UNIT – 1 Introduction to Machine Learning

By : Prof. Nootan Padia, Marwadi University, Rajkot

INTRODUCTION

- It has been more than 20 years – compute program defeated the world champion in a game –
- The computer program was

IBM's Deep Blue -> it defeated world chess champion Gary Kasparow.

- It requires a lot of intelligence to play.

- As of today, machine learning is a mature technology area finding its application in almost every sphere of life.
- It can recommend toys to toddlers much in the same way as it can suggest a technology book to a geek or a rich title in literature to a writer.
- It predicts the future market to help amateur (non-specialist) traders compete with seasoned stock traders.
- It helps an oncologist find whether a tumour is malignant or benign.
- It helps in optimizing energy consumption thus helping the cause of Green Earth.
- Google has become one of the front – runners focusing a lot of its research on machine learning and artificial intelligence – Google self – driving car and Google Brain being two most ambitious projects of Google in its journey of innovation in the field of machine learning.

EVOLUTION OF MACHINE LEARNING

1950	• Alan Turing proposes “learning machine”.
1952	• Arthur Samuel developed first machine learning program that could play Checkers.
1957	• Frank Rosenblatt designed the first neural network program simulating human brain.

1967
- Nearest neighbour algorithm created – start of basic pattern recognition

1979
- Stanford university students develop first self – driving cart that can navigate and avoid obstacles in a room.

1982
- Recurrent Neural network developed.

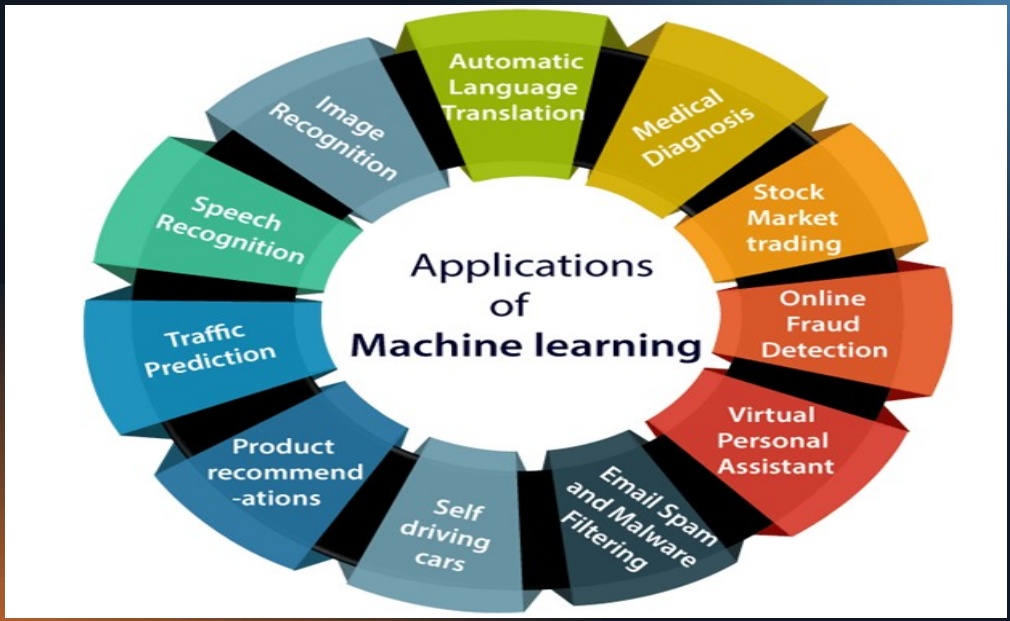
1989
- Reinforcement Learning conceptualized.
- Beginning of commercialization of Machine Learning

1995
- Random Forest and Support Vector machine algorithms developed.

1997
- IBM's Deep Blue beats the world chess champion Gary Kasparov.

2006	• First machine learning competition launched by Netflix. • Geoffrey Hinton conceptualizes Deep Learning.
2010	• Kaggle, a website for machine learning competitions, launched
2011	• IBM's Watson beats two human champions in Jeopardy
2016	• Google's AlphaGo program beats unhandicapped professional human player

- 2017 – A Machine learns how to stop online trolling.
  - the Alphabet's Jigsaw team built an intelligent system that was able to learn the online trolling. It used to read millions of comments of different websites to learn to stop online trolling.
- 2017 – What next in Machine learning?
  - How far this technology will take us remains to be seen, but applications in the works span everything from improving in – store retail experiences with IoT to boosting security with biometric data to predict and diagnosing disease.



WHAT IS HUMAN LEARNING?

- In cognitive science, learning is typically referred to as the process of gaining information through observation.
- And why do we need to learn?
- In our daily life, we need to carry out multiple activities.
- It may be a task as simple as walking down the street or doing the homework.
- Or it may be some complex task like deciding the angle in which a rocket should be launched so that it can have a particular trajectory.
- To do task in a proper way, we need to have prior information on one or more things related to the task.

- Also, as we keep learning more or in other words acquiring more information, the efficiency in doing the tasks keep improving.
- For e.g., with more knowledge, the ability to do homework with less number of mistakes increases.
- In the same way, information from past rocket launches helps in taking the right precautions and makes more successful rocket launch.
- Thus, with more learning, tasks can be performed more efficiently.

TYPES OF HUMAN LEARNING

- Learning under expert guidance
- Learning guided by knowledge gained from experts
- Learning by self

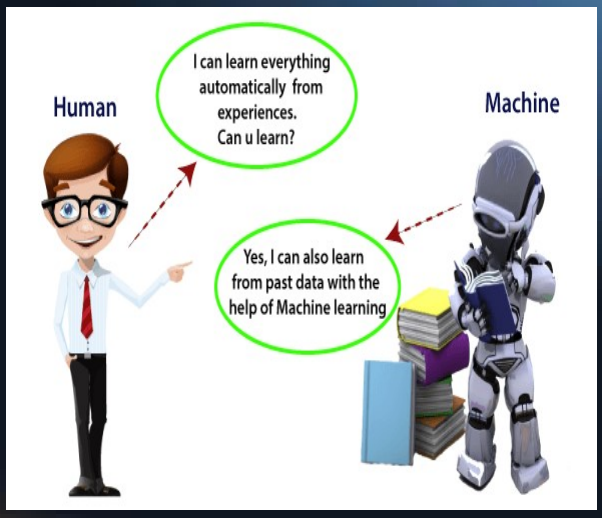
- Learning under expert guidance :
  - Infant – learn from parents – e.g. He calls his nose, a 'nose', because that is the information he gets from his parents.
  - Baby going to school – learns from teacher – e.g. alphabets and digits, how to form words, sentences, paragraphs, complex mathematics, science, etc.
  - Higher studies – experts or teachers – learns application oriented and complex skills. – e.g. engineering students get skilled in one of the disciplines like civil, computer science, electrical and mechanical, etc.
  - Working as a professional in some field – learn more about the hands-on application of the knowledge that he has acquired. The professional mentors help all new comers in the field to learn on – job.
  - In all phases of life of a human being, there is an element of guided learning.
  - This learning is imparted by someone, purely because of the fact that he/she has already gathered the knowledge by virtue of his/her experience in that field.
  - So guided learning is the process of gaining information from a person having sufficient knowledge due to the past experience.

- Learning guided by knowledge gained from experts :
  - An essential part of learning also happens with the knowledge which has been imparted by teacher or mentor at some point of time in some other form/context.
  - For e.g., a baby can group together all objects of same colour even if his parents have not specifically taught him to do so.
  - He is able to do so because at some point of time or other his parents have told him which colour is blue, which is green, etc.
  - A grown – up kid can select one odd word from a set of words because it is a verb and other words being all nouns.
  - He could do this because of his ability to label the words as verbs or nouns, thought by his English teacher long back.
  - In professional role, a person is able to make out to which customers he should market a campaign from the knowledge about preference that was given by his boss long back.
  - In all these situations, there is no direct learning.
  - It is some past information shared on some different context, which is used as a learning to make decisions.

- Learning by self :
  - In many situations, humans are left to learn on their own.
  - A classic example is a baby learning to walk through obstacles.
  - He bumps on the obstacles and falls down multiple times till he learns that whenever there is an obstacle, he needs to cross over it.
  - Not all things are taught by others.
  - A lot of things need to be learnt only from mistakes made in the past.
  - We tend to form a check list on things that we should do, and things that we should not do, based on our experiences.

What is Machine Learning?

- Some fundamental questions :
  - Do machines really learn?
  - If so, how do they learn?
  - Which problem can we consider as a well – posed learning problem? What are the important features that are required to well – define a learning problem?



- There are multiple ways to define machine learning.
- But the one which is perhaps most relevant, concise and accepted universally is the one stated by Tom M. Mitchell, Professor of Machine Learning Department, School of Computer Science, Cornegie Mellon University.
- The definition is :

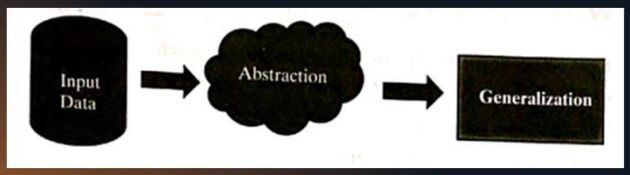
"A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with Experience E."

- What this essentially means is that a machine can be considered to learn if it is able to gather experience by doing a certain task and improve its performance in doing the similar task in the future.
- When we talk about past experience, it means past data related to the task.
- This data is an input to the machine from some source.
- In short,

"Machine learning enables a machine to automatically learn from data, improve performance from experiences, and predict things without being explicitly programmed."

How do machines learn?

- The basic machine learning process can be divided into three parts :
  - Data Input :
    - Past data or information is utilized as a basis for future decision – making.
  - Abstraction :
    - The input data is represented in a broader way through the underlying algorithm.
  - Generalization :
    - The abstracted representation is generalized to form a framework for making decisions.



- Example :

Object

Abstraction

Generalization

- Here, try to understand process of machine learning using human learning process :
- Situation : typical process of learning from classroom and books and preparing for the examination.
  - it is a tendency of many students to try and memorize (we often call it 'learn by heart') as many things as possible.
  - This may work well when the scope of learning is not so vast.
  - Also, the kinds of questions which are asked in the examination are pretty much simple and straightforward.
  - The questions can be answered by simply writing the same things which have been memorized.

- However, as the scope gets broader and the questions asked in the examination gets more complex, the strategy of memorizing doesn't work well.
- The no. of topics may get too vast for a student to memorize.
- Also, the capability of memorizing varies from student to student.
- Together with that, since the questions get more complex, a direct reproduction of the things memorized may not help.
- The situation continues to get worse as the student graduates to higher classes.

- So, what we seen in the case of human learning is that just by great memorizing and perfect recall, i.e. just based on knowledge input, students can do well in the examinations only till a certain stage.
- Beyond that, a better learning strategy needs to be adopted :
  - To be able to deal with the vastness of the subject matter and the related issues in memorizing it
  - To be able to answer questions where a direct answer has not been learnt

- A good option is to figure out the key points or ideas amongst a vast pool of knowledge.
- This helps in creating an outline of topics and a conceptual mapping of those out-lined topics with the entire knowledge pool.
- For example, a broad pool of knowledge may consist of all living animals and their characteristics such as whether they live in land or water, whether they lay eggs, whether they have scales or fur or none etc.
- It is a difficult task for any student to memorize the characteristics of all living animals —no matter how much photographic memory he/she may possess.
- It is better to draw a notion about the basic groups that all living animals belong to and the characteristics which define each of the basic groups.

- The basic groups of animals are invertebrates and vertebrates.
- Vertebrates are further grouped as mammals, reptiles, amphibians, fishes, and birds. Here, we have mapped animal groups and their salient characteristics.

1. Invertebrate: Do not have backbones and skeletons
2. Vertebrate
   (a) Fishes: Always live in water and lay eggs
   (b) Amphibians: Semi-aquatic i.e. may live in water or land; smooth skin; lay eggs
   (c) Reptiles: Semi-aquatic like amphibians; scaly skin; lay eggs; cold-blooded
   (d) Birds: Can fly; lay eggs; warm-blooded
   (e) Mammals: Have hair or fur; have milk to feed their young; warm-blooded

- This makes it easier to memorize as the scope now reduces to know the animal groups that the animals belong to.
- Moving to the machine learning paradigm, the vast pool of knowledge is available from the data input.
- However, rather than using it in entirety, a concept map much in line with the animal group to characteristic mapping explained above, is drawn from the input data.
- This is nothing but knowledge abstraction as performed by the machine.

- In the end, the abstracted mapping from the input data can be applied to make critical conclusions.
- For example, if the group of an animal is given, understanding of the characteristics can be automatically made.
- Reversely, if the characteristic of an unknown animal is given, a definite conclusion can he made about the animal group it belongs to.
- This is generalization in context of machine learning.

Abstraction

- Abstraction helps in deriving a conceptual map based on the input data.
- This map, or a model as it is known in the machine learning paradigm, is summarized knowledge representation of the raw data.
- The model may be in any one of the following forms :
  - Computational blocks like if/else rules
  - Mathematical equations
  - Specific data structures like trees or graphs
  - Logical groupings of similar observations
- The choice of the model used to solve a specific learning problem is a human task.

- The decision related to the choice of model is taken based on multiple aspects, some of which are listed below:
  - The type of problem to be solved: Whether the problem is related to forecast or prediction, analysis of trend, understanding the different segments or groups of objects, etc.
  - Nature of the input data: How exhaustive the input data is, whether the data has no values for many fields, the data types, etc.
  - Domain of the problem: If the problem is in a business critical domain with a high rate of data input and need for immediate inference(conclusion), e.g. fraud detection problem in banking domain.

- Once the model is chosen, the next task is to fit the model based on the input data.
- Let's understand this with an example.
- In a case where the model is represented by a mathematical equation, say

  $y = C_1 + C_2 x'$ – *simple linear regression,*

  based on the input data, we have to find out the values of C1 and C2.
- Otherwise, the equation (or the model) is of no use.
- So, fitting the model, in this case, means finding the values of the unknown coefficients or constants of the equation or the model.
- This process of fitting the model based on the input data is known as training.
- Also, the input data based on which the model is being finalized is known as training data.

Generalization

- The other key part is to tune up the abstracted knowledge to a form which can be used to take future decisions.
- This is achieved as a part of generalization.
- This part is quite difficult to achieve.
- This is because the model is trained based on a finite set of data, which may possess a limited set of characteristics.
- But when we want to apply the model to take decision on a set of unknown data, usually termed as test data, we may encounter two problems:

  - The trained model is aligned with the training data too much, hence may not portray the actual trend.
  - The test data possess certain characteristics apparently unknown to the training data.
- Hence, a precise approach of decision-making will not work.
- An approximate or heuristic approach, much like gut-feeling-based decision-making in human beings has to be adopted.
- This approach has the risk of not making a correct decision — quite obviously because certain assumptions that are made may not be true in reality.
- But just like machines, same mistakes can be made by humans too when a decision is made based on intuition or gut-feeling — in a situation where exact reason-based decision-making is not possible.

Well-posed learning problem

- For defining a new problem, which can be solved using machine learning, a simple framework, can be used.
- This framework also helps in deciding whether the problem is a right candidate to be solved using machine learning.
- The framework involves answering three questions:
  - What is the problem?
  - Why does the problem need to be solved?
  - How to solve the problem?

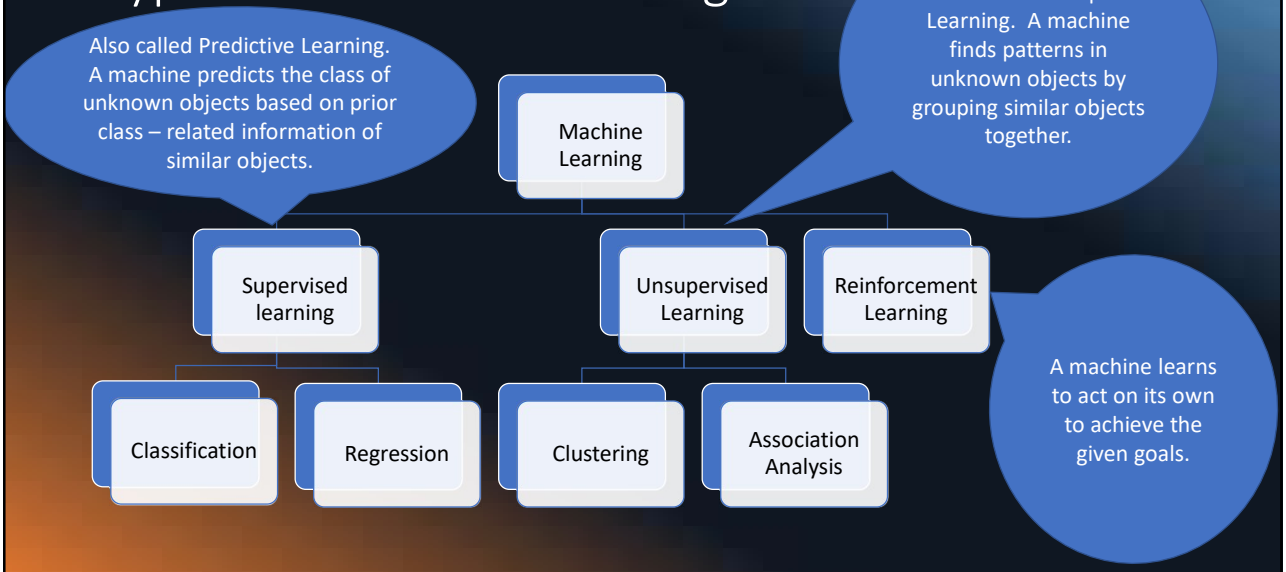
- Step – 1 : What is the Problem?
  - A no. of information should be collected to know what is the problem.
  - Informal description of the problem, e.g., I need a program that will prompt next word as and when I type a word.
  - Formalism :
    - Use Tom Mitchell's machine learning formalism stated above to define T. P, and E for the problem.
    - For e.g. :
      - Task (T) : Prompt the next word when I type a word.
      - Experience (E) : A corpus of commonly used English words and phrases.
      - Performance (P) : The no. of correct words prompted considered as a percentage (which in machine learning paradigm is known as learning accuracy).
    - **Assumptions** - Create a list of assumptions about the problem.

- Step 2: Why does the problem need to be solved?
  - Motivation
    - What is the motivation for solving the problem? What requirement will it fulfill?
    - For example, does this problem solve any long-standing business issue like finding out potentially fraudulent transactions? Or the purpose is more trivial like trying to suggest sonic movies for upcoming weekend.
  - Solution benefits
    - Consider the benefits of solving the problem. What capabilities does it enable?
    - It is important to clearly understand the benefits of solving the problem. These benefits can be articulated to sell the project.
  - Solution use
    - How will the solution to the problem be used and the life time of the solution is expected to have?

- Step 3: How would I solve the problem?
  - Try to explore how to solve the problem manually.
  - Detail out step-by-step data collection, data preparation, and program design to solve the problem. Collect all these details and update the previous sections of the problem definition, especially the assumptions.

- **Summary**
  - Step 1: What is the problem? Describe the problem informally and formally and list assumptions and similar problems.
  - Step 2: Why does the problem need to be solved? List the motivation for solving the problem, the benefits that the solution will provide and how the solution will be used.
  - Step 3: How would I solve the problem? Describe how the problem would be solved manually to flush domain knowledge.

Types of Machine Learning



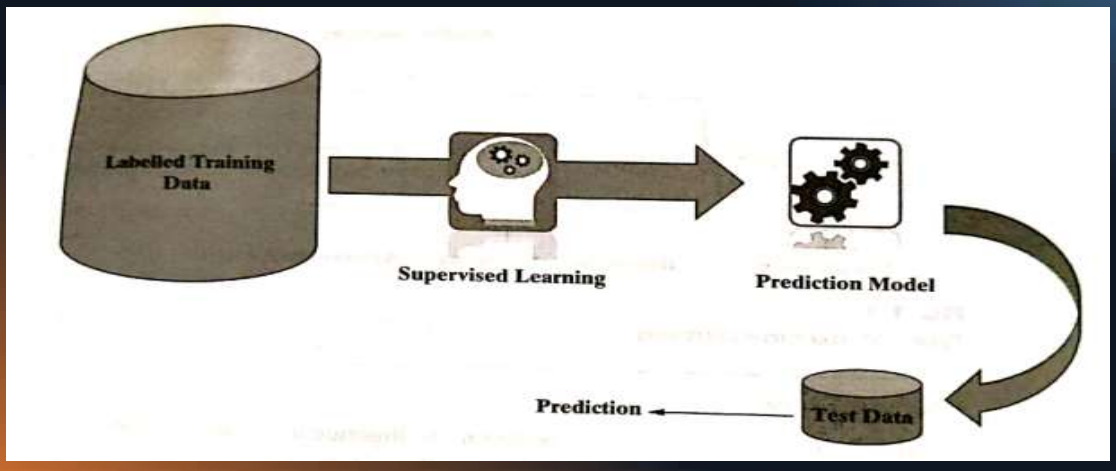
Supervised Learning

- The major motivation of supervised learning is to learn from past information.
- So what kind of past information does the machine need for supervised learning?
- It is the information about the task which the machine has to execute.
- In context of the definition of machine learning, this past information is the experience.
- **Example:**
  - Say a machine is getting images of different objects as input and the task is to segregate the images by either shape or color of the object.
  - If it is by shape, the images which are of round-shaped objects need to be separated from images of triangular-shaped objects, etc.
  - If the segregation needs to happen based on color, images of blue objects need to be separated from images of green objects.

- But how can the machine know what is round shape, or triangular shape?
- Same way, how can the machine distinguish image of an object based on whether it is blue or green in color?

- A machine is very much like a little child whose parents or adults need to guide him with the basic information on shape and color before he can start doing the task.
- A machine needs the basic information to be provided to it.
- This basic input, or the experience in the paradigm of machine learning, is given in the form of training data.
- Training data is the past information on a specific task.
- In context of the image segregation problem, training data will have past data on different aspects or features on a number of images, along with a tag on whether the image is round or triangular, or blue or green in color.

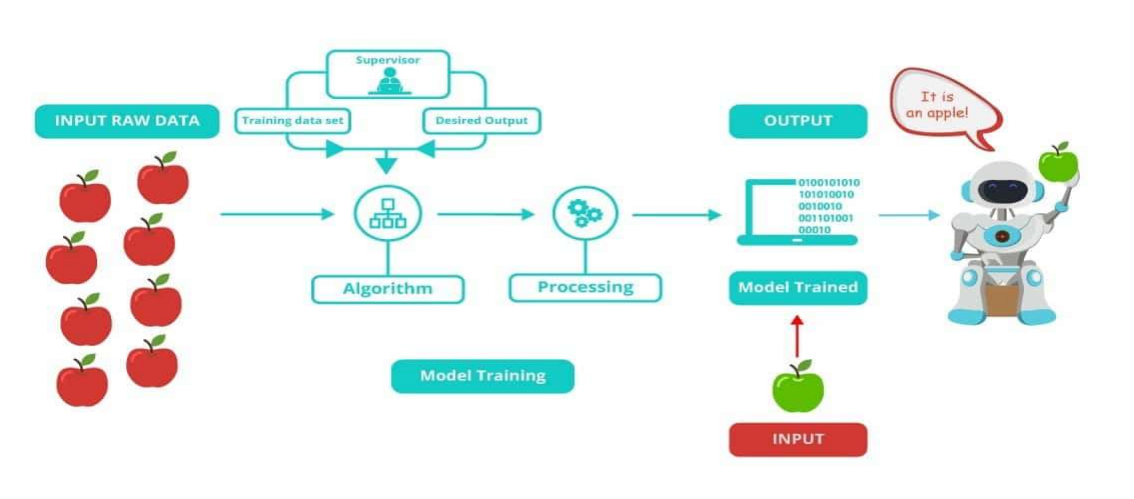
- The tag is called 'label' and we say that the training data is labeled in case of supervised learning.
- Figure - simple depiction of the supervised learning process.



- Labeled training data containing past information comes as an input.
- Based on the training data, the machine builds a predictive model that can be used on test data to assign a label for each record in the test data.
- Some examples of supervised learning are
  - Predicting the results of a game
  - Predicting whether a tumour is malignant or benign
  - Predicting the price of domains like real estate, stocks, etc.
  - Classifying texts such as classifying a set of emails as Spam or non-spam

- Now, let's consider two of the above examples, say 'predicting whether a tumor is malignant or benign' and 'predicting price of domains such as real estate'.
- Are these two problems same in nature?
- The answer is 'no'.
- Though, both of them are prediction problems, in one case we are trying to predict which category or class an unknown data belongs to whereas in the other case we are trying to predict an absolute value and not a class.
- When we are trying to predict a categorical or nominal variable, the problem is known as a classification problem.
- Whereas when we are trying to predict a real-valued variable, the problem falls under the category of regression.

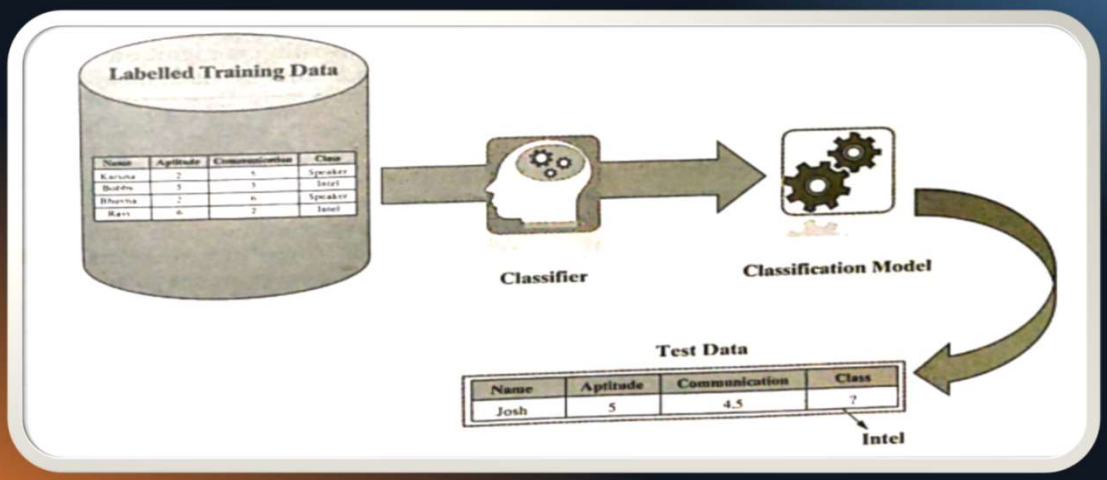
- Example :



Classification

- Classification is the process of predicting the class of given data points.
- Classes are sometimes called as targets/ labels or categories.
- Let's discuss how to segregate the images of objects based on the shape.
  - If the image is of a round object, it is put under one category, while if the image is of a triangular object, it is put under another category.
  - In which category the machine should put an image of unknown category, also called a test data in machine learning parlance, depends on the information it gets from the past data, which we have called as training data.
  - Since the training data has a label or category defined for each and every image, the machine has to map a new image or test data to a set of images to which it is similar to and assign the same label or category to the test data.
  - So we observe that the whole problem revolves around assigning a label or category or class to a test data based on the label or category or class information that is imparted by the training data.

- Since the target objective is to assign a class label, this type of problem as classification problem.
- Figure depicts the typical process of classification.



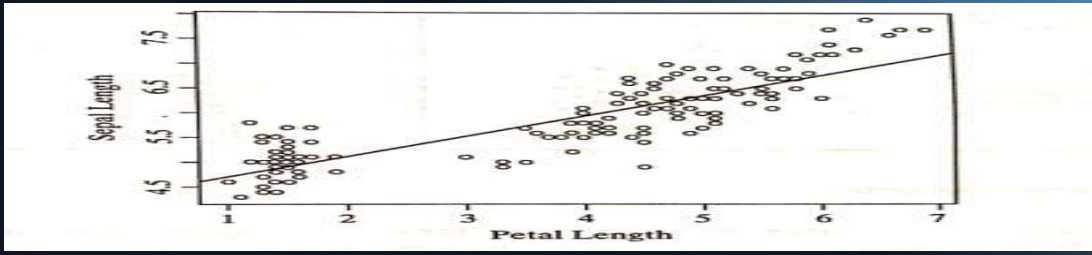
- There are number of popular machine learning algorithms which help in solving classification problems.
- To name a few, Naïve Bayes, Decision tree, Random forest and k-Near Neighbour algorithms are adopted by many machine learning practitioners.
- A critical classification problem in context of banking domain is identifying potential fraudulent transactions.
- Since there are millions of transactions which have to be scrutinized and assured whether it might be a fraud transaction, it is not any human being to carry out this task.
- Machine learning is effectively leveraged to do this task and this is a classic case of classification.
- Based on the past transaction data, specifically the ones labeled as fraudulent, all new incoming transactions are, marked or labeled as normal or suspicious.

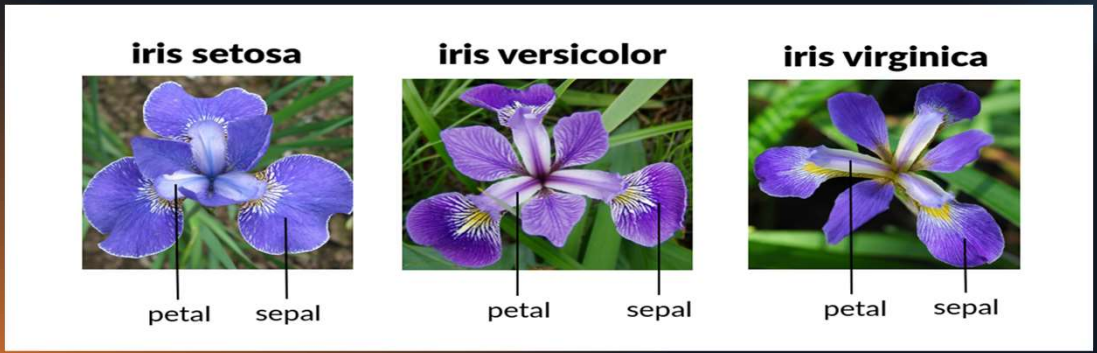
- The suspicious transactions are subsequently segregated for a closer review.
- In summary, classification is a type of supervised learning where a target feature, which is of type categorical, is predicted for test data based on the information imparted by training data.
- The target categorical feature is known as class.
- Some typical classification problems include:
  - Image classification
  - Prediction of disease
  - Win—loss prediction of games
  - Prediction of natural calamity like earthquake, flood, etc.
  - Recognition of handwriting
- ***Machine learning saves life — ML can spot 52 % of breast cancer cells, a year before patients are diagnosed.***
- ***US Postal Service uses machine learning for handwriting recognition.***

Regression

- In linear regression, the objective is to predict numerical features like real estate or stock price, temperature, marks in an examination, sales revenue, etc.
- The underlying predictor variable and the target variable are continuous in nature.
- In case of linear regression, a straight line relationship is 'fitted' between the predictor variables and the target variables, using the statistical concept of least squares method.
- As in the case of least squares method, the sum of square of error between actual and predicted values of the target variable is tried to be minimized.
- In case of simple linear regression, there is only one predictor variable whereas in case of multiple linear regression, multiple predictor variables can be included in the model.

- Let's take the example of yearly budgeting exercise of the sales managers.
- They have to give sales prediction for the next year based on sales figure of previous years vis-a-vis investment being put in.
- Obviously, the data related to past as well as the data to be predicted are continuous in nature.
- In a basic approach, a simple linear regression model can be applied with investment as predictor variable and sales revenue as the target variable.
- Figure in next slide shows a typical simple regression model, where regression line is fitted based on values of target variable with respect to different values of predictor variable.





- A typical linear regression model can he represented in the form
- $y = \alpha + \beta x$
  - where 'x' is the predictor variable and 'y' is the target variable.
- The input data come from a famous multivariate data set named Iris introduced by the British statistician and biologist Ronald Fisher.
- The data set consists of 50 samples from each of three species of Iris — Iris setosa, Iris virginica, and Iris versicolor.
- Four features were measured for each sample — sepal length, sepal width, petal length, and petal width.
- These features can uniquely discriminate the different species of the flower.
- The Iris data set is typically used as a training data for solving the classification problem of predicting the flower species based on feature values.

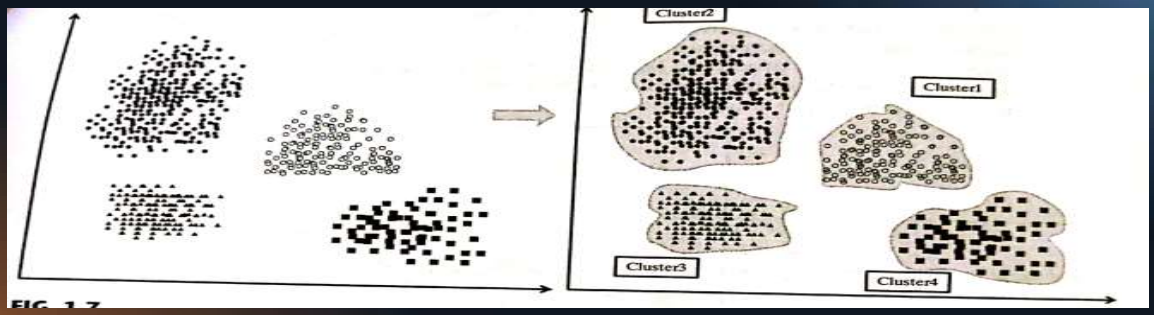
- However, we can also demonstrate regression using this data set, by predicting the value of one feature using another feature as predictor.
- In Figure of previous slide, petal length is a predictor variable which, when fitted in the simple linear regression model, helps in predicting the value of the target variable sepal length.
- Typical applications of regression can be seen in
  - Demand forecasting in retails
  - Sales prediction for managers
  - Price prediction in real estate
  - Weather forecast
  - Skill demand forecast in job market

Unsupervised learning

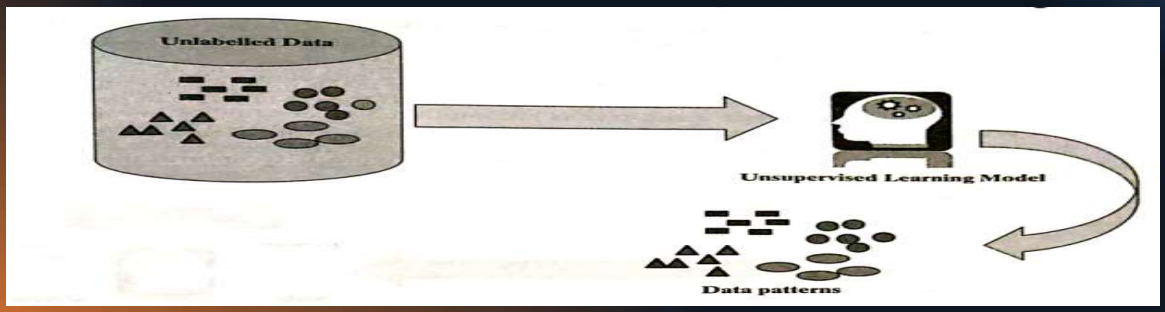
- In unsupervised learning, there is no labeled training data to learn from and no prediction to be made.
- In unsupervised learning, the objective is to take a dataset as input and try to find natural groupings or patterns within the data elements or records.

- Therefore, unsupervised learning is often termed as descriptive model and the process of unsupervised learning is referred as pattern discovery or knowledge discovery.
- One critical application of unsupervised learning is customer segmentation.
- Clustering is the main type of unsupervised learning.
- It intends to group or organize similar objects together.
- For that reason, objects belonging to the same cluster are quite similar to each other while objects belonging to different clusters are quite dissimilar.

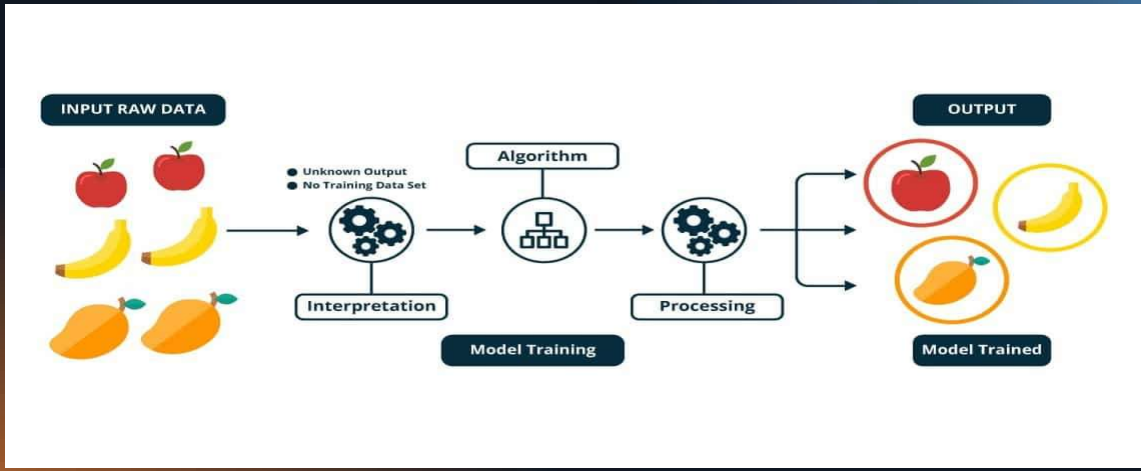
- Hence, the objective of clustering to discover the intrinsic grouping of unlabeled data and form clusters, as depicted in Figure.



- Different measures of similarity can be applied for clustering.
- One of the most commonly adopted similarity measure is distance.
- Two data items are considered as a part of the same cluster if the distance between them is less.
- In the same way, if the distance between the data items is high, the items do not generally belong to the same cluster.
- This is also known as distance-based clustering.
- Figure depicts the process of clustering at a high level.



- Example :



- Other than clustering of data and getting a summarized view from it, one more variant of unsupervised learning is association analysis.
- As a part of association analysis, the association between data elements is identified.
- Let's try to understand the approach of association analysis in context of one of the most common examples, i.e. market basket analysis as shown in Figure in next slide.
- From past transaction data in a grocery store, it may be observed that most of the customers who have bought item A , have also bought item B and item C or at least one of them.
- It means that there is a strong association of the event 'purchase of item 'A' with the event 'purchase of item 'B' or 'purchase of item 'C'.

- Identifying these sorts of associations is the goal of association analysis.
- This helps in boosting up sales pipeline, hence a critical input for the sales group.
- Critical applications of association analysis include market basket analysis and recommender systems.

TransID	Items Bought
1	{Butter, Bread}
2	{Diaper, Bread, Milk, Beer}
3	{Milk, Chicken, Beer, Diaper}
4	{Bread, Diaper, Chicken, Beer}
5	{Diaper, Beer, Cookies, Ice cream}
...	...

Market Basket transactions
Frequent itemsets → (Diaper, Beer)
Possible association: Diaper → Beer

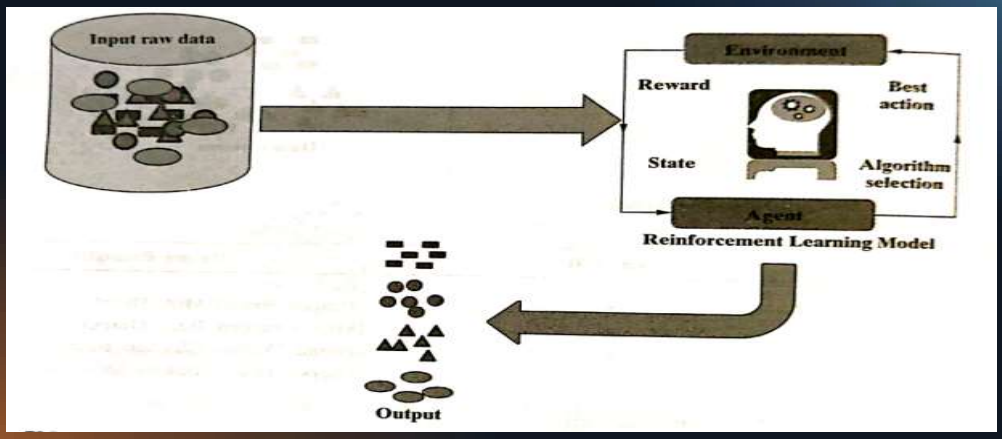
Market basket analysis

Reinforcement learning

- We have seen babies learn to walk without any prior knowledge of how to do it.
- Often we wonder how they really do it.
- They do it in a relatively simple way.
- First they notice somebody else walking around, for example parents or anyone living around.
- They understand that legs have to be used, one at a time, to take a step.
- While walking, sometimes they fall down hitting an obstacle, whereas other times they are able to walk smoothly avoiding bumpy obstacles.
- When they are able to walk overcoming the obstacle, their parents are elated and appreciate the baby with loud claps / or may be a chocolates.
- When they fall down while circumventing an obstacle, obviously their Parents do not give claps or chocolates.

- Slowly a time comes when the babies learn from mistakes and are able to walk with much ease.
- In the same way, machines often learn to do tasks autonomously.
- Let's try to understand in context of the example of the child learning to walk.
- The action tried to be achieved is walking, the child is the agent and the place with hurdles on which the child is trying to walk resembles the environment.
- It tries to improve its performance of doing the task.
- When a sub-task is accomplished successfully, a reward is given.
- When a sub-task is not executed correctly, obviously no reward is given.
- This continues till the machine is able to complete execution of the whole task.

- This process of learning is known as reinforcement learning.
- Figure captures the high-level process of reinforcement learning.



- One contemporary example of reinforcement learning is self-driving cars.
- The critical information which it needs to take care of are speed and speed limit in different road segments, traffic conditions, road conditions, weather conditions, etc.
- The tasks that have to be taken care of are start/stop, accelerate/ decelerate, turn to left/right, etc.
- *Reinforcement learning is getting more and more attention from both industry learning and academia. Annual publications count in the area of reinforcement in Google Scholar support this view.*
- *AlphaGo used RL to defeat the best human Go player.*
- *RL is an effective tool for personalized online marketing. It considers the demo-graphic details and browsing history of the user real-time to show most relevant advertisements.*

SUPERVISED	UNSUPERVISED	REINFORCEMENT
This type of learning is used when you know how to classify a given data, or in other words classes or labels are available.	This type of learning is used when there is no idea about the class or label of a particular data. The model has to find pattern in the data.	This type of learning is used when there is no idea about the class or label of a particular data. The model has to do the classification – it will get rewarded if the classification is correct, else get punished.
Labelled training data is needed. Model is built based on training data.	Any unknown and unlabelled data set is given to the model as input and records are grouped.	The model learns and updates itself through reward/ punishment.
The model performance can be evaluated based on how many misclassifications have been done based on a comparison between predicted and actual values.	Difficult to measure whether the model did something useful or interesting. Homogeneity of records grouped together is the only measure.	Model is evaluated by means of the reward function after it had some time to learn.
There are two types of supervised learning problems – classification and regression.	There are two types of unsupervised learning problems – clustering and association.	No such types.
Simplest one to understand.	More difficult to understand and implement than supervised learning.	Most complex to understand and apply.
Standard algorithms include • Naïve Bayes • *k*-nearest neighbour (kNN) • Decision tree • Linear regression • Logistic regression • Support Vector Machine SVM), etc.	Standard algorithms are • *k*-means • Principal Component Analysis (PCA) • Self-organizing map (SOM) • Apriori algorithm • DBSCAN etc.	Standard algorithms are • Q-learning • Sarsa
Practical applications include • Handwriting recognition • Stock market prediction • Disease prediction • Fraud detection, etc.	Practical applications include • Market basket analysis • Recommender systems • Customer segmentation, etc.	Practical applications include • Self-driving cars • Intelligent robots • AlphaGo Zero (the latest version of DeepMind's AI system playing Go)

Applications of machine learning

- Wherever there is a substantial amount of past data, machine learning can be used to generate actionable insight from the data.
- Here three areas have been covered :
  - Banking and finance
  - Insurance
  - Healthcare
- ***Banking and finance***
  - In the banking industry, fraudulent transactions, especially the ones related to credit cards, are extremely prevalent.
  - Since the volumes as well as velocity of the transactions are extremely high, high performance machine learning solutions are implemented by almost all leading banks across the globe.
  - The models work on a real-time basis, i.e. the fraudulent transactions are spotted and prevented right at the time of occurrence.

- This helps in avoiding a lot of operational hassles in settling the disputes that customers will otherwise raise against those fraudulent transactions.
- Customers of a bank are often offered lucrative (profitable) proposals by other competitor banks.
- Proposals like higher bank interest, to processing charge of loans, zero balance savings accounts, no overdraft penalty, etc. are offered to customers, with the intent that the customer switches over to the competitor bank.
- Also, sometimes customers get demotivated by the poor quality of services of the banks and shift to competitor banks.
- Machine learning helps in preventing or at least reducing the customer churn.
- Both descriptive and predictive learning can be applied for reducing customer churn.

- Using descriptive learning, the specific pockets of problem, i.e. a specific bank or a specific zone or a specific type of offering like car loan, may be spotted where maximum churn is happening.
- Quite obviously, these are troubled areas where further investigation needs to be done to find and fix the root cause.
- Using predictive learning, the set of vulnerable customers who may leave the bank very soon, can be identified.

- ***Insurance***
  - Insurance industry is extremely data intensive.
  - For that reason, machine learning is extensively used in the insurance industry.
  - Two major areas in the insurance industry where machine learning is used are risk prediction during new customer onboarding and claims management.

- During customer boarding, based on the past information, the risk profile of a new customer needs to be predicted.
- Based on the quantum of risk predicted, the quote is generated for the prospective customer.
- When a customer claim comes for settlement, past information related to historic claims along with the adjustor notes are considered to predict whether there is any possibility of the claim to be fraudulent.
- Other than the past information related to the specific customer, information related to similar customers, i.e. customer belonging to the same geographical location, age group, ethnic group, etc., are also considered to formulate the model.

- ***Healthcare***
  - Wearable device data form a rich source for applying machine learning and predict the health conditions of the person real time.

- In case there is some health issue which is predicted by the learning model, immediately the person is alerted to take preventive action.
- In case of some extreme problem, doctors or healthcare providers in the vicinity (area) of the person can be alerted.
- In case of sudden health problems in situation like morning walk of elderly person, which is tracked by the wearable device, the wearable data is sent to a remote server and a machine learning algorithm is constantly analyzing the streaming data.
- It also has the history of the elderly person and persons of similar age group.
- The model predicts some fatality unless immediate action is taken.
- Alert can be sent to the person to immediately stop walking and take rest.
- Doctors and healthcare providers can be alerted to be on standby.
- Machine learning along with computer vision also plays a crucial role in disease diagnosis from medical imaging.

Languages / tools for machine learning

- The algorithms related to different machine learning tasks are known to all and can be implemented using any language/platform.
- It can be implemented using a Java platform or C / C++ language or in .NET.
- However, there are certain languages and tools which have been developed with a focus for implementing machine learning.
- ***Python***
  - Python is one of the most popular, open source programming language widely adopted by machine learning community.
  - It was designed by Guido van Rossum and was first released in 1991.
  - The reference implementation of Python, i.e. CPython, is managed by Python Software Foundation.

- Python has very strong libraries for advanced mathematical functionalities (NumPy), algorithms and mathematical tools (SciPy) and numerical plotting there is a machine learning library named (matplotlib).
- Built on these libraries, scikit-learn, which has various classification, regression, and clustering algorithms embedded in it.

- ***R***
  - R is a language for statistical computing and data analysis.
  - It is an open source language, extremely popular in the academic community especially among statisticians and data miners.
  - R is considered as a variant of S, a GNU project which was developed at Bell Laboratories.
  - Currently, it is supported by the R Foundation for statistical computing.
  - R is a very simple programming language with a huge set of libraries available for different stages of machine learning.

- Some of the libraries standing out in terms of popularity are plyr/dplyr (for data transformation), caret (Classification and Regression Training' for classification), RJava (to facilitate integration with Java), tm (for text mining), ggplot2 (for data visualization).
- Other than the libraries, certain packages like Shiny and R Markdown have been developed around R to develop interactive web applications, documents and dashboards on R without much effort.

- ***Matlab***
  - MATLAB (matrix laboratory) is a licensed commercial software with a robust sup-port for a wide range of numerical computing.
  - MATLAB has a huge user base across industry and academia.
  - MATLAB is developed by MathWorks, a company founded in 1984.
  - Being proprietary software, MATLAB is developed much more professionally, tested rigorously, and has comprehensive documentation.

- MATLAB also provides extensive support of statistical functions and has a huge number of machine learning algorithms in-built.
- It also has the ability to scale up for large datasets by parallel processing on clusters and cloud.

- ***SAS***
  - SAS (earlier known as 'Statistical Analysis System') is another licensed commercial software which provides strong support for machine learning functionalities.
  - Developed in C by SAS Institute, SAS had its first release in the year 1976.
  - SAS is a software suite comprising different components.
  - The basic data management functionalities are embedded in the Base SAS component whereas the other components like SAS/INSIGHT, Enterprise Miner, SAS/STAT, etc. help in specialized functions related to data mining and statistical analysis.

- ***Other languages/tools***
  - There are a host of other languages and tools that also support machine learning functionalities.
  - Owned by IBM, SPSS (originally named as Statistical Package for the Social Sciences) is a popular package supporting specialized data mining and statistical analysis.
  - Originally popular for statistical analysis in social science (as the name reflects), SPSS now popular in other fields as well.
  - Released in 2012, Julia is an Open source, liberal license programming language for numerical analysis and computational science.
  - It has baked in all good things of MATLAB, Python, R and other programming languages used for machine learning for which it is gaining steady attention from machine learning development community.