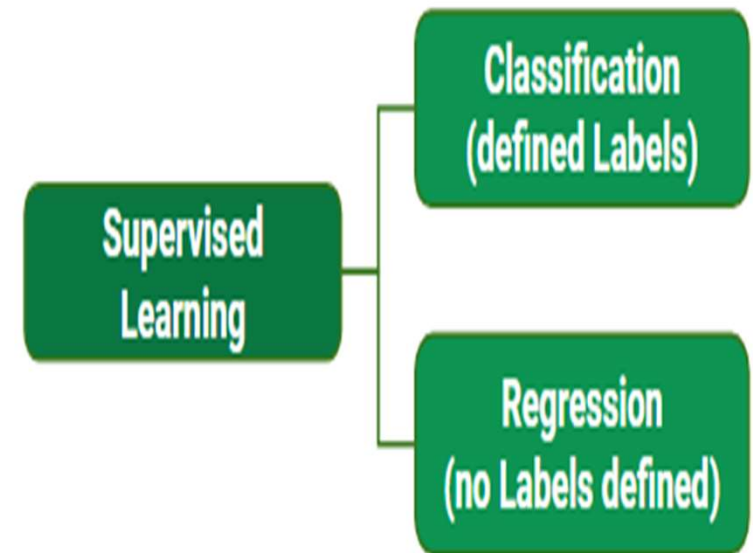
Classification algorithms are used when the output variable **is categorical**, which means there are two classes such as Yes-No, Male-Female, True-false, etc. Below are some popular Classification algorithms which come under supervised learning:

Random Forest

Decision Trees

Logistic Regression

Support vector Machines



- **Regression**

Regression algorithms are used if there is a relationship between the input variable and the output variable. It is used for the prediction of **continuous variables**, such as Weather forecasting, Market Trends, etc. Below are some popular Regression algorithms which come under supervised learning:

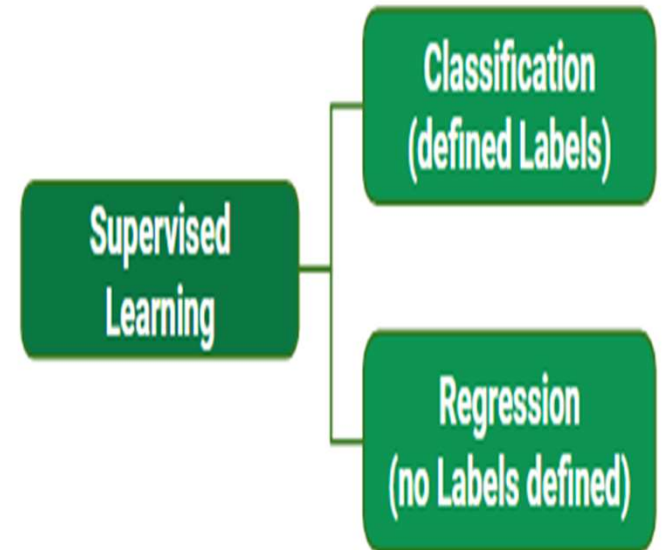
- Linear Regression

- Regression Trees

- Non-Linear Regression

- Bayesian Linear Regression

- Polynomial Regression



Advantages of Supervised learning:

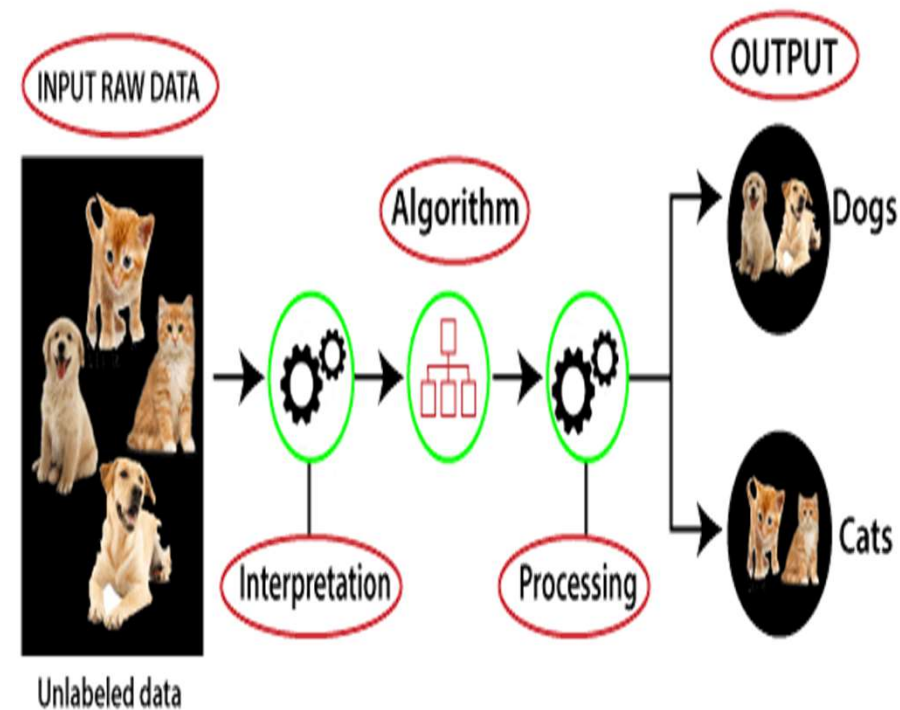
- With the help of supervised learning, the model can predict the output on the basis of prior experiences.
- In supervised learning, we can have an exact idea about the classes of objects.

Disadvantages of supervised learning:

- Supervised learning models are not suitable for handling the complex tasks.
- Supervised learning cannot predict the correct output if the test data is different from the training dataset.
- Training required lots of computation times.
- In supervised learning, we need enough knowledge about the classes of object.

Unsupervised

- Unsupervised learning is a type of machine learning in which models are trained using **unlabeled** dataset and are allowed to act on that data without any supervision.
- Unsupervised learning cannot be directly applied to a regression or classification problem because unlike supervised learning, we have the input data but **no corresponding output data**.
- The goal of unsupervised learning is to find the underlying structure of dataset, group that data according to similarities, and represent that dataset in a compressed format.

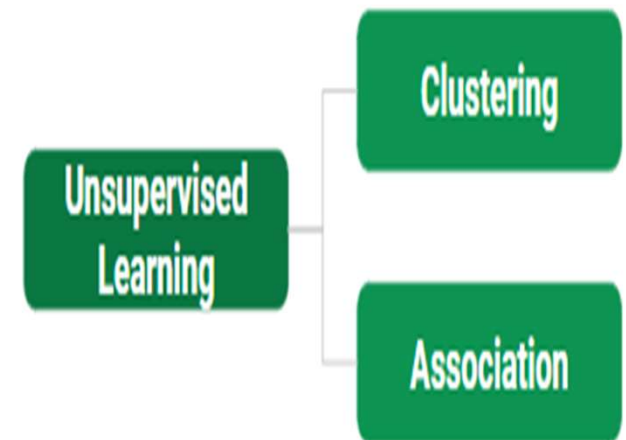


- **Clustering:**

Clustering is a method of grouping the objects into clusters such that objects with most similarities remains into a group and has less or no similarities with the objects of another group.

- **Association:**

An association rule is an unsupervised learning method which is used for finding the relationships between variables in the large database. It determines the set of items that occurs together in the dataset. Association rule makes marketing strategy more effective.



List of some popular unsupervised learning algorithms

- **K-means clustering**
- **Hierarchical clustering**
- **Anomaly detection**
- **Neural Networks**
- **Principle Component Analysis**
- **Independent Component Analysis**
- **Apriori algorithm (Association)**
- **Singular value decomposition**

Advantages of Unsupervised Learning

- Unsupervised learning is used for more complex tasks as compared to supervised learning because, in unsupervised learning, we don't have labeled input data.
- Unsupervised learning is preferable as it is easy to get unlabeled data in comparison to labeled data.

Disadvantages of Unsupervised Learning

- Unsupervised learning is intrinsically more difficult than supervised learning as it does not have corresponding output.
- The result of the unsupervised learning algorithm might be less accurate as input data is not labeled, and algorithms do not know the exact output in advance.

Semi-Supervised Learning

- Semi-supervised learning falls between unsupervised learning (with no labeled training data) and supervised learning (with only labeled training data).
- Semi-supervised learning is a kind of classification that combines labeled with unlabeled data.
- Semi-supervised learning is an approach to machine learning that combines a small amount of labeled data with a large amount of unlabeled data during training.
 - Usually, when you don't have enough labeled data to solve a problem using supervised learning, the alternative is to use unsupervised learning.

