Unit I:

Introduction- Artificial Intelligence, Machine Learning, Deep learning, Types of Machine Learning Systems, Main Challenges of Machine Learning.

Statistical Learning: Introduction, Supervised and Unsupervised Learning, Training and Test Loss, Tradeoffs in Statistical Learning, Estimating Risk Statistics, Sampling distribution of an estimator, Empirical Risk Minimization

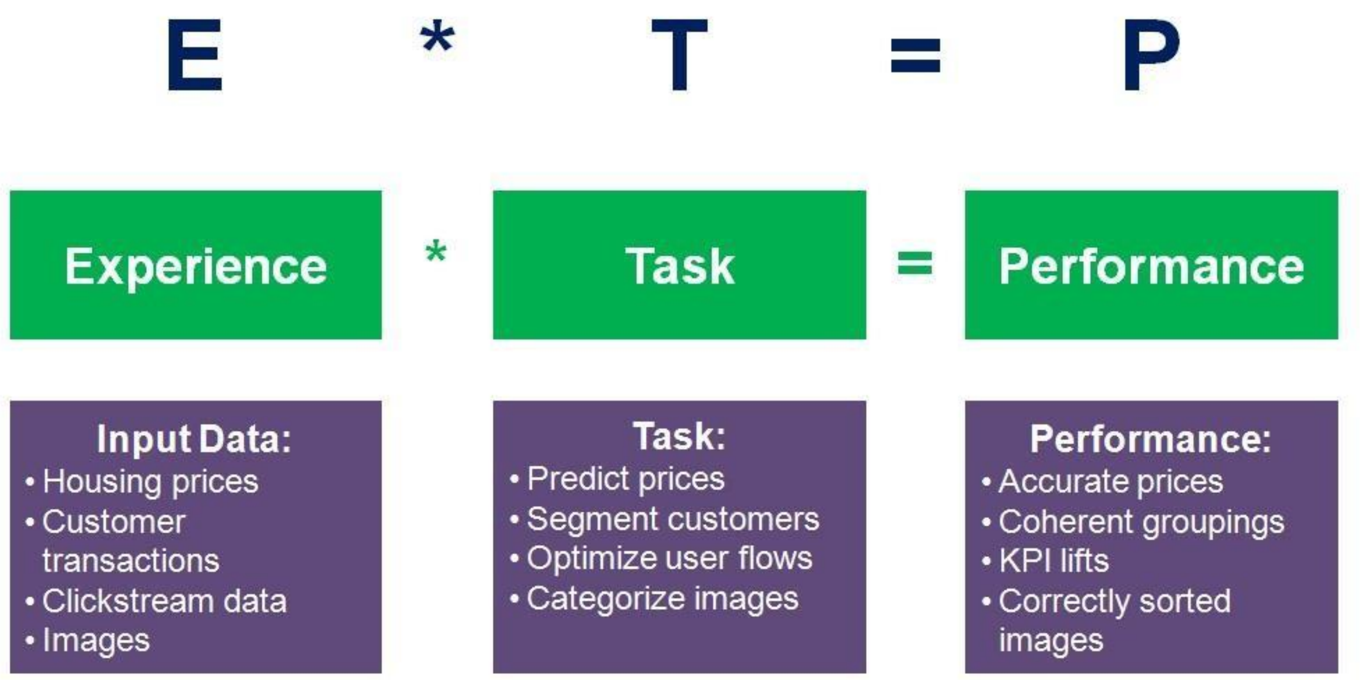
---

Introduction To Machine Learning

The term Machine Learning was first coined by Arthur Samuel in the year 1959. Looking back, that year was probably the most significant in terms of technological advancements.

If you browse through the net about 'what is Machine Learning', you'll get at least 100 different definitions. However, the very first formal definition was given by Tom M. 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."



E * T = P

Experience * Task = Performance

Input Data:
- Housing prices
- Customer transactions
- Clickstream data
- Images

Task:
- Predict prices
- Segment customers
- Optimize user flows
- Categorize images

Performance:
- Accurate prices
- Coherent groupings
- KPI lifts
- Correctly sorted images

In simple terms, Machine learning is a subset of Artificial Intelligence (AI) which provides machines the ability to learn automatically & improve from experience without being explicitly programmed to do so. In the sense, it is the practice of getting Machines to solve problems by gaining the ability to think.

But wait, can a machine think or make decisions? Well, if you feed a machine a good amount of data, it will learn how to interpret, process and analyze this data by using Machine Learning Algorithms, in order to solvereal-world problems.

Machine Learning Definitions

Algorithm: A Machine Learning algorithm is a set of rules and statistical techniques used to learn patterns from data and draw significant information from it. It is the logic behind a Machine Learning model. An example of a Machine Learning algorithm is the Linear Regression algorithm.

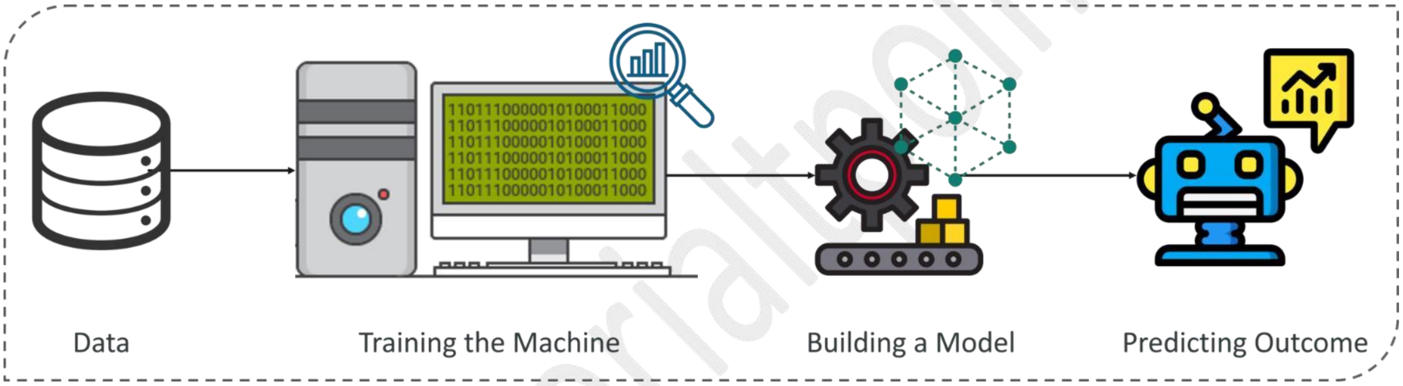
Model: A model is the main component of Machine Learning. A model is trained by using a Machine LearningAlgorithm. An algorithm maps all the decisions that a model is supposed to take based on the given input, inorder to get the correct output.

Predictor Variable: It is a feature(s) of the data that can be used to predict the output.

Response Variable: It is the feature or the output variable that needs to be predicted by using the predictorvariable(s).

Training Data: The Machine Learning model is built using the training data. The training data helps the modelto identify key trends and patterns essential to predict the output.

Testing Data: After the model is trained, it must be tested to evaluate how accurately it can predict an outcome. This is done by the testing data set.



Machine Learning Process

The Machine Learning process involves building a Predictive model that can be used to find a solution for a Problem Statement. To understand the Machine Learning process let’s assume that you have been given a problem that needs to be solved by using Machine Learning.

The problem is to predict the occurrence of rain in your local area by using Machine Learning.

The below steps are followed in a Machine Learning process:

Step 1: Define the objective of the Problem Statement

At this step, we must understand what exactly needs to be predicted. In our case, the objective is to predictthe possibility of rain by studying weather conditions. At this stage, it is also essential to take mental notes onwhat kind of data can be used to solve this problem or the type of approach you must follow to get to the solution.

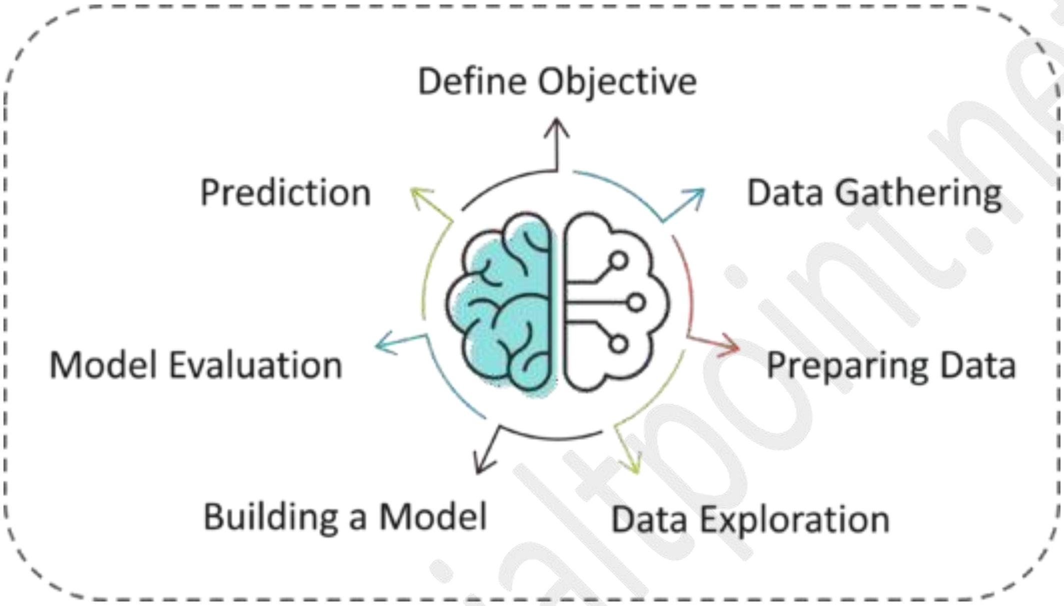
Step 2: Data Gathering

At this stage, you must be asking questions such as,

- What kind of data is needed to solve this problem?
- Is the data available?
- How can I get the data?

Once you know the types of data that is required, you must understand how you can derive this data. Datacollection can be done manually or by web scraping. However, if you're a beginner and you're just looking to learn Machine Learning you don't have to worry about getting the data. There are 1000s of data resourceson the web, you can just download the data set and get going.

Coming back to the problem at hand, the data needed for weather forecasting includes measures such as humidity level, temperature, pressure, locality, whether or not you live in a hill station, etc. Such data must becollected and stored for analysis.



Step 3: Data Preparation

The data you collected is almost never in the right format. You will encounter a lot of inconsistencies in the data set such as missing values, redundant variables, duplicate values, etc. Removing such inconsistencies is very essential because they might lead to wrongful computations and predictions. Therefore, at this stage, you scan the data set for any inconsistencies and you fix them then and there.

Step 4: Exploratory Data Analysis

Grab your detective glasses because this stage is all about diving deep into data and finding all the hidden data mysteries. EDA or Exploratory Data Analysis is the brainstorming stage of Machine Learning. Data Exploration involves understanding the patterns and trends in the data. At this stage, all the useful insights are drawn and correlations between the variables are understood.

For example, in the case of predicting rainfall, we know that there is a strong possibility of rain if the temperature has fallen low. Such correlations must be understood and mapped at this stage.

Step 5: Building a Machine Learning Model

All the insights and patterns derived during Data Exploration are used to build the Machine Learning Model. This stage always begins by splitting the data set into two parts, training data, and testing data. The trainingdata will be used to build and analyze the model. The logic of the model is based on the Machine Learning Algorithm that is being implemented.

In the case of predicting rainfall, since the output will be in the form of True (if it will rain tomorrow) or False (no rain tomorrow), we can use a Classification Algorithm such as Logistic Regression.

Choosing the right algorithm depends on the type of problem you're trying to solve, the data set and the levelof complexity of the problem. In the upcoming sections, we will discuss the different types of problems thatcan be solved by using Machine Learning.

Step 6: Model Evaluation & Optimization

After building a model by using the training data set, it is finally time to put the model to a test. The testing data set is used to check the efficiency of the model and how accurately it can predict the outcome. Once the accuracy is calculated, any further improvements in the model can be implemented at this stage. Methods like parameter tuning and cross-validation can be used to improve the performance of the model.

Step 7: Predictions

Once the model is evaluated and improved, it is finally used to make predictions. The final output can be a Categorical variable (eg. True or False) or it can be a Continuous Quantity (eg. the predicted value of a stock).

In our case, for predicting the occurrence of rainfall, the output will be a categorical variable.

So that was the entire Machine Learning process. Now it's time to learn about the different ways in which
Machines can learn.

2) Machine Learning Types

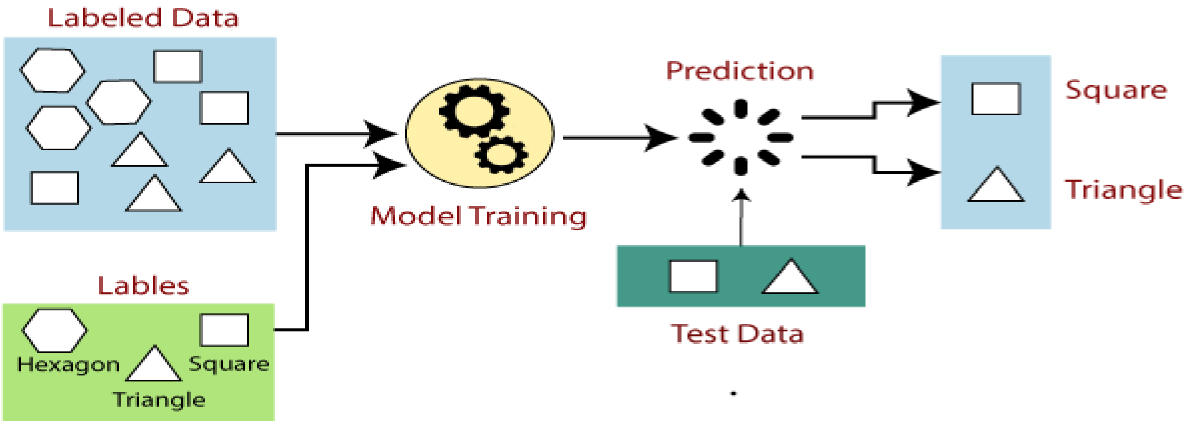
A machine can learn to solve a problem by following any one of the following three approaches. These are the ways in which a machine can learn:

1) Supervised Learning
2) Unsupervised Learning
3) Reinforcement Learning

Supervised Learning

In supervised learning, models are trained using labelled 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.

The working of Supervised learning can be easily understood by the below example and diagram:



suppose we have a dataset of different types of shapes which includes square, rectangle, triangle, and Polygon. Now the first step is that we need to train the model for each shape.

- If the given shape has four sides, and all the sides are equal, then it will be labelled as a **Square**.
- If the given shape has three sides, then it will be labelled as a **triangle**.
- If the given shape has six equal sides then it will be labelled as **hexagon**.

Now, after training, we test our model using the test set, and the task of the model is to identify the shape.

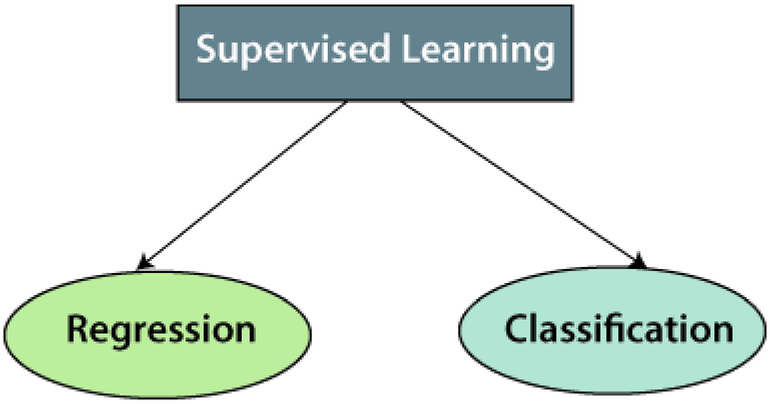
The machine is already trained on all types of shapes, and when it finds a new shape, it classifies the shape on the bases of a number of sides, and predicts the output.

Steps Involved in Supervised Learning:

- First Determine the type of training dataset
- Collect/Gather the labelled training data.
- Split the training dataset into training **dataset, test dataset, and validation dataset**.
- Determine the input features of the training dataset, which should have enough knowledge so that the model can accurately predict the output.
- Determine the suitable algorithm for the model, such as support vector machine, decision tree, etc.
- Execute the algorithm on the training dataset. Sometimes we need validation sets as the control parameters, which are the subset of training datasets.
- Evaluate the accuracy of the model by providing the test set. If the model predicts the correct output, which means our model is accurate.

Types of supervised Machine learning Algorithms:

Supervised learning can be further divided into two types of problems:



1. **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.

2. 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, Spam Filtering etc..

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.
- Supervised learning model helps us to solve various real-world problems such as **fraud detection, spam filtering**, etc.

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.

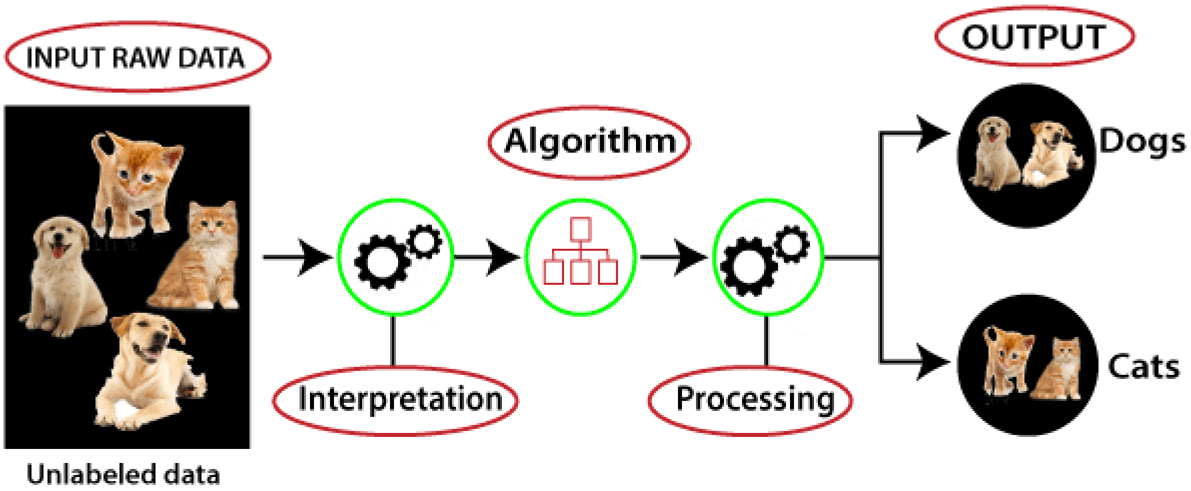
Unsupervised Learning

As the name suggests, unsupervised learning is a machine learning technique in which models are not supervised using training dataset. Instead, models itself find the hidden patterns and insights from the given data. It can be compared to learning which takes place in the human brain while learning new things. It can be defined as:

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.

Working of Unsupervised Learning

Working of unsupervised learning can be understood by the below diagram:

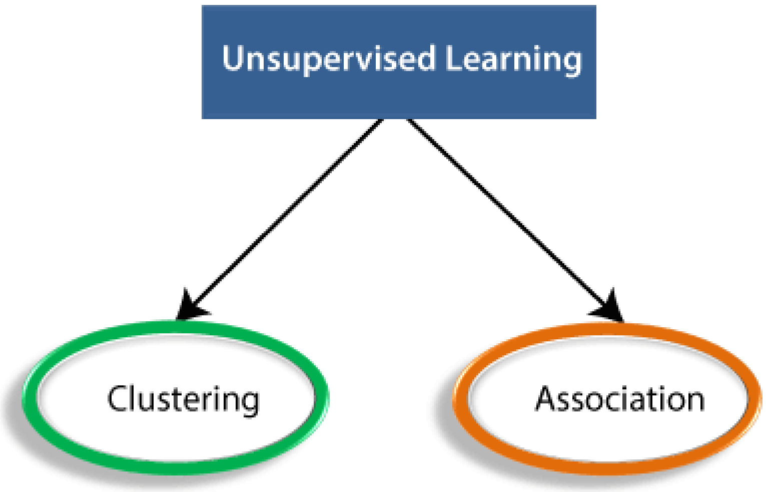


Here, we have taken an unlabeled input data, which means it is not categorized and corresponding outputs are also not given. Now, this unlabeled input data is fed to the machine learning model in order to train it. Firstly, it will interpret the raw data to find the hidden patterns from the data and then will apply suitable algorithms such as k-means clustering, Decision tree, etc.

Once it applies the suitable algorithm, the algorithm divides the data objects into groups according to the similarities and difference between the objects.

Types of Unsupervised Learning Algorithm:

The unsupervised learning algorithm can be further categorized into two types of problems:



- **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. Cluster analysis finds the commonalities between the data objects and categorizes them as per the presence and absence of those commonalities.
- **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. Such as people who buy X item (suppose a bread) are also tend to purchase Y (Butter/Jam) item. A typical example of Association rule is Market Basket Analysis.

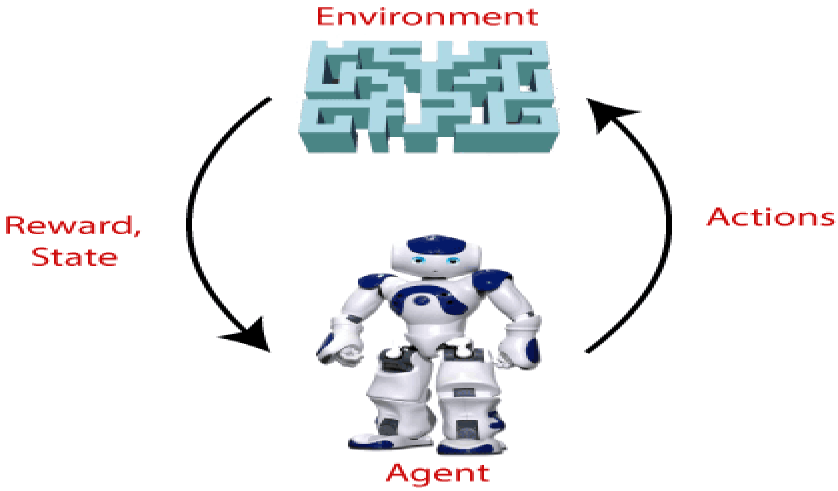
<u>Reinforcement Learning</u>

"Reinforcement learning is a type of machine learning method where an intelligent agent (computer program) interacts with the environment and learns to act within that."

- Reinforcement Learning is a feedback-based Machine learning technique in which an agent learns to behave in an environment by performing the actions and seeing the results of actions. For each good action, the agent gets positive feedback, and for each bad action, the agent gets negative feedback or penalty.
- In Reinforcement Learning, the agent learns automatically using feedbacks without any labeled data, unlike <u>supervised learning.</u>
- Since there is no labeled data, so the agent is bound to learn by its experience only.

Example: Suppose there is an AI agent present within a maze environment, and his goal is to find the diamond. The agent interacts with the environment by performing some actions, and based on those actions, the state of the agent gets changed, and it also receives a reward or penalty as feedback.

- The agent continues doing these three things (**take action, change state/remain in the same state, and get feedback**), and by doing these actions, he learns and explores the environment.
- The agent learns that what actions lead to positive feedback or rewards and what actions lead to negative feedback penalty. As a positive reward, the agent gets a positive point, and as a penalty, it gets a negative point.

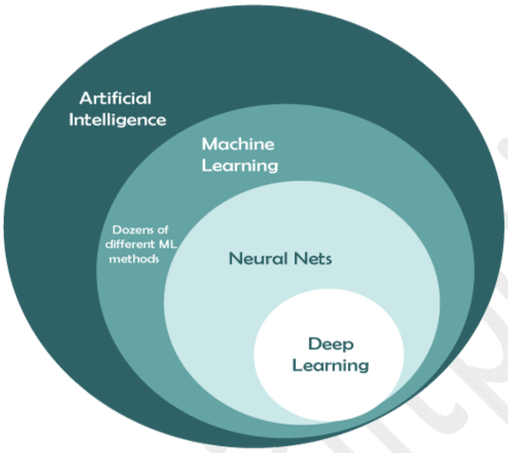


1. Artificial Intelligence, Machine Learning, Deep learning

Deep Learning, Machine Learning, and Artificial Intelligence are the most used terms on the internet for IT folks. However, all these three technologies are connected with each other. *Artificial Intelligence (AI) can beunderstood as an umbrella that consists of both Machine learning and deep learning.* Or We can say deeplearning and machine learning both are subsets of artificial intelligence.

As these technologies look similar, most of the persons have misconceptions about 'Deep Learning, Machinelearning, and Artificial Intelligence' that all three are similar to each other. But in reality, although all thesetechnologies are used to build intelligent machines or applications that behave like a human, still, they differby their functionalities and scope.

It means these three terms are often used interchangeably, but they do not quite refer to the same things.Let's understand thefundamental difference between deep learning, machine learning, and Artificial Intelligence with the below image.



With the above image, you can understand Artificial Intelligence is a branch of computer science that helps usto create smart, intelligent machines. Further, ML is a subfield of AI that helps to teach machines and build AI-driven applications. On the other hand, Deep learning is the sub-branch of ML that helps to train ML modelswith a huge amount of input and complex algorithms and mainly works with neural networks.

What is Artificial Intelligence (AI)?

Artificial Intelligence is defined as a field of science and engineering that deals with making intelligent machines or computers to perform human-like activities.

Mr. John McCarthy is known as the godfather of this amazing invention. There are some popular definitionsof AI, which are as follows:

"AI is defined as the capability of machines to imitate intelligent human behavior."

"A computer system able to perform tasks that normally require human intelligence, such as visual perception, speech recognition, decision-making, and translation between languages."

What is Deep Learning?

"Deep learning is defined as the subset of machine learning and artificial intelligence that is based on artificial neural networks". In deep learning, the deep word refers to the number of layers in a neural network.

Deep Learning is a set of algorithms inspired by the structure and function of the human brain. It uses a huge amount of structured as well as unstructured data to teach computers and predicts accurate results. The maindifference between machine learning and deep learning technologies is of presentation of data. Machine learning uses structured/unstructured data for learning, while deep learning uses neural networks for learningmodels.

3. Main Challenges of Machine Learning

During the development phase our focus is to select a learning algorithm and train it on some data, the twothings that might be a problem are a bad algorithm or bad data, or perhaps both of them.

The following are some of the challenges of ML

1. **Not enough training data.**

Machine Learning is not quite there yet; it takes a lot of data for most Machine Learning algorithms to workproperly. Even for very simple problems you typically need thousands of examples, and for complex problemssuch as image or speech recognition you may need millions of examples.

2. **Poor Quality of data:**

Obviously, if your training data has lots of errors, outliers, and noise, it will make it impossible for your machine learning model to detect a proper underlying pattern. Hence, it will not perform well.

So put in every ounce of effort in cleaning up your training data. No matter how good you are in selecting and hyper tuning the model, this part plays a major role in helping us make an accurate machine learning model.

“Most Data Scientists spend a significant part of their

time in cleaning data”.There are a couple of examples

when you’d want to clean up the data :

- If you see some of the instances are clear outliers just discard them or fix them manually.
- If some of the instances are missing a feature like (E.g., 2% of user did not specify their age), you can either ignore these instances, or fill the missing values by median age, or train one model with the feature and train one without it to come up with a conclusion.

3. Irrelevant Features:

Remove Garbage Data

The credit for a successful machine learning project goes to coming up with a good set of features on which it has been trained (often referred to as feature engineering), which includes feature selection, extraction, and creating new features which are other interesting topics to be covered in upcoming blogs.

4. Non representative training data:

To make sure that our model generalizes well, we have to make sure that our training data should be representative of the new cases that we want to generalize to.

If train our model by using a nonrepresentative training set, it won’t be accurate in predictions it will be biasedagainst one class or a group.

For E.G., Let us say you are trying to build a model that recognizes the genre of music. One way to build your training set is to search it on youtube and use the resulting data. Here we assume that youtube’s search engine is providing representative data but in reality, the search will be biased towards popular artists and maybe even the artists that are popular in your location(if you live in India you will be getting the music of Arijit Singh, Sonu Nigam or etc).

So use representative data during training, so your model won’t be biased among one or two classes when it
works on testing data.

5. Overfitting the Training Data

Overfitting happens when the model is too complex relative to the amount and noisiness of the training data.The possible solutions are:

To simplify the model by selecting one with fewer parameters (e.g., a linear model rather than a high-degree polynomial model), by reducing the number of attributes in the training data or by constraining the model

- To gather more training data

- To reduce the noise in the training data (e.g., fix data errors and remove outliers)

6. Underfitting the Training Data

Underfitting is the opposite of overfitting: it occurs when your model is too simple to learn the underlying structure of the data. For example, a linear model of life satisfaction is prone to underfit; reality is just more complex than the model, so its predictions are bound to be inaccurate, even on the training examples.

Chapter-II
Statistical Learning

Introduction, supervised and unsupervised learning, Training and Testing loss, Tradeoffs in statistical learning, Risk statistics, sampling distribution of an estimator, Empirical Risk minimization.

Introduction

→ structuring and visualizing data are important aspects of data science.

→ When The goal is to interpret The model and quantify The uncertainty in The data, This analysis is usually refered to as statistical learning.

→ There are two major goals for modeling data:

(1) To accurately predict some future quantity of interest, given some observed data

(2) To discover usual (or) interesting patterns in The data.

To achive These goals, one must rely on knowledge from Three important pillars of The mathematical sciences.

(1) Function approximation

(2) Optimization

(3) probability and statistics.

Function approximation:- A mathematical function is used to represent The relationships between The variables. As a data sceintist, you need to understand how best to approximate and represent function using least amount of computer processing and memory.

Optimization:-

→ Given a set of mathematical models, we wish to find best possible model in that set.

→ This step usually requires knowledge of optimization algorithms and efficient computer coding or programming.

probability and statistics

The knowledge of probability and statistics is needed to fit (or) train an algorithm and generate a model.

1. Supervised and Unsupervised Learning

→ Given an input 'x', one of the main goals of ML is to predict an output variable 'y'.

Examples:

1) x: digitized signature

y: Whether the signature is genuine (or) false.

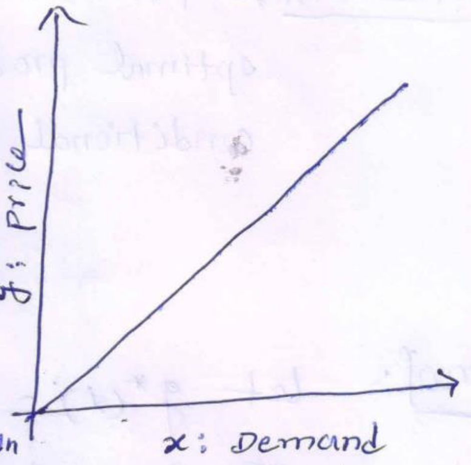
2) x: weight and smoking habits of an expecting mother.

y: The birth weight of the baby.

→ The data sceintist attempt at this prediction is encoded in a mathematical function 'g', called the prediction-function, which takes an input 'x' and outputs a guss $g(x)$ for 'y' denoted by $\hat{y}$.

→ In regression problems, the response variable 'y' can take any real value.

→ In contract, when 'y' can only lie in a finite set say $y \in \{0, \ldots, c-1\}$ then predicting 'y' is conceptually the same as classifying the input 'x' into one of 'c' categories, and so prediction becomes a classification problem.



→ We measure the accuracy of a prediction $\hat{y}$ with respect to a given response 'y' by using some loss function $\text{loss}(y, \hat{y})$

→ It is unlikely that any mathematical function 'g' ~~that~~ will be able to make predictions for all possible pairs (x, y) one may encounter in Nature.

→ One reason for this is that, even with the same input x, the output y may be different, depending on chance circumstances or randomness.

→ For this reason, we adopt a probabilistic approach and assume that each pair (x, y) is the outcome of a random pair (X, Y) that has some joint probability density $f(x, y)$.

→ We then assess the predictive performance via the expected loss, usually called the risk, for g:

$$l(g) = \mathbb{E}\,\text{Loss}(Y, g(X))$$

2. Training and Testing loss

In machine learning, training and testing loss are used to evaluate the performance of a model.

Training loss is the error rate on the training data during the training process. The goal of the training process is to minimize the loss, meaning the model is learning to make accurate predictions on the training data.

Testing loss is the error rate on the testing data, which is a separate dataset that is not used during during the training process. This is used to evaluate the generalization performance of the model, meaning how well it can make accurate predictions on new, unseen data.

The training loss typically decreases during the training process, as the model becomes better at making predictions the training data. However, if the model is overfitting to the training data, the testing loss may increase even as the training loss continues to decrease. This indicates that the model is not able to generalize well on to new data.

The goal of ML is to minimize both the training loss and testing loss, while also preventing overfitting. This is achieved by using techniques such as regularization, early stopping, and cross validation.

3. Risk statistics

Risk statistics are used in machine learning to evaluate the performance and generalization of a model. some commonly used risk stastistics include:

(1) Training Error:- This is the error rate of the model on the training data. The training error is used to evaluate how well the model fits the training data.

(2) Testing Error:- This is the error rate of the model on the test data. The test error is used to evaluate how well the model generalizes to new, unseen data

(3) cross-validation Error:- This the error rate of the model on a validation set that is created by partitioning the data into multiple subsets.

cross validation is used to estimate the generalization performance of the model and to prevent over fitting.

(4) Bias:- This the difference between the expected value of the predictions made by the model and the true value of the target variable. Bias measures how well the model captures the true relationship between the input features and the target variable.

(5) Variance : This is The variability of The model's predictions for different training sets. Variance measures how sensitive The model is to small changes in training data

(6) Mean sure Error (MSE):-

This is The average of The square differences between The predicted value and The true value. MSE is used to evaluate The overall performance of The model.

7. Root mean squared Error:-

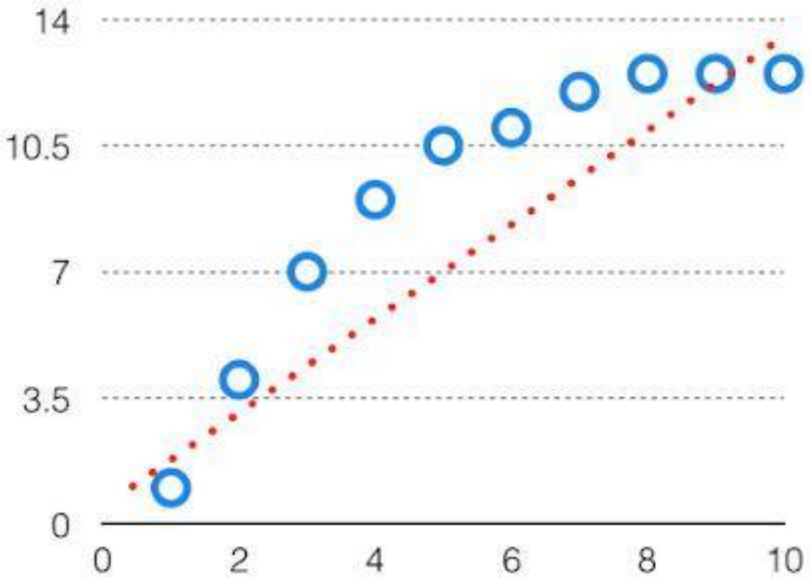
This is The square root of The MSE. RMSE is used to measure The average magnitude of The error made by The model.

In summary, risk stastistics are used to evaluate The performance and generalization of ML model. By using multiple risk stastistics, we can gain a more comprehensive understanding of The model's strengths and weaknesses.

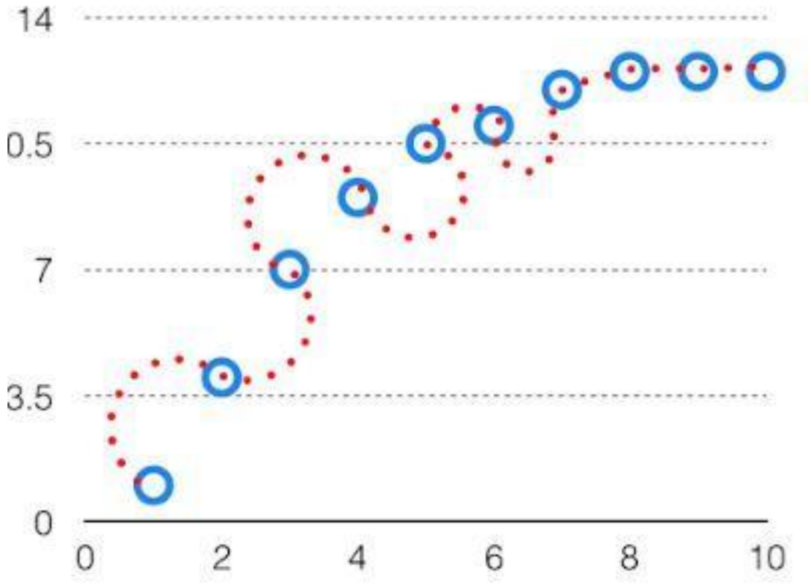
Tradeoffs in Statistical Learning :

The tradeoffs in statistical learning are mostly from bias and variance tradeoff which gives over fitting and uderfitting of data.

Bias: The bias is known as the difference between the prediction of the values by the Machine Learning model and the correct value. Being high in biasing gives a large error in training as well as testing data. It recommended that an algorithm should always be low-biased to avoid the problem of under fitting. By high bias, the data predicted is in a straight line format, thus not fitting accurately in the data in the data set. Such fitting is known as the Under fitting of data.

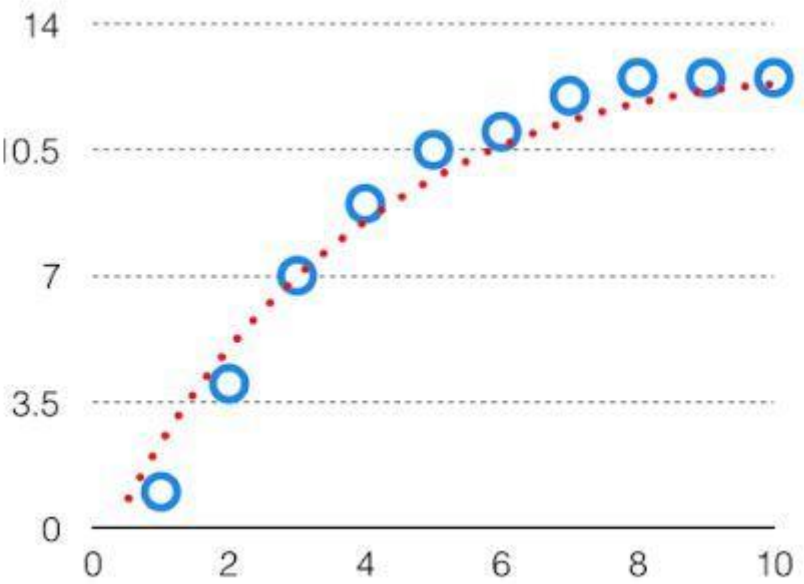


Variance: The variability of model prediction for a given data point which tells us the spread of our data is called the variance of the model. The model with high variance has a very complex fit to the training data and thus is not able to fit accurately on the data which it hasn't seen before. As a result, such models perform very well on training data but have high error rates on test data. When a model is high on variance, it is then said to as Over fitting of Data**.**



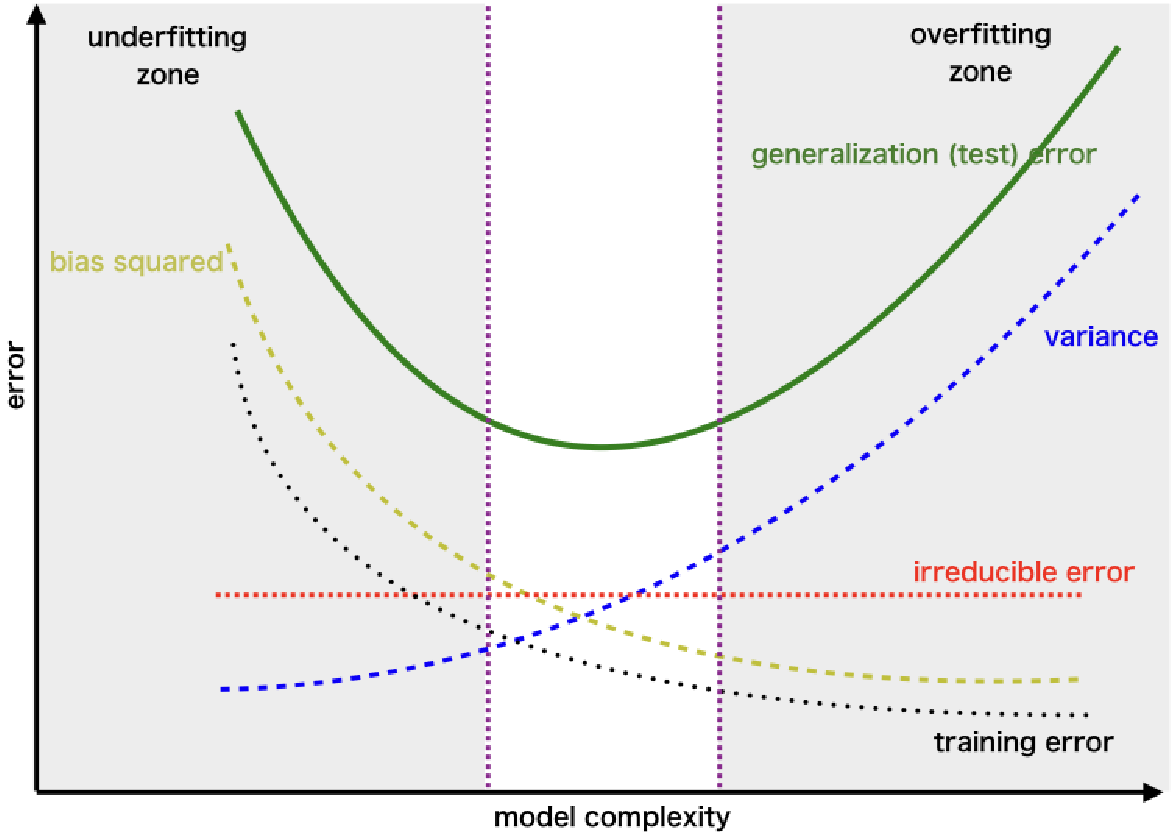
Bias Variance Tradeoff

If the algorithm is too simple (hypothesis with linear equation) then it may be on high bias and low variance condition and thus is error-prone. If algorithms fit too complex (hypothesis with high degree equation) then it may be on high variance and low bias. In the latter condition, the new entries will not perform well. Well, there is something between both of these conditions, known as a Trade-off or Bias Variance Trade-off. This tradeoff in complexity is why there is a tradeoff between bias and variance. An algorithm can't be more complex and less complex at the same time. For the graph, the perfect tradeoff will be like this.



We try to optimize the value of the total error for the model by using the bias and variance Tradeoff.

The best fit will be given by the hypothesis on the tradeoff point. The error to complexity graph to show trade-off is given as –



This is referred to as the best point chosen for the training of the algorithm which gives low error in training as well as testing data.

Sampling Distribution:

The sampling distribution of an estimator is a theoretical probability distribution that shows the possible values that the estimator can take when calculated from different random samples of the same size from the population.

Example:

- She wants to analyze the number of teens riding a bicycle between two regions of 13-18
- Instead of considering such individual in the population of 13-18 years of age in the 2 regions, she selected 200 samples randomly from each area.

  Here, the average count of the bicycle usage here is the sample mean
- Each chosen sample has its own generated mean, and the distribution for average mean is the sample distribution
- The deviation obtained is termed the standard error
- She plots the data gathered from the sample on a graph to go a clear view of the finite sample distribution

Empirical Risk Minimization

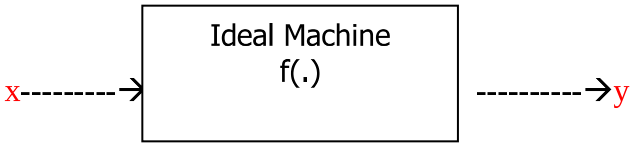
The Empirical Risk Minimization (ERM) principle is a learning paradigm which consists in selecting the model with minimal average error over the training set. This so-called training error can be seen as an estimate of the risk (due to the law of large numbers), hence the alternative name of empirical risk.

By minimizing the empirical risk, we hope to obtain a model with a low value of the risk. The larger the training set size is, the closer to the true risk the empirical risk

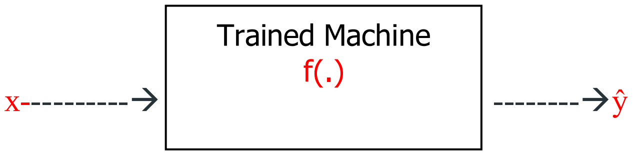
Consider the training data D= {(*Xi*, *Yi*)} for i=1, 2, 3... N

Assume p(x,y) be the probability density of the distribution from which training samples are drawn.

Assume a machine defined by a decision function *f*:X→Y
Where x is the input space and y is output space.
Here f maps from x∈X to y∈Y



Trained Machine:A machine when f(.) selected gives the output ŷ=f(x)



There may be multiple options available for f(.)
We need to selecty optimal f(.) that minimizes the loss in the predictions.

Let f(x,w) be the set of possible functions of training machines.

Where w are the adjustable parameters.

Loss function L(y,f(x,w)):To measure the error between the actual output y and predicted output ŷ=f(x,w)

Risk Function R(w):Refers to the risk /expected loss associated with the decision f(x,w)

$$R(w)= E[L(y,f(x,w))] = \int_{XxY} L(y,f(x,w))\, p(x,y)\, dxdy$$

Our goal is to train a machine with decision function f(x,w) against p(x,y) that minimizes the risk function R(w)

Here the issue is that the joint probability density function p(x,y) is not know explicitly.

Then the true risk function R(w) is approximated as empirical risk function Remp(W) from the training samples {*Xi*,*Yi*}.

$$Remp(w)=1/N \sum_{i=1}^{N} L(Yi,f(xi,W))$$

Empirical Risk value=Average loss over all training samples