Reasoning and Machine Learning

Unit 3

Syllabus Content

- **Reasoning:** Probability and Bayes Theorem, Certainty factors, Probabilistic Graphical Models, Bayesian Networks, Markov Networks, Fuzzy Logic.
- **Introduction to Machine Learning:** Idea of Machines learning from data, Classification of problem – Regression and Classification, Types of machine learning, Supervised and Unsupervised learning.

Uncertainty

- In real life, it is not always possible to determine the state of the environment as it might not be clear. Due to partially observable or non-deterministic environments, agents may need to handle uncertainty
- When an agent knows enough fats about its environment. The logical plans and actions produces a guaranteed Work
- Unfortunately, agent never have access to the whole environment. Agents acts uncertainty.

CAR

Cause of Uncertainty

Uncertainty due to

- Uncertain data: Data that is missing, unreliable, inconsistent or noisy
- Uncertain knowledge: When the available knowledge has multiple causes leading to multiple effects or incomplete knowledge of causality in the domain
- Uncertain knowledge representation: The representations which provides a restricted model of the real system, or has limited expressiveness
- Inference: In case of incomplete or default reasoning methods, conclusions drawn might not be completely accurate.

Uncertainty

- Knowledge representation in propositional logic and predicate logic is based on certainty, means we are sure about predicates.
- Example : A-> B means if A is true then B is true
- But what about the situation where we are not sure about whether A is true or not then we can not express this statement, this situation comes under uncertainty
- So to characterize uncertainty knowledge, where we are not sure about the predicates, we need uncertain reasoning or probabilistic reasoning .

Nature of Uncertain Knowledge

- The Diagnosis : medicine, automobile repair, or whatever-is a task that almost always involves uncertainty
- Let us try to write rules for dental diagnosis using first-order logic, so that we can see how the logical approach breaks down. Consider the following rule:
  - $\forall p\ Symptom(p, Toothache) \Rightarrow Disease(p, Cavity)$.
- The problem is that this rule is wrong.
- Not all patients with toothaches have cavities; some of them have gum disease, swelling, or one of several other problems
  - $\forall p\ Symptom(p, Toothache) \Rightarrow Disease(p, Cavity) \lor Disease(p, GumDisease) \lor Disease(p, Swelling) \ldots$

Nature of Uncertain Knowledge

- to make the rule true, we have to add almost unlimited list of possible causes.
- We could try a causal rule:
  - $\forall p\ Disease(p, Cavity) \Rightarrow Symptom(p, Toothache)$.
- But this rule is also not right either; not all cavities cause pain
- Toothache and a Cavity are unconnected, so the judgement may go wrong.

Nature of Uncertain Knowledge

- This is a type of the medical domain, as well as most other judgmental domains: law, business, design, automobile repair, gardening, dating, and so on.
- The agent take action, only a **degree of belief** in the relevant sentences.
- Our main tool for dealing with degrees of belief will be **probability theory**
- **The Probability** assigns to each sentence a numerical degree of belief between 0 and 1.

Need of probabilistic reasoning

- Unpredictable outcomes
- Predicates are too large to handle
- Unknown error occurs

In probabilistic reasoning, there are two methods to solve difficulties with uncertain knowledge:

✓ Bayes' rule

✓ Bayesian Statistics

Probability

- Probability can be defined as chance of occurrence of an uncertain event.
- It is the numerical measure of the likelihood that an event will occur.
- The value of probability always remains between 0 and 1.

$0 \leq P(X) \leq 1$, where $P(X)$ is the probability of an event X.

✓ $P(X) = 0$, indicates total uncertainty in an event X.

✓ $P(X) = 1$, indicates total certainty in an event X.

Sample Space

- In probability theory, the sample space is also called as a sample description space or possibility space.
- The sample space (Ω) is the set of possible outcomes of an experiment or random trial. Points ω in Ω are called sample outcomes, realizations, or elements. Subsets of Ω are called Events.

- **Example**. If we toss a coin twice then
  - Sample space (Ω) = {HH,HT, TH, TT}.
  - **The event that the first toss is heads is A** = {HH,HT}

- **Complement** - The event that A does not occur, denoted as **A´** , is called the complement of event A. **P(A´) = 1- P(A)**

- **Intersection** - The intersection of two events A and B, denoted by **A ∩ B**, is the event containing all elements that are common to A and B.

Sample Space

- **Union -** The union of the two events A and B, denoted by **A** ∪ **B,** is the event containing all the elements that belong to A or B or both.

- **Mutually Exclusive -** Events that have no outcomes in common are said to be ***disjoint or mutually exclusive.***

➢ Clearly, A and B are mutually exclusive or disjoint if and only if $A \cap B$ is a null set

➢ We say that events A1 and A2 are disjoint (mutually exclusive) if $A_i \cap A_j = \{\}$

  - Example: first flip being heads and first flip being tails

Sample Space

•**Collectively exhaustive -** A set of events is jointly or collectively exhaustive **if at least one of the events must occur**.

•The outcomes must be **collectively exhaustive**, i.e. on every experiment (or random trial) there will always take place some outcome.

- E1,E2 ,E3, …En are called **exhaustive events** if at least one of them necessarily occurs whenever the experiment is performed.

Sample Space

- **Sample space S = {1, 2, 3, 4, 5, 6}**
- Assume that A, B and C are the events associated with this experiment. Also, let us define these events as:

❑ A be the event of getting a number greater than 3

❑ B be the event of getting a number greater than 2 but less than 5

❑ C be the event of getting a number less than 3

- We can write these events as:
- A = {4, 5, 6}
- **B = {3, 4}**
- **and C = {1, 2}**
- We observe that
- $A \cup B \cup C = \{4, 5, 6\} \cup \{3, 4\} \cup \{1, 2\} = \{1, 2, 3, 4, 5, 6\} = S$
- **Therefore, A, B, and C are called exhaustive events.**

Probability

- A **probability** is a number that reflects the chance or likelihood that a particular event will occur.
- **Probabilities** can be expressed as proportions that range from 0 to 1, and they can also be expressed as percentages ranging from 0% to 100%.
- A probability of 0 indicates that there is no chance that a particular event will occur, whereas a probability of 1 indicates that an event is certain to occur.
- A probability of 0.45 (45%) indicates that there are 45 chances out of 100 of the event occurring.

Axioms of Probability

- We will assign a real number P(A) to every event A, called the probability of A.
- To qualify as a probability, P must satisfy three axioms:
  - Axiom 1: $P(A) \geq 0$ for every A
  - Axiom 2: $P(\Omega) = 1$
  - Axiom 3: If A1,A2, . . . are disjoint then

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$$

Marginal Probabilities

- Marginal probabilities are those of individual events.
- The probability of an event occurring P(A)
- **Example:**

Event A - The probability that a card drawn is a 4

$$P(A)=1/13$$

Joint Probability

- The Joint probability is a statistical measure that is used to calculate the probability of two events occurring together at the same time
- The probability of event A and event B occurring
- It is the probability of the intersection of two or more events.

  $P(A, B)$, $P(A \cap B)$, or $P(A \text{ and } B)$

- **Example:**
- The probability that a card is a four and red

  =P(four and red)

  = 2/52=1/26.

Conditional Probability

If A and B are two dependent events, then the probability of occurrence of A given that B has already occurred and is denoted by P(A/B) is given by

$$P(A|B) = P(A \wedge B) / P(B)$$

$$P(B|A) = P(A \wedge B) / P(A)$$

Bayes Theorem

- **Bayes' rule, Bayes' law**, or **Bayesian reasoning**, which determines the probability of an event with uncertain knowledge.
- it relates the **conditional probability** and **marginal probabilities** of two random events.
- Bayes' theorem named after the British mathematician **Thomas Bayes**.
- It is a way to calculate the value of **P(B|A)** with the knowledge of **P(A|B).**
- Allows **updating the probability prediction of an event** by observing new information of the real world.

Bayes Theorem

- Bayes' theorem can be derived using **product rule** and **conditional probability** of event A with known event B:

$P(A \wedge B) = P(A|B)\, P(B)$

or

$P(B \wedge A) = P(B|A)\, P(A)$

$$P(A|B) = \frac{P(B|A)\, P(A)}{P(B)} \quad \text{....(a)}$$

The above equation (a) is called as **Bayes' rule** or **Bayes' theorem**.

Bayes Theorem

- This equation is basic of most modern AI systems for **probabilistic inference**.
- It shows the simple relationship between **joint and conditional probabilities.**
- P(A|B) is known as **posterior,** Probability of hypothesis A when we have occurred an evidence B.
- P(B|A) is called the **likelihood**, in which we consider that hypothesis is true, then we calculate the probability of evidence.
- P(A) is called the **prior probability,** probability of hypothesis before considering the evidence
- P(B) is called **marginal probability,** pure probability of an evidence.

Bayes Theorem Applications

- It is used to calculate the next step of the robot when the already executed step is given.
- Bayes' theorem is helpful in weather forecasting.

Bayes Theorem

- A doctor is aware that disease meningitis causes a patient to have a stiff neck, and it occurs 80% of the time. He is also aware of some more facts, which are given as follows:
- ➢The Known probability that a patient has meningitis disease is 1/30,000.
- ➢The Known probability that a patient has a stiff neck is 2%.

Find the probability that a patient with the stiff neck has meningitis.

Bayes Theorem

Let a be the proposition that patient has stiff neck and b be the proposition that patient has meningitis.

- $P(a|b) = 0.8$
- $P(b) = 1/30000$
- $P(a) = .02$

$$P(b|a) = \frac{P(a|b)P(b)}{P(a)} = \frac{0.8*(\frac{1}{30000})}{0.02} = 0.001333333.$$

Bayesian Belief Network

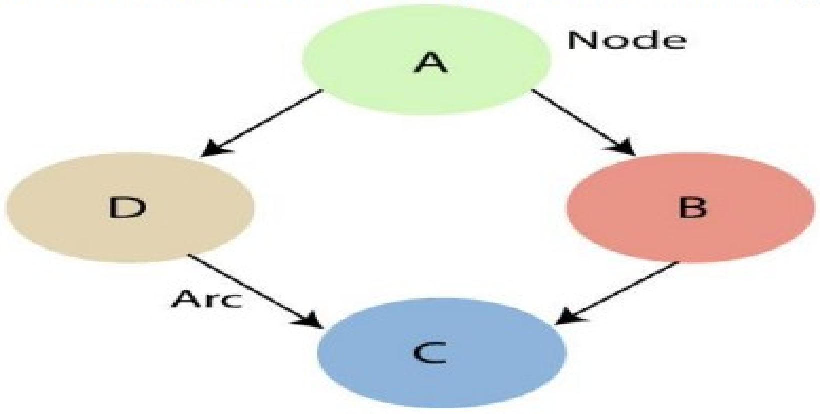
Bayesian Network can be used for building models from data and experts opinions, and it consists of two parts:

- **Directed Acyclic Graph**
- **Table of conditional probabilities**

The generalized form of Bayesian network that represents and solve decision problems under uncertain knowledge is known as an **Influence diagram**.

Note: It is used to represent conditional dependencies.

A Bayesian network graph is made up of nodes and Arcs (directed links), where:



- Each **node** corresponds to the random variables, and a variable can be **continuous** or **discrete**.
- **Arc or directed arrows** represent the causal relationship or conditional probabilities between random variables. These directed links or arrows connect the pair of nodes in the graph. These links represent that one node directly influence the other node, and if there is no directed link that means that nodes are independent with each other
- **Note: The Bayesian network graph does not contain any cyclic graph. Hence, it is known as a directed acyclic graph or DAG.**
- The Bayesian network has mainly two components:
    1. **Causal Component**
    2. **Actual numbers**
- Each node in the Bayesian network has condition probability distribution $\mathbf{P(X_i \mid Parent(X_i))}$, which determines the effect of the parent on that node.
- Bayesian network is based on Joint probability distribution and conditional probability.

- **Example:** Harry installed a new burglar alarm at his home to detect burglary. The alarm reliably responds at detecting a burglary but also responds for minor earthquakes. Harry has two neighbors David and Sophia, who have taken a responsibility to inform Harry at work when they hear the alarm. David always calls Harry when he hears the alarm, but sometimes he got confused with the phone ringing and calls at that time too. On the other hand, Sophia likes to listen to high music, so sometimes she misses to hear the alarm. Here we would like to compute the probability of Burglary Alarm.

- **Problem: Calculate the probability that alarm has sounded, but there is neither a burglary, nor an earthquake occurred, and David and Sophia both called the Harry.**

Note: List of all events occurring in this network:

Burglary (B)

Earthquake(E)

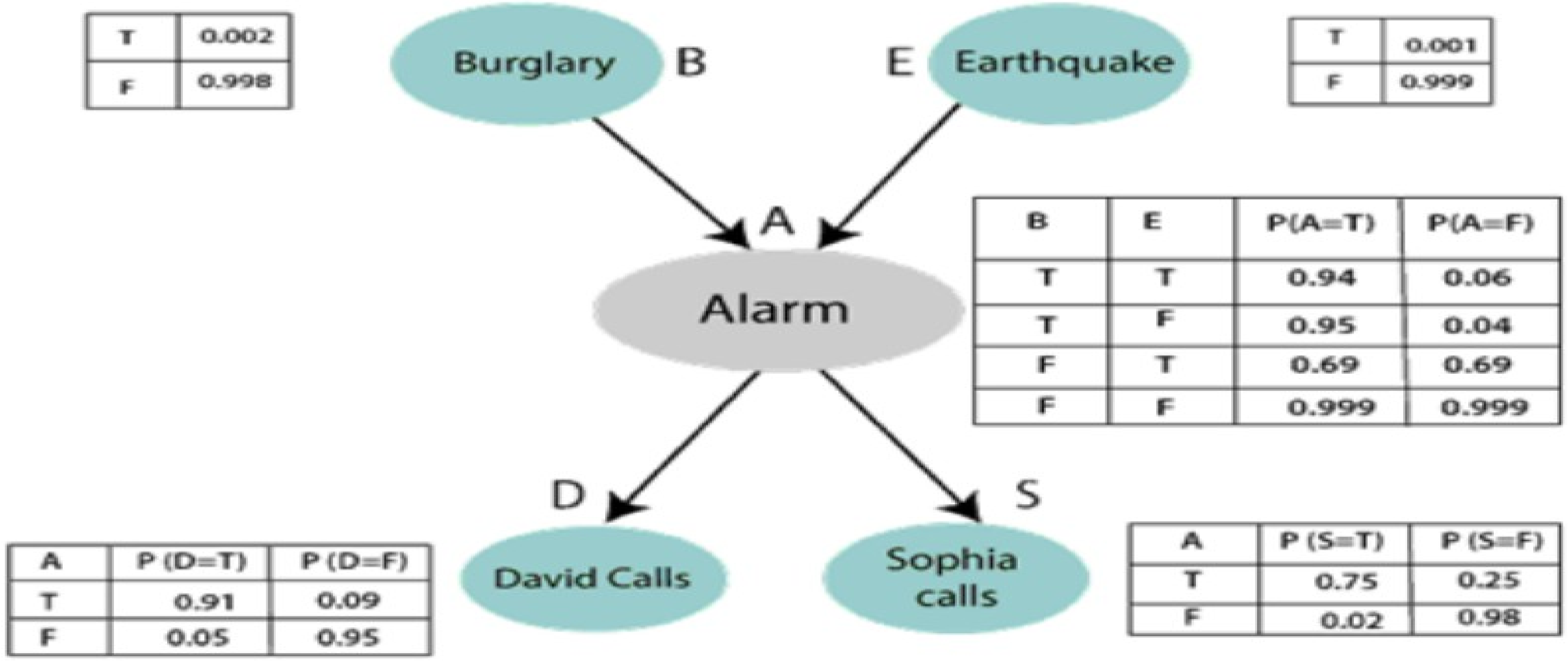
Alarm(A)

David Calls(D)

Sophia calls(S)

- From the formula of joint distribution, we can write the problem statement in the form of probability distribution:

$$P(S, D, A, \neg B, \neg E) = P(S|A) * P(D|A) * P(A|\neg B \wedge \neg E) * P(\neg B) * P(\neg E)$$



T	0.002
F	0.998

T	0.001
F	0.999

B	E	P(A=T)	P(A=F)
T	T	0.94	0.06
T	F	0.95	0.04
F	T	0.69	0.69
F	F	0.999	0.999

A	P (D=T)	P (D=F)
T	0.91	0.09
F	0.05	0.95

A	P (S=T)	P (S=F)
T	0.75	0.25
F	0.02	0.98

P(S, D, A, ¬B, ¬E) = P (S|A) *P (D|A)*P (A|¬B ^ ¬E) *P (¬B) *P (¬E).

= 0.75* 0.91* 0.001* 0.998*0.999

= 0.00068045.

Hence, a Bayesian network can answer any query about the domain by using Joint distribution.

Some Applications of BN

- Medical diagnosis
- Troubleshooting of hardware/software systems
- Fraud/uncollectible debt detection
- Data mining
- Analysis of genetic sequences
- Data interpretation, computer vision, image understanding

Markov Model

- Incorporate the principles of the Markov property, as defined by Russian mathematician Andrey Markov in 1906
- In short, the prediction of an outcome is based solely on the information provided by the current state, not on the sequence of events that occurred before.
- The four main forms of Markov models are the Markov chain, Markov decision process, hidden Markov model, and the partially observable Markov decision process.
- The specific uses of each of these models are dependent on two factors; whether or not the system state is fully observable, and if the system is controlled or fully autonomous.

Markov Model

	State is fully observable	State is partially observable
System is autonomous	Markov Chain	Hidden Markov Model
System is controlled	Markov decision process	Partially observable Markov decision process

Hidden Markov Model

- Statistical model which is also used in machine learning.
- In this any event will only depend on its previous event.
- The Hidden Markov model is a probabilistic model which is used to explain or derive the probabilistic characteristic of any random process
- Terminologies in HMM

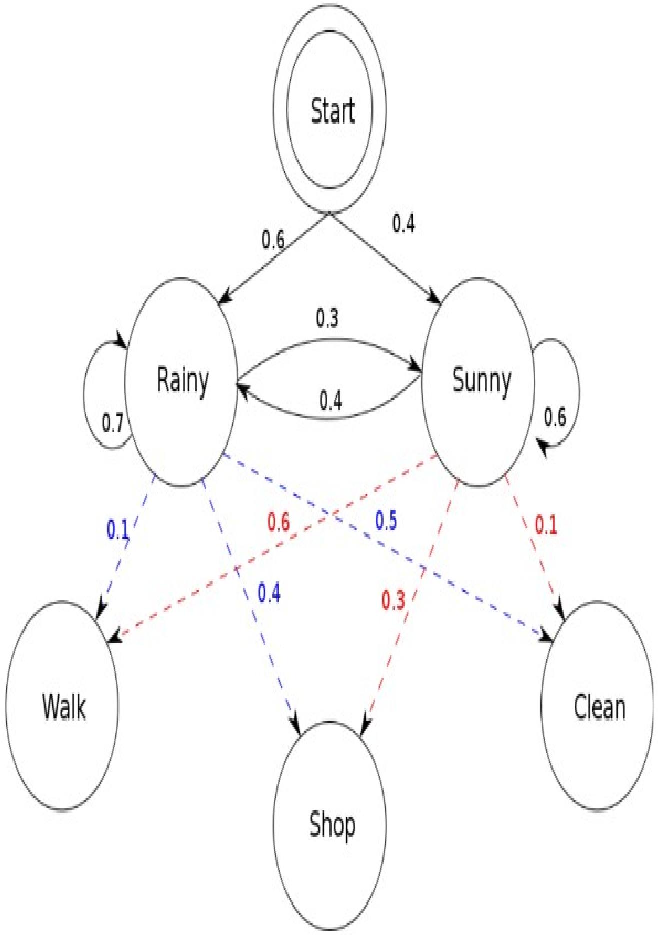
1)Hidden

2)Observation

3)Transition Probability

4)Emissions

- An observation is termed as the data which is known and can be observed. The below diagram depicts the interaction between two ‘HIDDEN’ states, ‘Rainy’ and ‘Sunny’ in this case. ‘Walk’, ‘Shop’, and ‘Clean’ in the below diagram are known as data, referred to as




```
states = ('Rainy', 'Sunny')

observations = ('walk', 'shop', 'clean')

start_probability = {'Rainy': 0.6, 'Sunny': 0.4}

transition_probability = {
   'Rainy' : {'Rainy': 0.7, 'Sunny': 0.3},
   'Sunny' : {'Rainy': 0.4, 'Sunny': 0.6},
   }

emission_probability = {
   'Rainy' : {'walk': 0.1, 'shop': 0.4, 'clean': 0.5},
   'Sunny' : {'walk': 0.6, 'shop': 0.3, 'clean': 0.1},
   }
```

Hidden Markov Model

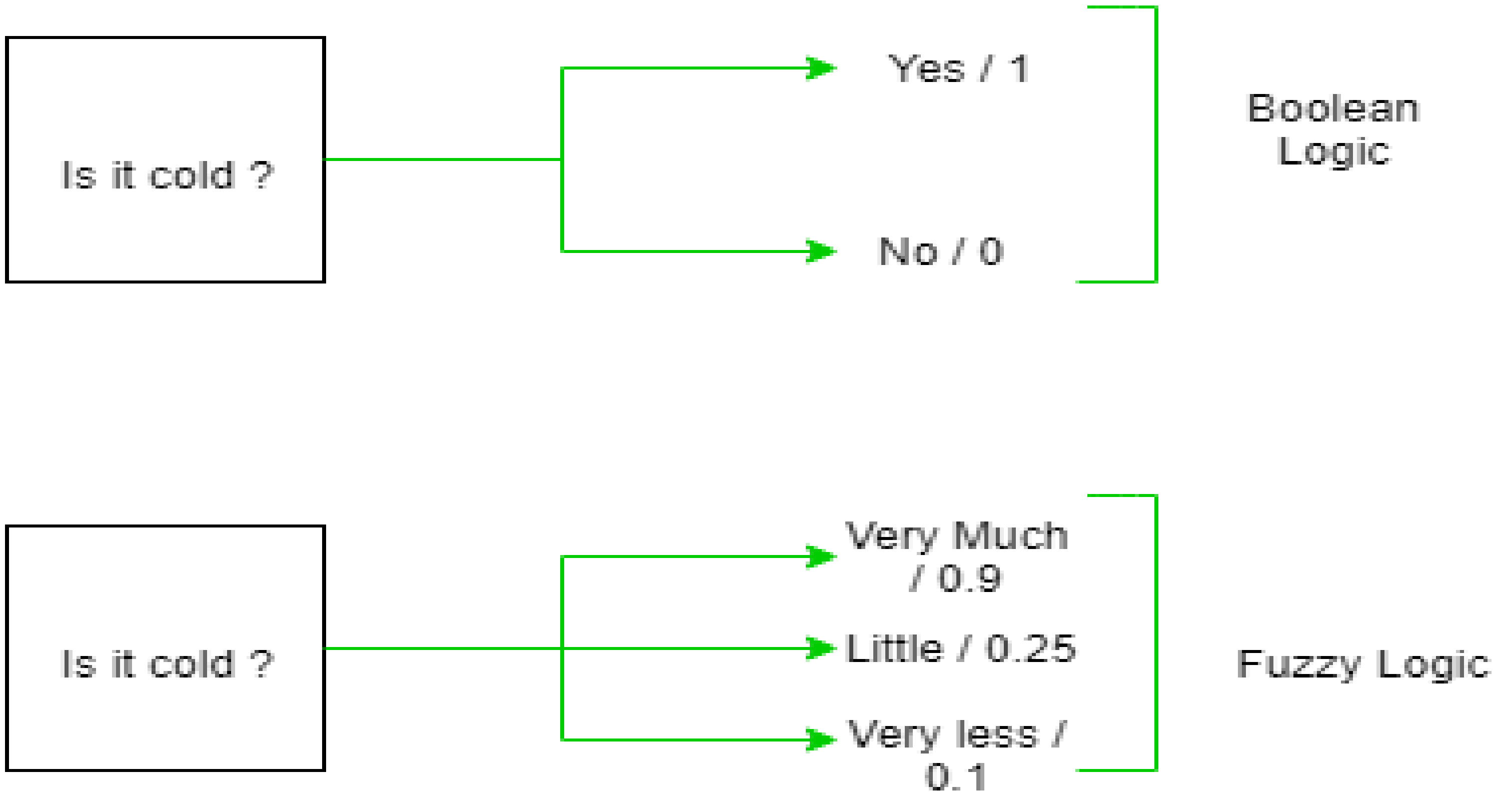
- **Hidden states (rainy and sunny)** - These states are hidden because what is observed as the process output is whether the person is shopping, walking, or cleaning
- **Observations** – Shop, walk, Clean
- **Start probability** - initial probability distribution
- **Emission probability** - probability of observing the output, shop, clean and walk given the states, rainy or sunny

Markov Models and Machine Learning

- Computational finance
- speed analysis
- Speech recognition
- Speech synthesis
- Part-of-speech tagging

Other Approaches to Uncertain reasoning- Fuzzy sets and Fuzzy logic.

What is Fuzzy Logic?



Fuzzy logic

- Fuzzy means uncertain, indefinite, vague, or unclear.
- It is a computing technique that is based on the degree of truth.
- A fuzzy logic system uses the input's degree of truth and linguistic variables to produce a certain output.
- The state of this input determines the nature of the output.
- This technique is different from *boolean logic*
- In *boolean logic*, two categories (0 and 1) are used to describe objects
- **Example – the temperature in water served in glass may be**

BL - *High* (1) or *Low* (0)

fuzzy logic- *very cold*, *very warm*, or *warm*.

Fuzzy logic

- **Example -** Suppose we have a question that we need to answer

BL – *Yes or No*

Fuzzy Logic - *possibly yes*, *possibly no*, or *certainly no*.

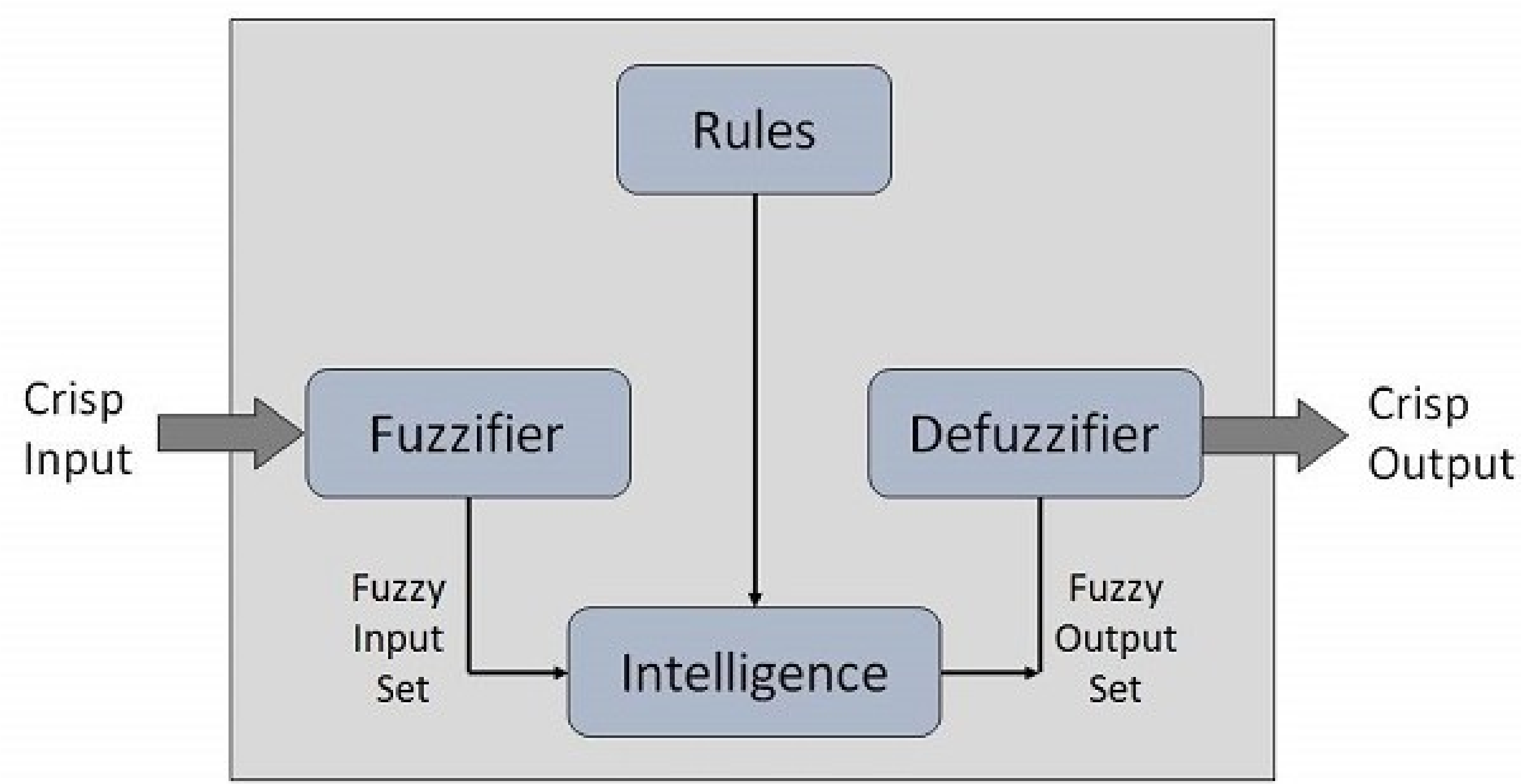
What is Fuzzy Logic?

- The inventor/father of fuzzy logic is Lotfi Zadeh
- Fuzzy Logic (FL) is a method of reasoning that resembles human reasoning.
- The approach of FL imitates the way of decision making in humans that involves all intermediate possibilities between digital values YES and NO.
- The conventional logic block that a computer can understand takes precise input and produces a definite output as TRUE or FALSE, which is equivalent to human's YES or NO.
- Fuzzy Logic in AI provides valuable flexibility for reasoning

Why Fuzzy Logic?

- It solves the problem of uncertainty in the engineering field.
- When accurate reasoning is not available, it provides an accurate level of reasoning.
- Fuzzy logic has a simple structure that is easy to understand.
- It is an effective way of controlling machines.
- It provides solutions to various industrial problems (especially decision making).
- It requires little data to be executed.

Fuzzy Logic Systems Architecture



Fuzzy Logic Systems Architecture

- Fuzzy Logic Systems (FLS) produce acceptable but definite output in response to incomplete, ambiguous, distorted, or inaccurate (fuzzy) input.
- FSL has four main parts as shown −
  1. **Fuzzification Module** − It transforms the system inputs, which are crisp numbers, into fuzzy sets. It splits the input signal into five steps such as −

LP	x is Large Positive
MP	x is Medium Positive
S	x is Small
MN	x is Medium Negative
LN	x is Large Negative

Fuzzy Logic Systems Architecture

2. **Knowledge Base** – It stores IF-THEN rules provided by experts.
3. **Inference Engine** – It simulates the human reasoning process by making fuzzy inference on the inputs and IF-THEN rules.
4. **Defuzzification Module** – It transforms the fuzzy set obtained by the inference engine into a crisp value.

The **membership functions work on** fuzzy sets of variables

Membership Function

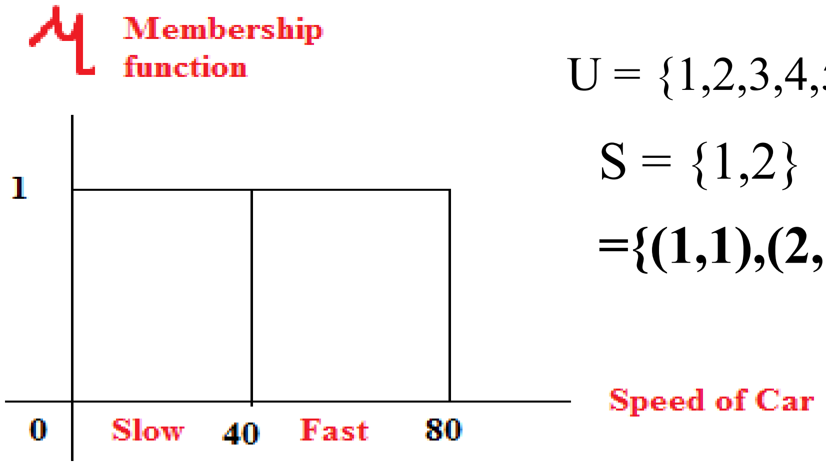
- It is a graph that defines how each point in the input space is mapped to membership value between 0 and 1.

- Allow you to quantify linguistic term and represent a fuzzy set graphically.

- A **membership function** for a fuzzy *set A* on the universe of discourse X is defined as $\mu_A : X \rightarrow [0,1]$.

- Here, each element of *X* is mapped to a value between 0 and 1. It is called **membership value** or **degree of membership**. It quantifies the degree of membership of the element in *X* to the fuzzy set *A*.
  - x axis represents the universe of discourse.
  - y axis represents the degrees of membership in the [0, 1] interval.

Membership Function



U = {1,2,3,4,5}

S = {1,2}

={(1,1),(2,1),(3,0),(4,0),(5,0) }

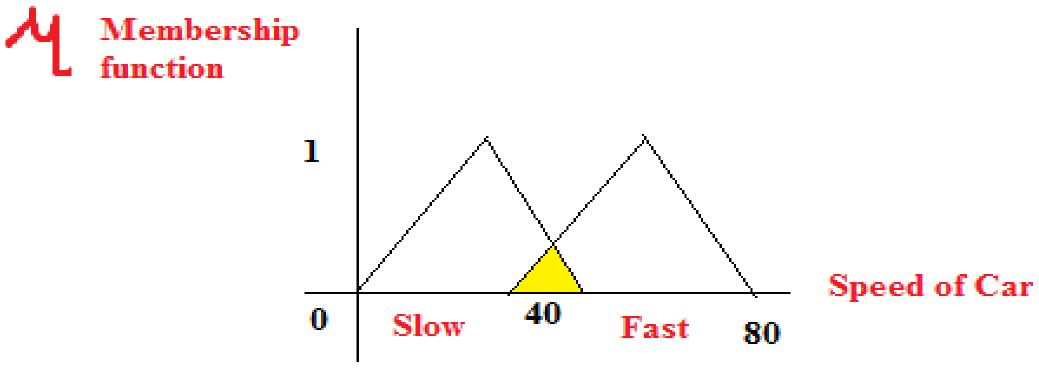
Degree of fastness =

0, if speed(x) <= 40

speed(x)-40/10, if 40 < speed (x) < 50

1, if speed(x) >= 50

Example –



x = 30 (30,0)

x = 60 (60,1)

x = 42 = 42 – 40/10 = 2/10 =0.2

x = 45 = 45-40/10 = 5/2 = 0. 5

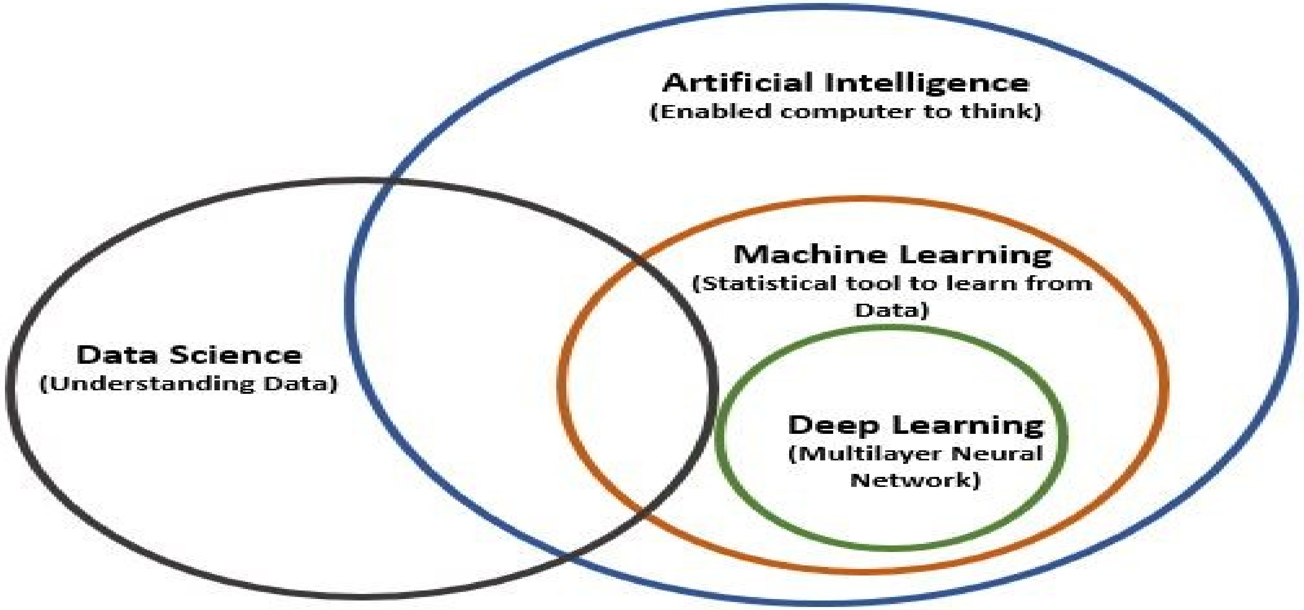
Application Areas of Fuzzy Logic

- The key application areas of fuzzy logic are as given −
- **Automotive Systems**
  - Automatic Gearboxes, Four-Wheel Steering, Vehicle environment control,etc
- **Consumer Electronic Goods**
  - Photocopiers, Still and Video Cameras, Television,etc
- **Domestic Goods**
  - Microwave Ovens, Refrigerators, Toasters, Vacuum Cleaners, Washing Machines, etc
- **Environment Control**
  - Air Conditioners/Dryers/Heaters, Humidifiers, etc

Introduction to Machine Learning

Machine Learning

- ML is nothing but making machines or computers learn
- Machine Learning is a branch of Artificial Intelligence (AI) and computer science
- Machine learning uses algorithms and data to learn, predict everything



Machine Learning

- **Arthur Samuel**, a pioneer in the field of artificial intelligence and computer gaming, coined the term “**Machine Learning**”.

✓ He defined machine learning as – a **“Field of study that gives computers the capability to learn without being explicitly programmed”**.

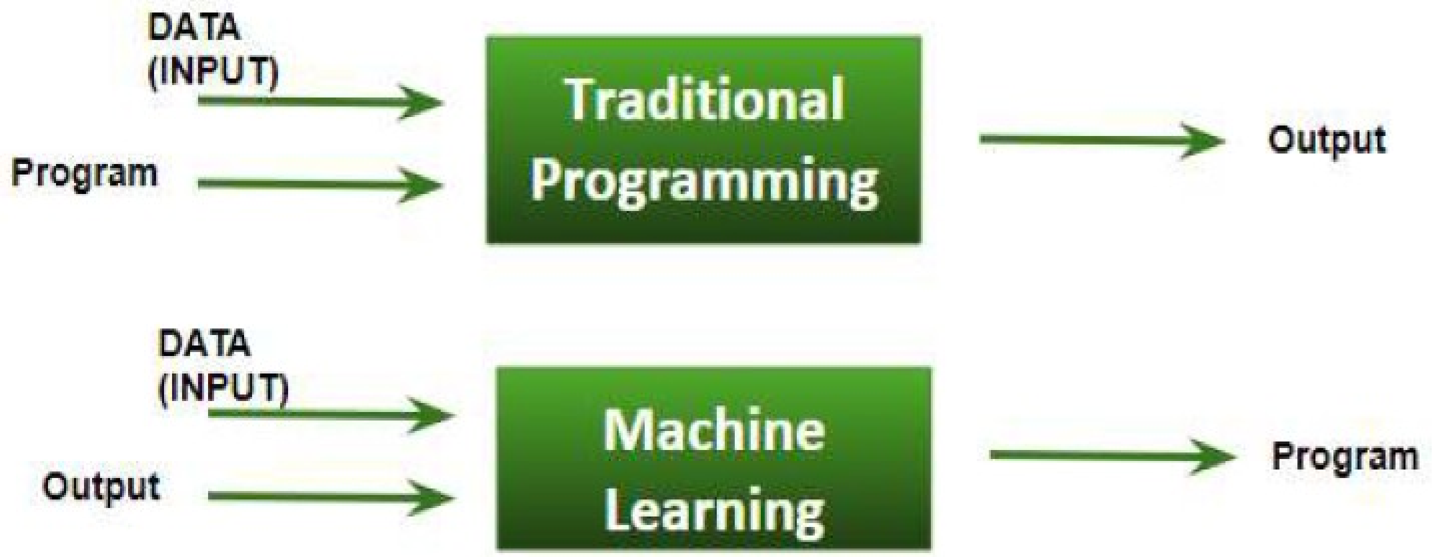
- In a very layman’s manner, Machine Learning(ML) can be explained as **automating and improving the learning process of computers** based on their experiences without being actually programmed i.e. without any human assistance.

✓ The process starts with feeding good quality data and then training our machines(computers) by building machine learning models using the data and different algorithms.

✓ The choice of algorithms depends on what type of data do we have and what kind of task we are trying to automate.

- **Example** - Training of students during exams

- **Basic Difference in ML and Traditional Programming?**
  - ✓ **Traditional Programming:** We feed in DATA (Input) + PROGRAM (logic), run it on the machine, and get the output.
  - ✓ **Machine Learning:** We feed in DATA(Input) + Output, run it on the machine during training and the machine creates its own program(logic), which can be evaluated while testing.



- **What does exactly learning mean for a computer?**
- 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**
- **Example -** Playing checkers

✓ **E** = the experience of playing many games of checkers

✓ **T** = the task of playing checkers.

✓ **P** = the probability that the program will win the next game

- **Example -** Classify Email as spam or not spam
- **Example** - Recognizing hand written digits/ characters

Example 1

Classify Email as spam or not spam

- **Task (T):** Classify email as spam or not spam
- **Experience(E):** watching the user to mark/label the email as spam or not spam
- **Performance (P):** The number or fraction of email to be correctly classified as spam or not spam

Example 2

Recognizing handwritten digits/ characters

- **Task(T):** Recognizing handwritten digit
- **Experience (E):** watching the user to mark/ label the handwritten digit to 10 classes(0-9) & identify underling pattern
- **Performance(P):** The number of fractions of hand-written digits correctly classified

Applications of Machine Learning

- Face detection
- Stock prediction
- Spam Email Detection
- Machine Translation
- Self-parking Cars
- Airplane Navigation Systems
- Medicine
- Data Mining

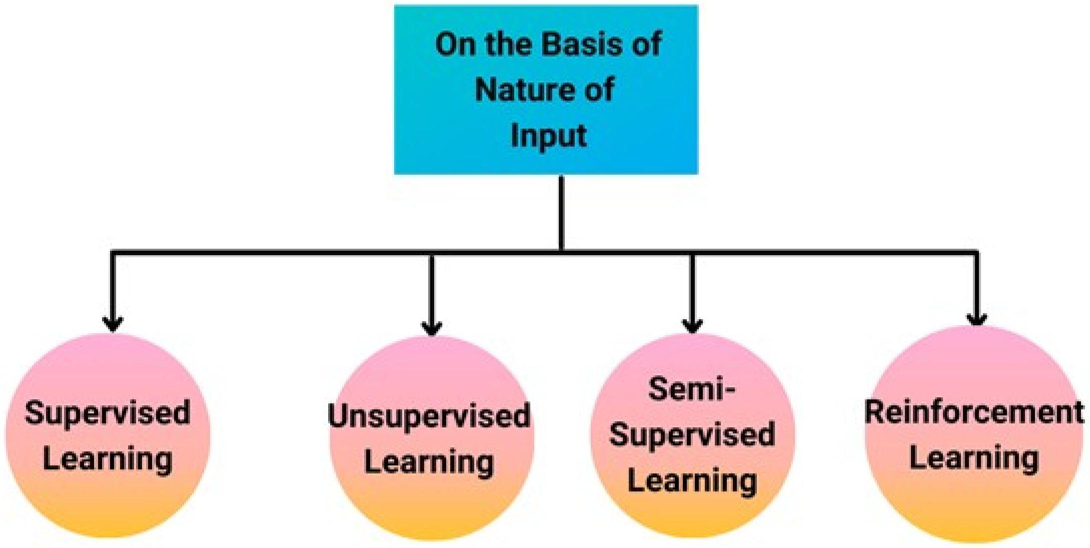
Speech recognition
Hand-written digit recognition
Computational Biology
Recommender Systems
Guiding robots
Space Exploration
Supermarket Chain

How does ML work?

- Gathering past data in any form suitable for processing. The better the quality of data, the more suitable it will be for modeling
- Data Processing – Sometimes, the data collected is in raw form and it needs to be pre-processed.
- Divide the input data into training, cross-validation, and test sets. The ratio between the respective sets must be 6:2:2
- Building models with suitable algorithms and techniques on the training set.
- Testing our conceptualized model with data that was not fed to the model at the time of training and evaluating its performance using metrics such as F1 score, precision, and recall.

Classification of Machine Learning

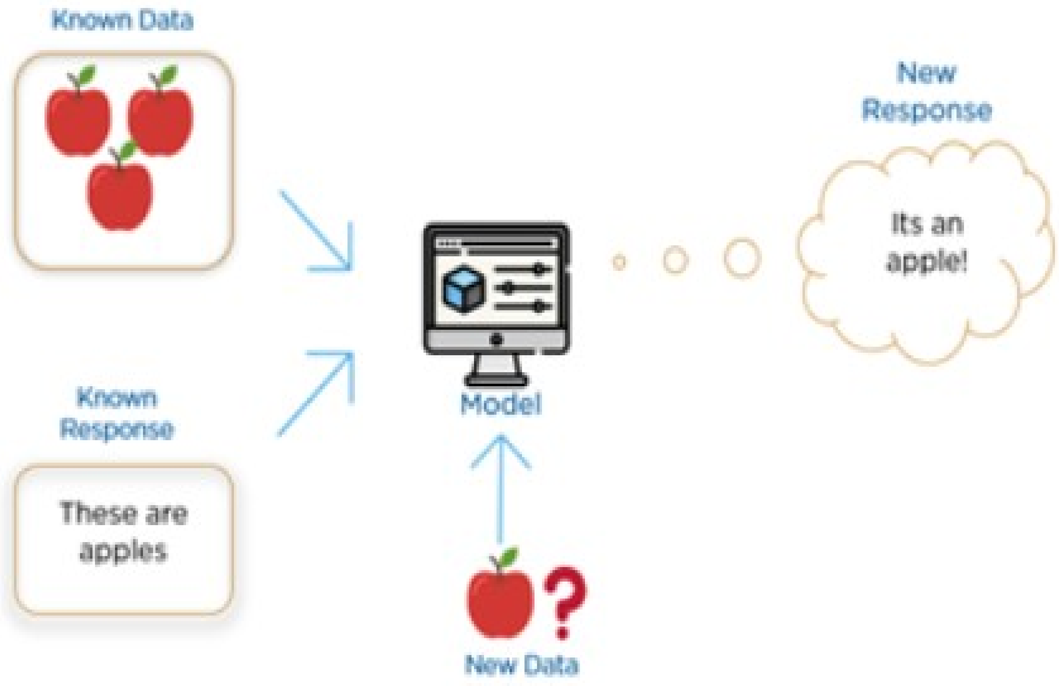
Machine learning implementations are classified into four major categories, depending on the nature of the learning "signal" or "response" available to a learning system



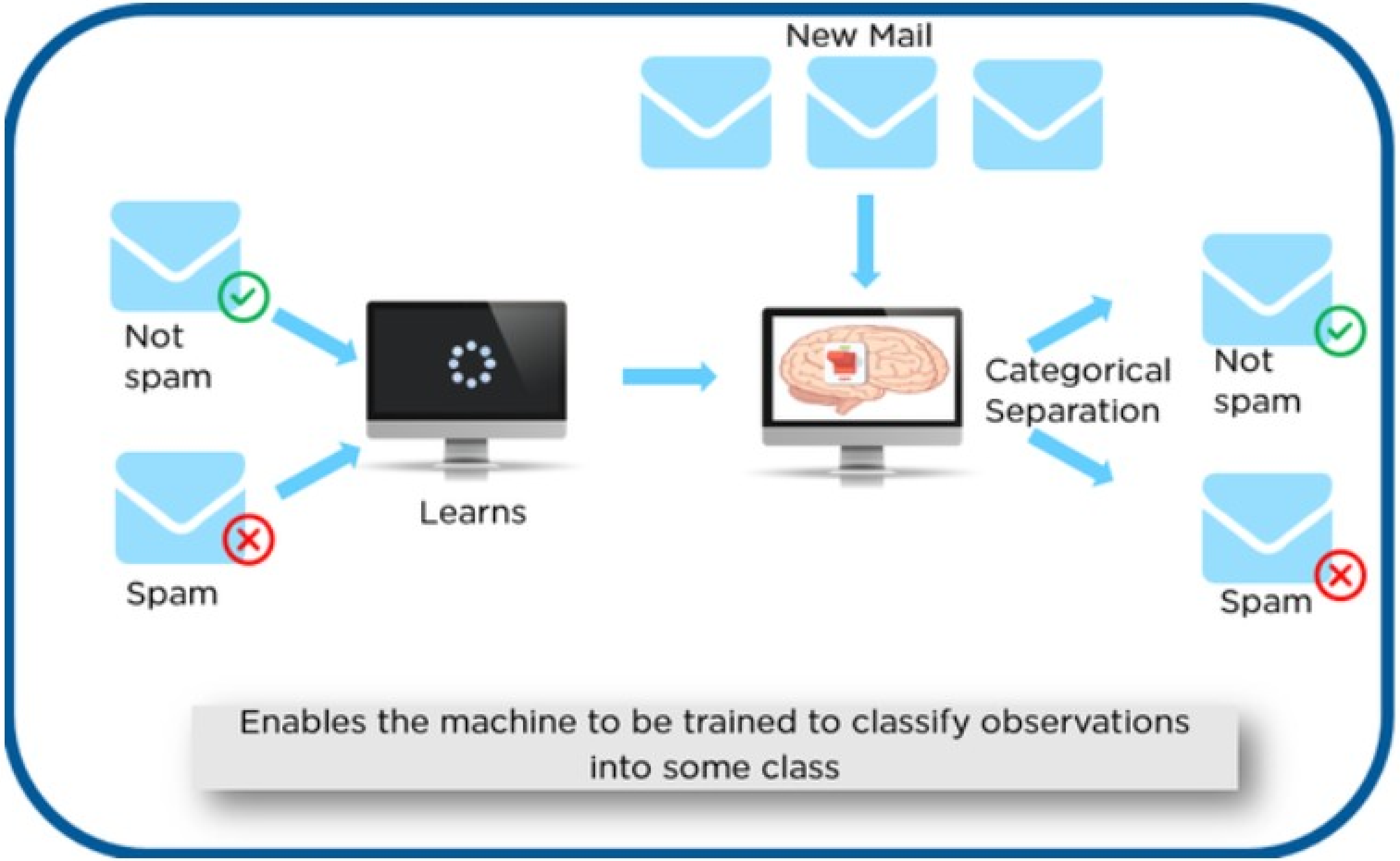
Supervised Learning

- It is a type of machine learning in which the algorithm is trained on the labeled dataset.
- It learns to map input features to targets based on labeled training data.
- The algorithm is provided with input features and corresponding output labels, and it learns to generalize from this data to make predictions on new, unseen data.
- Two main types - Regression and Classification

Supervised Learning



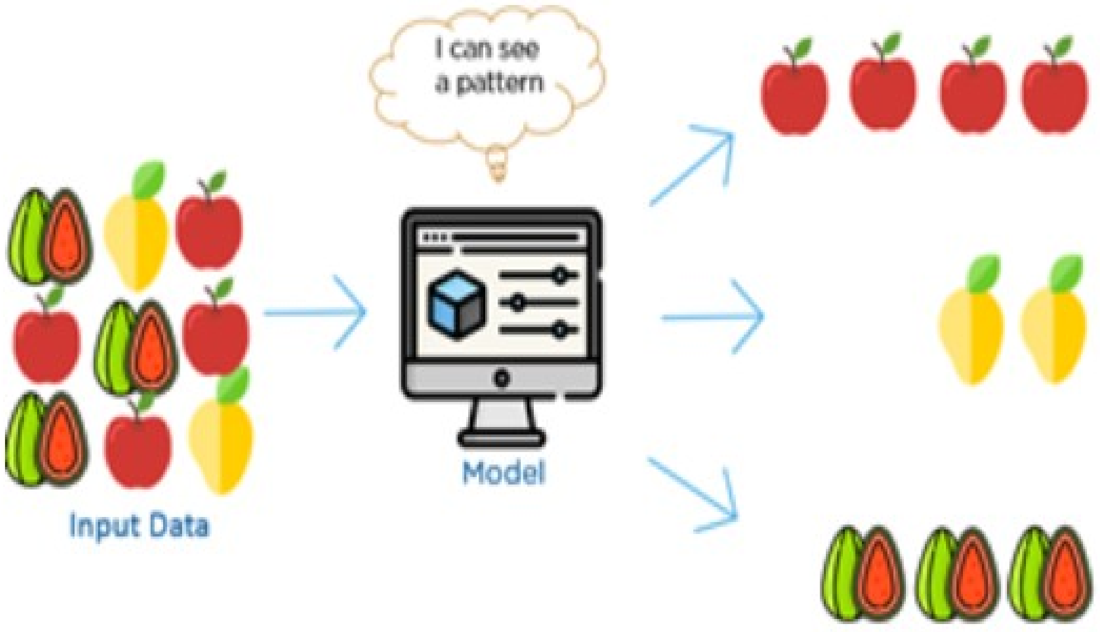
Supervised Learning



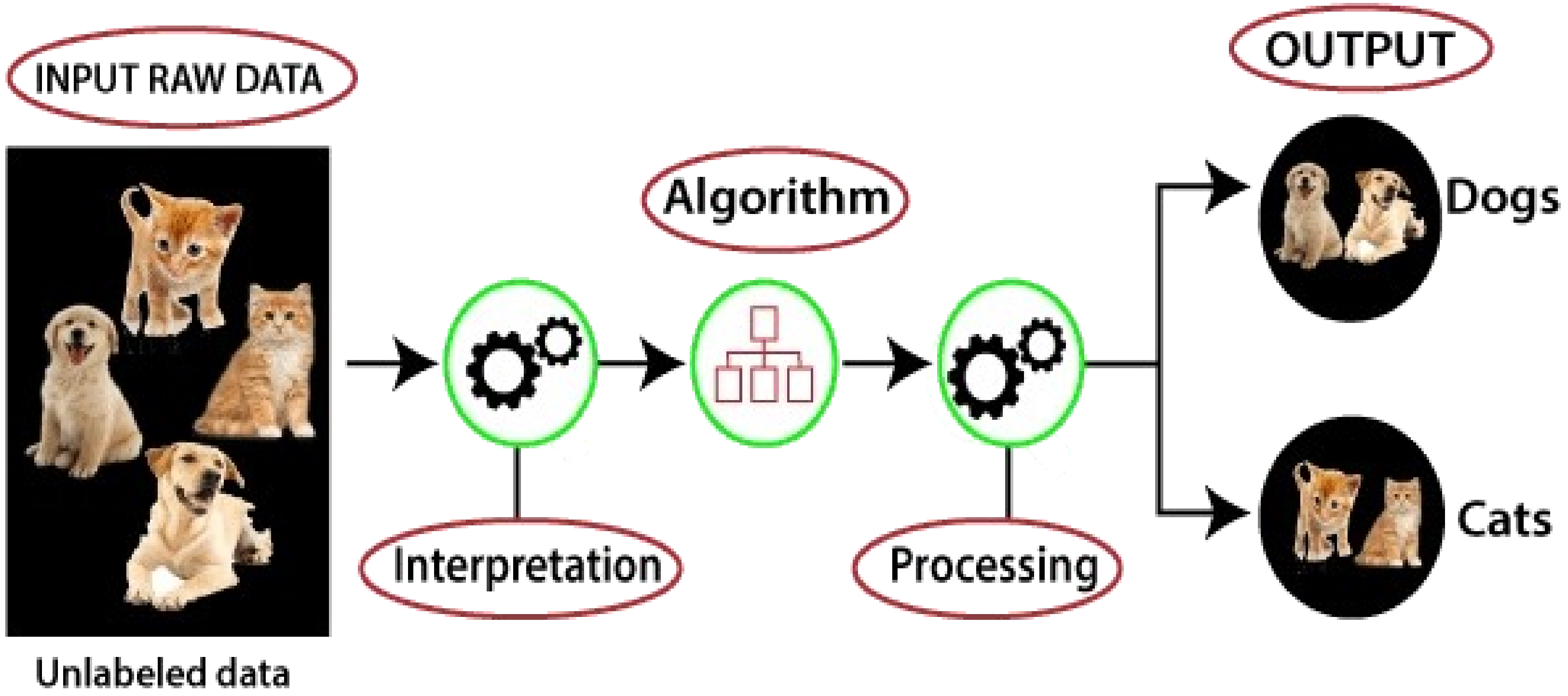
Unsupervised Machine Learning

- It is a type of machine learning where the algorithm learns to recognize patterns in data without being explicitly trained using labeled examples.
- The goal of unsupervised learning is to discover the underlying structure or distribution in the data.
- Two main types – Clustering, Dimensionality Reduction

Unsupervised Machine Learning



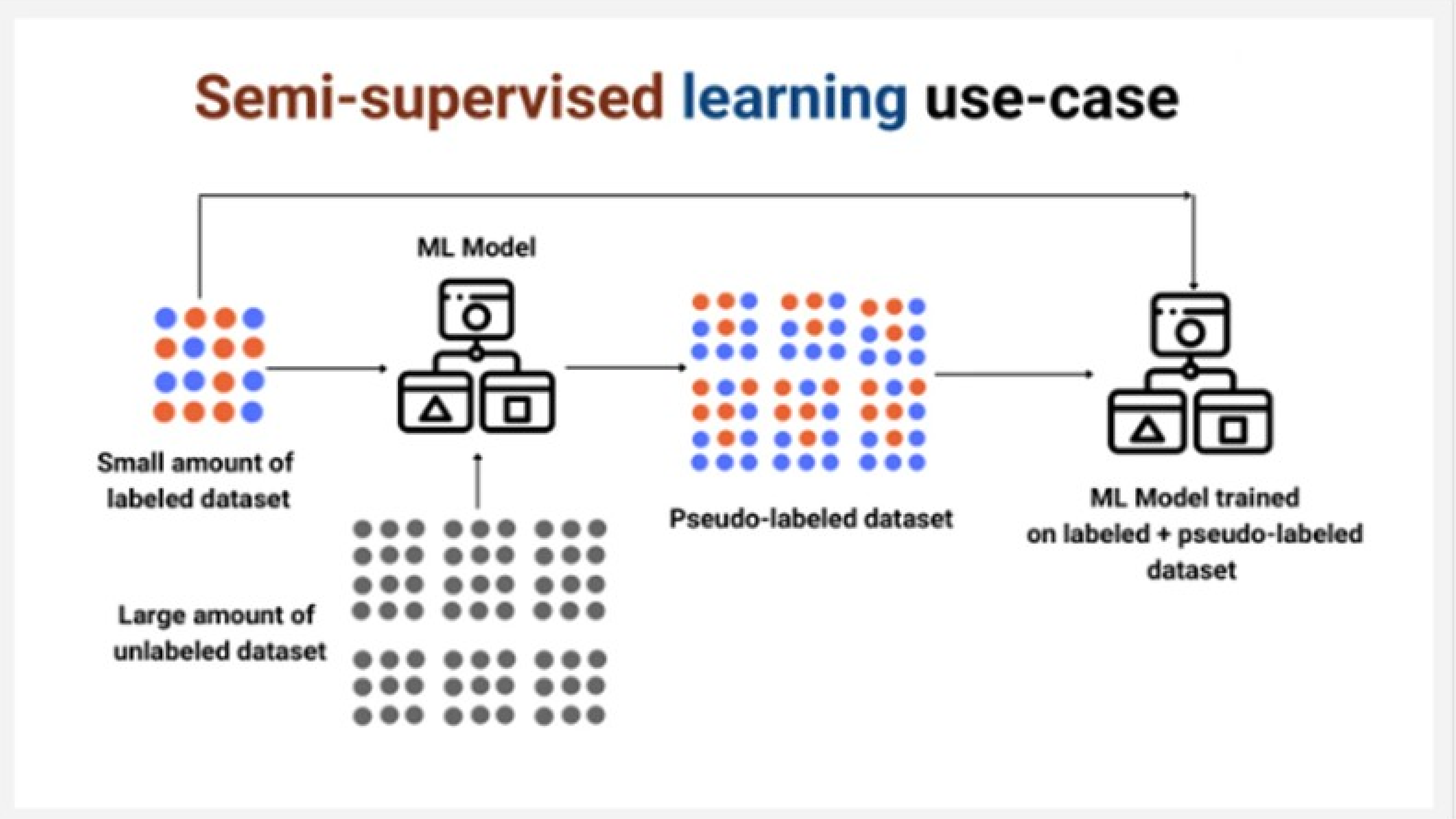
Example of Unsupervised learning



Semi-supervised learning

- Where an incomplete training signal is given: a training set with some (often many) of the target outputs missing.
- There is a special case of this principle known as Transduction where the entire set of problem instances is known at learning time, except that part of the targets are missing
- It is an approach to machine learning that combines small labeled data with a large amount of unlabeled data during training
- Semi-supervised learning falls between unsupervised learning and supervised learning.

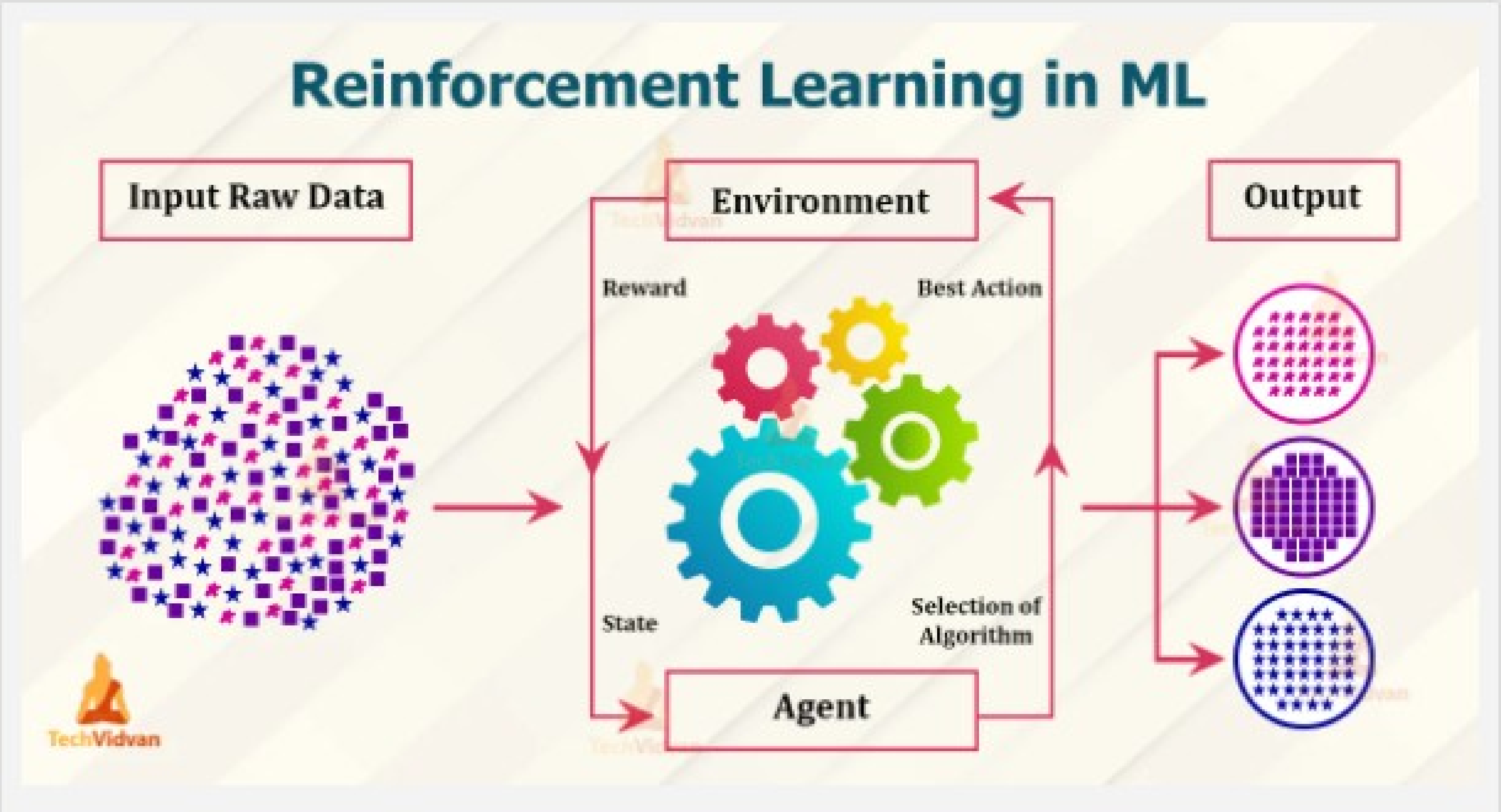
Semi-supervised learning



Reinforcement learning

- Reinforcement learning is the problem of getting an agent to act in the world so as to maximize its rewards.
- A learner is not told what actions to take as in most forms of machine learning but instead must discover which actions yield the most reward by trying them.
- the goal of reinforcement learning is to learn a policy, which is a mapping from states to actions, that maximizes the expected cumulative reward over time.
- Tow main types – Model Based Reinforcement Learning, Model Free reinforcement Learning

Reinforcement learning



Reinforcement learning

Where is reinforcement learning in the real world?

- **Video Games**
- **Industrial Simulation**
- **Resource Management**

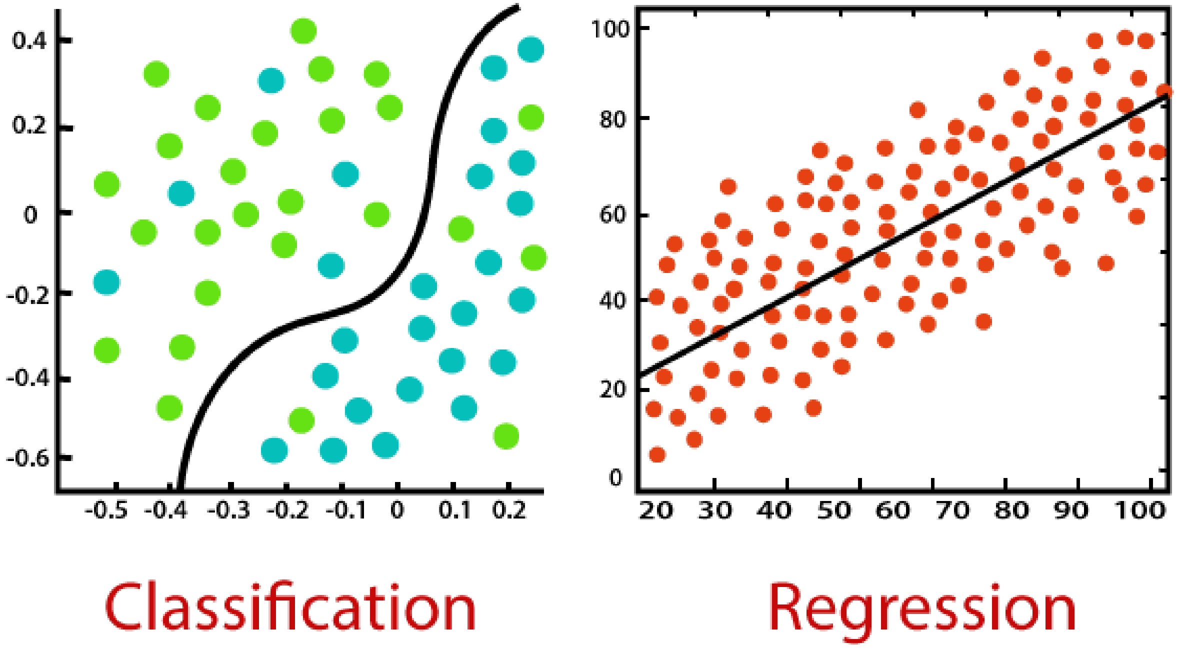
Classification of Problem

- ML problem divided into different categories

1. Classification
2. Regression

- Both are Supervised Learning algorithms
- Both the algorithms are used for prediction in Machine learning and work with the labeled datasets.
- Difference - How they are used for different machine learning problems.

- Regression algorithms are used to **predict the continuous** values such as price, salary, age, etc.
- Classification algorithms are used to **predict/Classify the discrete values** such as Male or Female, True or False, Spam or Not Spam, etc.



Classification

- It is a process of finding a function which helps in dividing the dataset into classes based on different parameters.
- A computer program is trained on the training dataset and based on that training, it categorizes the data into different classes.
- **Task of algorithm** - find the mapping function to map the input(x) to the discrete output(y).
- **Example** - Email Spam Detection

The model is trained on the basis of millions of emails on different parameters and and whenever it receives a new email, it identifies whether the email is spam or not. If the email is spam, then it is moved to the Spam folder.

- **Types of ML Classification Algorithms:**

✓ Logistic Regression
✓ K-Nearest Neighbours
✓ Support Vector Machines
✓ Kernel SVM
✓ Naïve Bayes
✓ Decision Tree Classification
✓ Random Forest Classification

Regression

- Regression is a process of finding the correlations between dependent and independent variables.
- It helps in predicting the continuous variables such as prediction of Market Trends, prediction of House prices, etc.
- **Task of Algorithm -** The task of the Regression algorithm is to find the mapping function to map the input variable(x) to the continuous output variable(y).
- **Example:** Weather forecasting

In weather prediction, the model is trained on the past data, and once the training is completed, it can easily predict the weather for future days.

- **Types of Regression Algorithm:**

✓ Simple Linear Regression

✓ Multiple Linear Regression

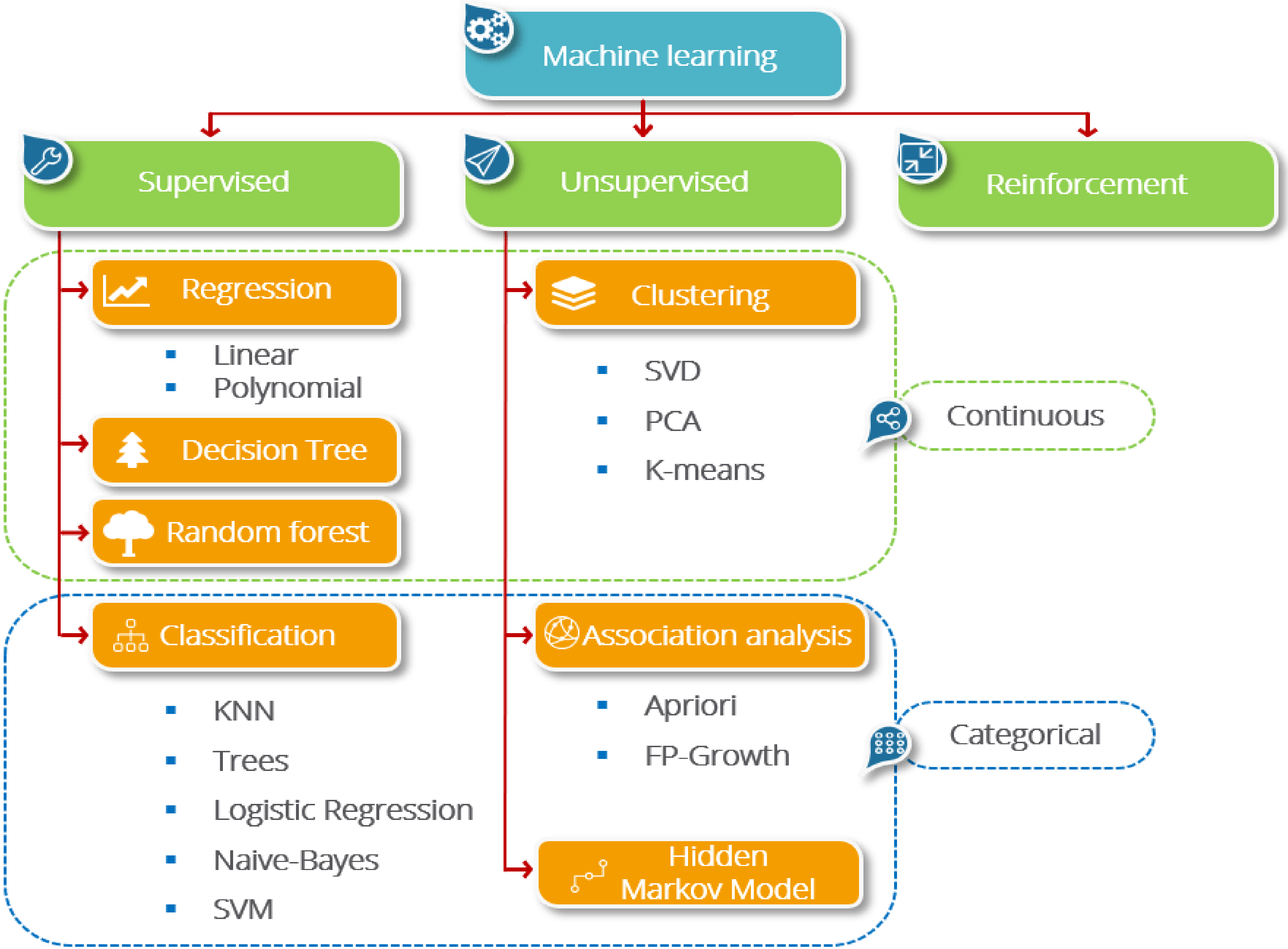
✓ Polynomial Regression

✓ Support Vector Regression

✓ Decision Tree Regression

✓ Random Forest Regression

Regression Algorithm	Classification Algorithm
In Regression, the output variable must be of continuous nature or real value.	In Classification, the output variable must be a discrete value.
The task of the regression algorithm is to map the input value (x) with the continuous output variable(y).	The task of the classification algorithm is to map the input value(x) with the discrete output variable(y).
Regression Algorithms are used with continuous data.	Classification Algorithms are used with discrete data.
In Regression, we try to find the best fit line, which can predict the output more accurately.	In Classification, we try to find the decision boundary, which can divide the dataset into different classes.
Regression algorithms can be used to solve the regression problems such as Weather Prediction, House price prediction, etc.	Classification Algorithms can be used to solve classification problems such as Identification of spam emails, Speech Recognition, Identification of cancer cells, etc.
The regression Algorithm can be further divided into Linear and Non-linear Regression.	The Classification algorithms can be divided into Binary Classifier and Multi-class Classifier.


Machine learning
Supervised
Unsupervised
Reinforcement
Regression
Linear
Polynomial
Decision Tree
Random forest
Classification
KNN
Trees
Logistic Regression
Naive-Bayes
SVM
Clustering
SVD
PCA
K-means
Association analysis
Apriori
FP-Growth
Hidden Markov Model
Continuous
Categorical