
**********HOW DO MACHINE LEARN?**********

ML Process can be devided into 3 sequencial parts:

(1).DATA INPUT

=>Past data or information which is utilize as a basis for future decision making(prediction).

(2).ABSTRACTION

=>The Input data is processed using underling algoritham.

=>Knowledge is inputed in form of data which may not be used in original shape and form.

=>ABSTARCTION helps in deriving a "CONCEPTUAL MAP/MODEL" based on input data.

=>The model may be in any one of the following form:

(1).A Block of "IF...ELSE..." rule

(2).Mathemetical Equations

(3).A Data structure like TREE/GRAPH

(4).Logical Grouping of Similar Observations

=>A decision related to choice of model is based on multiple aspects.

=>For Example,

(1).A type of problem to be solved

=>EX.forcastion or prediction, trend analysis,understanding different sagments.

(2).Nature of data

=>type of data, values inside data etc.

(3).Domain of the problem

=>whether the problem belongs to critical domain?

EX. Frod detection in Banking domain

=>Once model is decided the next task is to build a Model based on input data,

=>This process is known as "TRAINING".

(3).GENERALIZATION

=>The Abstracted representation is generalized to form a framework for making decisions.

=>The model is trained based on finite set of data which may have limited set of charactristics.

=>To apply the model,to take decision on "set of UNKNOWN DATA(Testing Data)" .

=>TWO types of problem may occure:

(1).The trained model is aligned with training data too much Hence, may not give actual trend.

(2).Test data have certain charactristics which may be unknown to the training data.

=>Due to above two problems the decision making won't work, In That case a "HEURISTIC APPROACH" can be used.[HEURISTIC=NEAREST NEIGHBOUR]

Figure: {{InputData}}--->{{Abstraction}}--->{{Generalization}}

**********WELL_PAUSED LEARNING PROBLEM**********

=>Whether a problem can be solved using machine learning or not can be answered using following frameworks:
=>There are 3 Questions to be answered:
   (1).WHAT IS THE PROBLEM?
      =>For Example, A Problem is a programme that will prompt the next word as and when user types a word.
         -Task(T)::Prompt Next Word.
         -Experience(E)::Corpus/A Data set of commonly used words or phrases.
         -Performance(P)::The number of correct words cosidered by user(Learning Accurecy).

   (2).WHY DOES THE PROBLEM NEED TO BE SOLVED?
      =>For example,Does the problem solve any long standing issue or problem?,
                  Suggesting some movies for upcoming weekend.
      =>It is important to clearly understand the benefit of solving the problem.

   (3).HOW TO SOLVE THE PROBLEM?
      =>Explore how this problem can be manually solved?
      =>Detailed out step by step process to solve the problem.

**********TYPES OF MACHINE LEARNING**********

=>ML can be classified into 3 categories:

   ==========
   (1).SUPERVISED LEARNING:~
   ==========
   =>Also known as "PREDICTIVE LEARNING" where a ____MACHINE____ predicts the class of unknown objects based on previous
     class related information of similar objects.
   =>Major objective is to learn from previous information.
   =>Ex.-Saperating images by either shape or color.
       -How machine is able to differentiate between different shapes?
       -Some information needs to be provided so that machine will understand different shapes(TRAINING).
   =>Providing information to machine is known as "TRAINING DATA" which is a past information on specific task.
   =>Along with features a tag is also provided,tag is called "LABEL".
   =>In case of supervised learning, "TRAINING DATA ARE LABELED".
   =>Figure:
   {{LABELED TRAINING DATA}}--->{{Supervised Learning}}-->{{Prediction Model}}--->{{TestData}}--->{{Prediction Output}}
   =>Examples:
      (1).Predicting the result of a game.
      (2).Predicting whether tumour is malignent of benign.
      (3).Prediting the price of domain like realastet,stock-market Etc.
      (4).Classified text such as whther email is sparm or not.
   NOTE::If training data are of poor quality than prediction will not be precise.

(I).CLASSIFICATION

=>Labeled training data are given to the classifier.
=>As training data has labeled or category defined the task of a classifier is to map testing dataset and assign some labeled or category.
=>Figure:
{{LABELED TRAINING DATA(class/category)}}--->{{Classifier}}-->{{Prediction Model}}--->{{TestData}}--->{{Prediction Output}}
=>Examples:
(1).In a banking domain,to identifie fraudlent transactions,millions of transactions needs to be scrutinized.It is not possible for any human being to carry out this task.ML is effectivelty applied to do this task.Based on the past transaction data labeled as produlent ,new transactions are to be labeled as "Normal" or "Suspisious".
NOTE:-Classification is a type of supervised learning where a target feature is of type "Categorical".The target categorical feature is known as "CLASS".
-Typicall classification problems includes:
(1).Image Classification
(2).Deises Prediction
(3).Win/Lose Prediction in Game
(4).Handwritting Recognization

(II).REGRESSION

=>In Linear Regression,the goal is to predict ___Numerical___ features like Realatstate or stock price,temprature, marks in examination,sales revenue Etc.
=>The target veriable are "Continuous" in a nature.
=>In Linear regression, a streight line relationship is fitted between __Targeted__ veriables and __Predictor__ veriables.[Target=Dependent || Predictor=Independent].
=>The target is to __Minimize__ the error between __Actual__ value and predictor value.
=>In Simple Linear regression,there is only 1 predictor veriable where as In Multiple Linear Regression, Multiple predictor veriable can be included.
=>Example:
(1).-In a yearly budgeting exersice of the sales managers, they have to give sales prediction for the next year based on sales figure of previous year.
-A Simple linear regression model can be applied with "Investment as PREDICTOR veriable" and "Sales revenue as TARGET veriable".
=>A typical linear regression model can be presented in form of:
"Y=A+Bx"; X=Predictor,Y=Target
=>Examples:
(1).Demand Forcasting in retail
(2).Sales Prediction for Managers
(3).Price Prediction in Realastate
(4).Whether Forcasting
(5).Skill Demand Forcasting in Job Market
=>Figure:

{{LABELED TRAINING DATA(Continous/Numerical)}}--->{{Regression Algo.}}-->{{Prediction Model}}--->{{TestData}}---> {{Prediction Output}}

==========
(2).UNSUPERVISED LEARNING:~
==========
=>It is also known as "DESCRIPTIVE LEARNING" where machine finds a pattern in unknown object by grouping similar objects together.
=>In this,there is NO LABELED training data to learn from.
=>The objective is to take dataset as input and find 'NATURAL GROUPINGS' or 'PATTERNS' from dataset.
=>Therefor it is often known as "DISCRIPTIVE MODEL" and the process is also reffered as "PATTERN DISCOVERY" or "KNOWLEDGE DISCOVERY".
=>A critical application of unsupervised learning is CUSTOMER SAGMENTATION.

(I).CLUSTERING

=>Clustering is a main type of unsupervised learning which "Groups Simillar Objects Together ".
=>Object belonging to different cluster are quite ___Desimilar___.
=>Clustering discovers 'Patterns' from ___Unlabeled___ data and forms a cluster.
=>Different measures of similarilty can be applied for clustering.
=>ONE MOST COMMONLY ADDPOTED SIMILARITY MEASURE IS "DISTANCE".
=>2 data items are considered as a part of same cluster, if distance between them is __Less__.
=>If the distance between data items is high, it means it doesn't belongs to the same cluster.
=>This Process is known as "DISTANCE BASED CLUSTERING".

(II).ASSOCIATION ANALYSIS

=>In this, association between data elements is identified.
=>For Example,In Market Basket Analysis, from past transaction data in grocery store the observation of customer who have purchased item A have also purchased item B.
=>In majority transaction, this observation is true which means there is association of event "purchasing item A" with " purchasing item B".
=>Indentifying this association is the goal of Association Analysis.
=>Examples:
(1).Market Basket Analysis
(2).Recommandation System

==========
(3).REINFORCEMENT LEARNING:~
==========
=>A machine LEARN TO ACT its own to achieve given goals.
=>Machine often learns to do task autonomously.
=>For example,If the action is to walk than the programme is an agent and in an environment agent has to walk where there

are certain herdles/obstacles.
=>It improves its performence by doing the task if it does successfully, A REWARD IS GIVEN.
=>The machine countinuous this this process for the entire task which is known as "REINFORCEMENT LEARNING".
=>Figure:

=>For example,SELF DRIVING CARS where critical information needs to be taken care like speed and speed limit in different road sagements, traffic conditions, road conditions, wether condition Etc.
=>Some other task needs to be taken care are start or stop, acccerate-disslerate, turning left-right Etc.

**********SUPERVISED v/s UNSUPERVISED v/s REINFORCEMENT**********

1.	Labeled Training Data	Unknown-unlabeled Data	No-predifioned Data
2.	Predictive Learning	Discriptive Learning	Reward Based Leaning
3.	Future Prediction	Patterns Finding	Learn Series of Actions.
4.	Algoritham: 1.Linear regression 2.Logistic 3.SVM 4.naive bayes 5.dicision tree	Algoritham: 1.K-Means 2.C-means	Algoritham: 1.Q-Learning 2.Sarsa
5.	model bulding: Model based on Labeled training through reward and punishment.	model builed on unlabeled training data	Model learns and Update itself through reward and punishment.
6.	2 types of problems: regression & classification	2 types of problem: clustring & association	reward based problem
7.	Simplest one to understand	More difficult to undestand & implement	Most Complex
8.	Applications: handwritting recognition, stock market prediction, dieses prediction fraud-detection....	Applications: Market-basket analysis, recommander system, sagementations...	Applications: self driving cars, inteligent robots games...

performance-Model:
->can be evaluated on basis of how many misclassifications are done(Predicted value-actual value)
->it is difficult to measure whether model did something useful? but if similar records are grouped together tha it is the only measurement
->model is evaluated on basis of the rewards

algoithams:
->NaN
->SOM(Self organisng map),PCA(Principal Component Analysis),A-priori Algo
->NaN

**********APPLICATIONS OF MACHINE LEARNING**********

1.Banking and finance
2.Healthcare
3.Insurence

**********MACHINE LEARNING TOOLS and LANGUAGE**********

1.Python(Numpy,scyp,sikit-learn)
2.R
   -It is Language for statstical computing and data analysis.
   -Open-source
   -Libraries are PLYR, DPLYR(Data Transformation),CARET(Classification and Regression training),TM(TextMining),ggplot2(data visulization),shiny etc....
3.Matlab(Matrix laboratary)
   -It is a licenced commercial soft.
   -Robust support for wide range of numerical computing and statstical functions
   -Provides ability to scale up for large datasets by parralel processing on clusters and cloud.
4.SAS(Stastical Analysis System)
   -Another licensed commercial soft.
   -provides basic data managment functionality and data mining and statstical analysis techniques

---

**********ISSUES IN MACHINE LEARNING**********

1.The biggest issue is related to data privacy and brich of it.