INTRODUCTION TO MACHINE LEARNING

Machine Learning is an idea to learn from examples and experience, without being explicitly programmed. Instead of writing code, you feed data to the generic algorithm, and it builds logic based on the data given.

For example, one kind of algorithm is a classification algorithm. It can put data into different groups. The classification algorithm used to detect handwritten alphabets could also be used to classify emails into spam and not-spam.

Examples of Machine Learning:

There are many examples of machine learning. Here are a few examples of classification problems where the goal is to categorize objects into a fixed set of categories.

Face detection: Identify faces in images (or indicate if a face is present).

Email filtering: Classify emails into spam and not-spam.

Medical diagnosis: Diagnose a patient as a sufferer or non-sufferer of some disease.

Weather prediction: Predict, for instance, whether or not it will rain tomorrow.

Salary Prediction: Predict the salary of a particular employee based on this role, experience, skill set…

NEED of Machine Learning:

Machine Learning is a field which is raised out of Artificial Intelligence (AI). Applying AI, we wanted to build better and intelligent machines. But except for few mere tasks such as finding the shortest path between point A and B, we were unable to program more complex and constantly evolving challenges. There was a realization that the only way to be able to achieve this task was to let machine learn from itself. This sounds similar to a child learning from its self. So machine learning was developed as a new capability for computers. And now machine learning is present in so many segments of technology, that we don't even realize it while using it.

Finding patterns in data on planet earth is possible only for human brains. The data being very massive, the time taken to compute is increased, and this is where Machine Learning comes into action, to help people with large data in minimum time.

If big data and cloud computing are gaining importance for their contributions, machine learning as technology helps analyze those big chunks of data, easing the task of data scientists in an automated process and gaining equal importance and recognition.

The techniques we use for data mining have been around for many years, but they were not effective as they did not have the competitive power to run the algorithms. If you run deep learning with access to better data, the output we get will lead to dramatic breakthroughs which is machine learning.

Kinds of Machine Learning

There are three kinds of Machine Learning Algorithms.

a. Supervised Learning

b. Unsupervised Learning

c. Reinforcement Learning

Supervised Learning

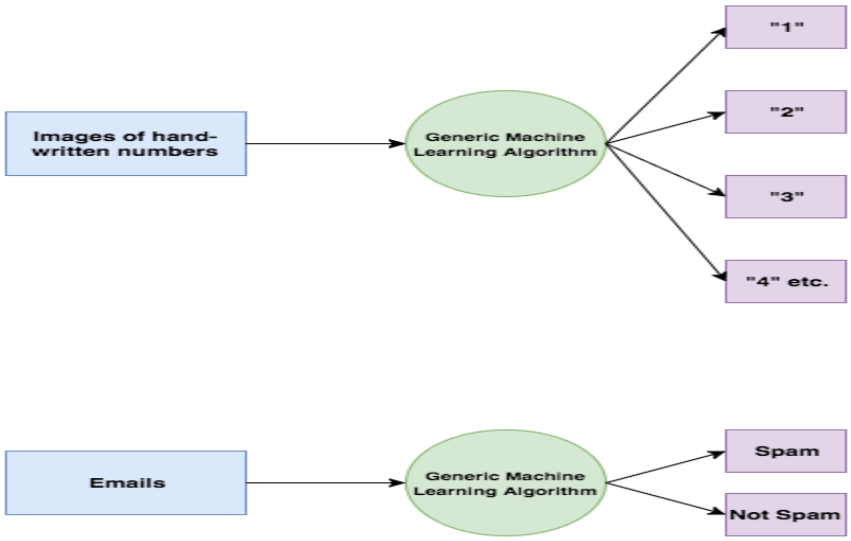
A majority of practical machine learning uses supervised learning.

Supervised learning is a method used to enable machines to classify objects, problems or situations based on related data fed into the machines. Machines are fed with data such as characteristics, patterns, dimensions, color and height of objects, people or situations repetitively until the machines are able to perform accurate classifications. Supervised learning is a popular technology or concept that is applied to real-life scenarios. Supervised learning is used to provide product recommendations, segment customers based on customer data, diagnose disease based on previous symptoms and perform many other tasks.

Speaking mathematically, supervised learning is where you have both input variables (x) and output variables(Y) and can use an algorithm to derive the mapping function from the input to the output.

The mapping function is expressed as **Y = f(X).**

Example:



Supervised learning problems can be further grouped into **Regression** and **Classification** problems. Both problems have as goal the construction of a succinct model that can predict the value of the dependent attribute from the attribute variables. The difference between the two tasks is the fact that the dependent attribute is numerical for regression and categorical for classification.

Classification:

A classification problem is when the output variable is a category, such as “red” or “blue” or “disease” and “no disease”. A classification model attempts to draw some conclusion from observed values. Given one or more inputs a classification model will try to predict the value of one or more outcomes.

For example, when filtering emails “spam” or “not spam”, when looking at transaction data, “fraudulent”, or “authorized”. In short Classification either predicts categorical class labels or classifies data (construct a model) based on the training set and the values (class labels) in classifying attributes and uses it in classifying new data. There are a number of classification models.

Classification models include:

i) Logistic Regression
ii) Naive Bayes
iii) Stochastic Gradient Descent
iv) K-Nearest Neighbours
v) Decision Tree
vi) Random Forest
vii) SupportVectorMachines

We will see in detail about these algorithms in the upcoming sections.

Regression:

A regression problem is when the output variable is a real value, such as “Rupees” or “height.”

Regression models include:

i) Simple Linear Regression
ii) Multiple Linear Regression

We will see in detail about these algorithms in the upcoming sections.

Unsupervised Learning:

Unsupervised Learning is a class of Machine Learning techniques to find the patterns in data. The data given to unsupervised algorithm are not labelled, which means only the input variables(X) are given with no corresponding output variables. In unsupervised learning, the algorithms are left to themselves to discover interesting structures in the data.

Unsupervised models include:

iii) K-Means Clustering:
iv) Mean-Shift Clustering:
v) DBSCAN:
vi) Hierarchical clustering:
vii) Gaussian Mixture Models(GMM)

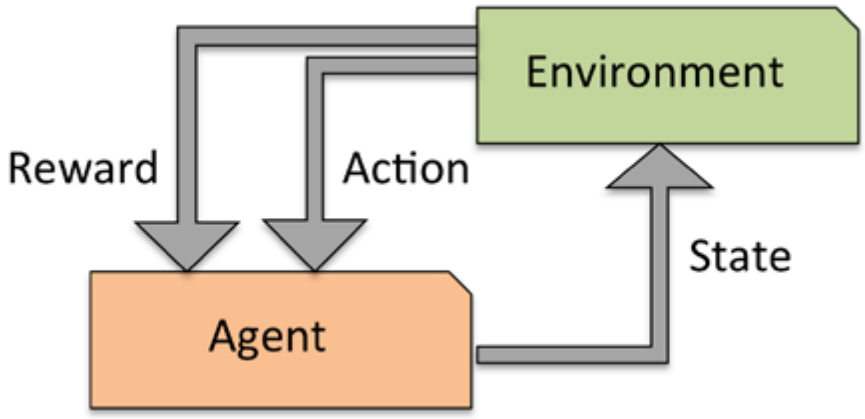
We will see in detail about these algorithms in the upcoming sections.

Reinforcement Learning:

Another type of machine learning is reinforcement learning. In reinforcement learning, the goal is to develop a system (agent) that improves its performance based on interactions with the environment. Since the information about the current state of the environment typically also includes a so-called reward signal, we can think of reinforcement learning as a field related to supervised learning. However, in reinforcement learning this feedback is not the correct ground truth label or value, but a measure of how well the action was measured by a reward function. Through the interaction with the

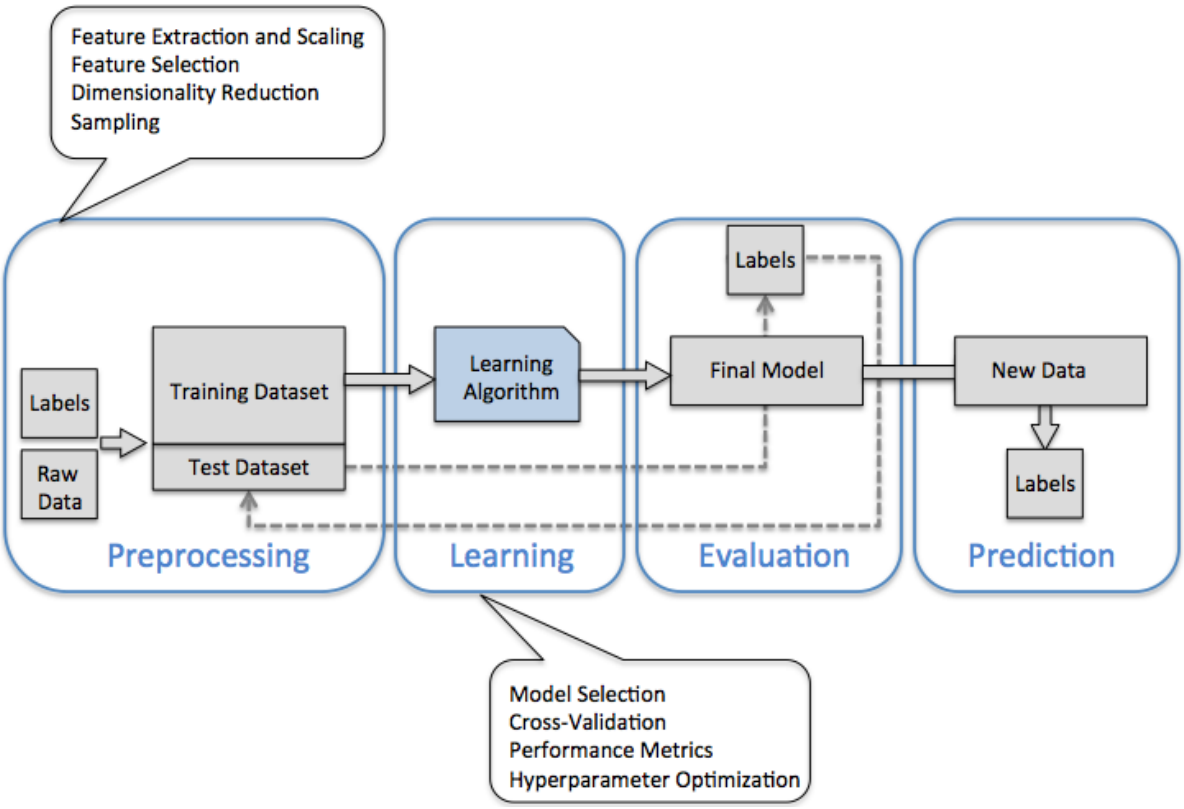
environment, an agent can then use reinforcement learning to learn a series of actions that maximizes this reward via an exploratory trial-and-error approach or deliberative planning.

A popular example of reinforcement learning is a chess engine. Here, the agent decides upon a series of moves depending on the state of the board (the environment), and the reward can be defined as win or lose at the end of the game:



Workflow for Machine learning model:

In the previous sections, we discussed the basic concepts of machine learning and the three different types of learning. In this section, we will discuss other important parts of a machine learning system accompanying the learning algorithm. The diagram below shows a typical workflow diagram for using machine learning in predictive modeling.



The above diagram can be summarized to the below points:

i) Data Preprocessing
ii) Choosing best model for your use case
iii) Compressing Data via Dimensionality Reduction
iv) Model Evaluation
v) Hyper parameter Tuning
vi) Predicting unseen data instances
vii) Saving the models for future use.