Introduction, Concepts and Fundamentals of AI

Dr. P. Kalpana, M.E., PhD.

Faculty of Mechanical Engineering

kalpana@iiitdm.ac.in

INDIAN INSTITUTE OF INFORMATION TECHNOLOGY,
DESIGN AND MANUFACTURING,
KANCHEEPURAM

Introduction

*"Emerging fields, including **Artificial intelligence, Machine learning**, internet of things, blockchain, and big data can take India to new heights of development, and improve the quality of life of its citizens," PM Modi said, adding for India, this is not just an industrial transformation, but a social transformation.*

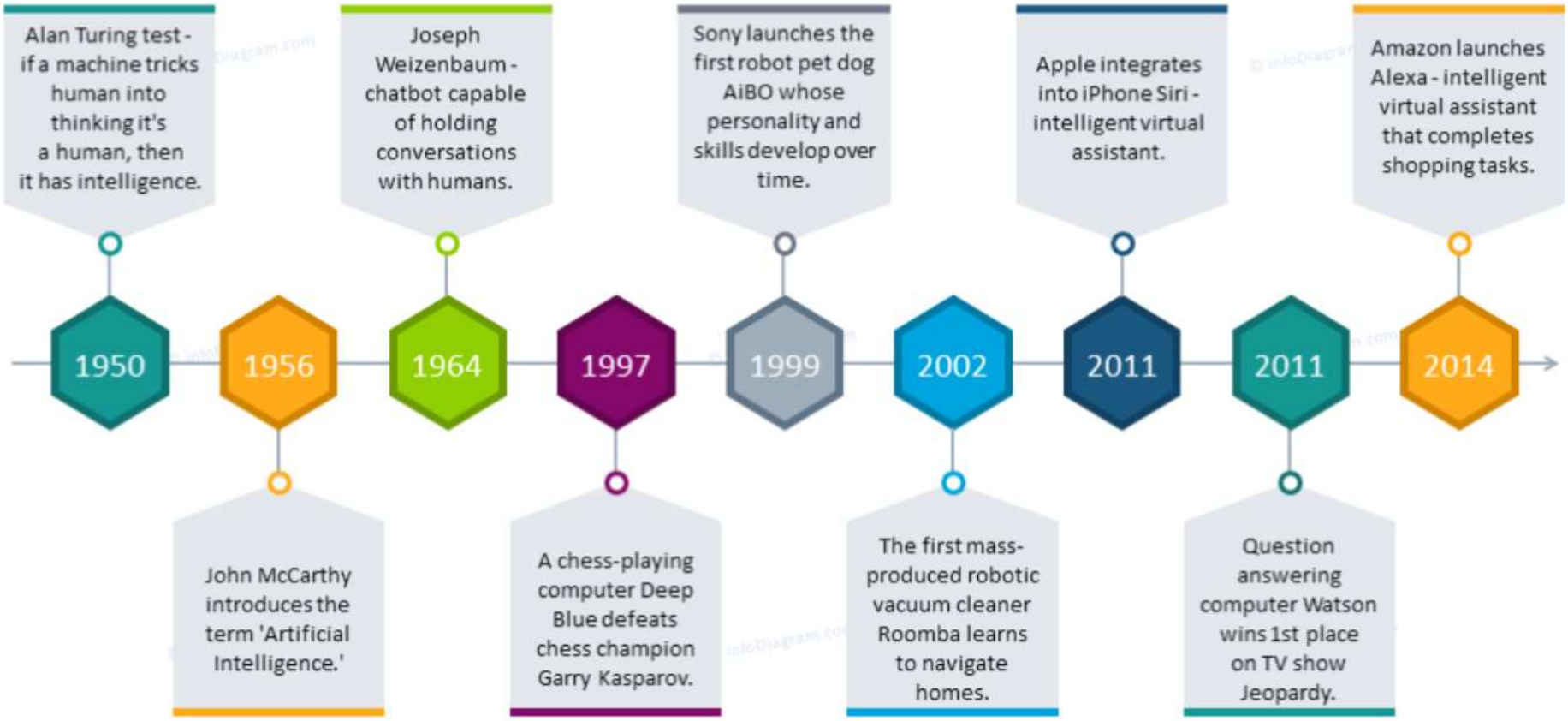
The new technologies, including Internet of Things and blockchain, will help India to address challenges in education, healthcare, agriculture, job creation and environmental stainability, PM Modi said.

- Source: https://gadgets.ndtv.com/internet/news/pm-modi-says-ai-blockchain-will-change-the-nature-of-jobs-1931108

Agenda

- History of AI
- What is artificial intelligence?
- AI Applications
- Types of AI
- Programming languages for AI
- Machine Learning
- Machine Learning Algorithms
- Limitations of Machine Learning
- Deep Learning
- Neural Networks
- Natural Language processing

History of AI



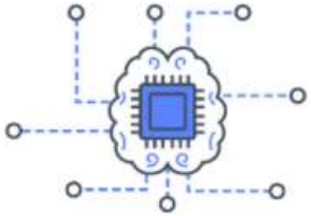
Source: https://www.infodiagram.com/diagrams/ai-diagrams-machine-learning-ppt-template.html

History of AI

Year	Evolution of AI	Year	Evolution of AI
1950	Alan Turing introduced Turing Test for evaluation of intelligence	1969	*Shakey* a robot, equipped with locomotion, perception, and problem solving.
1951	checkers program , Program for chess	1973	Freddy, Robot, capable of using vision to locate and assemble models
1956	John McCarthy coined the term Artificial Intelligence	1979	The first computer-controlled autonomous vehicle- standford cart
1958	LISP programming language for AI	1985	The drawing program, Aaron.
1959	The first AI laboratory (MIT Lab)	1990	Major advances in all areas of AI
1960	The first robot by General Motors assembly line	1997	The Deep Blue Chess Program
1961	The first chat bot ELIZA was invented	2000	Interactive robot pets
1964	A Dissertation showed at MIT that computers can understand natural language	2005	A robotic car named Stanley built by Stanford's racing team
		2011	IBM's question answering system, Watson

Domains of AI

To process, analyze and interpret, data in order to solve real-world problems



Machine Learning

Neural Networks on high dimensional data to get insights and form solutions

Neural Networks

artificial agents acting in a real-world environment to produce results

Robotics

AI-based computer system that learns and reciprocates the decision-making ability of a human expert

Expert Systems

Fuzzy Logic

Natural Language Processing

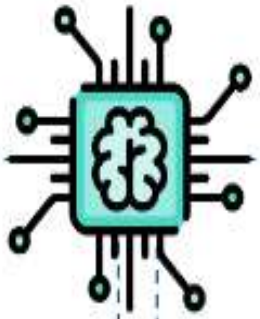
Drawing insights from natural human language in order to communicate with machines

computing approach based on the principles of "degrees of truth"

Reasons for Demand of AI

More Computational Power

Perform high-level computations and implement complex algorithms (GPU's)



More Data

IOT and Social Media

Better algorithms

Machine learning and deep learning algorithms

Broad Investment

Tesla, Netflix and Facebook, Alibaba, Baidu etc Ecommerce to autonomous cars

Intelligence

- The ability of a system to calculate, reason, perceive relationships and analogies, learn from experience, store and retrieve information from memory, solve problems, comprehend complex ideas, use natural language fluently, classify, generalize, and adapt new situations.

Types of Intelligence	Description	Roles
Linguistic intelligence	The ability to speak, recognize, and use mechanisms of phonology (speech sounds), syntax (grammar), and semantics (meaning)	Narrators, Orators
Musical intelligence	The ability to create, communicate with, and understand meanings made of sound, understanding of pitch, rhythm	Musicians, Singers, Composers
Logical-mathematical intelligence	The ability of use and understand relationships in the absence of action or objects. Understanding complex and abstract ideas	Mathematicians, Scientists
Spatial intelligence	The ability to perceive visual or spatial information, change it, and re-create visual images without reference to the objects, construct 3D images, and to move and rotate them	Map readers, Astronauts, Physicists
Bodily-Kinesthetic intelligence	The ability to use complete or part of the body to solve problems or fashion products, control over fine and coarse motor skills, and manipulate the objects	Players, Dancers
Interpersonal intelligence	The ability to recognize and make distinctions among other people's feelings, beliefs, and intentions	Mass Communicators, Interviewers

What is Artificial Intelligence

- In 1956 John McCarthy defined AI as ***"The science and engineering of making intelligent machines, especially intelligent computer programs"***

- The theory and development of computer systems able to perform tasks normally requiring human intelligence such as visual perception, speech recognition, decision making and translation between languages

- Artificial Intelligence is a way of making a computer, a computer-controlled robot, or a software think intelligently, in the similar manner the intelligent humans think

- Behaviors associated with human intelligence:

planning, learning, reasoning, problem solving, knowledge representation, perception, motion, and manipulation and, to a lesser extent, social intelligence and creativity

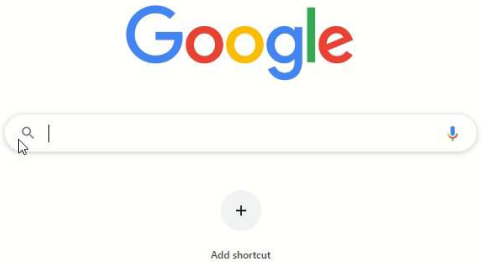
What is Artificial Intelligence

- Patrick Winston, the Ford professor of artificial intelligence and computer science at MIT, defines AI as "algorithms enabled by constraints, exposed by representations that support models targeted at loops that tie thinking, perception and action together."
- DataRobot CEO Jeremy Achin defined AI as "AI is a computer system able to perform tasks that ordinarily require human intelligence. Many of these artificial intelligence systems are powered by machine learning, some of them are powered by deep learning and some of them are powered by very boring things like rules."
- Artificial intelligence can be viewed from a variety of perspectives
  - From the perspective of **intelligence** artificial intelligence is making machines 'intelligent'.
  - From a **business** perspective AI is a set of very powerful tools, and methodologies for using those tools to solve business problems.

- Goals of AI
  - To Create Expert Systems
  - To Implement Human Intelligence in Machines

AI Applications

- ***Predictive search engine***
  - Recommendation Engine
  - Based on data that google collects about you such as browse history, age, location and personal details
  - involves Natural language processing, Deep Learning and Machine Learning



AI Applications as Expert Systems

- ***Financial sector***
  - J.P.Morgan Chase's contract Intelligent platform
    - Uses AI,ML and Image recognition to analyze legal documents
    - Manually reviewing around 12000 documents will take 36000hours
    - Computing complex problems in a matter of seconds
- ***Medicine***
  - *IBM Watson technology*:
    - More than 230 health care organizations use IBM AI technology.
    - Able to cross reference 20 million records quickly
    - Correctly diagnose a rare leukemia condition in a patient
  - *Google's AI eye doctor*
    - An initiative taken by google with an Indian Eye care chain
    - To examine retinal scans and identify the diabetic retinopathy which can cause blindness

AI Applications

- ***Vision Systems***
  - These systems understand, interpret, and comprehend visual input on the computer.
  - Face book
    - Face verification and auto tagging
  - Twitter
    - Identify terroristic, offensive and reportable contents in tweets
    - Identified 300000 terroristic link accounts
    - 95% were non human AI machine generated links
- ***Speech Recognition***
  - Capable of hearing and comprehending the language
  - Handle different accents, slang words, noise in the background, change in human's noise due to cold
  - Google Duplex
    - Google Virtual Assistant
    - Not only respond to calls and book appointments
    - Adds human touch
    - Difficult to understand human AI speaking over the phone

AI Applications

- ***Handwriting Recognition***
  - Reads the text written on paper by a pen or on screen by a stylus
  - Recognize the shapes of the letters and convert it into editable text
  - Example: writing math equations in One note

- ***Intelligent Robots***
  - Robots are able to perform the tasks given by a human.
  - They have sensors to detect physical data from the real world
    - light, heat, temperature, movement, sound, and pressure.
  - Eg. Self-driving vehicles – Tesla
  - Robo taxi-one that can ferry passengers without anyone behind the wheel
- Other Applications
  - Movie recommendation by Netflix
  - Spam filtering in Gmail

Types Of AI

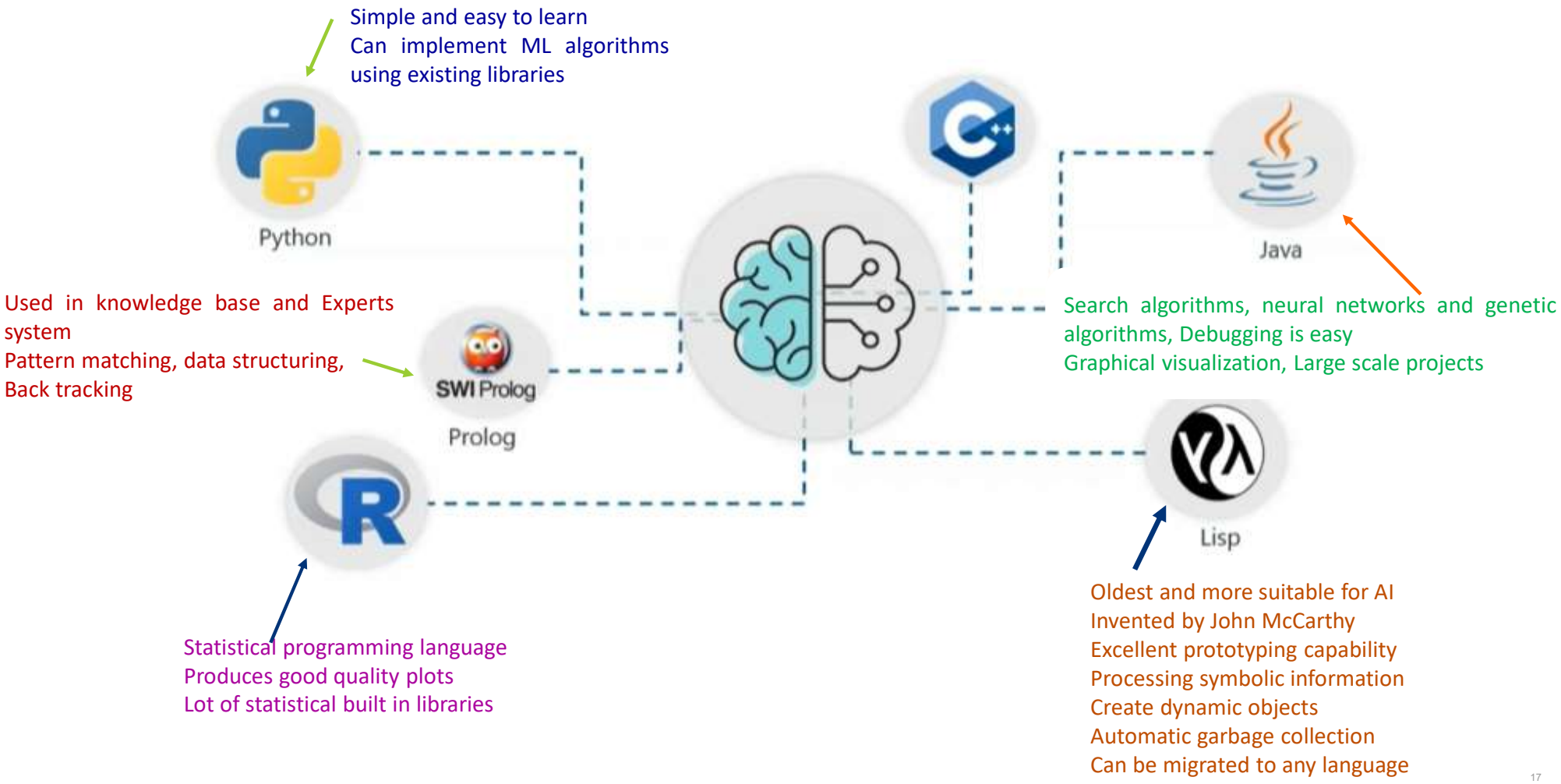
- ***Artificial Narrow Intelligence***
- It is also known as weak AI.
- It involves applying AI only to specific task.
- Many currently existing systems are actually operating as weak AI focused on a narrowly defined specific problem.
- Eg. Alexa
  - It operates within unlimited pre-defined range of functions.
  - There is no genuine intelligence or self-awareness.
- Other examples:
  - Google search engine, Sophia the humanoid, self-driving cars, AlphaGo

Types Of AI

- ***Artificial General Intelligence***
- It is also known as strong AI
- It involves machines that possess the ability to perform any intelligent task that a human being can
- In general, machines don't possess human-like abilities
- Though, they can perform high level computations with strong processing units, machines don't have the reasonable way of thinking

- ***Artificial Super Intelligence***
- It is referring to the time when the capabilities of computers will surpass humans
- Presently, it is seen as a hypothetical situation as depicted in science fiction books, where machines have taken over the world
- Eg. Science fiction movies like terminator depict artificial super intelligence

Programming Languages For AI



Why is Python best for AI?

Windows, MacOS, Linux, Unix
Transferring code from one platform to the other by PyInstaller

support pre-defined packages
"check as you code"

install and load the necessary packages
NumPy, Keras, Tensorflow, Pytorch

very simple syntax

Interpreted object oriented high level language, No memory allocation issue, Also invoke c, c++, java libraries

AI based python libraries

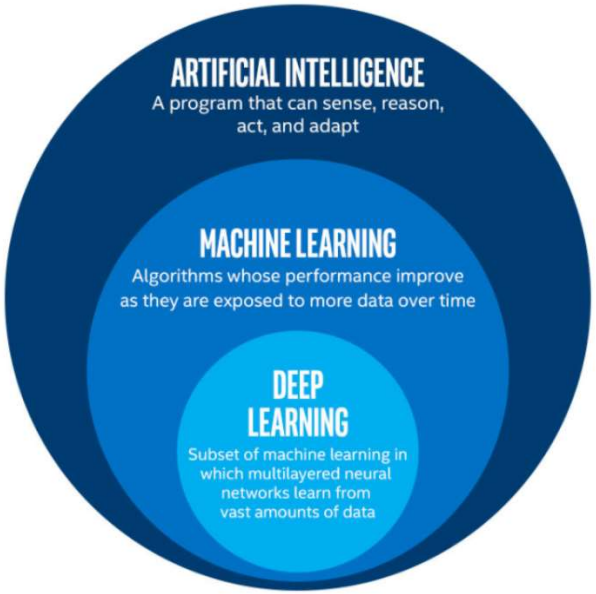
Library	Description
Tensorflow	Developed by Google, this library is popularly used in writing Machine Learning algorithms and performing heavy computations that involve Neural Networks.
Scikit-Learn	It is associated with NumPy and SciPy. It is considered as one of the best libraries for working with complex data.
NumPy	It is specifically used for computing scientific/mathematical data.
Theano	It is a functional library that effectively calculates and computes mathematical expressions involving multi-dimensional arrays.
Keras	It simplifies the implementation of neural networks. It also has the best functionalities for computing models, evaluating data-sets, visualizing graphs, and much more.
NLTK	NLTK or Natural Language Tool Kit is an open source Python library specifically built for Natural Language Processing, text analysis, and text mining.
Pandas	It is built on the Numpy package and its key data structure is called the DataFrame. DataFrames allow you to store and manipulate tabular data in rows of observations and columns of variables. There are several ways to create a DataFrame.

Sub-fields of Artificial Intelligence

- Neural Networks – e.g. brain modelling, time series prediction, classification
- Evolutionary Computation – e.g. genetic algorithms, genetic programming
- Vision – e.g. object recognition, image understanding
- Robotics – e.g. intelligent control, autonomous exploration
- Expert Systems – e.g. decision support systems, teaching systems
- Speech Processing– e.g. speech recognition and production
- Natural Language Processing – e.g. machine translation
- Planning – e.g. scheduling, game playing
- Machine Learning – e.g. decision tree learning, version space learning

AI Vs ML Vs DL

Artificial Intelligence	Machine learning	Deep learning
It is basically the science of getting machines to mimic the behavior of human beings.	It is a subset of AI that focuses on getting machines to make decisions by feeding them data.	It is a subset of machine learning that uses the concept of neural networks to solve complex problems



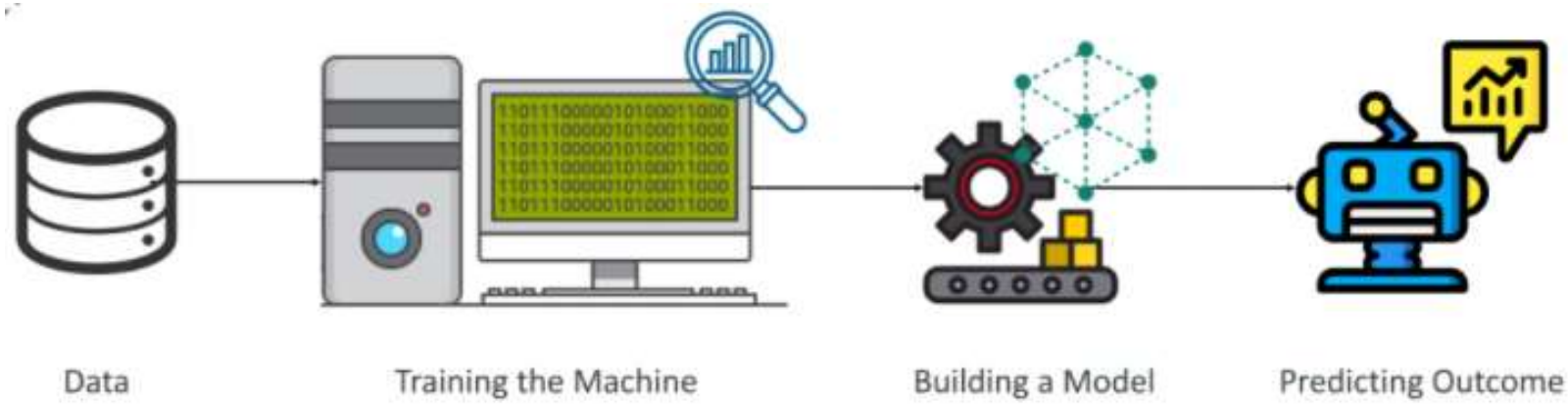
Siri is the application of AI, ML and DL

Introduction To Machine Learning

- A method in artificial intelligence
- Machine learning was first coined by Arthur Samuel in the year 1959

Tom M. Mitchell defines Machine Learning as

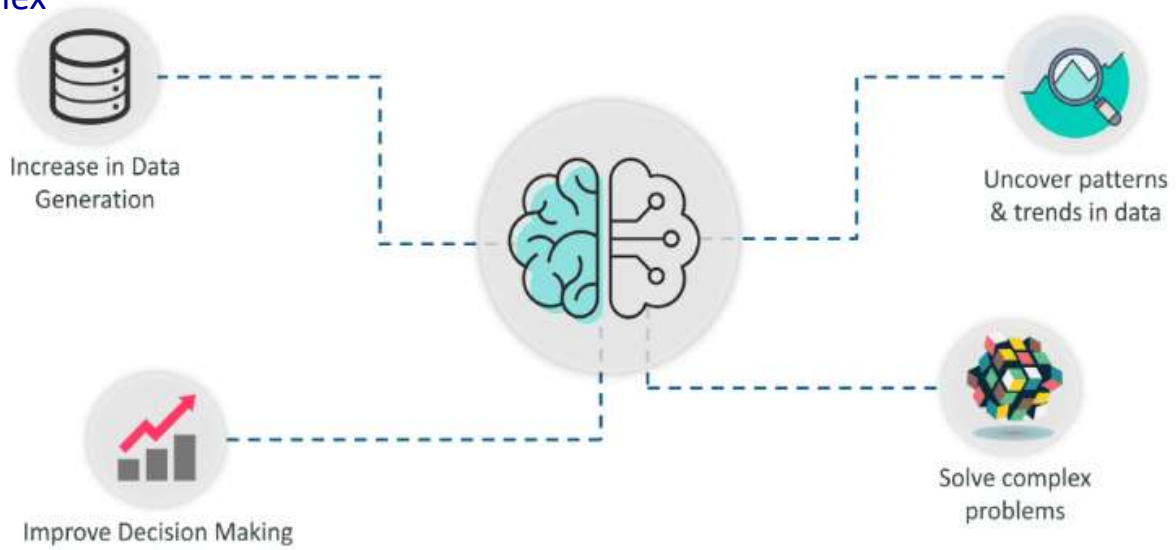
"A computer program is said to learn from experience E with respect to some class of T and performance measure P if its performance at task in T, as measured by P, improves with experience E."



Need for Machine learning

To structure, analyze and draw useful insights from data
To solve problems and find solutions to the most complex tasks

Finding hidden patterns and extracting key insights
to dig beneath the surface and explore the data



To make better business decisions
To forecast sales, predict downfalls in the stock market, identify risks and anomalies

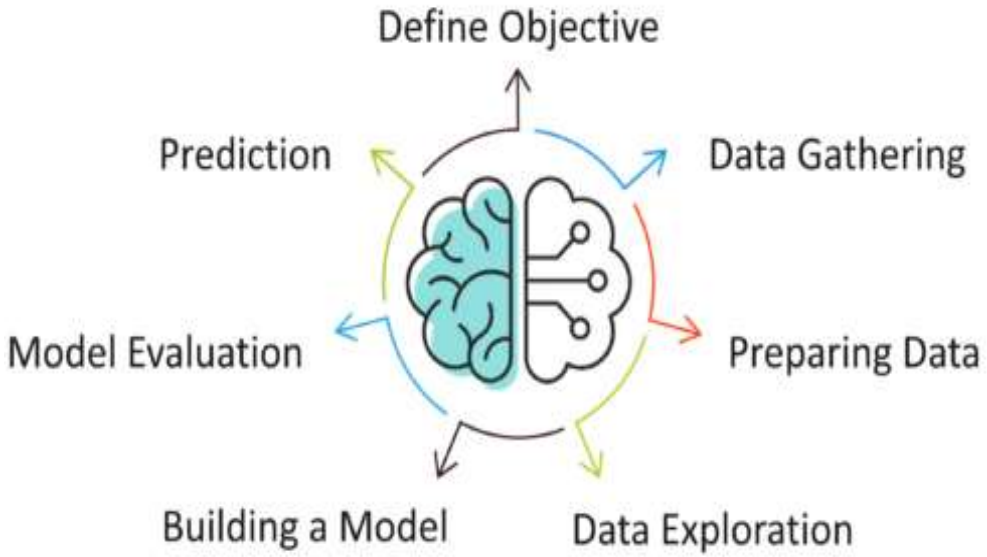
detecting the genes linked to the deadly ALS disease
building self-driving cars

Machine learning definitions

- **Algorithm**
  - 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 trained by using a machine learning algorithm. So, the model uses the algorithm in order to draw useful insights from the input and gives the outcome that is very precise.
- **Predictor variable**
  - It is a feature 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 predictor variable.
- **Training data**
  - The machine learning model is built using the training data. It helps the model to identify key trends and patterns which are essential to predict the output
- **Testing data**
  - The machine learning model is evaluated using the testing data.
  - It helps to test the efficiency of the model

Machine learning process

- The machine learning process involves building a predictive model that can be used to find the solution for a problem statement



Machine learning process

- **Define the Objective**
  - What are we trying to predict?
  - What is the target variable?
  - Is that response variable continuous or discrete?
  - What the predictor variable?
  - What kind of problem we are solving, classification, regression or clustering?
  - The objective is to predict the possibility of rain by studying weather conditions
- **Data gathering**
  - What kind of data is needed to solve this problem?
  - Is the data available?
  - How can I get the data?
  - humidity level, temperature, pressure, locality etc
- **Data preparation**
  - Transform data into desired format
  - Inconsistencies such as missing values, redundant variables, duplicate values, corrupted data
  - Removing these inconsistencies is very essential

Machine learning process

- **Exploratory data analysis**
  - The brainstorming stage of Machine learning
  - Involves understanding the patterns and trends in the data
  - correlations between the variables are understood
  - Eg: There is a strong possibility of rain if the temperature has fallen low
- **Building a Machine learning model**
  - splitting data set into two parts, training data and testing data
  - The training data will be used to build and analyze the model
  - The logic of the model is based on the Machine Learning Algorithm
  - Choosing the right algorithm
    - the type of problem,
    - the data set
    - the level of complexity
  - Algorithm for rain prediction is classification algorithm
    - Logistic Regression or Decision Tree

Machine learning process

- Model evaluation & optimization
  - The testing data set is used to check the efficiency of the model
  - any further improvements in the model based on the accuracy of the model
  - Accuracy measures
    - MSE-Mean square error
    - MAPE-Mean absolute percentage error
    - RMSE-Root mean square error
    - Tracking signal, a method to control the forecast accuracy
- Predictions
  - Predict the out for the given input
  - Output can be Categorical variable (eg. True or False) or it can be a Continuous Quantity (eg. the predicted value of a stock)

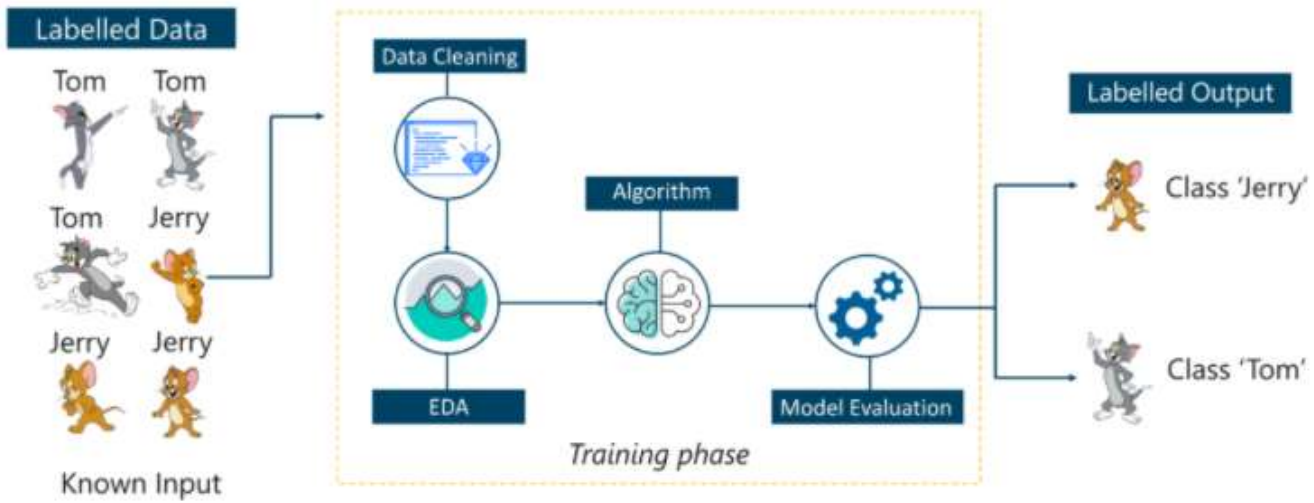
Types of Machine learning

- Supervised Learning
- Unsupervised Learning
- Reinforcement Learning

Types of Machine learning

➢Supervised learning

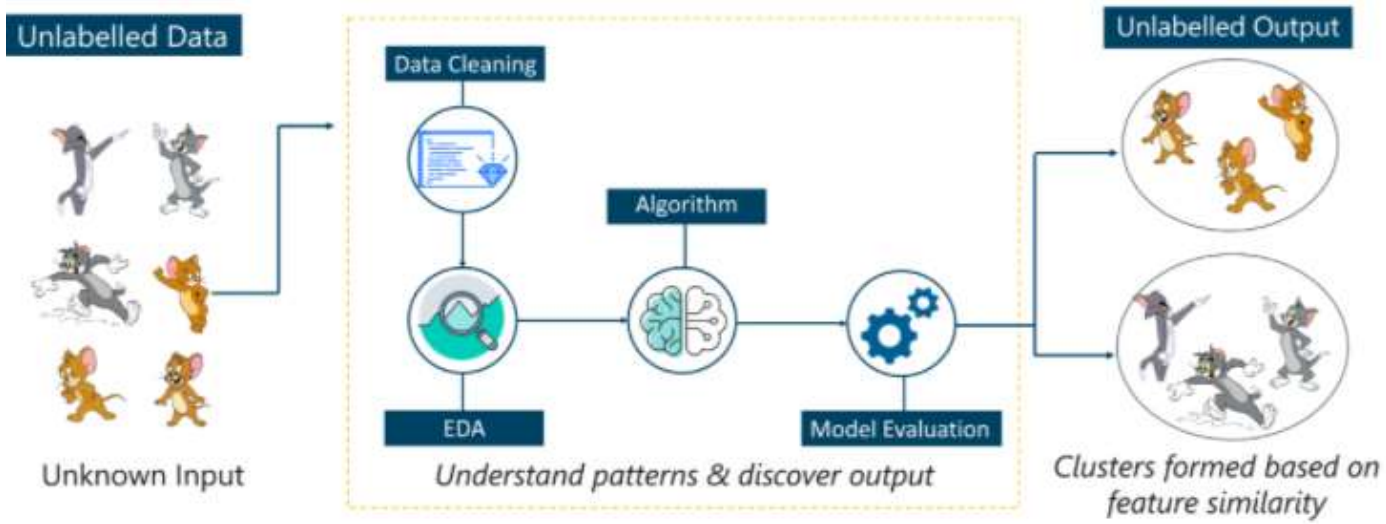
- ➢It is a technique in which we teach or train the machine by using the data, which is well labelled. The label data is a teacher that will train to understand the patterns in the data
- ➢The label data is a teacher that will train to understand the patterns in the data



Types of Machine learning

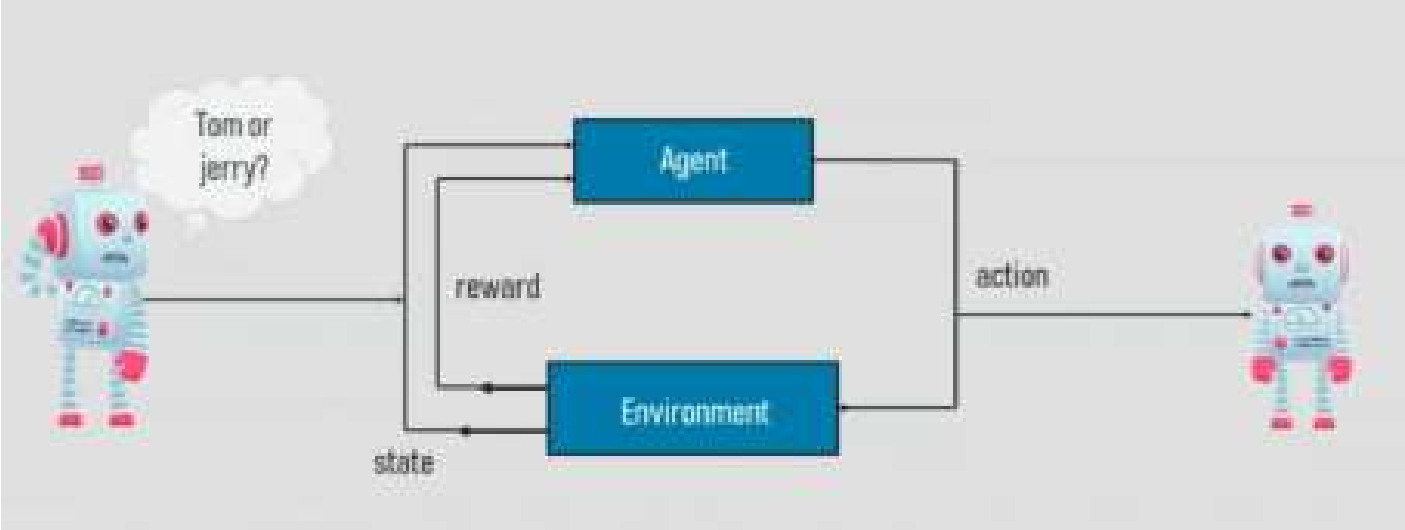
➢Unsupervised learning

➢It involves training by using unlabeled data and allowing the model to act on that information without any guidance. The machine has to understand the patterns and discover the output on its own.



Types of Machine learning

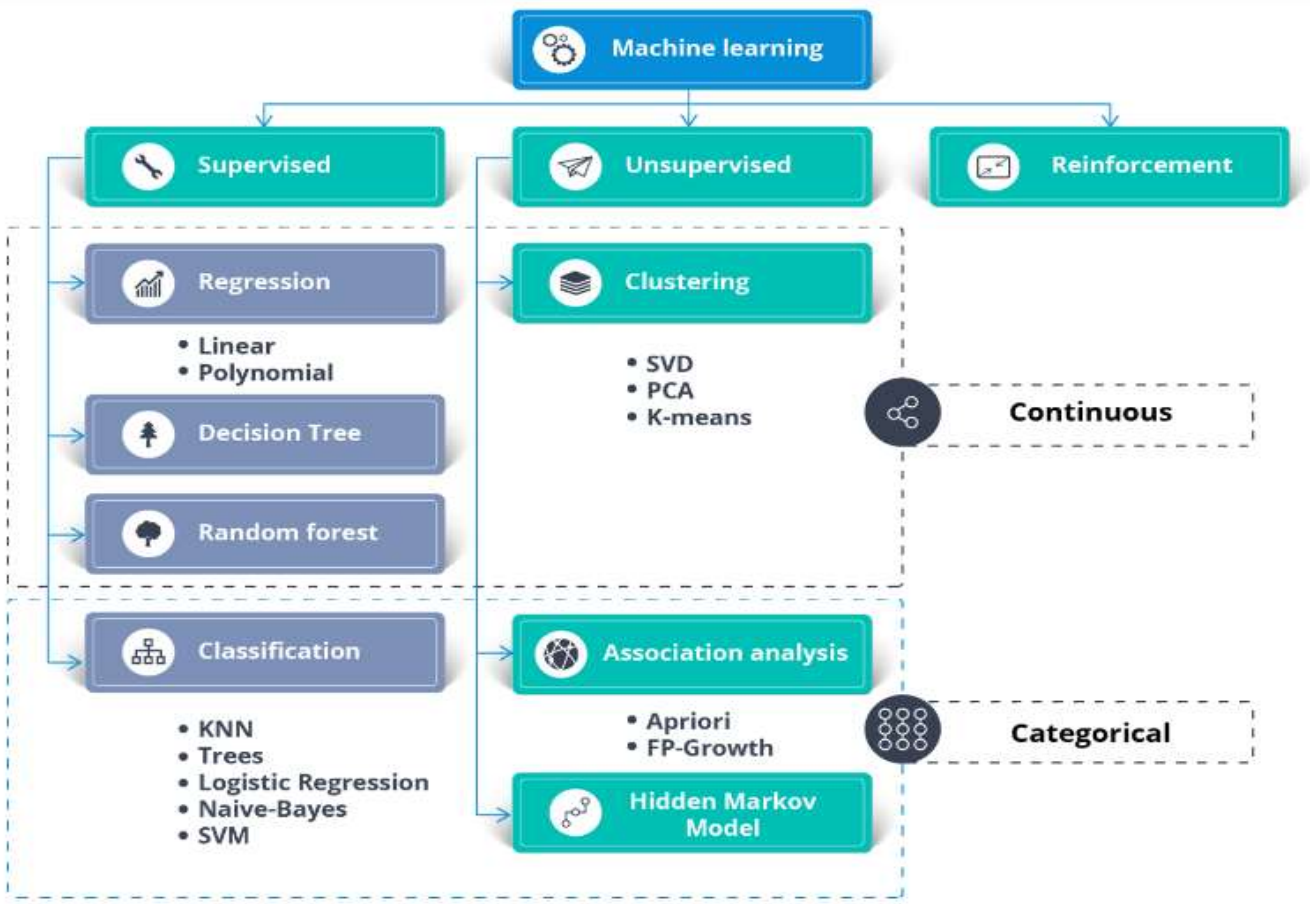
- **Reinforcement learning**
- An agent is put in an environment and he learns to behave in this environment by performing certain actions and observing the rewards which it gets from the actions. It is mainly used in advanced machine learning areas such as self-driving cars and AlphaGo.



Supervised learning vs Unsupervised learning vs Reinforcement learning

Description	Supervised learning	Unsupervised learning	Reinforcement learning
Definition	The machine learns by using labelled data	The machine is trained on unlabeled data without any guidance	An agent interacts with its environment by producing actions and discovers errors or rewards
Type of problems	Regression and Classification	Association and Clustering	Reward based
Type of data	Labelled data	Unlabeled data	No pre-defined data
Training	External supervision	No supervision	No supervision
Approach	Map labelled input to known output	Understand patterns and discover output	Follow trial and error method
Popular algorithm	Linear regression, Logistic regression, Support vector machine, KNN etc.	K-means, C-means etc.	Q-learning, SARSA etc.

Machine Learning algorithms



Types of problems solved using Machine learning

Regression	Classification	Clustering
Supervised learning	Supervised learning	Unsupervised learning
Output is a continuous quantity	Output is a categorical quantity	Assigns data points into clusters
Main aim is to forecast or predict	Main aim is to compute the category of the data	Main aim is to group the similar items clusters
Eg. Predict stock market price	Eg. Classify emails as spam or non-spam	Eg. Find all transactions which are fraudulent in nature
Algorithm: Linear regression	Algorithm: Logistic regression	Algorithm: K-means

Examples:

Regression problem statement: To study the house sales dataset and build a machine learning model that predicts the housing price index

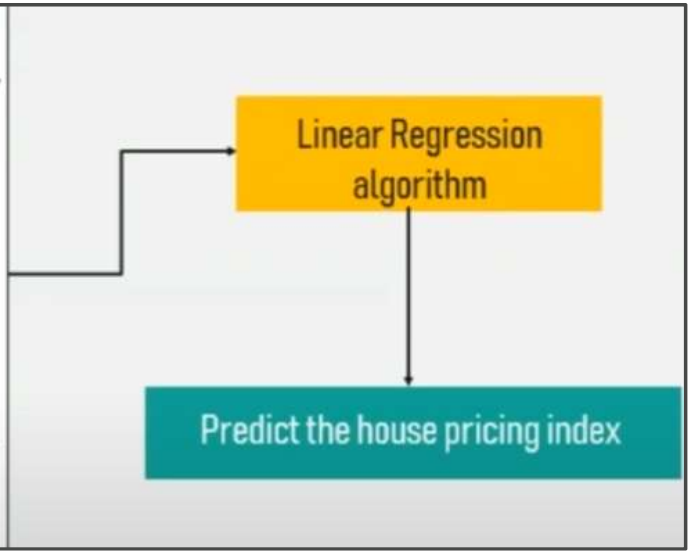
Classification problem statement: Study a bank credit dataset and make a decision about whether to approve the loan of an applicant based on his profile

Clustering problem statement: To cluster a set of movies as either good or average based on their social media out reach

Regression Problems

Problem statement: to study the house sales data set and build a machine learning model that predicts the house pricing index

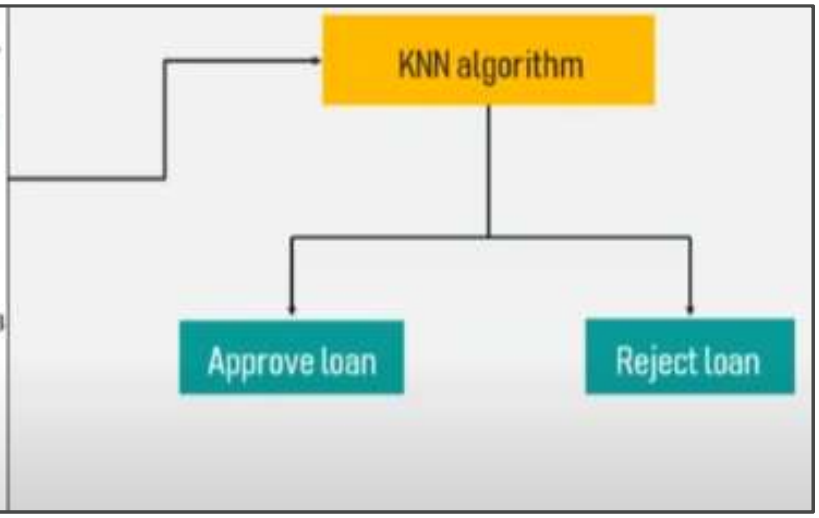
```
> str(data)
'data.frame':   21613 obs. of  21 variables:
 $ id           : num  7.13e+09 6.41e+09 5.63e+09 2.49e+09 1.95e+09 ...
 $ date         : Factor w/ 372 levels "20140502T000000",..: 165 221 291 221 284 11 57
 $ price        : num  221900 538000 180000 604000 510000 ...
 $ bedrooms     : int  3 3 2 4 3 4 3 3 3 3 ...
 $ bathrooms    : num  1 2.25 1 3 2 4.5 2.25 1.5 1 2.5 ...
 $ sqft_living  : int  1180 2570 770 1960 1680 5420 1715 1060 1780 1890 ...
 $ sqft_lot     : int  5650 7242 10000 5000 8080 101930 6819 9711 7470 6560 ...
 $ floors       : num  1 2 1 1 1 1 2 1 1 2 ...
 $ waterfront   : int  0 0 0 0 0 0 0 0 0 0 ...
 $ view         : int  0 0 0 0 0 0 0 0 0 0 ...
 $ condition    : int  3 3 3 5 3 3 3 3 3 3 ...
 $ grade        : int  7 7 6 7 8 11 7 7 7 7 ...
 $ sqft_above   : int  1180 2170 770 1050 1680 3890 1715 1060 1050 1890 ...
 $ sqft_basement: int  0 400 0 910 0 1530 0 0 730 0 ...
 $ yr_built     : int  1955 1951 1933 1965 1987 2001 1995 1963 1960 2003 ...
 $ yr_renovated : int  0 1991 0 0 0 0 0 0 0 0 ...
 $ zipcode      : int  98178 98125 98028 98136 98074 98053 98003 98198 98146 98038 ...
 $ lat          : num  47.5 47.7 47.7 47.5 47.6 ...
 $ long         : num  -122 -122 -122 -122 -122 ...
 $ sqft_living15: int  1340 1690 2720 1360 1800 4760 2238 1650 1780 2390 ...
 $ sqft_lot15   : int  5650 7639 8062 5000 7503 101930 6819 9711 8113 7570 ...
```



Classification Problems

Problem statement: study the bank credit data set and make a decision about whether to approve the loan of an applicant based on his profile

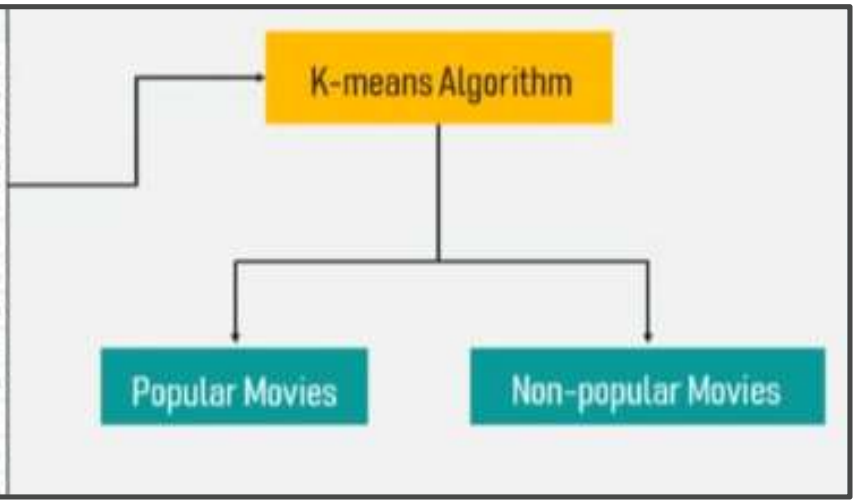
```
$ Account.Balance                    : int  1 1 2 1 1 1 1 1 4 2 ...
$ Duration.of.Credit..month.         : int  18 9 12 12 12 10 8 6 18 24 ..
$ Payment.Status.of.Previous.Credit: int  4 4 2 4 4 4 4 4 4 2 ...
$ Purpose                            : int  2 0 9 0 0 0 0 0 3 3 ...
$ Credit.Amount                      : int  1049 2799 841 2122 2171 2241
$ Value.Savings.Stocks               : int  1 1 2 1 1 1 1 1 1 3 ...
$ Length.of.current.employment       : int  2 3 4 3 3 2 4 2 1 1 ...
$ Instalment.per.cent                : int  4 2 2 3 4 1 1 2 4 1 ...
$ Sex...Marital.Status               : int  2 3 2 3 3 3 3 3 2 2 ...
$ Guarantors                         : int  1 1 1 1 1 1 1 1 1 1 ...
$ Duration.in.Current.address        : int  4 2 4 2 4 3 4 4 4 4 ...
$ Most.valuable.available.asset      : int  2 1 1 1 2 1 1 1 3 4 ...
$ Age..years.                        : int  21 36 23 39 38 48 39 40 65 23
$ Concurrent.Credits                 : int  3 3 3 3 1 3 3 3 3 3 ...
$ Type.of.apartment                  : int  1 1 1 1 2 1 2 2 2 1 ...
$ No.of.Credits.at.this.Bank         : int  1 2 1 2 2 2 2 1 2 1 ...
$ Occupation                         : int  3 3 2 2 2 2 2 2 1 1 ...
$ No.of.dependents                   : int  1 2 1 2 1 2 1 2 1 1 ...
$ Telephone                          : int  1 1 1 1 1 1 1 1 1 1 ...
$ Foreign.Worker                     : int  1 1 1 2 2 2 2 2 1 1 ...
```



Clustering Problems

Problem statement: to cluster a set of movies as either good or average based on the social media outreach

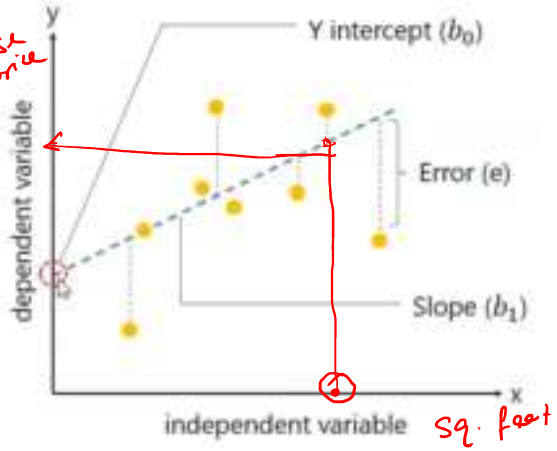
	director_facebook_likes	actor_3_facebook_likes	actor_1_facebook_likes	cast_total_facebook_likes
AvatarÂ	0	855	1000	4834
Pirates of the C...	563	1000	40000	48350
SpectreÂ	0	161	11000	11700
The Dark Knigh...	22000	23000	27000	106759
John CarterÂ	475	530	640	1873
Spider-Man 3Â	0	4000	24000	46055
TangledÂ	15	284	799	2036
Avengers: Age ...	0	19000	26000	92000
Harry Potter an...	282	10000	25000	58753
Batman v Super...	0	2000	15000	24450
Superman Retur...	0	903	18000	29991
Quantum of Sol...	395	393	451	2023
Pirates of the C...	563	1000	40000	48486

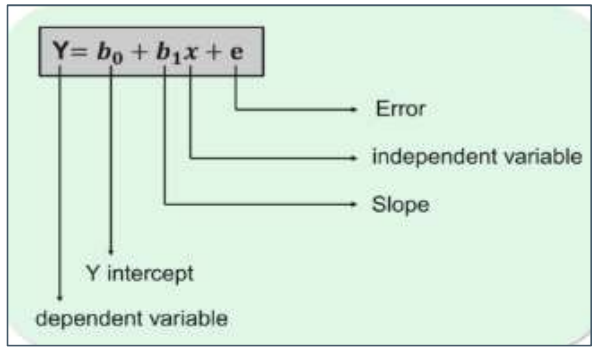


Supervised learning algorithms

Linear regression

- It is a predictive modelling technique used to predict a continuous dependent variable (Y), given one or more independent variables (X). It can be used for the cases where we want to predict some continuous quantity.





- **Example:** Predicting stock price
- **Simple linear regression:** A linear regression algorithm is called simple linear regression if it is having only one independent variable.
- **Multiple linear regression:** A linear regression algorithm is called multiple linear regression if it is having more than one independent variable.

Demo using python

$$\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_n \end{pmatrix} \qquad \beta = (X^T X^{-1})(X^T y)$$

$$\begin{pmatrix} x_{11} & \cdots & x_{1n} \\ & & \\ x_{n1} & \cdots & x_{nn} \end{pmatrix}$$

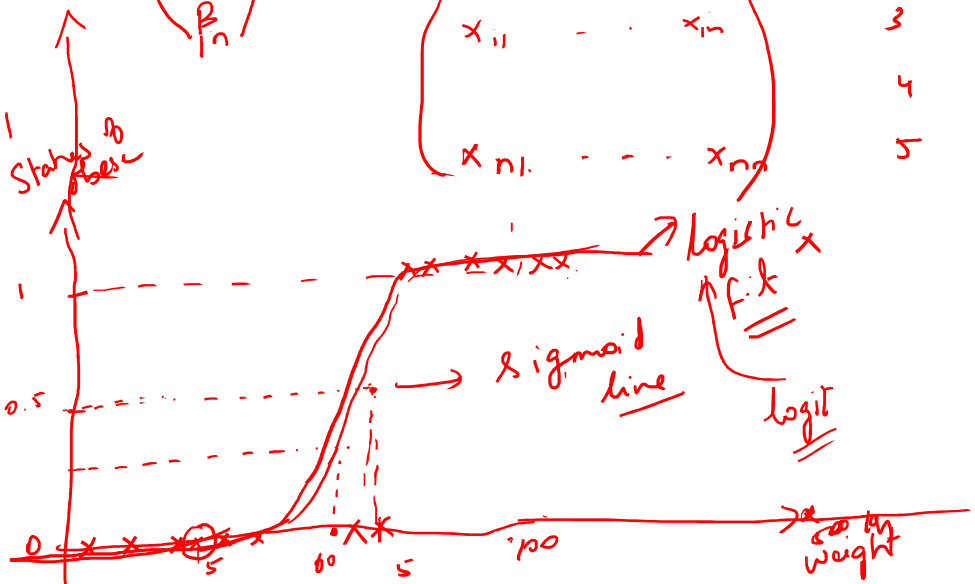
	Weight	Status of Obese
1	30	0
2	25	0
3	50	0
4	80	1
5	75	1
	90	1
	60	0
	45	0

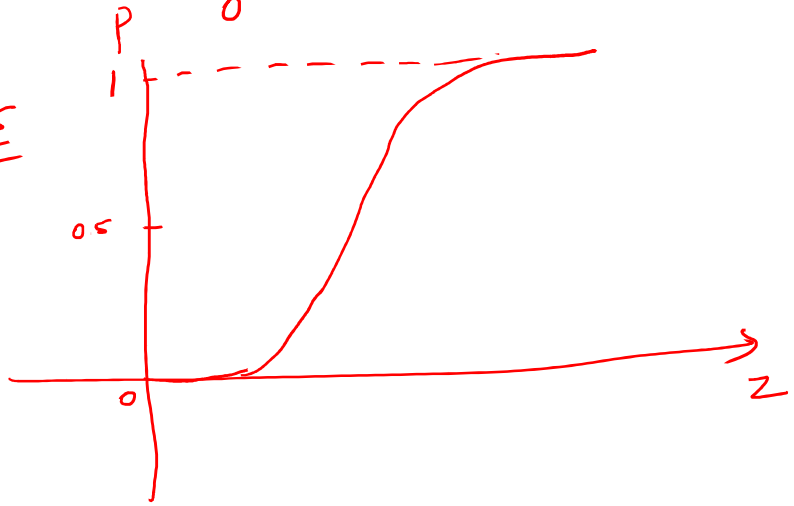
Gain Gotting

Sigmoid function

$$P(\text{class}=1) = \frac{1}{1+e^{-z}}$$

$$Z = \beta_0 + \beta_1 x$$





→ -5

① → $y < 0$, $y > 1$

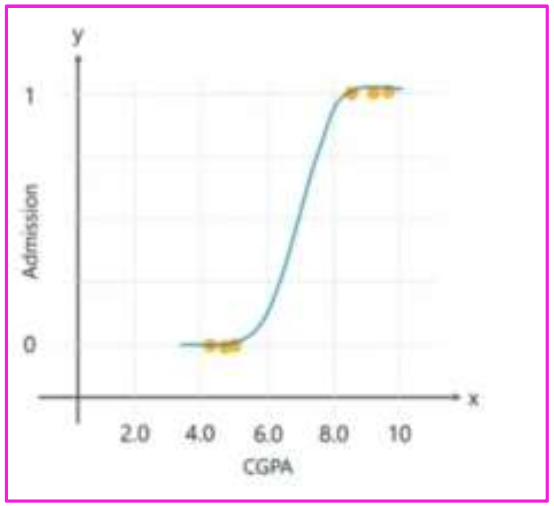
② Out

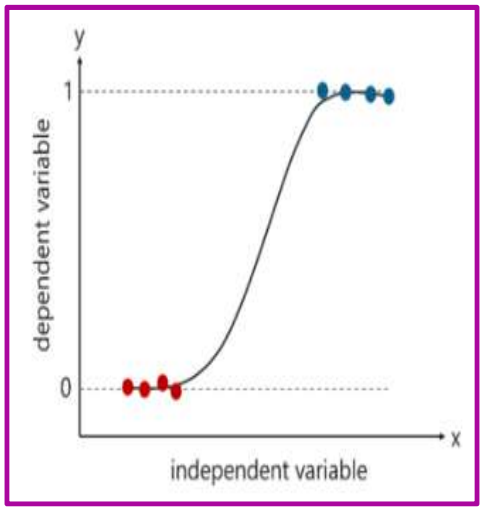
$P \geq 0.5$ Class = 1

$P < 0.5$ Class 0

Logistic regression

- It is a method used to predict a dependent variable, given a set of independent variables, such that the dependent variable is categorical.
- It is a classification algorithm that is used to estimate the discrete values like 0 or 1, true or false, yes or no based on a given set of independent variables.
- it predicts the probability hence its output lies in between 0 and 1





Logistic regression

- The linear regression equation: $Y = \beta 0 + \beta 1x + \epsilon$
- Represent a relationship between p(X)=Pr (Y=1/X) and X?
- Take the exponent of the equation, since the exponential of any value is a positive number.
- Secondly, a number divided by itself + 1 will always be less than 1.

$$P(X) = \frac{e^{(\beta 0+\beta 1x)}}{1+ e^{(\beta 0+\beta_1)}}$$

$$P(X) = \frac{e^{(\beta_0+\beta_1 x)}}{e^{(\beta_0+\beta x)} + 1}$$

$$\Rightarrow p(e^{(\beta_0+\beta x)}+1) = e^{(\beta_0+\beta_1 x)}$$

$$\Rightarrow p.e^{(\beta_0+\beta x)} + p = e^{(\beta_0+\beta_1 x)}$$

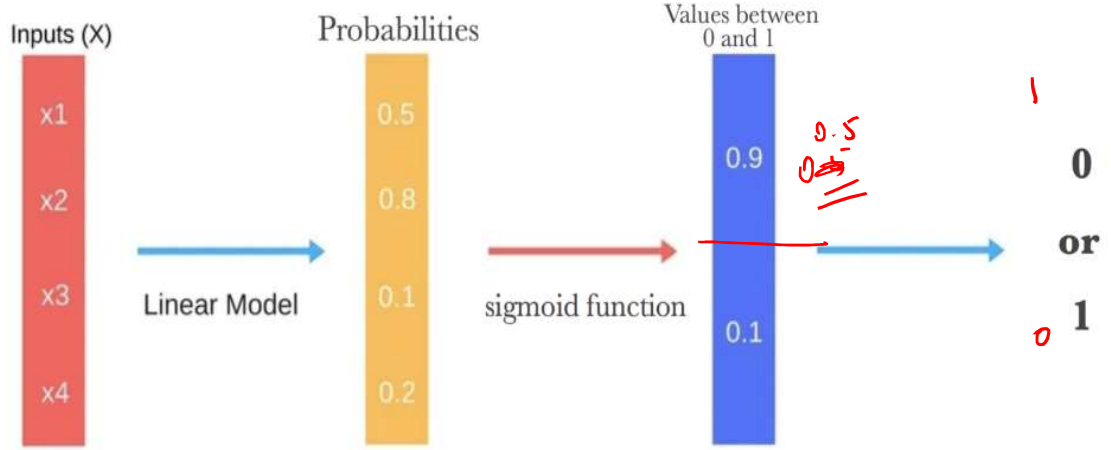
$$\Rightarrow p = e^{(\beta_0+\beta_1 x)} - p.e^{(\beta_0+\beta x)}$$

$$\Rightarrow p = e^{(\beta_0+\beta_1 x)}(1-p)$$

$$\Rightarrow \frac{p}{(1-p)} = e^{(\beta_0+\beta_1 x)}$$

$$\Rightarrow \ln[\frac{p}{(1-p)}] = (\beta_0 + \beta_1 x)$$

logit function



	Linear	Logistic Reg
Def	Continuous data	Binary, categorical var
Eqn	$y = \beta_0 + \beta_1 x + e$	$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x + e$
Best fit line	Straight line	Sigmoidal curve
Relationship Between dv & I-v	Linear	Linear is mandatory
o/p	Integer/continuous var	Binary var (0, 1)
App	Stock price, House price etc	Classification purpose

Decision Tree

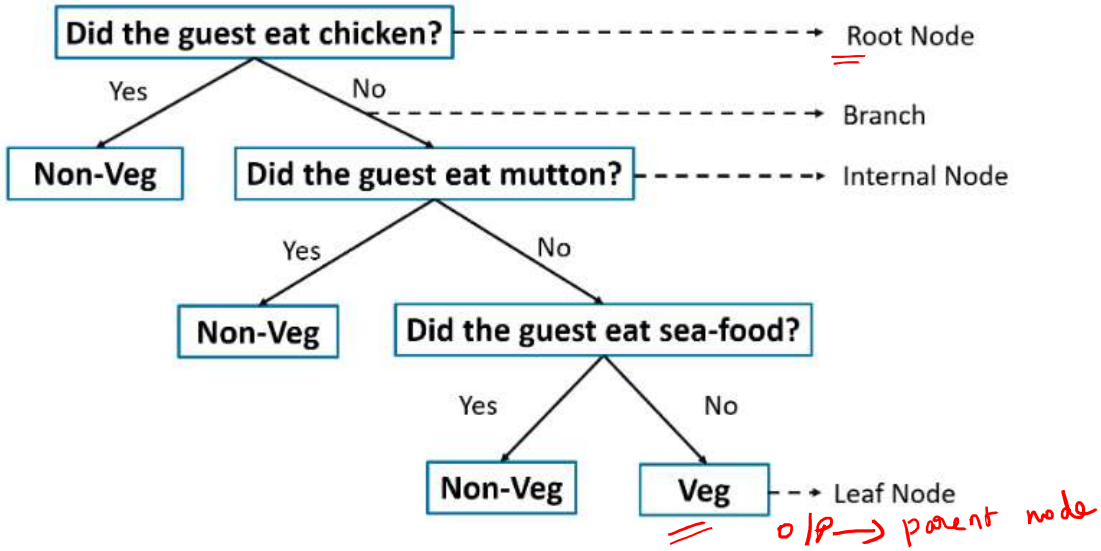
- It is a supervised machine learning algorithm which looks like an inverted tree, wherein each node represents a predictor variable (feature), the link between the nodes represent a decision and each leaf node represents an outcome (response variable).
- reasons for choosing this algorithm are
  - It is considered to be the most understandable Machine Learning algorithm and it can be easily interpreted.
  - It can be used for classification and regression problems.
  - Unlike most Machine Learning algorithms, it works effectively with non-linear data.
  - Constructing a Decision Tree is a very quick process since it uses only one feature per node to split the data.

Decision Tree

- It has the following structure.
- **Root Node:** The root node is the starting point of a tree. At this point, the first split is performed.
- **Internal Nodes:** Each internal node represents a decision point (predictor variable) that eventually leads to the prediction of the outcome.
- **Leaf/ Terminal Nodes:** Leaf nodes represent the final class of the outcome and therefore they're also called terminating nodes.
- **Branches:** Branches are connections between nodes, they're represented as arrows. Each branch represents a response such as yes or no.

more information gain

Guest	Chicken	Mutton	Fish	Type
1	Y	Y	Y	Y
2	N	N	Y	Y
3				



o/p → parent node

Decision Tree

ID3 or the Iterative Dichotomiser 3 algorithm is one of the most effective algorithms used to build a Decision Tree. It uses the concept of *Entropy* and *Information Gain* to generate a Decision Tree for a given set of data.

ID3 algorithm follows the below workflow in order to build a Decision Tree

Step 1: Select Best attribute (A)

Step 2: Assign A as a decision variable for the root node

Step 3: For each value of A, build a descendant of the node

Step 4: Assign classification labels to the leaf node

Step 5: If data is correctly classified: Stop.

Step 6: Else: Iterate over the tree

Decision Tree –Information Gain and Entropy

Problem statement: to study the dataset and create a Decision tree that classifies the speed of a car either slow or fast

Road type	Obstruction	Speed limit	Speed
steep	yes	yes	slow
steep	no	yes	slow
flat	yes	no	fast
steep	no	no	fast

Step1: Select Best Attribute "A"

"How do you know which variable best separates the data?

The variable which has highest information gain best divides the data into desired output classes

- Two measures are used to decide the best attribute.
  - Entropy measures the impurity or uncertainty present in the data.
  - Information gain indicates how much 'information' a particular feature/variable gives us about the final outcome

Decision Tree –Information Gain and Entropy

- $Entropy = \sum -p(x) log p(x)$
- $Information\ gain = entropy(parent) - [weighted\ average] * entropy(children)$
- **Calculating IG of parent node (Speed of car)**

find out the fraction of the two classes (slow & fast) present in the parent node

- P(Slow) – fraction of 'slow' outcomes in the parent node
- P(Fast) – fraction of 'fast' outcome in the present node

The formula to calculate P(Slow) is:

P(Slow) = No. of outcomes in the parent node / total number of outcomes

P(Slow) = 2 / 4 = 0.5

P(Fast) = 2 / 4 = 0.5

Therefore, the entropy of the parent node is,

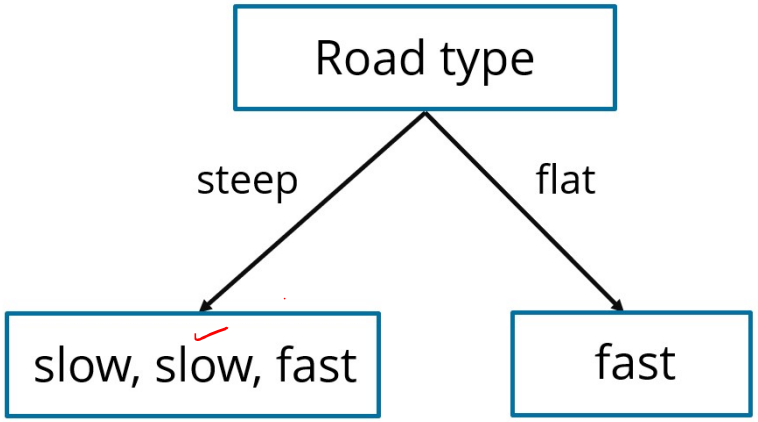
$Entropy_{parent} = -(P_{slow} \log_2 (P_{slow}) + P_{fast} \log_2 (P_{fast}))$

Entropy(parent) = - [(0.5 log2(0.5) + 0.5 log2(0.5)]

= -(-.05+-0.5)=1

Decision Tree –Information Gain and Entropy

- Calculate the Information Gain of 'Road type' variable



The entropy of the right-side child node (fast) = 0

The entropy of the left side child node (slow, slow, fast) is,

P(slow) = 2 / 3 = 0.667

P(fast) = 1 / 3 = 0.334

Therefore the entropy is:

Entropy (left child node) = -[(0.667 log2(0.667) + 0.334log2(0.334)] = 0.9

=-(-0.38-0.52)=0.9

Decision Tree –Information Gain and Entropy

- calculate the Entropy(children) with weighted average:
- Total number of outcomes in parent node: 4
- Total number of outcomes in left child node: 3
- Total number of outcomes in right child node: 1

- The formula for Entropy(children) with weighted average:
- **[Weighted avg] Entropy(children) = (no. of outcomes in left child node) / (total no. of outcomes in parent node) * (entropy of left node) + (no. of outcomes in right child node)/ (total no. of outcomes in parent node) * (entropy of right node)**

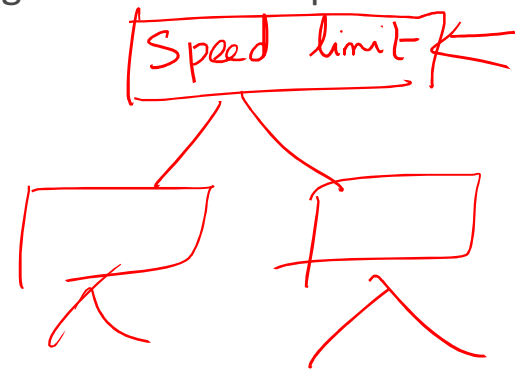
- Entropy (children)with weighted average is =0.675 (Road type variable)

Decision Tree –Information Gain and Entropy

- IG of the 'Road type' variable:
- $\boldsymbol{Information\ gain = entropy(parent) - [weighted\ average] * entropy(children)}$

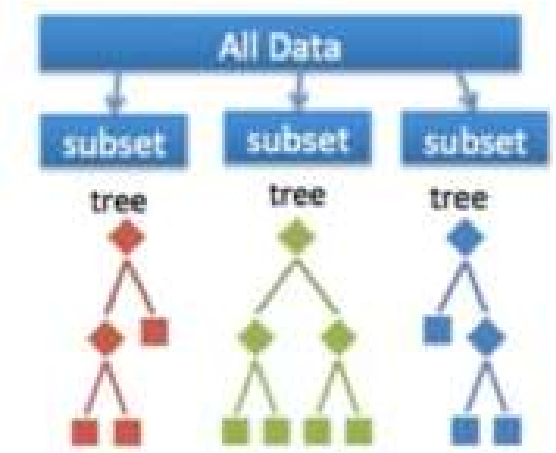
- the final IG of the 'Road type' variable:
- *Information gain (Road type) = 1 – 0.675 = 0.325*

- So, by using the above methodology, we can get the following values for each predictor variable:
- Information gain (Road type) = 1 – 0.675 = 0.325
- Information gain (Obstruction) = 1 – 1 = 0
- Information gain (Speed limit) = 1 – 0 = 1



Random Forest

- Random forest algorithm is a supervised classification and regression algorithm
- it builds multiple decision trees (called the forest) and glues them together to get a more accurate and stable prediction
- Advantages of Random Forest algorithm:
- Random forest classifier can be used for both classification and regression tasks
- They can handle the missing values and hence avoids the bias in the results
- It won't overfit the model even if we have a greater number of trees in the forest



Random Forest

• Use the following data set that predicts if a person has a heart disease or not

Blood Flow	Blocked Arteries	Chest Pain	Weight	Heart Disease
Abnormal	No	No	130	No
Normal	Yes	Yes	195	Yes
Normal	No	Yes	218	No
Abnormal	Yes	Yes	180	Yes

Step 1: Creating a bootstrapping data set

Bootstrapping is an estimation method used to make predictions on a data set by resampling it

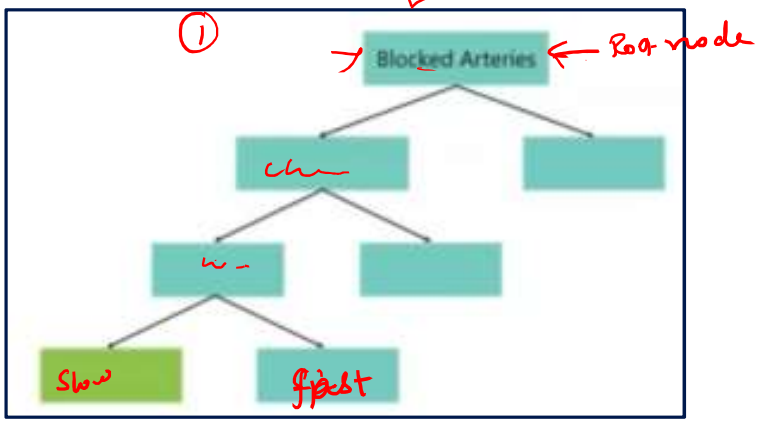
Blood Flow	Blocked Arteries	Chest Pain	Weight	Heart Disease
Normal	Yes	Yes	195	Yes
Abnormal	No	No	130	No
Abnormal	Yes	Yes	180	Yes
Abnormal	Yes	Yes	180	Yes

Random Forest

Step 2: Creating decision trees

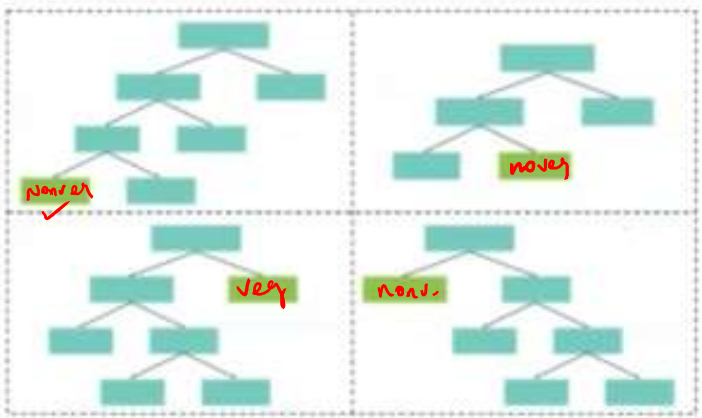
- Build a decision tree by using the bootstrapped data set
- Begin the root node and select the best attribute to split the data set
- Repeat the same process for each of the upcoming branch nodes

Blood Flow	Blocked Arteries	Chest Pain	Weight	Heart Disease
Normal	Yes	Yes	195	Yes
Abnormal	No	No	130	No
Abnormal	Yes	Yes	180	Yes
Abnormal	Yes	Yes	180	Yes



Random Forest

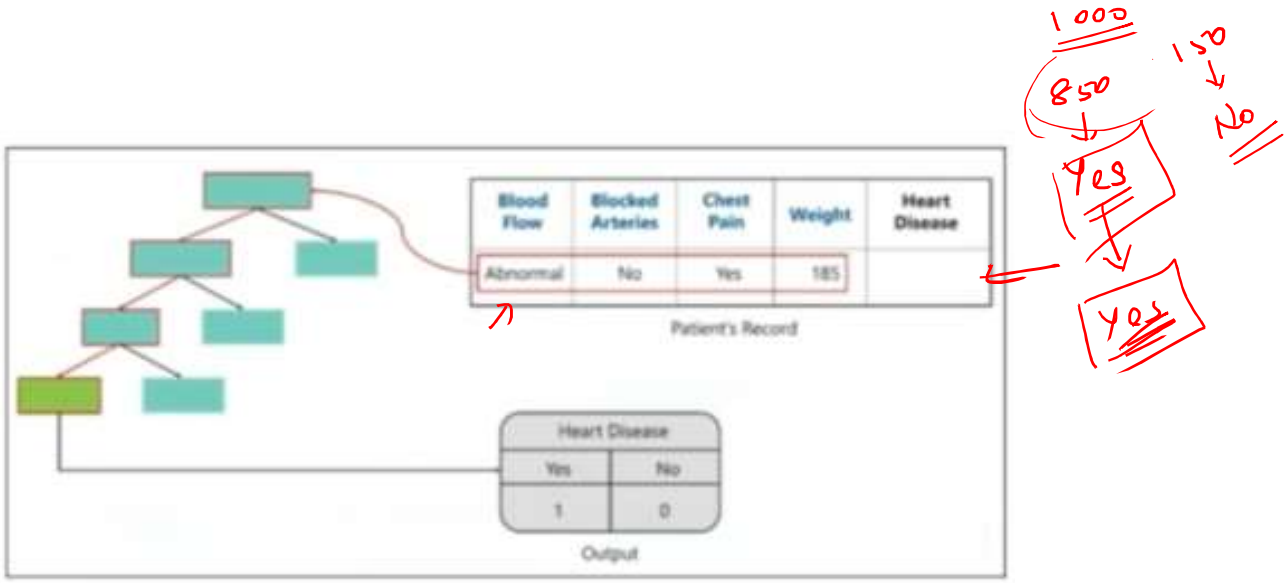
- Step 3: go back to step 1 and repeat
  - Each decision tree predicts the output class based on the respective predictor variables used in tree
  - Go back to step 1,create a new bootstrapped data set and then build a decision tree by considering only sub set of variables at each step
  - The iteration is performed 100's of time, creating multiple decision trees



Random Forest

➢Step 4:Predicting the outcome of a new data point

- To predict whether the new patient has heart disease or not, run the new data down the decision trees
- After running the data down all the trees in the random forest, check which class got the majority votes
- In our case the class “Yes” received the most number of votes, hence it’s clear that the new patient has heart disease



Random Forest

➢Step 5: Evaluate the model

- In real world problem about 1/3 rd of the original data set is not included in the original data set
- This sample data set that does not include in the bootstrapped data set is known as the out of bag (OOB)data set
- We can measure the accuracy of a random forest by the proportion of OOB samples that are correctly classified

Blood Flow	Blocked Arteries	Chest Pain	Weight	Heart Disease
Normal	No	Yes	218	No

Naive Bayes classifier → No dependency between predictor variable

- It is based on the Bayes theorem that is used to solve classification problems by following a probabilistic approach
- It is based on the idea that the predictor variables in a machine learning model are independent of each other A, B. events

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

- P(A/B) conditional probability of event A occurring, given the event B
- P(A) probability of event A occurring
- P(B) probability of event A occurring
- P(B/A) conditional probability of event B occurring, given the event A

$P(C|x) = \frac{P(x|C)\ P(C)}{P(x)}$

C → class

x → Features → Input Variables / predictor Variables

x_1 x_2 x_3 x_4

Feature: F_1 F_2 F_3 F_4

o/p: Y / N

$P(C|x_1, x_2, x_3, x_4) = \frac{P(x_1|C) * P(x_2|C) * P(x_3|C) \dots P(x_n|C) * P(C)}{P(x_1)\ P(x_2)\ P(x_3) \dots P(x_n)}$

normal spam, dear, friend

$P(C|x_1, x_2, x_3 \dots x_n) = P(x_1|C) * P(x_2|C) * \dots P(x_n|C) * P(C)$

C - Y = 5/12

C - N = 7/12

Naive Bayes classifier

Type	Swim	Wings	Green	Sharp teeth
Cat	450 / 500	0	0	500 / 500
Parrot	50 / 500	500 / 500	400 / 500	0
Turtle	500 / 500	0	100 / 500	50 / 500

From the above table it can be summarized that:

- 90% cats can swim, no cats have wings, no cats are of green color, all cats have sharp teeth.
- 10% parrots have a true value for swim, all parrots have wings, 80% parrots are green in color, no parrots have sharp teeth.
- All turtles can swim, no turtles have wings, 20% turtles are green in color, 10% turtles have sharp teeth.

Naive Bayes classifier

- Classify the following observation into one of the output classes (Cat, Parrot or Turtle) by using the Naïve Bayes classifier

	Swim	Wings	Green	Sharp teeth
Observation	True	False	True	False

- To solve this we will use the Naïve Bayes approach:
- P(H/multiple evidences)={P(C1/H)*P(C2/H), …….*P(Cn/H)*P(H)}/P(Multiple Evidences)
- Swim and Green are true and the outcomes can be any one of the animals (cat, Parrot, Turtle)

Naive Bayes classifier

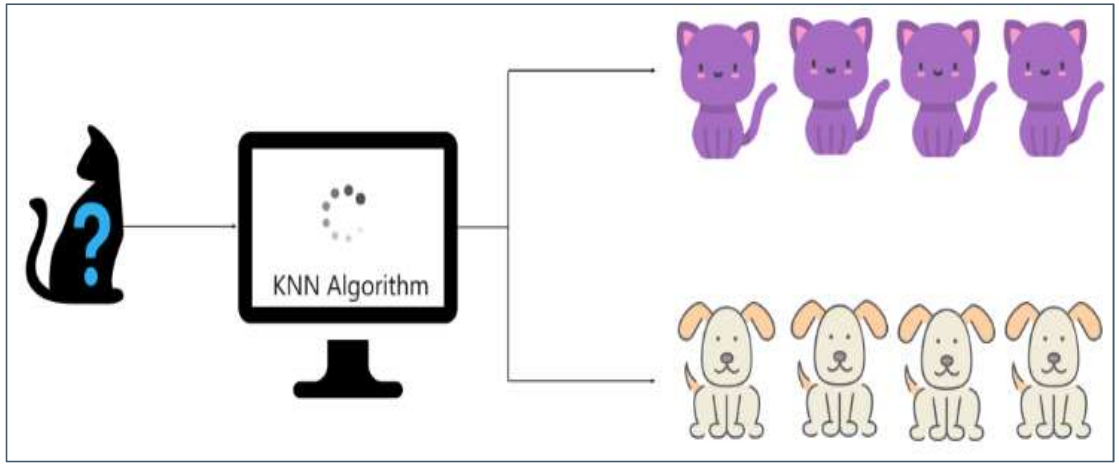
- **To check if the animal is a cat**
- P(cat/swim,green)=p(swim/cat)*P(Green/Cat)*P(Cat)/P(swim, green)=0.9*0*0.333/ P(swim, green)=0

- **To check if the animal is a parrot**
- P(parrot/swim,green)=p(swim/parrot)*P(Green/parrot)*P(parrot)/P(swim,green)=0.1*0.8*0.333/ P(swim, green)=0.0264/ P(swim, green)

- **To check if the animal is a Turtle**
- P(Turtle/swim,green)=p(swim/Turtle)*P(Green/Turtle)*P(Turtle)/P(swim,green)=1*0.2*0.333/P(swim, green)=0.0666/ P(swim, green)

K Nearest Neighbor (KNN)

- It is a supervised learning algorithm that classifies a new data point into the target class, depending on the features of its neighboring data points.
- It is widely used to solve classification problems.
- The main concept of this algorithm is that it is used to store all the available cases and classifies new cases by majority votes of its k neighbors.
- The case being then assigned to the class which is the most common amongst its K-nearest neighbors, measured by a distance function.



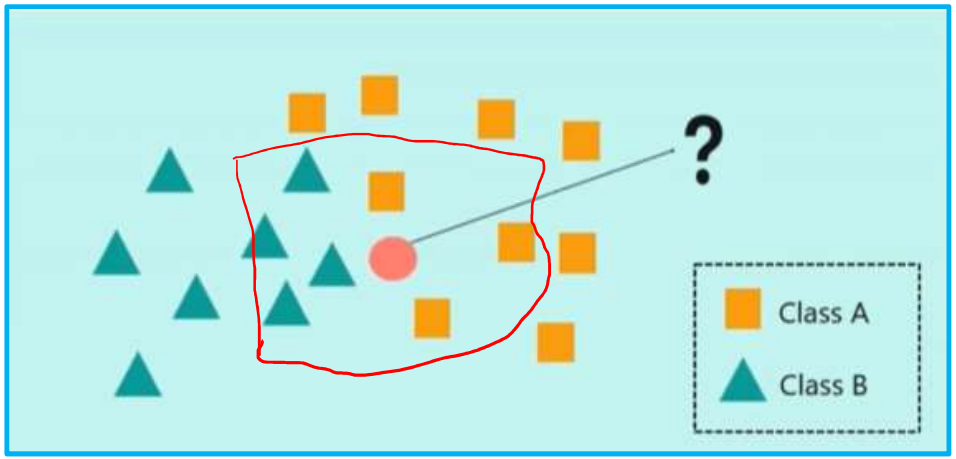
K Nearest Neighbor (KNN)

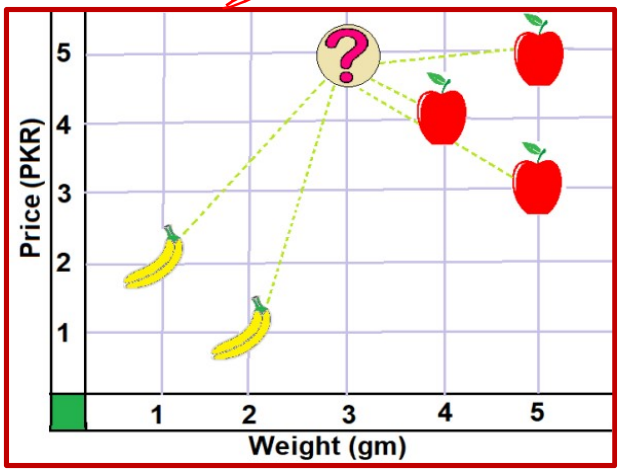
1) Find out the distance between the given data pt, all other points

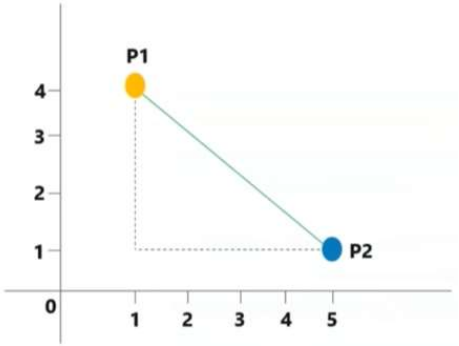
2) K = 3, 5, 7 . . . . .

K = 3 O/P class=A K = 7 O/P Class B

- Problem statement: Assign the new data point into one of the clusters







Calculations

Point P1 = (1,1) x_1, y_1

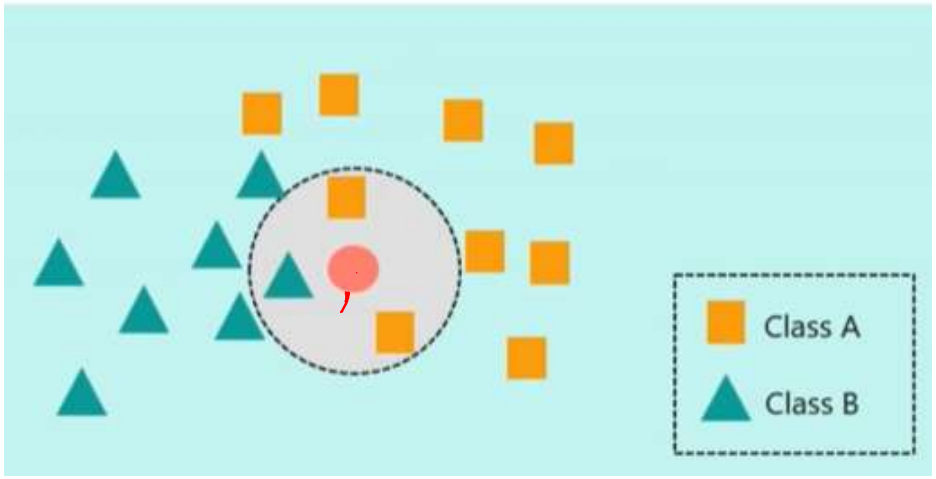
Point P2 = (5,4) x_2, y_2

$\sqrt{(x_1-x_2)^2+(y_1-y_2)^2}$

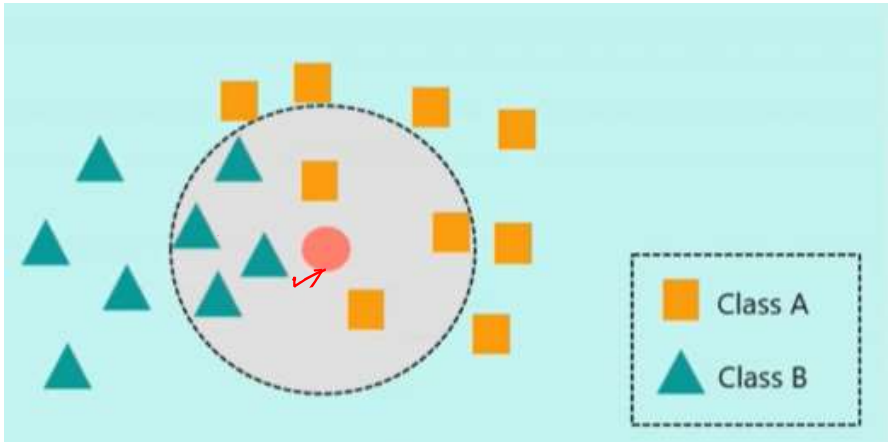
Euclidian distance = $\sqrt{(5-1)^2+(4-1)^2} = 5$

K Nearest Neighbor (KNN)

K=3

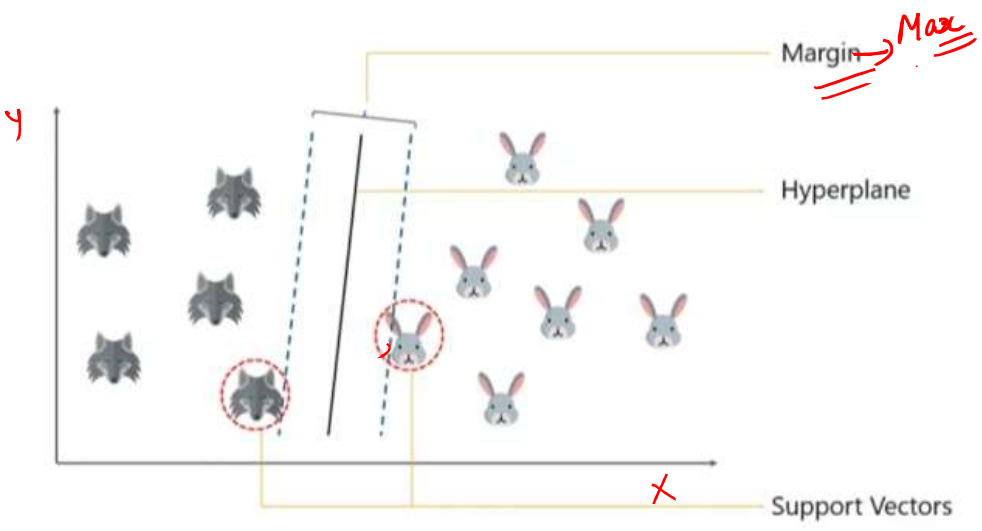


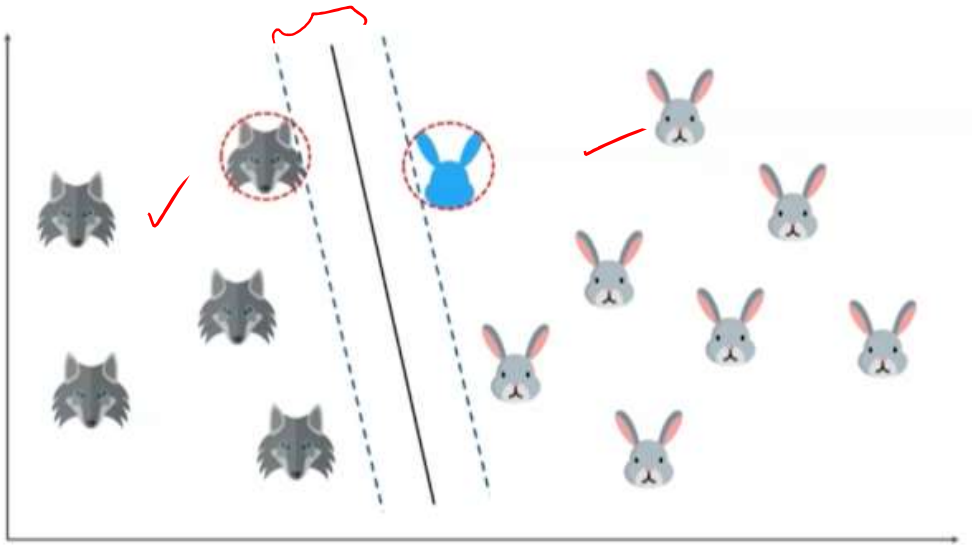
K=7



Support Vector Machines (SVM)

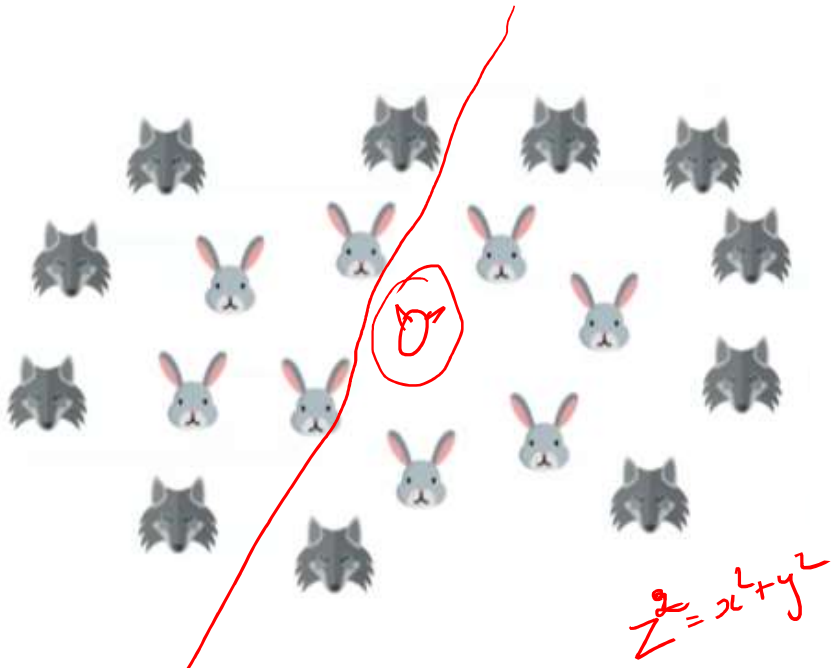
- It is a supervised classification method that separates data using hyperplanes.
- It is one of the simplest algorithms in supervised learning.
- It is used for both classification and regression problems.
- It can be used to generate multiple separating hyperplanes such that the data is divided into segments and each segment contains only one kind of data.

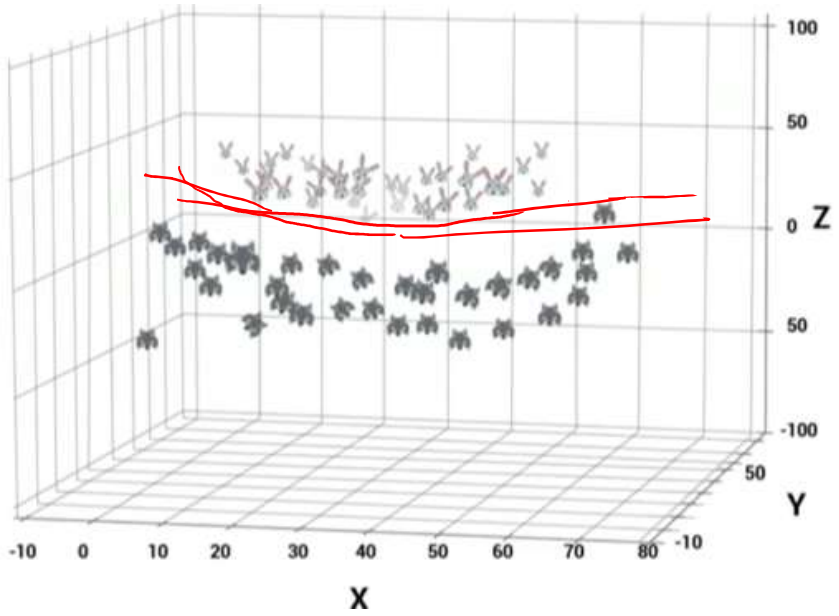




Non-linear SVM

- dataset cannot be classified by using a straight line if the data is non linear

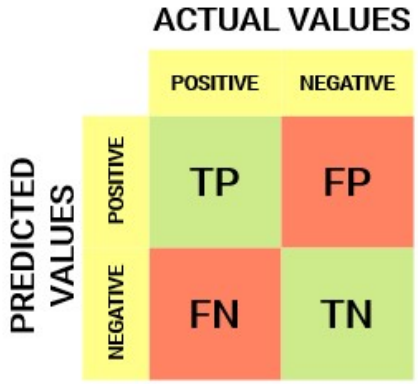




- Kernel function transforms the non linear data into linear data
- Data is drawn in 3D instead 2D which helps to classify the data

Confusion Matrix

- A Confusion matrix is an N x N matrix used for evaluating the performance of a classification model
- N is the number of target classes.
- The matrix compares the actual target values with those predicted by the machine learning model.
- This gives us a holistic view of how well our classification model is performing and what kinds of errors it is making.

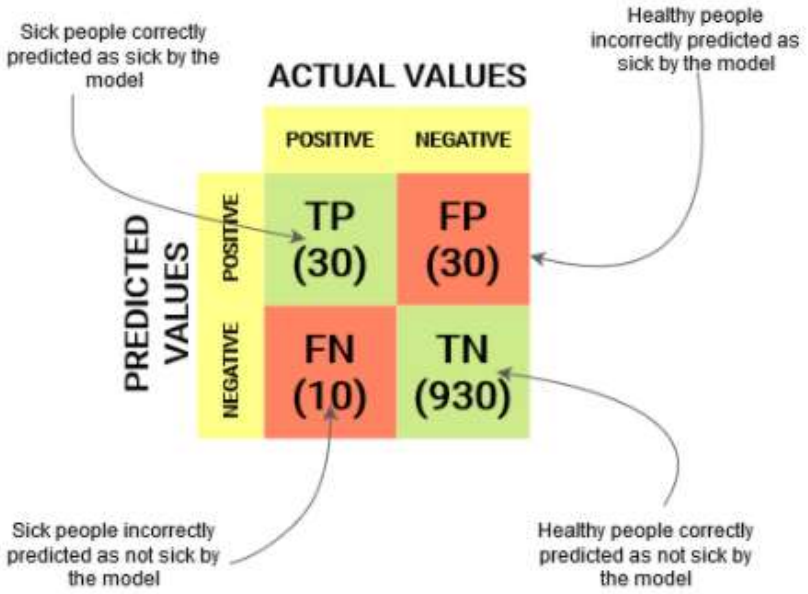


- The target variable has two values: **Positive** or **Negative**
- The **columns** represent the **actual values** of the target variable
- The **rows** represent the **predicted values** of the target variable

Confusion Matrix

- **True Positive (TP)**
  - The predicted value matches the actual value
  - The actual value was positive and the model predicted a positive value
- **True Negative (TN)**
  - The predicted value matches the actual value
  - The actual value was negative and the model predicted a negative value
- **False Positive (FP) – Type 1 error**
  - The predicted value was falsely predicted
  - The actual value was negative but the model predicted a positive value
  - Also known as the **Type 1 error**
- **False Negative (FN) – Type 2 error**
  - The predicted value was falsely predicted
  - The actual value was positive but the model predicted a negative value
  - Also known as the **Type 2 error**

Confusion Matrix



- The different values of the Confusion matrix would be as follows:
- True Positive (TP) = 30
- True Negative (TN) = 930
- False Positive (FP) = 30
- False Negative (FN) = 10

Confusion Matrix

ID	Actual Sick?	Predicted Sick?	Outcome
1	1	1	TP
2	0	0	TN
3	0	0	TN
4	1	1	TP
5	0	0	TN
6	0	0	TN

- The total outcome values are:
- TP = 30, TN = 930, FP = 30, FN = 10

1) $$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

$$Accuracy = \frac{30 + 930}{30 + 30 + 930 + 10} = 0.96$$

5) *Specificity is the correctly -ve labeled by the program to all who are healthy in reality.*

$$Specificity = TN/(TN+FP)$$

2) *Precision tells us how many of the correctly predicted cases actually turned out to be positive.*

$$Precision = \frac{TP}{TP + FP}$$

3) *Recall tells us how many of the actual positive cases we were able to predict correctly with our model.*

$$Recall = \frac{TP}{TP + FN}$$

$$Precision = \frac{30}{30 + 30} = 0.5$$

$$Recall = \frac{30}{30 + 10} = 0.75$$

4) *The F1-score captures both the trends in a single value:*

$$F1 - score = \frac{2}{\frac{1}{Recall} + \frac{1}{Precision}}$$

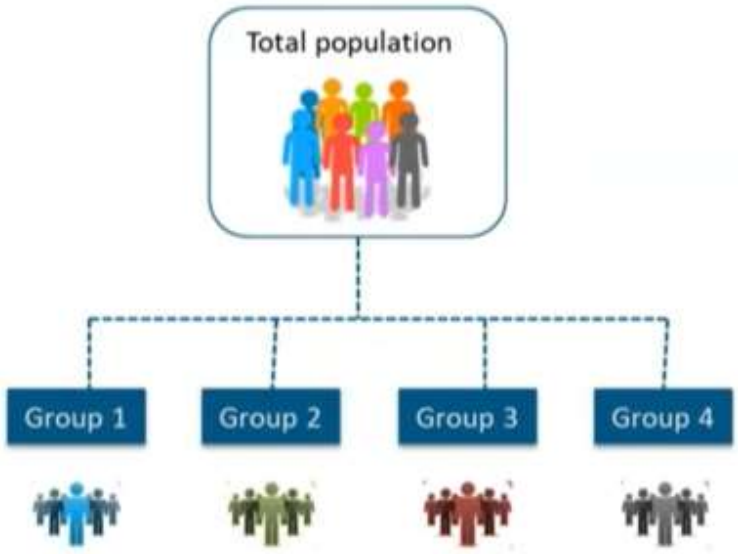
It is maximum when Precision is equal to Recall

Unsupervised learning algorithms

K-means clustering

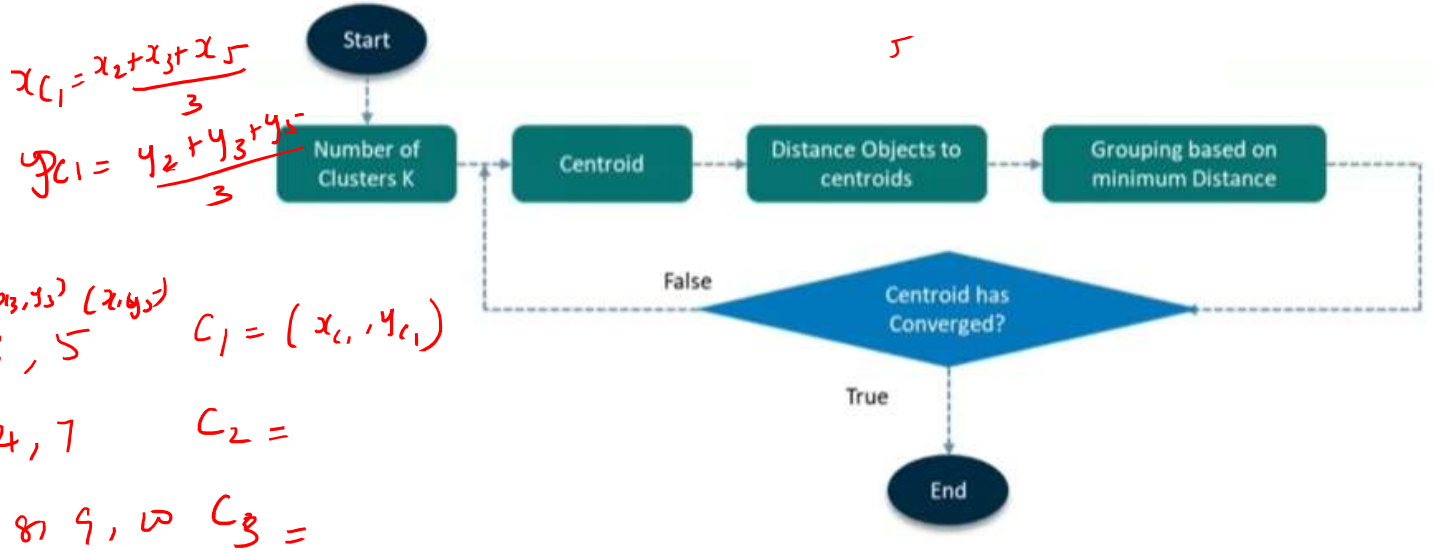
The process by which objects are classified into a pre-defined number of groups so that they are as much dissimilar as possible from one group to another group, but as much similar as possible within each group

- It is basically a type of unsupervised learning. The main logic of K-Means clustering algorithm is to classify the data set through a number of clusters.
- It requires following two inputs.
  - k = number of clusters
  - Training set (m) = {x1, x2, x3, ……., xm}

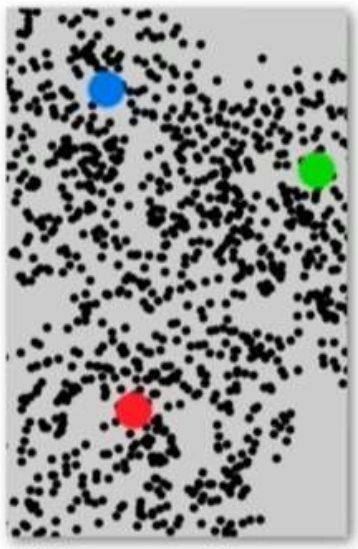


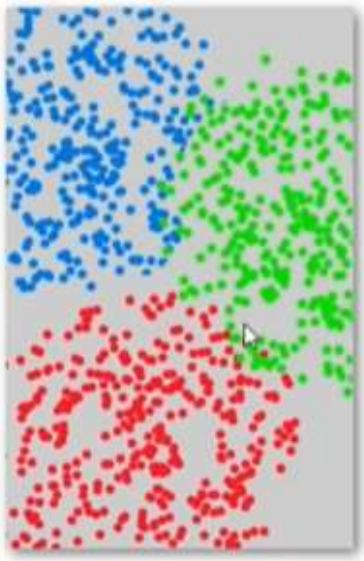
K-means clustering

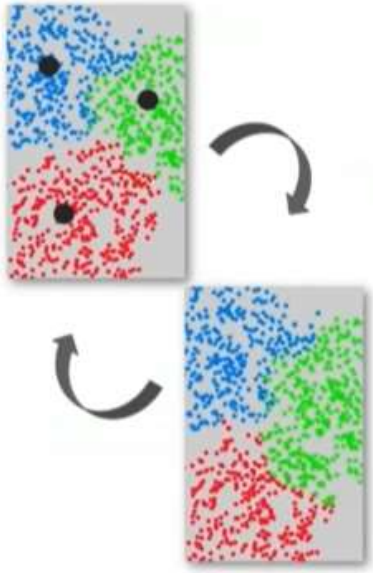
- form clusters by K-means:
  - K-means picks k number of points for each cluster known as centroids.
  - Now each data point forms a cluster with the closest centroids, i.e., k clusters.
  - Now, it will find the centroids of each cluster based on the existing cluster members.
  - We need to repeat these steps until convergence occurs.
- The execution of this algorithm involves the following steps
  - Initialization
  - Cluster assignment
  - Move centroid
  - Optimization
  - Convergence



K-means clustering







$k = 3$

$k = 1, \ldots, 8$

Group $= \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$

	$A_1(2,10)$ C_1	$A_4(5,8)$ C_2	$A_7(1,2)$ C_3	Group
$A_1 = (2,10)$	0	3.6	8.06	C_1
$A_2 = (2,5)$	5	4.24	3.16	C_3
$A_3 = (8,4)$	8.48	4.12	7.2	C_2
$A_4 = (5,8)$	3.6	0	7.2	C_2
$A_5 = (7,5)$	7.07	3.6	6.7	C_2
$A_6 = (6,4)$	7.2	4.12	5.38	C_2
$A_7 = (1,2)$	8.06	7.1	0	C_3
$A_8 = (4,9)$	2.23	1.41	7.6	C_2

New centroid

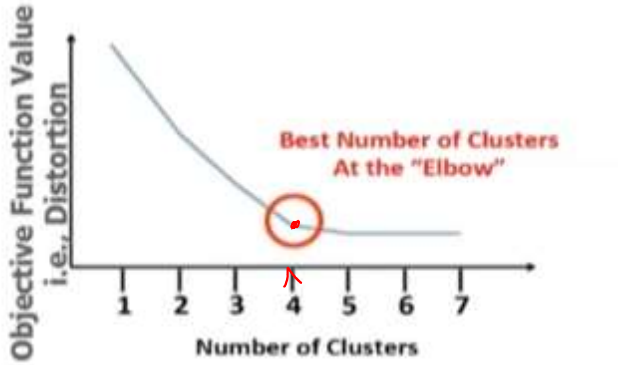
Cluster 1 $\rightarrow A_1$ $\rightarrow C_1 \rightarrow (2,10)$ $C_1 \rightarrow (2,10)$

Cluster 2 $\rightarrow A_3, A_4, A_5, A_6, A_8 \rightarrow C_2 \rightarrow (6,6)$ $C_2 (6,6)$

Cluster 3 $\rightarrow A_2, A_7$ $\rightarrow C_3 \rightarrow (1.5, 3.5)$ $C_3 = (1.5, 3.5)$

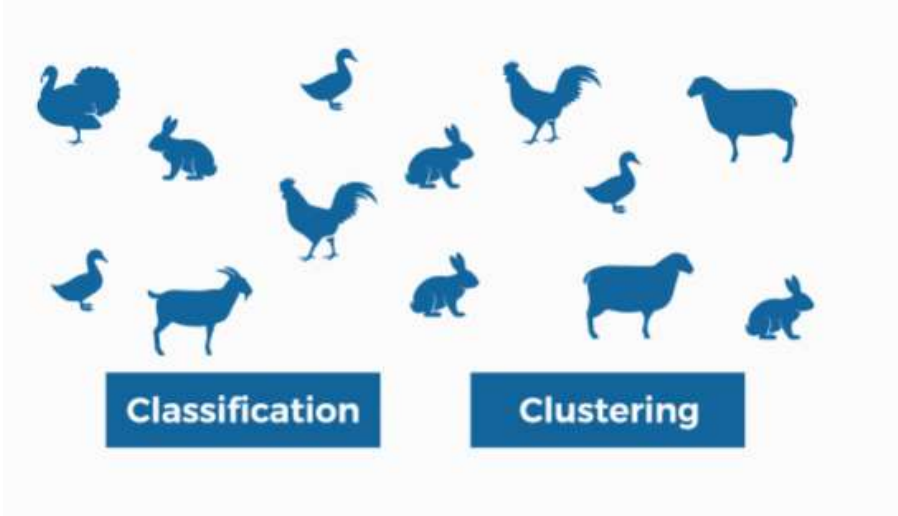
K-means clustering

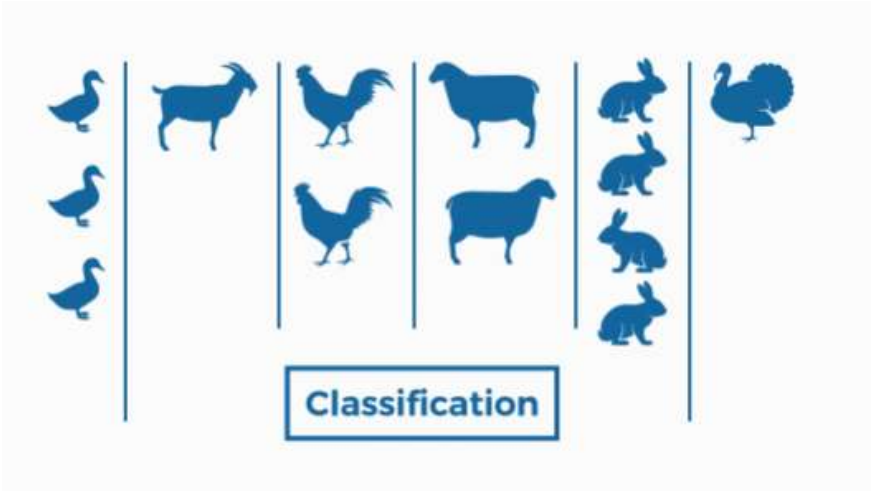
- To find out the optimal Number of clusters
- Elbow method:
  - compute the sum of squared error(SSE)for some values of K(ex 2,4,6,…)
  - The SSE is defined as the sum squared distance between each member of the cluster and its centroid

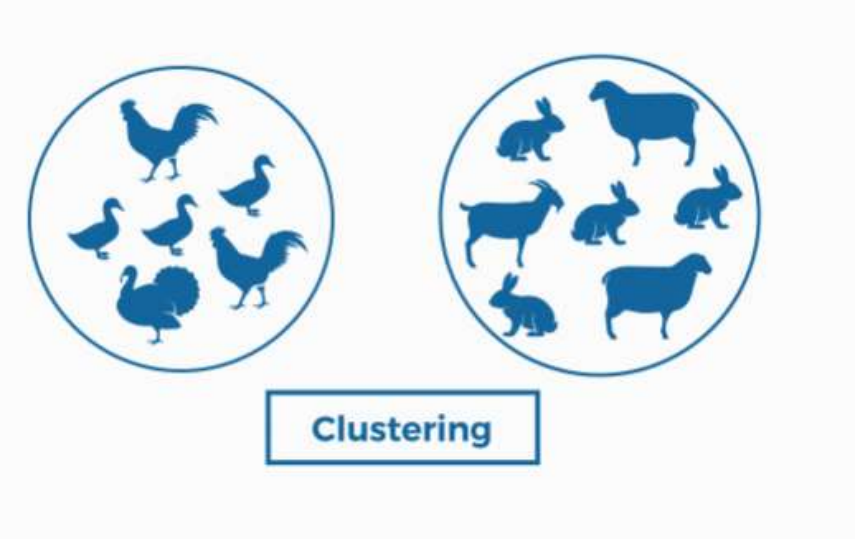


$$SSE = \sum_{i=1}^{K} \sum_{x \in c_i} dist(x, c_i)^2$$

Classification Vs Clustering

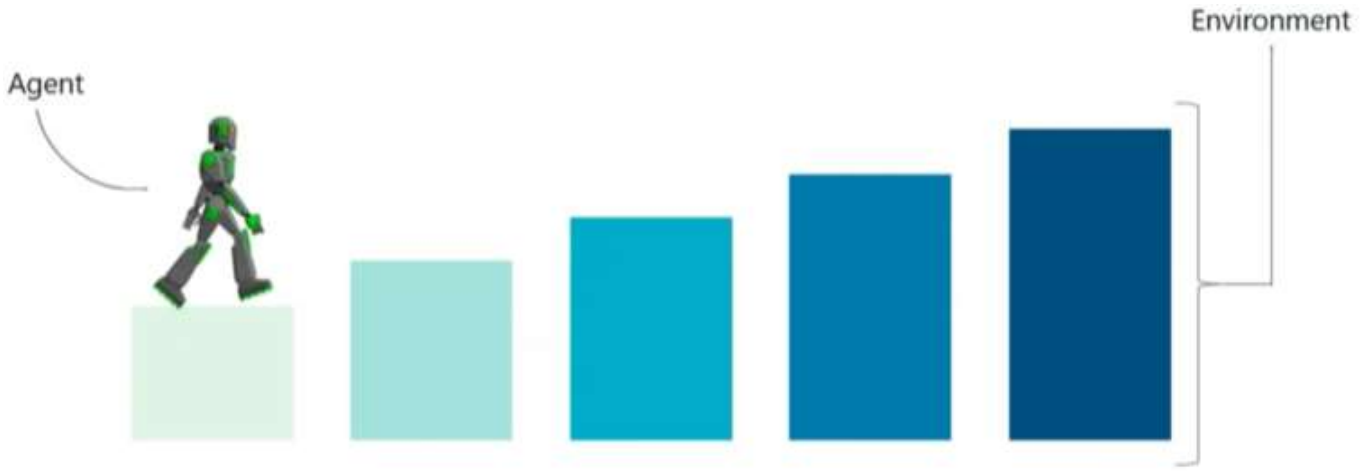






Reinforcement learning algorithm

- Reinforcement learning system is comprised of two main components
  - Agent
  - Environment

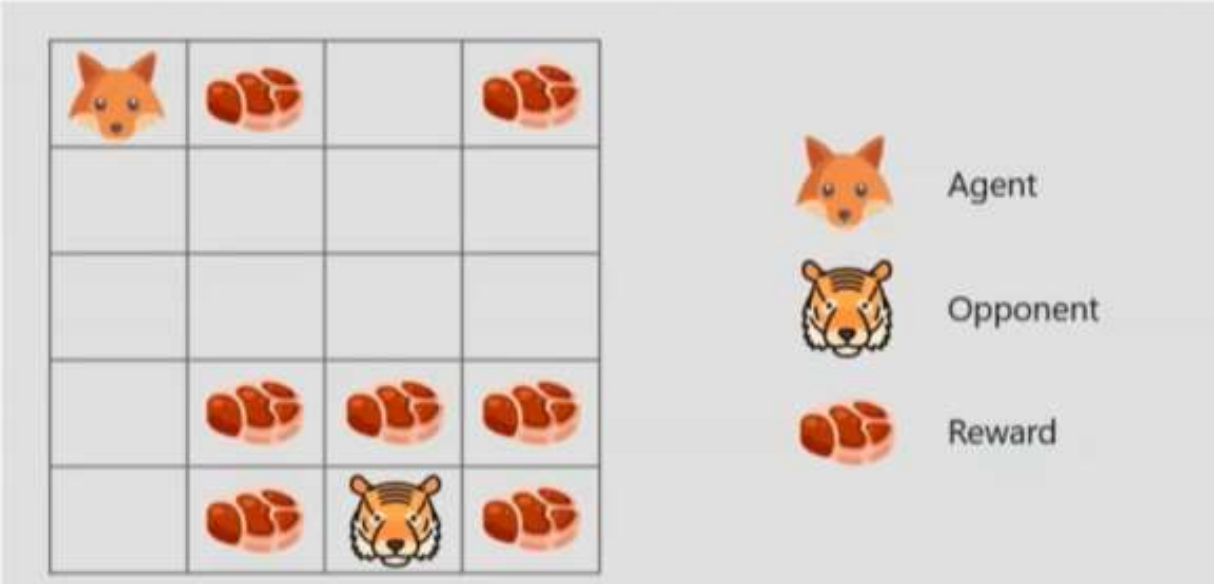


Terminologies in RL

- Agent: the algorithm that learn from trail and error
- Environment: the world through which the agent moves
- Action(A):All the possible actions that the agent can take
- State (S):current condition returned by the environment
- Reward(R): An instant return from the environment to appraise the last action
- Policy(π): the approach that agent uses to determine the next action based on the current status
- Value (V):the expected long term return with discount as opposed to the short term reward R
- Action Value(Q):This similar to value, except, it takes an extra parameter, the current action (A)

Reward Maximization theory

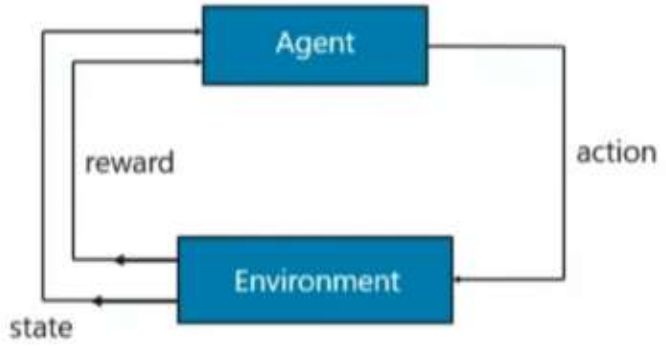
- It states that a RL agent must be trained in such a way that he takes the best action so that the reward is maximum
- Exploitation: using the already known exploited information to increase the reward
- Exploration: Exploring and capturing more information about an environment



Markov's Decision process

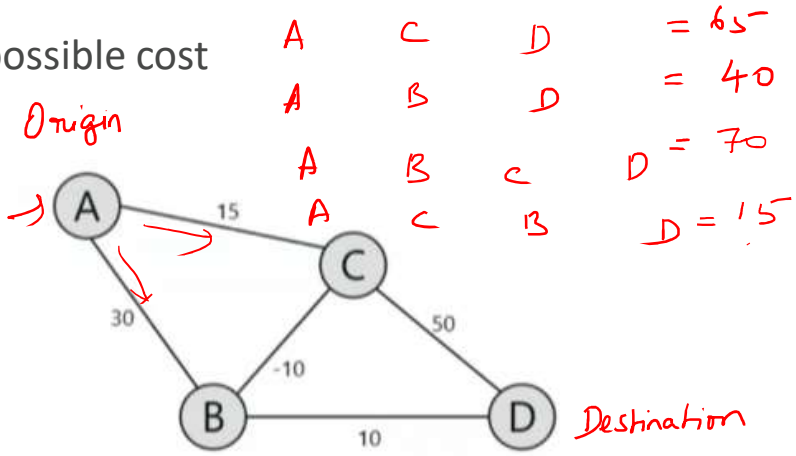
- The mathematical approach for mapping a solution in reinforcement learning is called Markov's decision process

- The following parameters are used to attain the solution
- Set of Actions (A)
- set of states (S)
- Reward (R)
- Policy(π)
- Value (V)



Markov's Decision process

- Goal: Find the shortest path between A and D with minimum possible cost



- In this problem
  - Set of states are denoted by nodes ie (A,B,C,D)
  - Action is to traverse from one node to another (A->B,C->D)
  - Reward is the cost represented by each edge
  - Policy is the path taken to reach the destination (A->B->D)
  - This is known as policy based learning

- Other learning process
- Value based learning-Reward maximization
- Action based learning-emphasize is given on the actions taken

Q-learning

- It is a model-free reinforcement learning algorithm to learn a policy telling an agent what action to take under what circumstances.
- It does not require a model of the environment, and it can handle problems with stochastic transitions and rewards, without requiring adaptations.
- It involves an agent, a set of state S, and a set of actions A per state.
- Executing an action in a specific state provides the agent with a reward (a numerical score). Before learning begins, Q matrix is initialized using the formula,
- Q (state, action) = R (state, action) + Gamma * Max [Q (next state, all actions)]

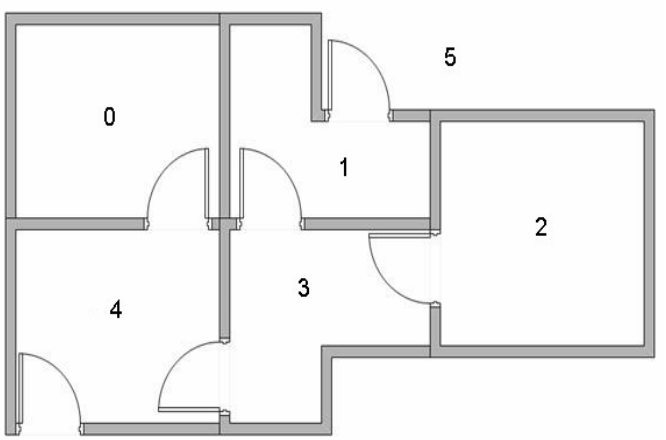
- The Gamma parameter is a discount factor has a range of 0 to 1. If Gamma closer to zero, the agent will tend to consider only immediate rewards. If Gamma is closer to one, the agent will consider future rewards with greater weight.

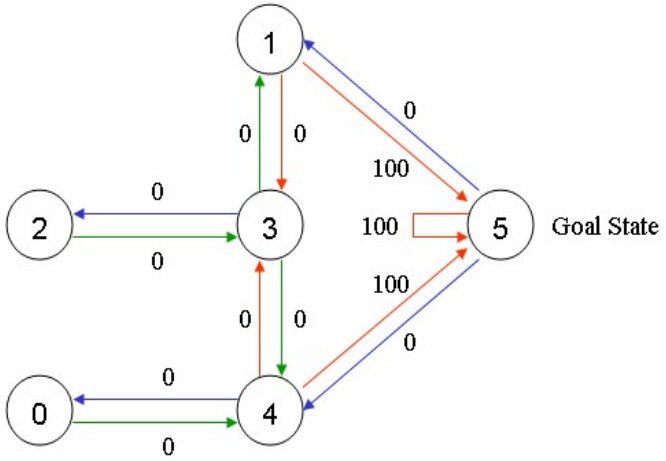
Q-learning

- The steps involved in the algorithm are
  - Set the gamma parameter and environment rewards in matrix
  - Initialize matrix Q to zero
  - Select a random initial state
  - Set initial state = current state
  - Select one among all possible actions for the current state
  - Using the possible action, consider going to the next state
  - Get maximum Q value for this next state based on all possible actions
  - Compute Q matrix
  - Repeat above steps until current state = goal state

- Example: Place an agent on any one of the rooms (0,1,2,3,4) and the goal is to reach outside the building (room5)

Q-learning





Action

State	0	1	2	3	4	5
0	-1	-1	-1	-1	0	-1
1	-1	-1	-1	0	-1	100
2	-1	-1	-1	0	-1	-1
3	-1	0	0	-1	0	-1
4	0	-1	-1	0	-1	100
5	-1	0	-1	-1	0	100

$R =$

The -1's in the table represent null values

Reward Matrix

Example (Agent traverse from room 2 to room5):

1. Initial state = state 2
2. State 2 -> state 3
3. State 3 -> state (2, 1, 4)
4. State 4 -> state 5

Q matrix represents the memory of what the agent has learned through experience

Q-learning Example

First set the Value of learning parameter Gamma =0.8 and the initial state as room 1

Next initialize Q matrix as zero matrix

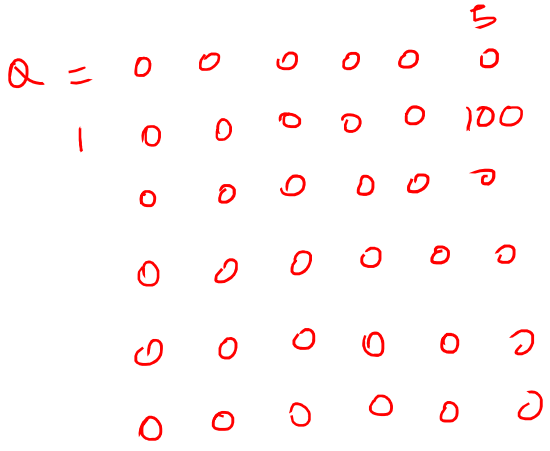
From 1 it can go to either 3 or 5, let's select room number 5

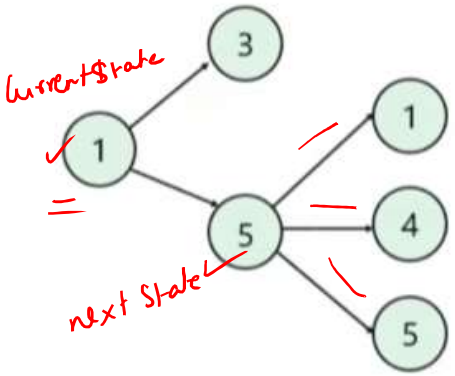
From room 5 calculate maximum Q value for this next state based on all possible actions : go to state 1, 4 or 5

Q (state, action) = R (state, action) + Gamma * Max [Q (next state, all actions)]

Q (1, 5) = R (1, 5) + 0.8 * Max [Q (5, 1), Q (5, 4), Q (5, 5)] = 100 + 0.8 * 0 = 100

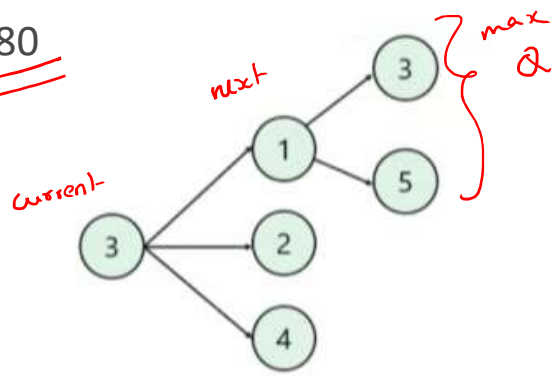
Since matrix Q is still initialized to zero, Q (5, 1), Q (5, 4), Q (5, 5), are all zero





Q-learning Example

- This time, we have state 3 as our initial state.
- it has 3 possible actions: go to state 1, 2 or 4. By random selection, we select to go to state 1 as our action.
- Now we imagine that we are in state 1. Look at the second row of reward matrix R (i.e. state 1). It has 2 possible actions: go to state 3 or state 5. Then, we compute the Q value:
- Q (state, action) = R (state, action) + Gamma * Max [Q (next state, all actions)]
- Q (3,1) = R (3,1) + 0.8 * Max [Q (1, 3), Q (1, 5)] = 0 + 0.8 * Max (0, 100) = 80



$Q =$

	0	1	2	3	4	5
0	0	0	0	0	0	0
1	0	0	0	0	0	100
2	0	0	0	0	0	0
3	0	80	0	0	0	0
4	0	0	0	0	0	0
5	0	0	0	0	0	0

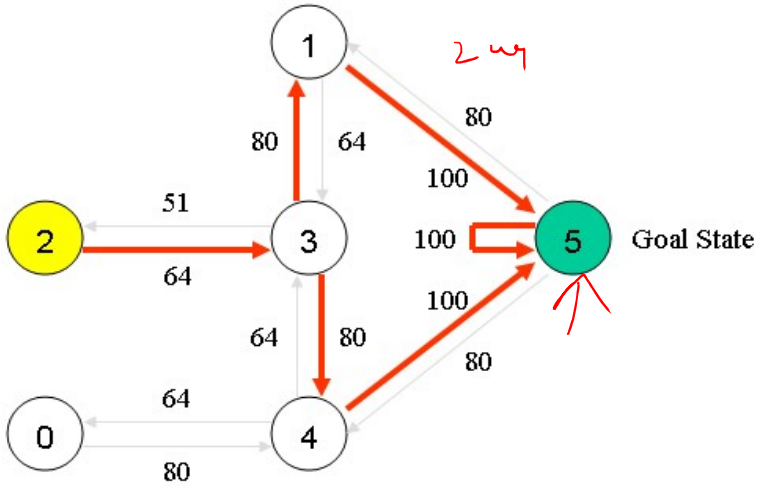
$R =$

State \ Action	0	1	2	3	4	5
0	-1	-1	-1	-1	0	-1
1	-1	-1	-1	0	-1	100
2	-1	-1	-1	0	-1	-1
3	-1	0	0	-1	0	-1
4	0	-1	-1	0	-1	100
5	-1	0	-1	-1	0	100

Q-learning Example

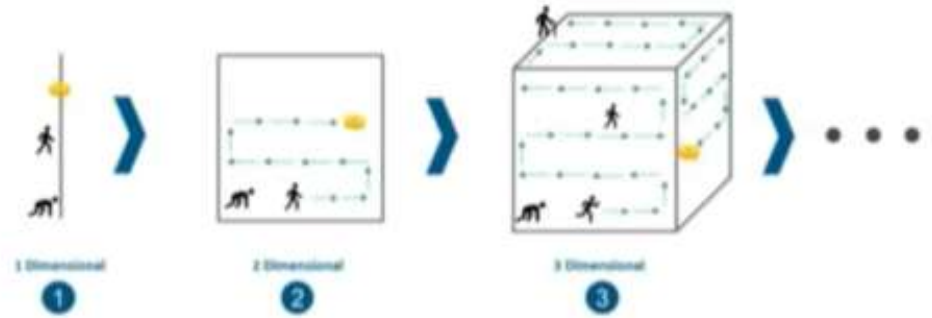
- Repeating this procedure up to Q gets close enough to a state of convergence

$$Q = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{cccccc} 0 & 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{bmatrix} 0 & 0 & 0 & 0 & 80 & 0 \\ 0 & 0 & 0 & 64 & 0 & 100 \\ 0 & 0 & 0 & 64 & 0 & 0 \\ 0 & 80 & 51 & 0 & 80 & 0 \\ 64 & 0 & 0 & 64 & 0 & 100 \\ 0 & 80 & 0 & 0 & 80 & 100 \end{bmatrix} \end{array}$$



Limitations of ML

- It is not capable enough to handle high dimensional data which leads to handling and processing such type of data becomes very complex and it takes up lot of resources. This is also known as curse of dimensionality.



High dimensional data

Limitations of ML

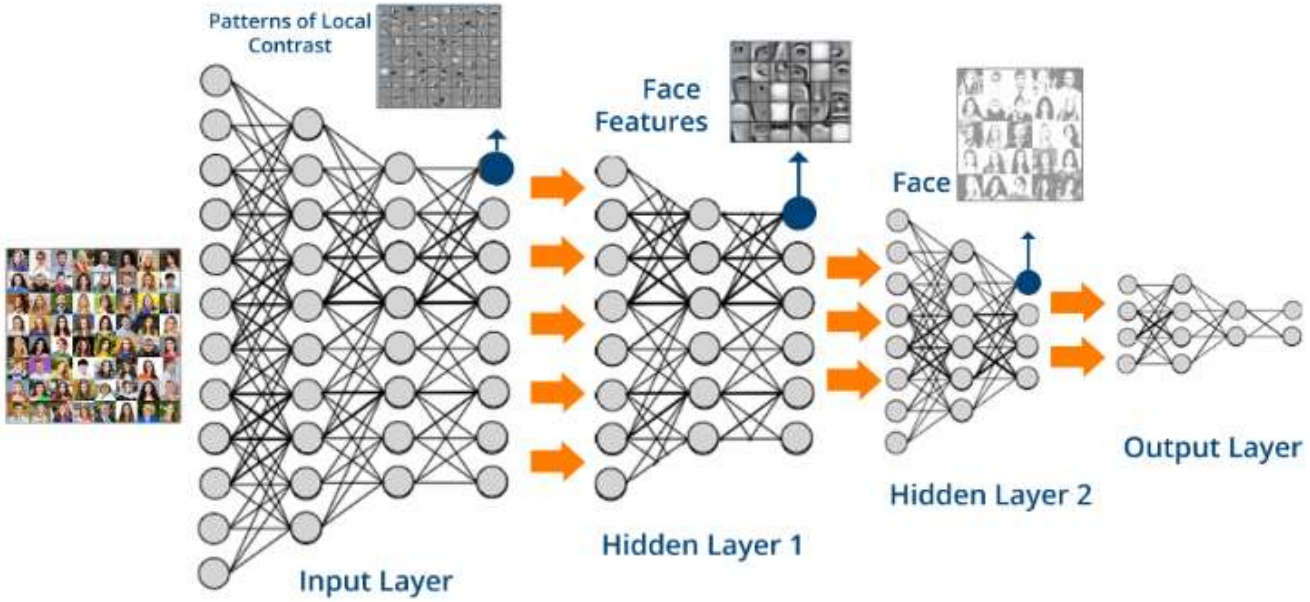
- It cannot be used in image recognition, because image recognition and images have lot of pixels and they have a lot of high dimensional data.

Limitations of ML

- Another big challenge with traditional machine learning models is a process of feature extraction. For complex problems such as object recognition or handwriting recognition, this is a huge challenge. This feature extraction problem can be overcome by deep learning models which requires very little guidance from the programmer

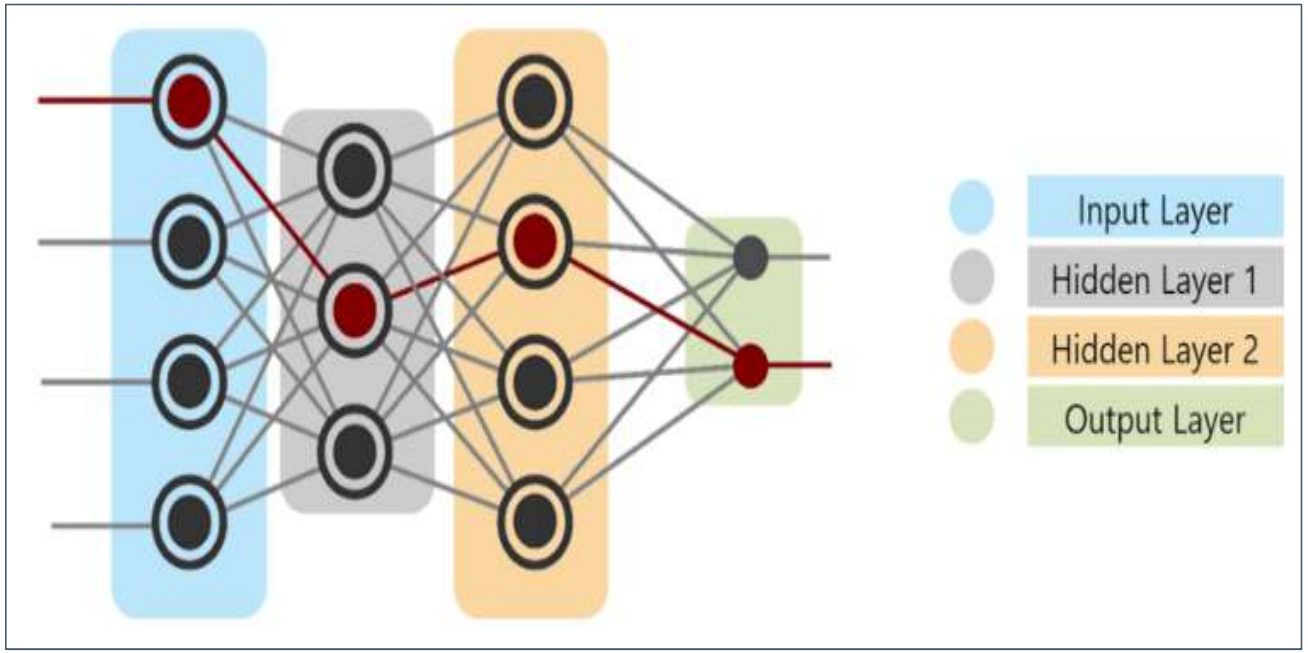
Why Deep Learning?

- Deep learning models are capable to focus on the right features by themselves, requiring little guidance from the programmer. These models also partially solve the dimensionality problem. The idea behind the deep learning is to build learning algorithms that mimic brain functions.



What is Deep Learning?

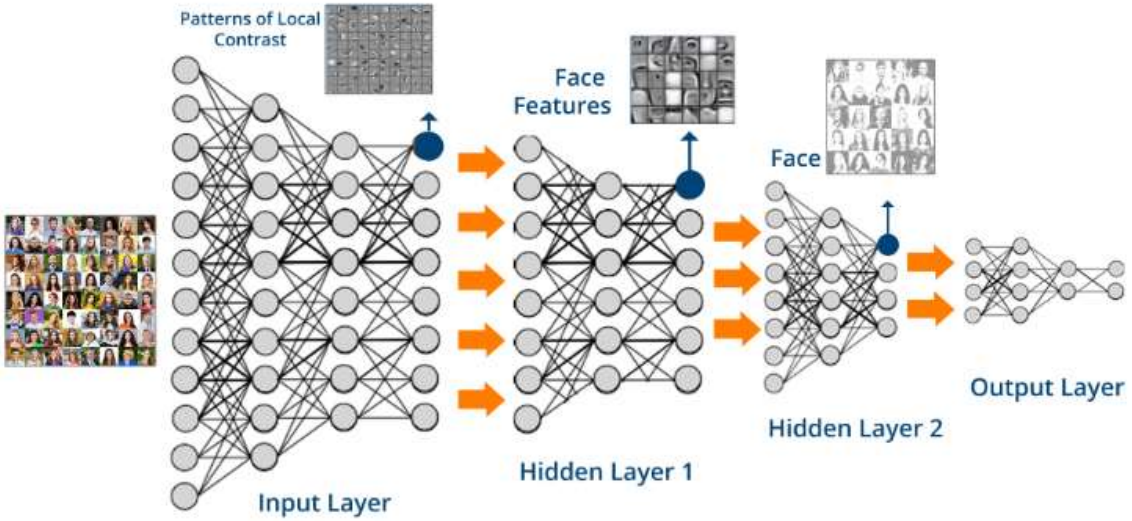
- Deep Learning is a collection of statistical machine learning techniques used to learn feature hierarchies based on the concept of artificial neural networks.



Understanding Deep Learning

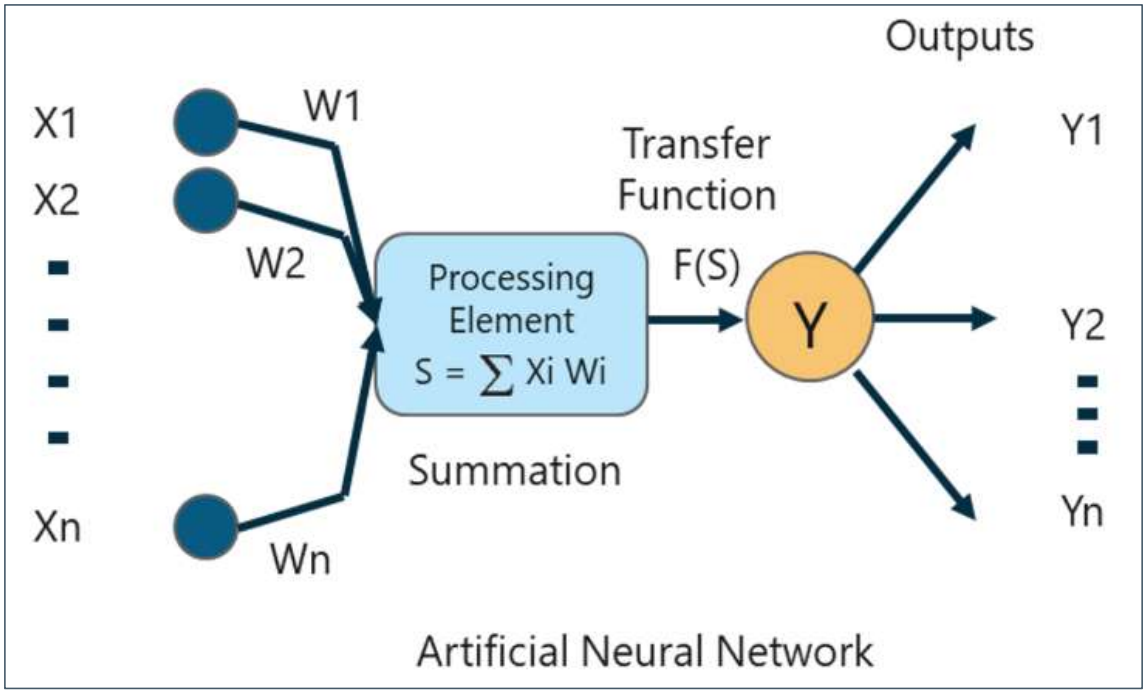
• Image recognition using Deep Networks:

1. Pass the high dimensional data to the input layer
2. Output received from the input layer contains patterns which are extracted
3. Output will be fed to the Hidden layer 1
4. Hidden layer will able to form the entire faces
5. The output layer performs classification



Perceptron

- An artificial neuron or perceptron is a single layer neural network that is used to classify linear data. It models a neuron which has a set of inputs, each of which is given a specific weight. The neuron computes some function on these weighted inputs and gives the output.

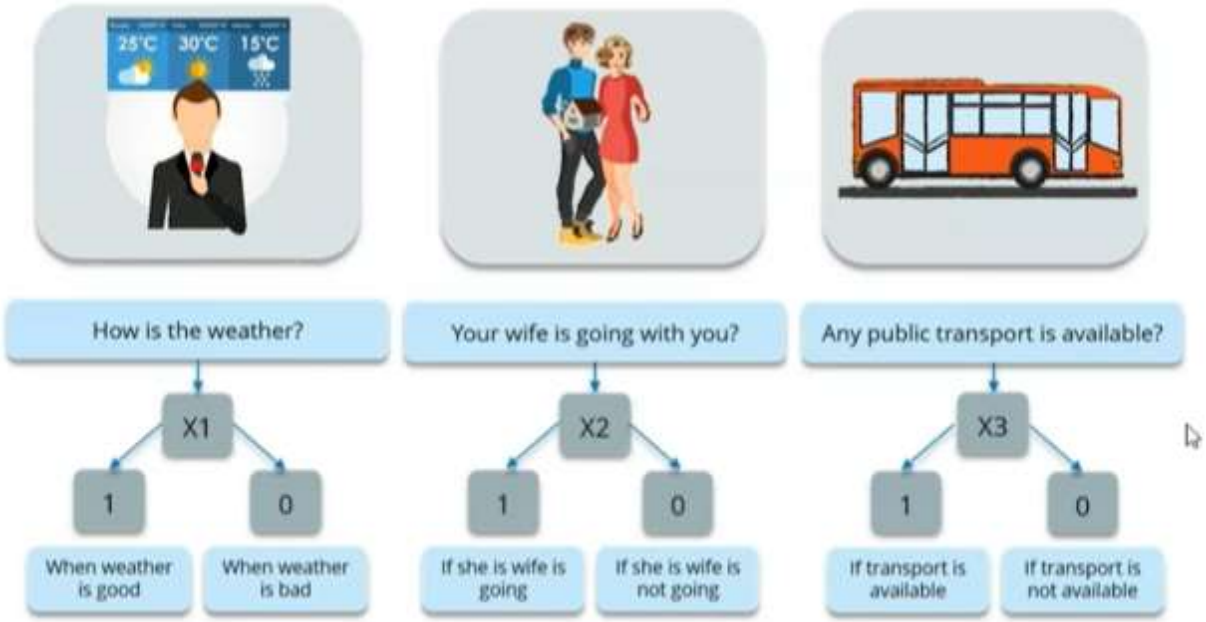


Perceptron Learning Analogy

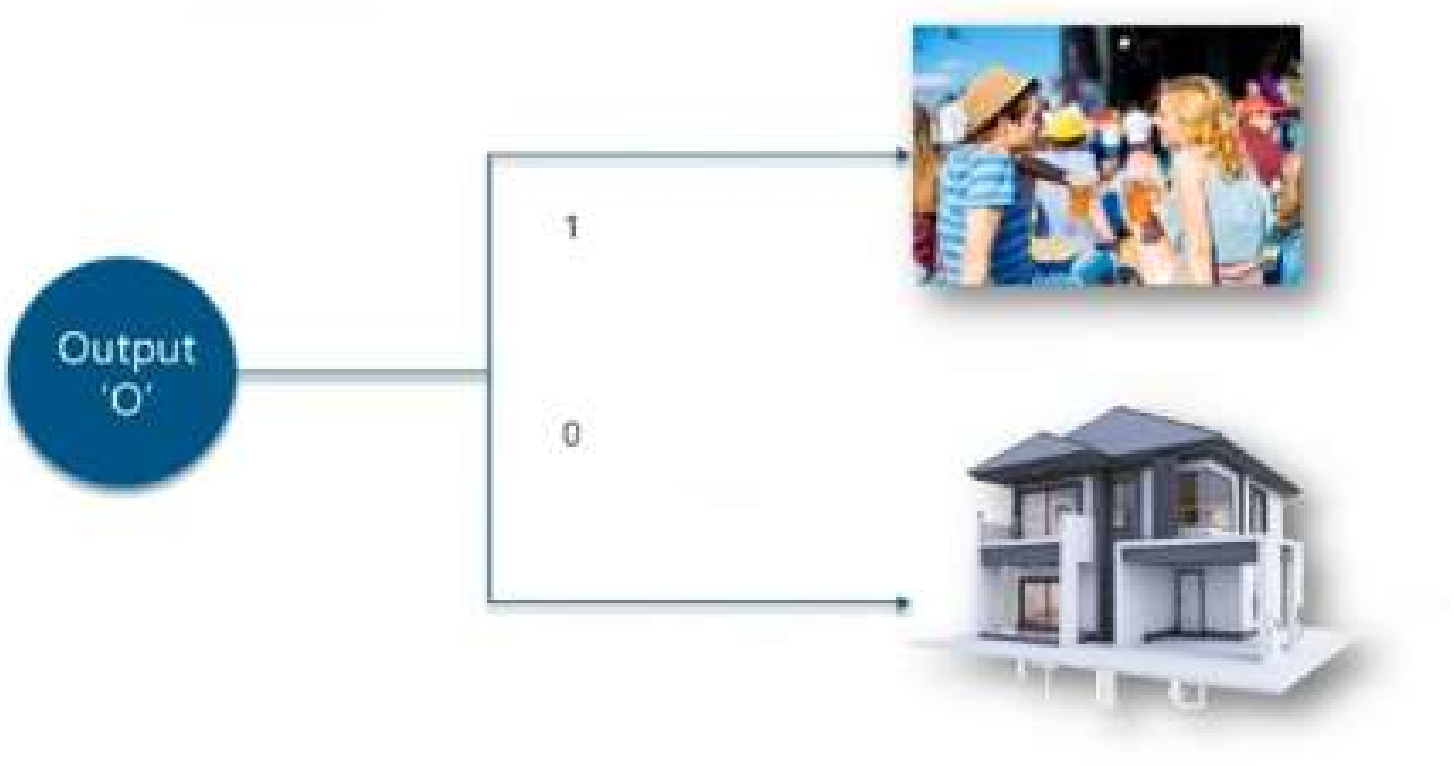
- Suppose you want to go to a party happening near your house. And your decision depends on multiple factors:

1. How is weather?
2. Your wife is going with you?
3. Any public transport is available?

Perceptron Learning Analogy



Perceptron Learning Analogy

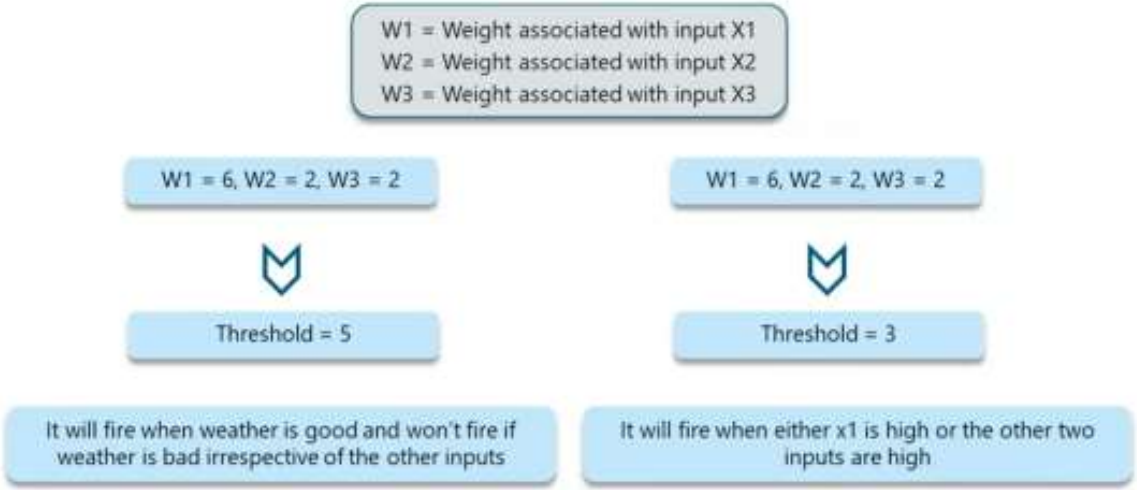


Perceptron Learning Analogy

- Suppose for you the most important factor is weather.

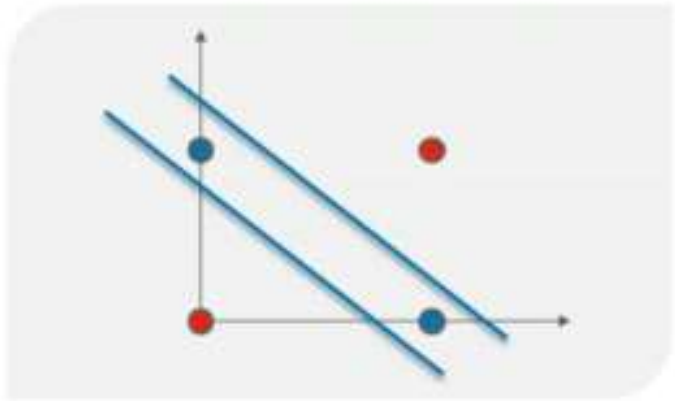
- Weather ---- Good = Party
- Weather ----- Not Good = Sit home

Perceptron Learning Analogy



Limitations of a Perceptron

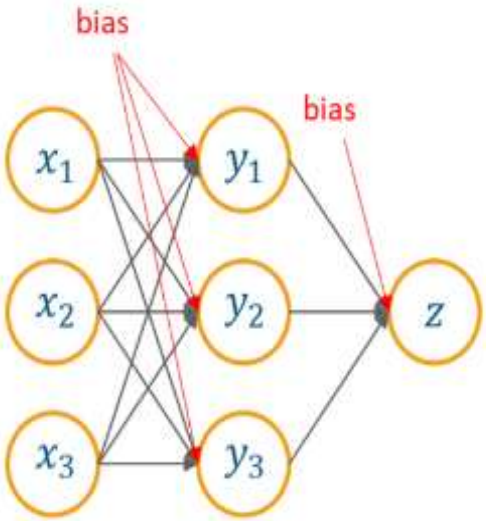
Dealing with non linearly separable data

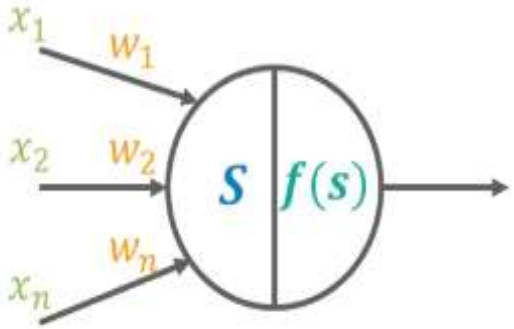


A Multi-Layer Perceptron with backpropagation can be used to solve this problem

Multi- Layer Perceptron

- A multi-layer perceptron has the same structure of a single layer perceptron but with one or more hidden layers and is thus considered a deep neural network.





Summation:

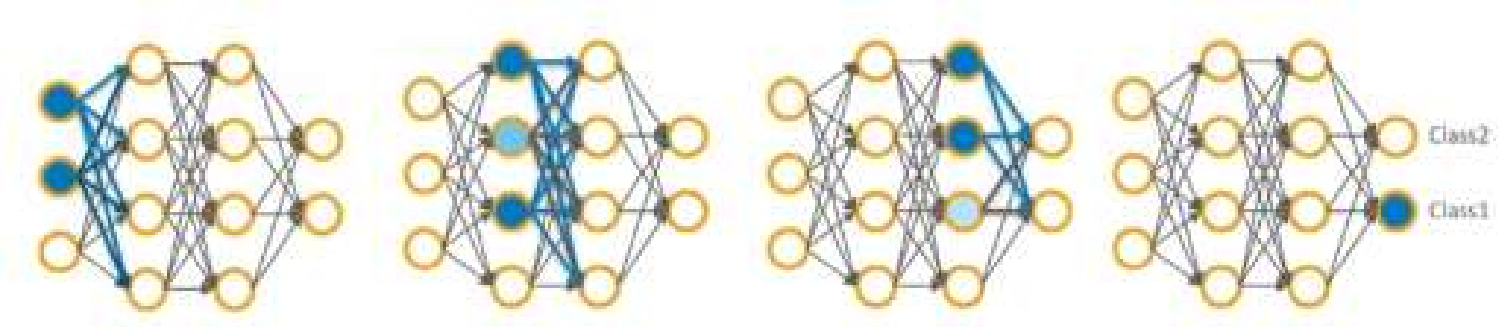
$$S = \sum_{i=1}^{n} w_i * x_i$$

Transformation:

$$f(x) = \frac{1}{1+e^{-\beta x}}$$

Multi- Layer Perceptron

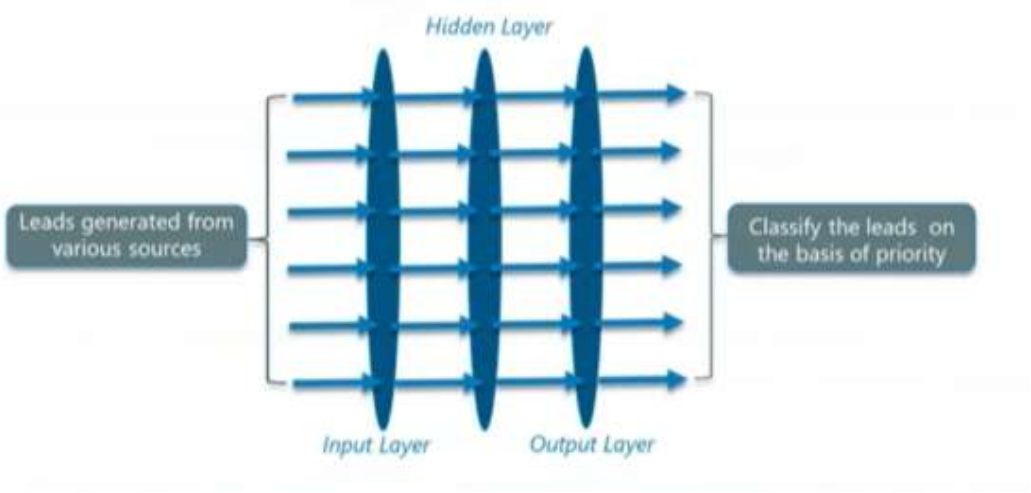
- The weights between the units are the primary means of long-term information storage in neural networks.
- Updating the weights is the primary way the neural network learns new information.



A set of inputs are passed to the first hidden layer, the activations from that layer are passed to the next layer and so on, until you reach the output layer.

Backpropagation

- The Backpropagation algorithm is a supervised learning method for Multilayer Perceptron.



Maximum weight is assigned to the most important lead/input.

Training a Neural network

• The most common deep learning algorithm for supervised training of the multi-layer perceptron is known as backpropagation. After calculating the weighted sum of inputs and passing them through the activation function we propagate backwards and update the weights to reduce the error (desired output – model output). Consider the below example.

Input	Desired Output
0	0
1	1
2	4

Training a Neural network

- Let's consider the initial value of the weight as 3 and see the model output

Input	Desired Output	Model Output (W = 3)
0	0	0
1	2	3
2	4	6

- Now, we will see the error (Absolute and Square):

Input	Desired Output	Model Output (W = 3)	Absolute error	Square error
0	0	0	0	0
1	2	3	1	1
2	4	6	2	4

Training a Neural network

- Let's update the weight value and make it as 4

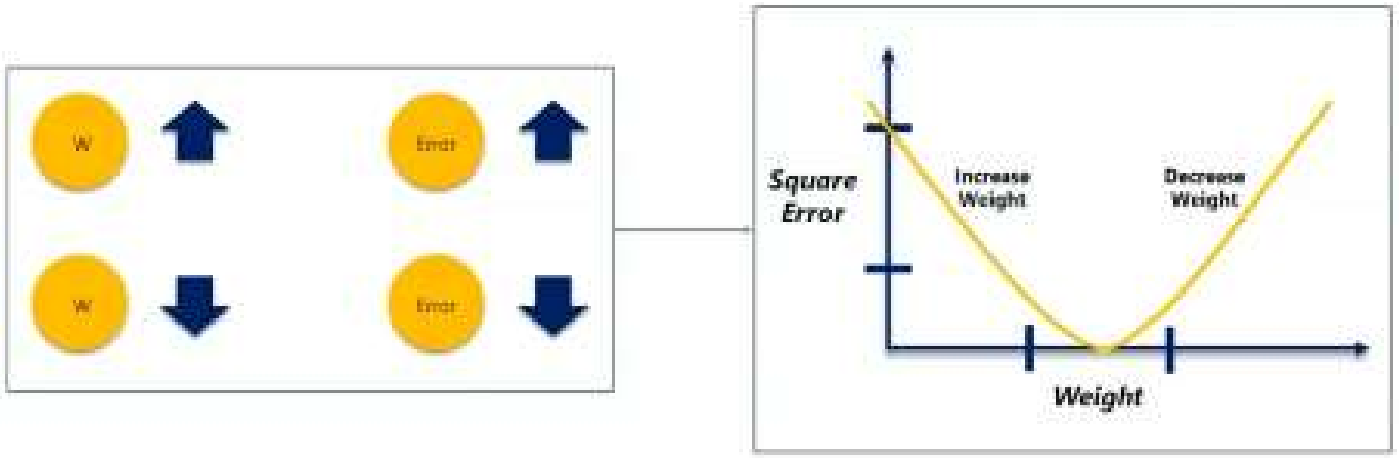
Input	Desired Output	Model Output (W = 3)	Absolute error	Square error	Model Output (W = 4)
0	0	0	0	0	0
1	2	3	1	1	4
2	4	6	2	4	8

- Notice the error has increased

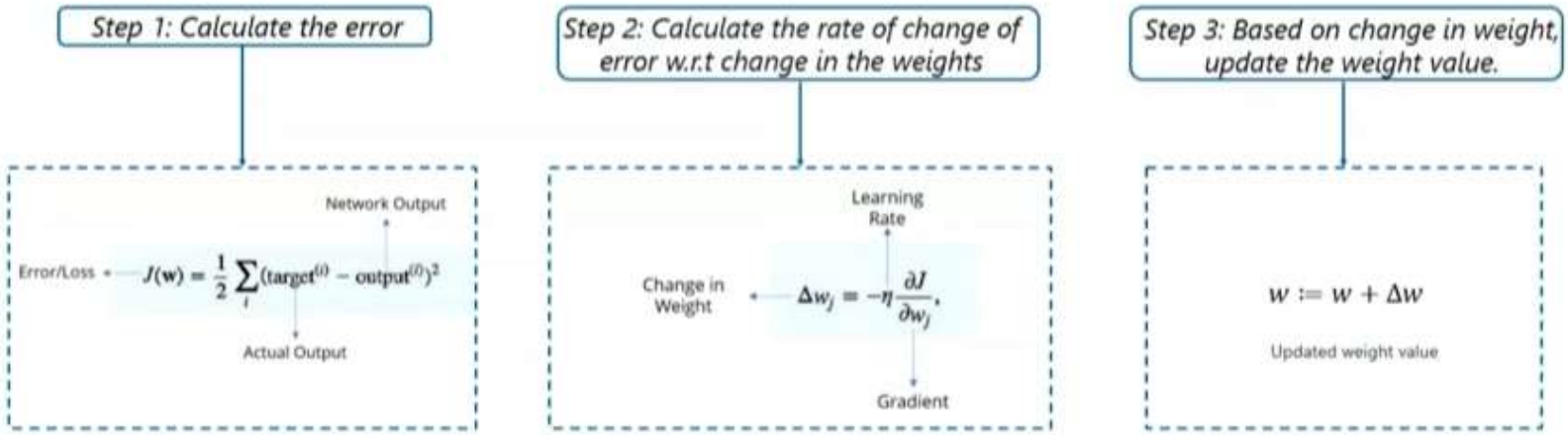
Input	Desired Output	Model Output (W = 3)	Absolute error	Square error	Model Output (W = 4)	Square error
0	0	0	0	0	0	0
1	2	3	1	1	4	4
2	4	6	2	4	8	16

Training a Neural network

- Relationship between the assigned weight and the error

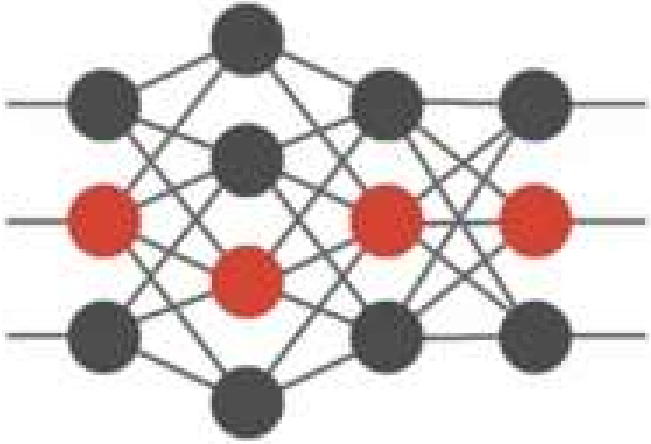


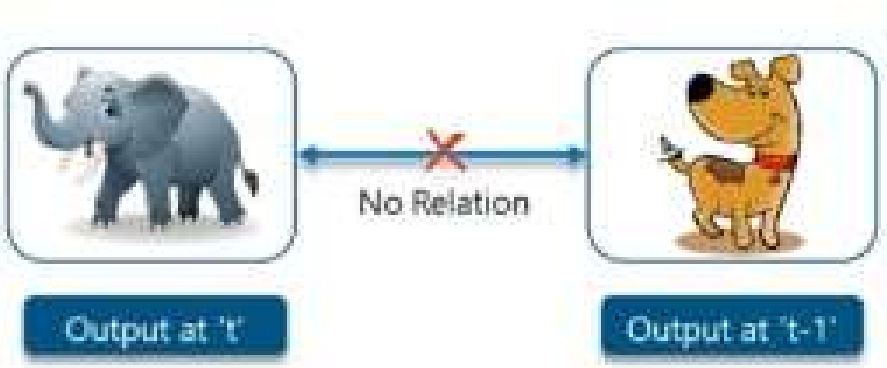
Reducing the Error / Loss



Why not Feed Forward Network?

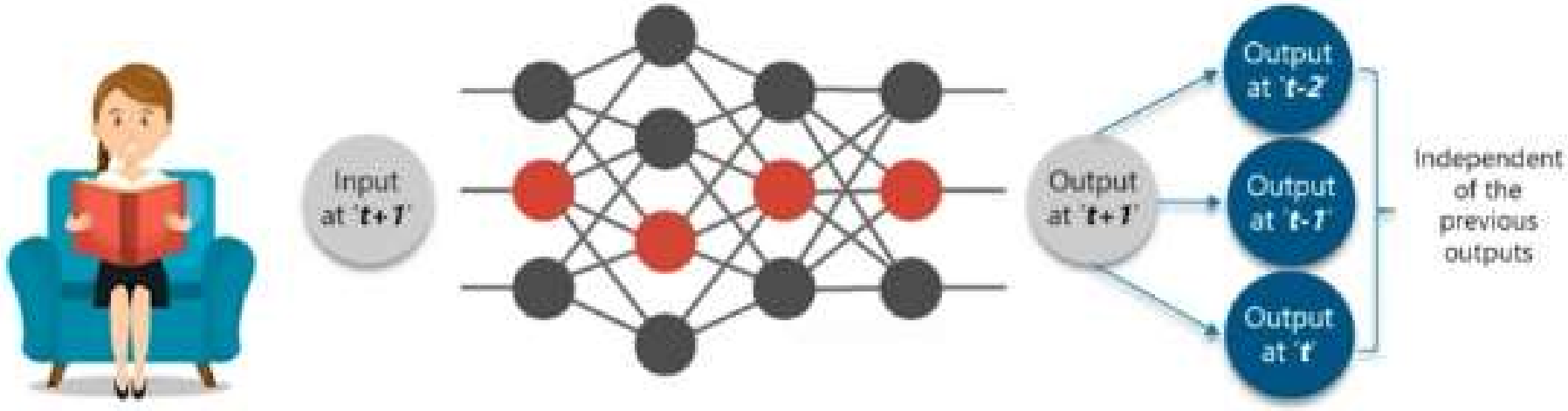
- A trained feedforward network can be exposed to any random collection of photographs, and the first photograph it is exposed to will not necessarily alter how it classifies the second one.



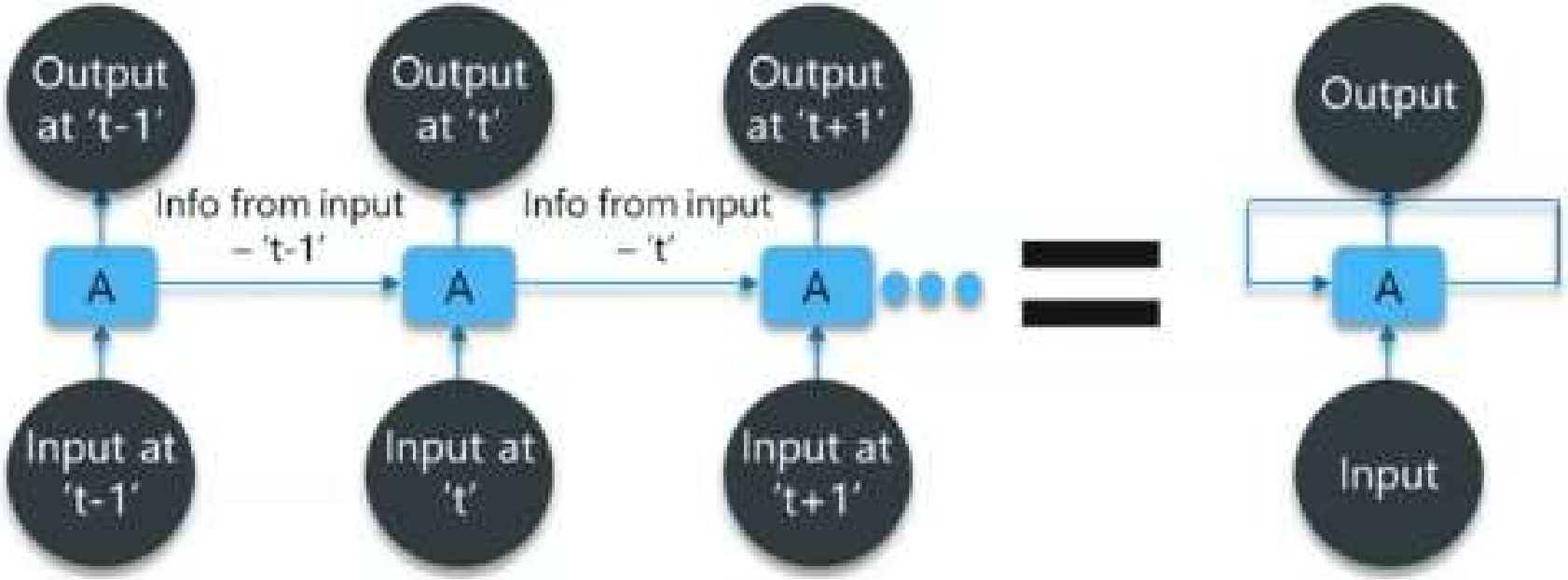


Why not Feed Forward Network?

- When you read a book, you understand it based on your understanding of previous words



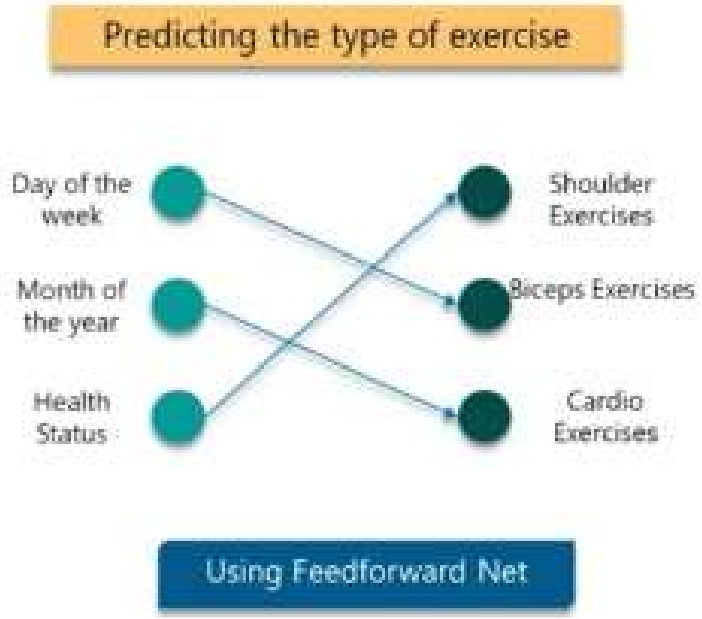
Solution


Output at 't-1'
Output at 't'
Output at 't+1'
Info from input – 't-1'
Info from input – 't'
A
A
A
Input at 't-1'
Input at 't'
Input at 't+1'
Output
A
Input

Recurrent Neural Network

- A Recurrent neural networks are a type of artificial neural network designed to recognize patterns in sequence of data, such as text, genomes, handwriting, the spoken word, or numerical times series data emanating from sensors, stock markets and government agencies.

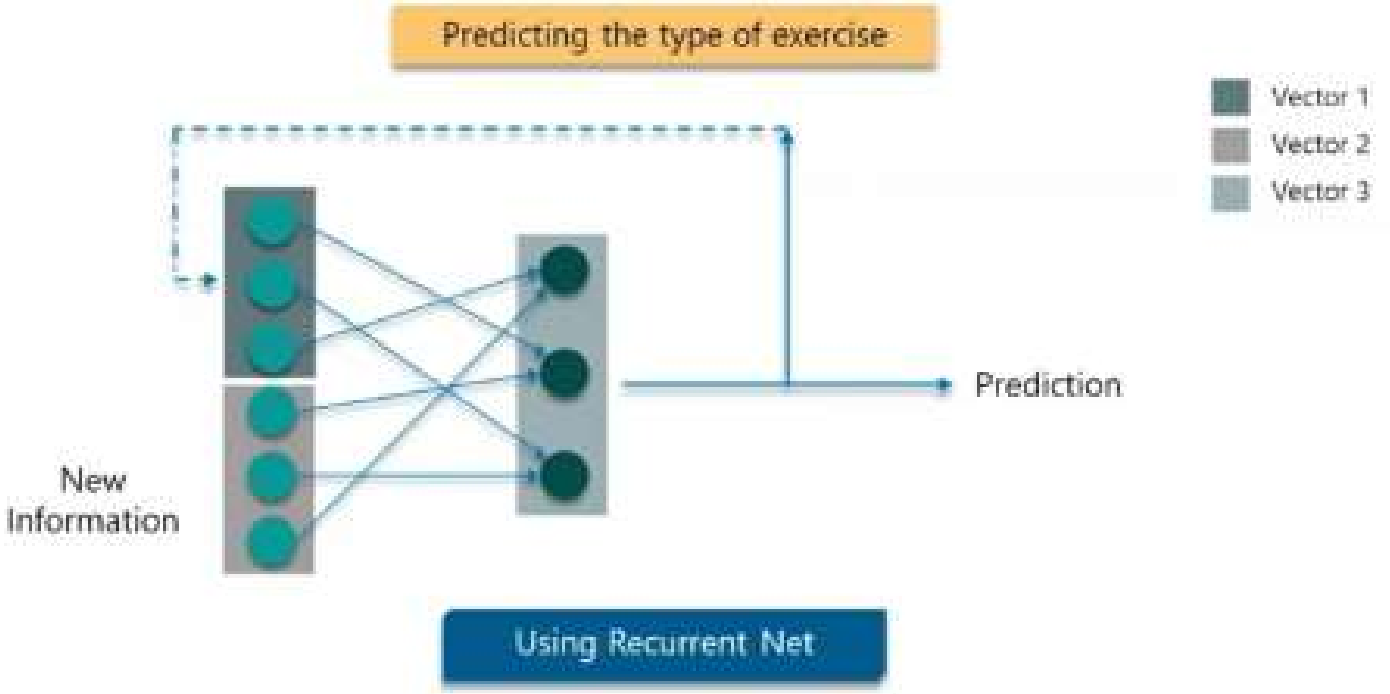
Recurrent Neural Network



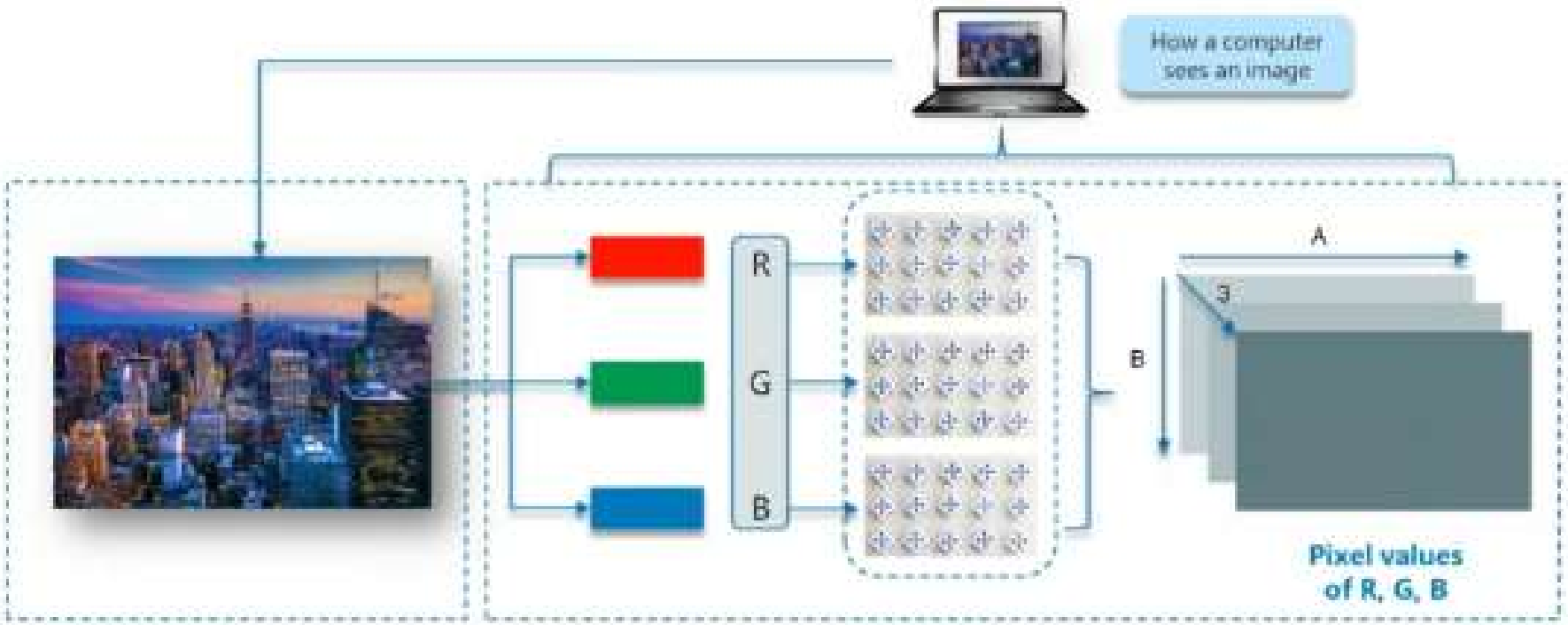
Recurrent Neural Network

Recurrent Neural Network

Recurrent Neural Network



Why CNN?



- Size of the image will be = B x A x 3

Why CNN?

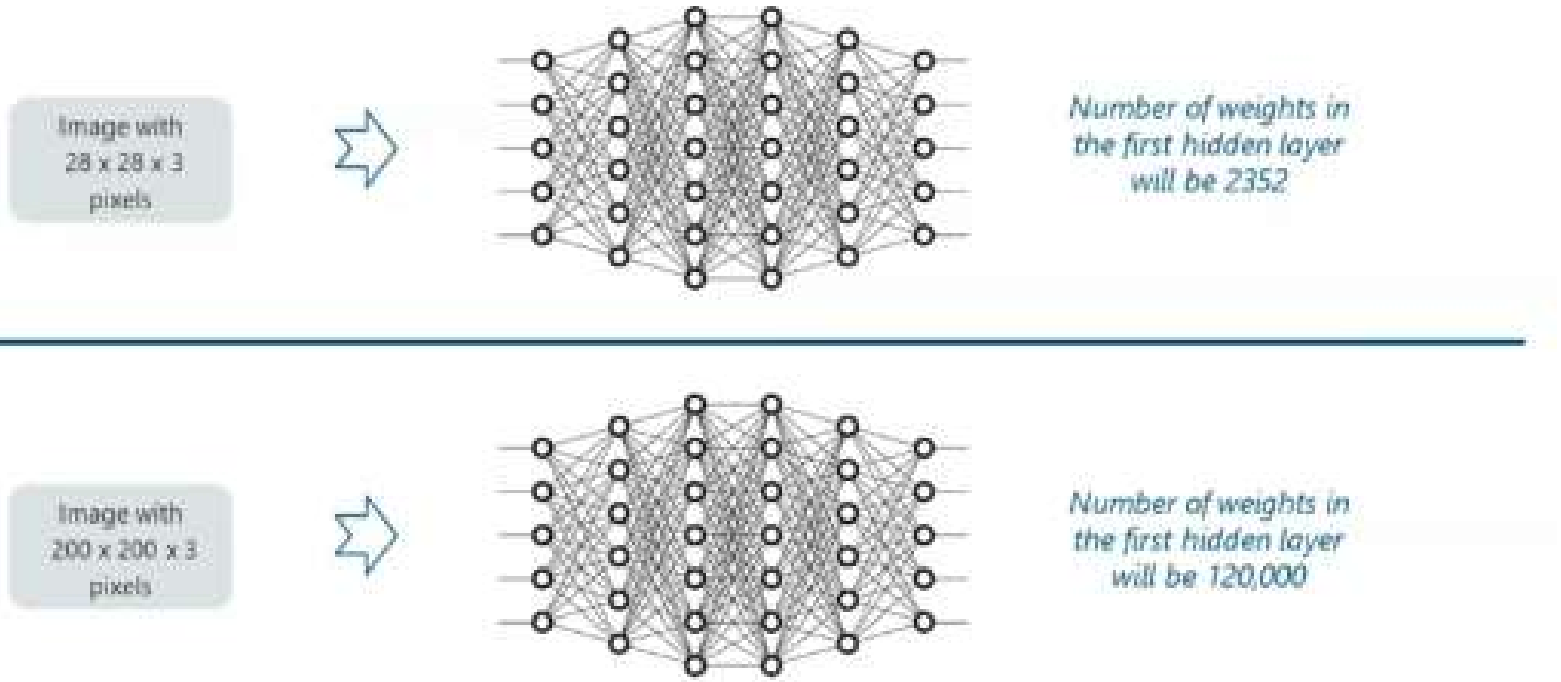
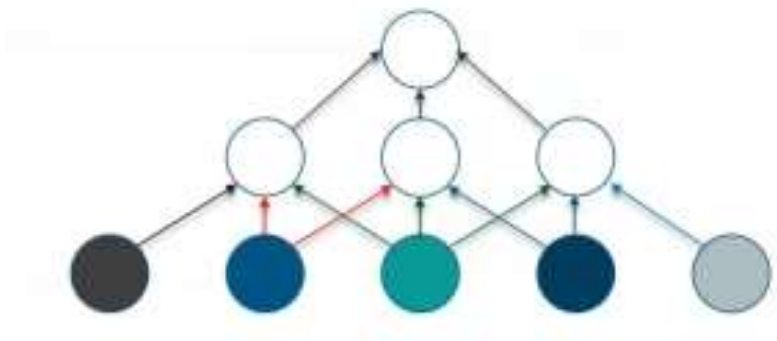


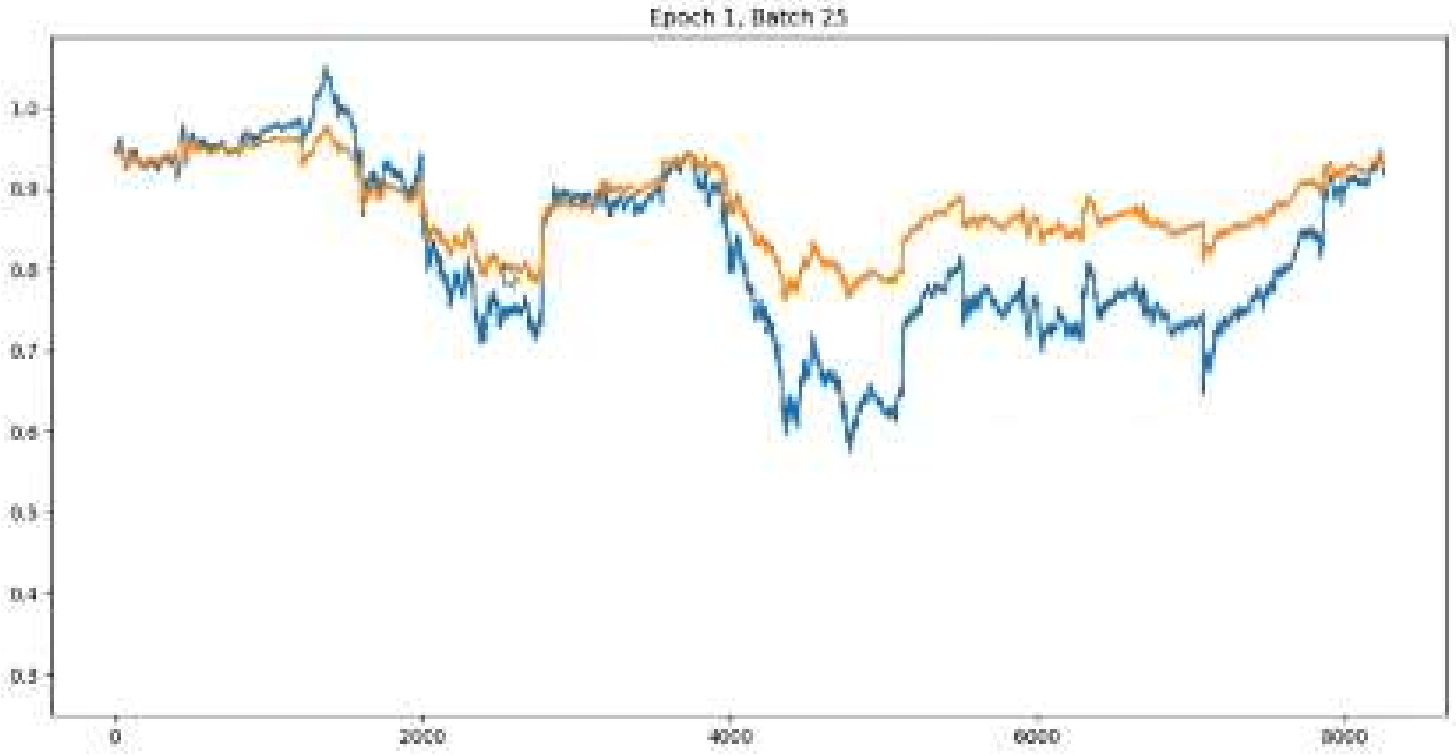
Image with
28 x 28 x 3
pixels
Number of weights in
the first hidden layer
will be 2352
Image with
200 x 200 x 3
pixels
Number of weights in
the first hidden layer
will be 120,000

Convolutional Neural Networks

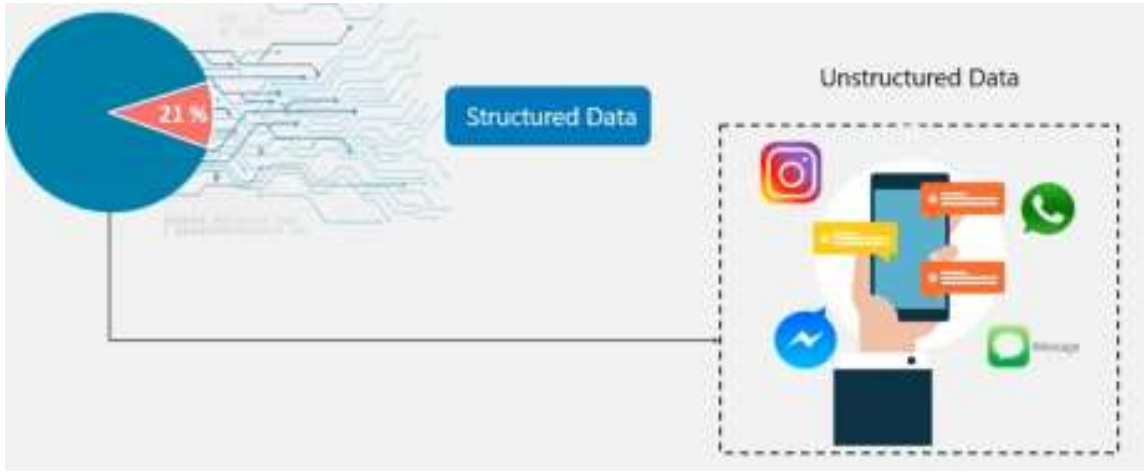
- In case of CNN, the neuron in a layer will only be connected to a small region of the layer before it, instead of all of the neurons in a fully-connected manner



Convolutional Neural Networks



Need for Text Mining & NLP

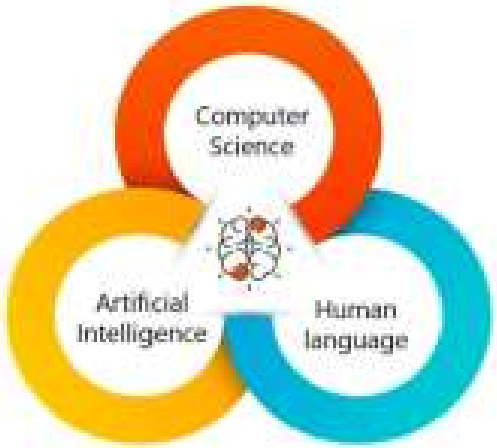


What is Text Mining?

Text mining / Text analytics is the process of deriving meaningful information from natural language text

What is NLP?

- Natural language processing is a part of computer science and artificial intelligence which deals with human languages



Where is Text Mining is used?

Applications of NLP

Sentiment analysis is contextual mining of text which identifies and extracts subjective information in source material, and helping a business to understand the social **sentiment** of their brand, product or service while monitoring online conversations

A chatbot is a software application used to conduct an on-line chat conversation via text or text-to-speech, in lieu of providing direct contact with a live human agent.

Speech recognition, or speech-to-text, is the ability for a machine or program to identify words spoken aloud and convert them into readable text

Tokenization

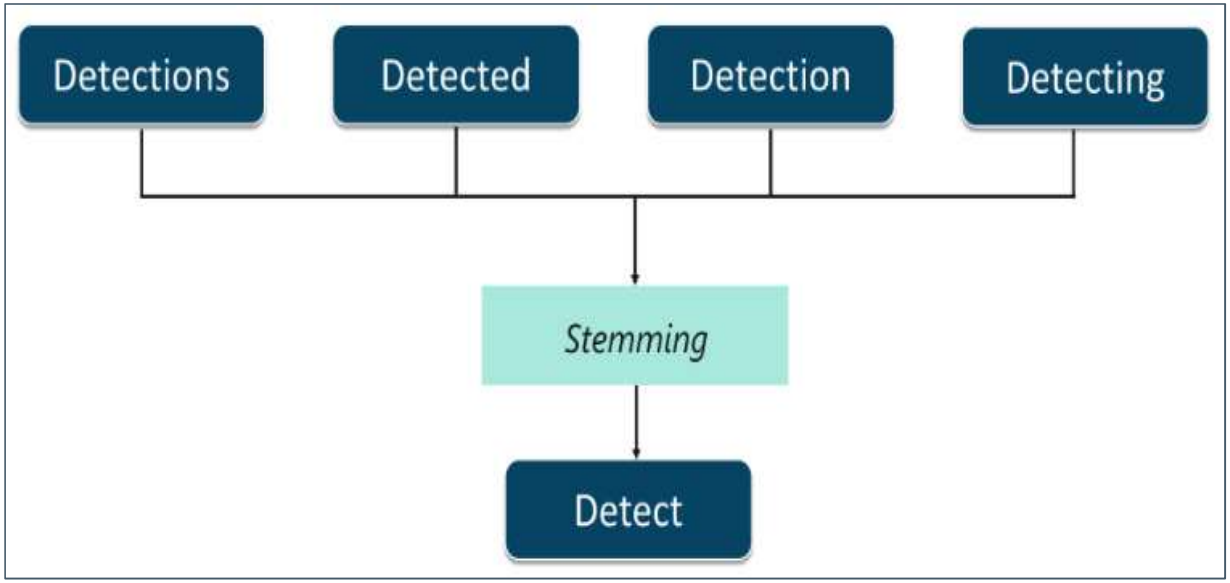
- The process of splitting data into smaller chunks or tokens is known as tokenization

1.Break a complex sentence into words
2.Understand the importance of each of the words with respect to the sentence
3.Produce a structural description on an input sentence

Tokens are simple

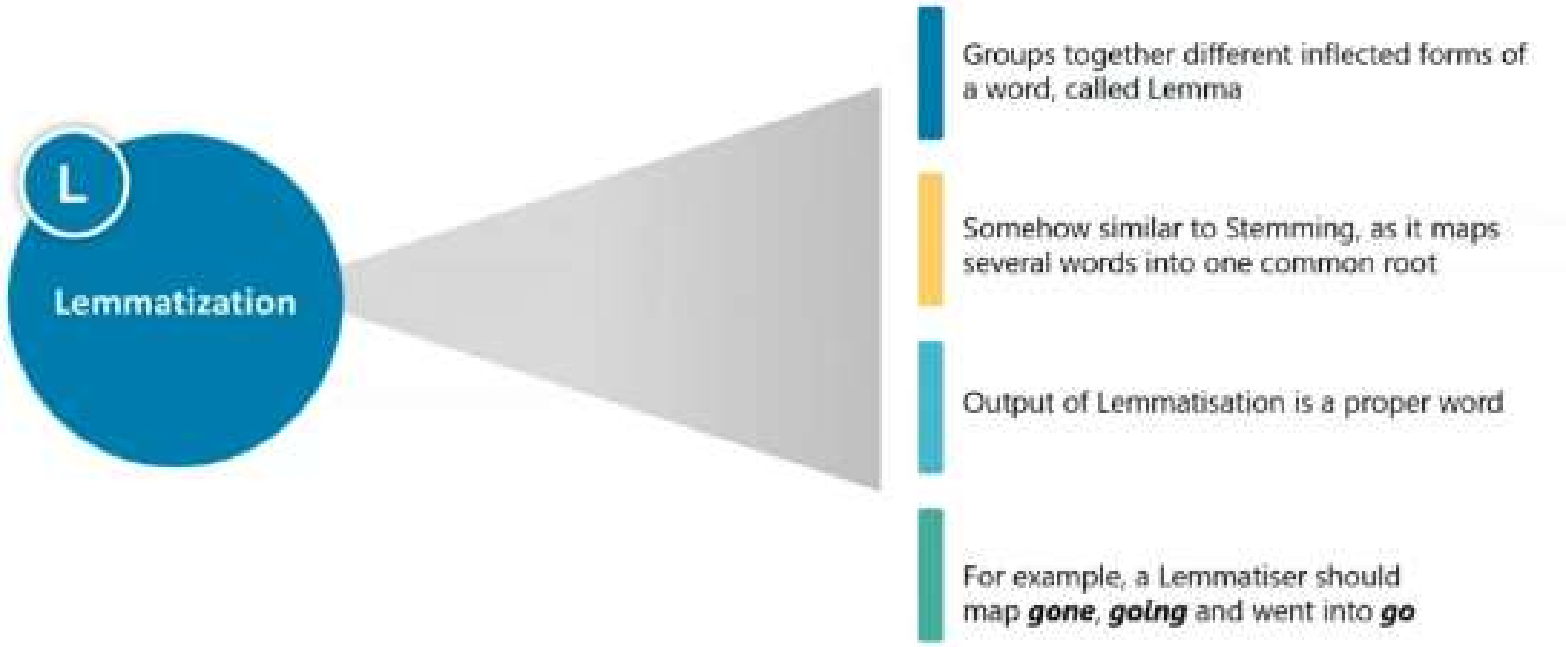
Stemming

Normalize words into its base form or root form.



Lemmatization

- It is similar to Stemming; however, it is more effective because it takes into consideration the morphological analysis of the words.



Stop Words

- These are a set of commonly used words in any language. Stop words are critical for text analysis and must be removed in order to better understand any document.

Document Term Matrix

- It is a matrix with documents designated by rows and words by columns.

Documents

	This	is	fun
Doc 1	1	1	1
Doc 2	1	1	0
Doc 3	0	1	0
Doc 4	1	0	0

Document Term Matrix


thank you
danke
謝謝
ngiyabonga
teşekkür ederim
gracias
спасибо
Баярлалаа
рахмат
dank je
mochchakkeram
go raibh maith agat
dziękuję
bedankt
obrigado
sukriya
kop khun krap
terima kasih
grazie
arigatô
takk
dakujem
мерси
merci
ευχαριστώ
감사합니다
ありがとう
tapadh leat
хвала
obrigada
சாகுமுறு
நன்றி

	Linear Regression	Logistic Regression
1 Definition	To predict a continuous dependent variable based on values of independent variables	*To predict a categorical dependent variable based on values of independent variables*
2 Variable Type	Continuous dependent variable	Categorical dependent variable
3 Estimation method	Least square estimation	Maximum like-hood estimation
4 Equation	$Y = b_0 + b_1 x + e$	$\log\left(\frac{Y}{1-Y}\right) = C + B1X1 + B2X2 + \ldots$
5 Best fit line	Straight line	Curve
6 Relationship between DV & IV	Linear relationship between the dependent and independent variable	Linear relationship is not mandatory
7 Output	Predicted integer value	Predicted binary value (0 or 1)
8 Applications	Business domain, forecasting sales	Classification problems, cybersecurity, image processing