SIX WEEKS TRAINING REPORT

On

Machine Learning Engineer Intern's

Submitted by

Sahil Panchal

Employee ID – MIC098

Under the Guidance of

Sanilbhai Diwanaji

Computer Engineering

LDRP Institute of Technology and

Research. (Nov-Jan, 2020-21)

COMPUTER DEPARTMENT

LDRP Institute of Technology and Research, Gandhinagar

Kadi Sarva Vishwavidyalaya

APRIL, 2021

LDRP INSTITUTE OF TECHNOLOGY AND RESEARCH

GANDHINAGAR

CE-IT Department

CERTIFICATE

-

This is to certify that **"Internship Report"** has been submitted out by **Sahil Panchal(1721BECE30073)** under my guidance in fulfilment of the degree of Bachelor of Engineering in COMPUTER DEPARTMENT Semester-8 of Kadi Sarva Vishwavidyalaya University during the academic year 2020-2021.

Ashish B. Patel

Internal Guide

LDRP ITR

Dr. Shivangi Surati

Head of the Department

LDRP ITR

DECLARATION

I hereby declare that I have completed my six weeks summer training at **Moba Mobile Automation Pvt Ltd.** from **3rd. November 2020** to **3rd. January 2021.** under the guidance of **Sanil Diwanji.** I have declared that I have worked with full dedication during these six weeks of training and my learning outcomes fulfill the requirements of training for the award of degree of **Bachelor of Engineering (B.E.) in CE** at **LDRP-ITR.**

Date:………………

ACKNOWLEDGEMENT

The success and final outcome of learning Machine Learning required a lot of guidance and assistance from many people and I am extremely privileged to have got this all along the completion of my course and few of the projects. All that I have done is only due to such supervision and assistance and I would not forget to thank them.

I respect and thank **Moba Mobile Automation Pvt Ltd**, for providing me an opportunity to do the course and project work and giving me all support and guidance, which made me complete the course duly. I amextremely thankful to the course advisor **Mr. Sanil Diwanji.**

I am thankful to **Moba Mobile Automation Pvt Ltd** and fortunate enough to get constant encouragement, support and guidance from all Teaching staffs of which helped us in successfully completing my courseand project work.

Date:…………..

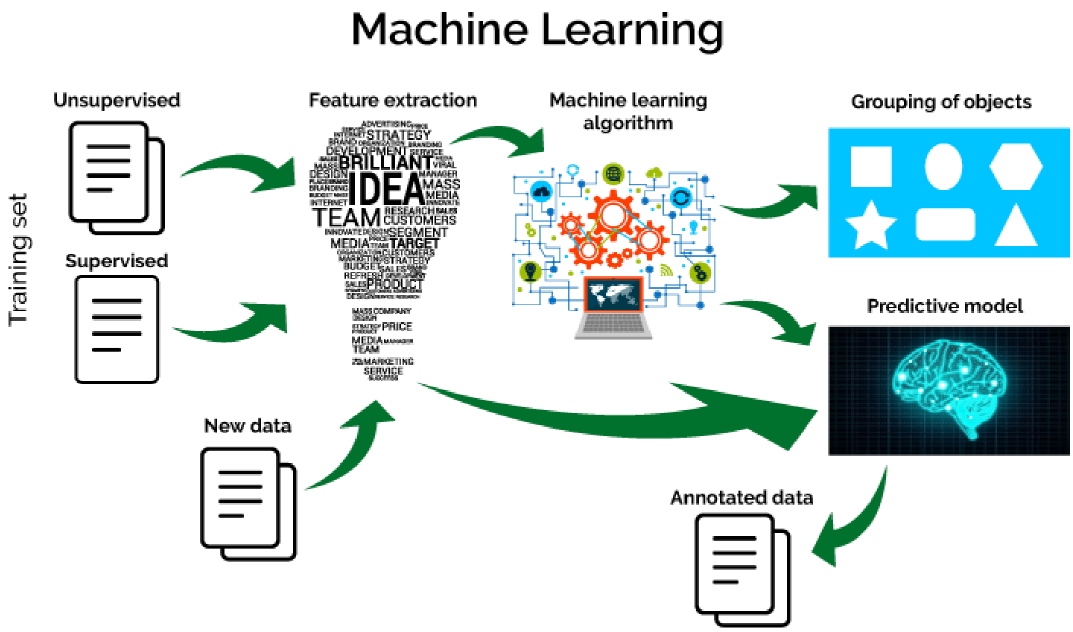
TABLE OF CONTENTS

1. Introduction

1.1. A Taste of Machine Learning

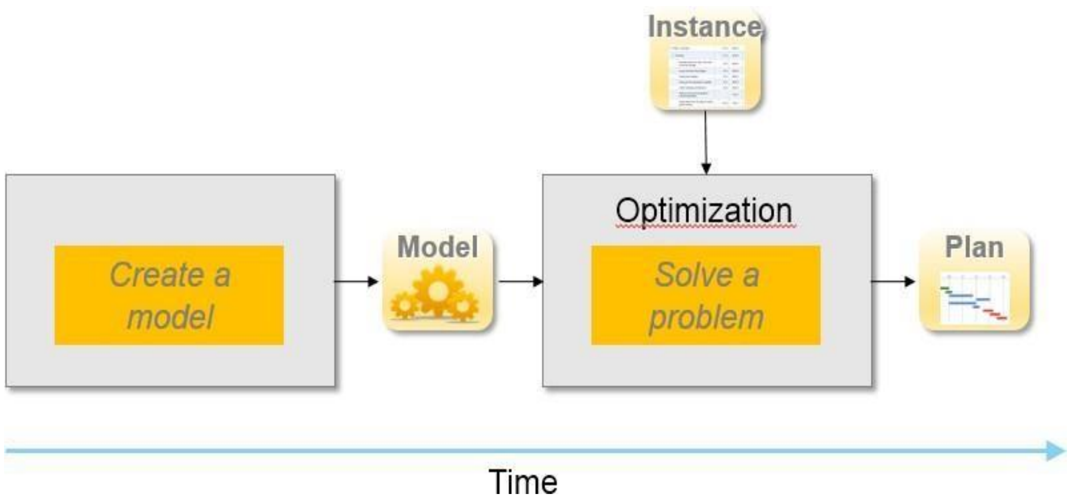
- ✓ Arthur Samuel, an American pioneer in the field of computer gaming and artificial intelligence, coined the term "Machine Learning" in 1959.
- ✓ Over the past two decades Machine Learning has become one of the mainstays of information technology.
- ✓ With the ever-increasing amounts of data becoming available there is good reason to believe that smart data analysis will become even more pervasive as a necessary ingredient for technological progress.

1.2. Relation to Data Mining



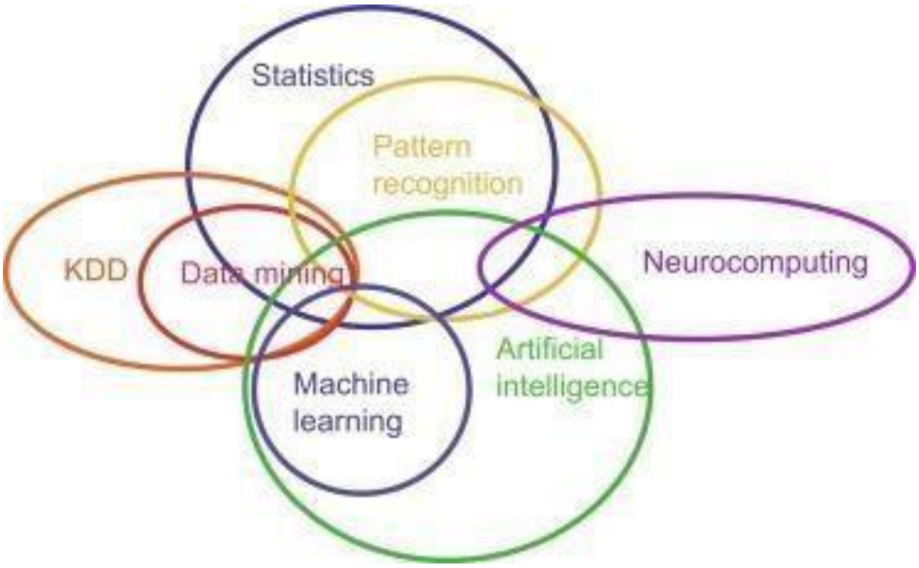
- Data mining uses many machine learning methods, but with different goals; on the other hand, machine learning also employs data mining methods as "unsupervised learning" or as a preprocessing step to improve learner accuracy.

1.3. Relation to Optimization



- Machine learning also has intimate ties to optimization: many learning problems are formulated as minimization of some loss function on a training set of examples.
- Loss functions express the discrepancy between the predictions of the model being trained and the actual problem instances.

1.4. Relation to Statistics



- Michael I. Jordan suggested the term data science as a placeholder to call the overall field.
- Leo Breiman distinguished two statistical modelling paradigms: data model and algorithmic model, wherein "algorithmic model" means more or less the machine learning algorithms like Random forest.

1.5. Future of Machine Learning

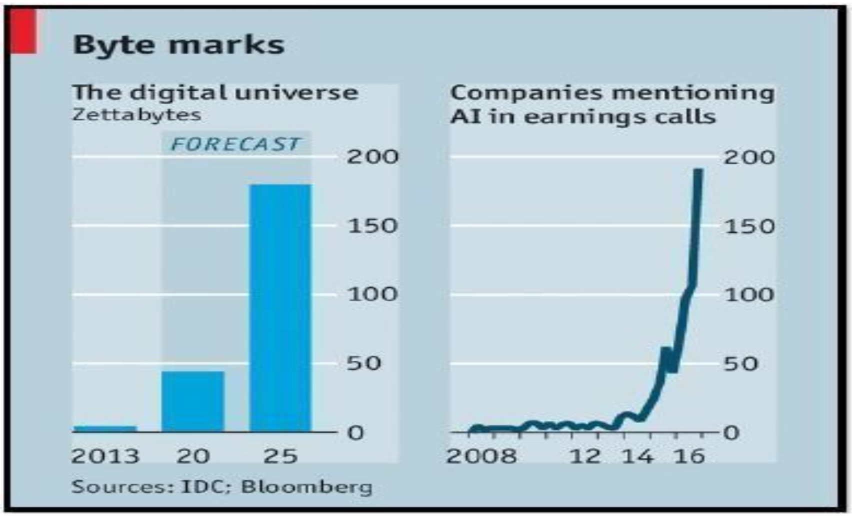
- Machine Learning can be a competitive advantage to any company be it a top MNC or a startup as things that are currently being done manually will be done tomorrow by machines.
- Machine Learning revolution will stay with us for long and so will be the future of Machine Learning.

2. Technology Learnt

2.1. Introduction to AI & Machine Learning

2.1.1. Definition of Artificial Intelligence

- **Data Economy**
  - ✓ World is witnessing real time flow of all types structured and unstructured data from social media, communication, transportation, sensors, and devices.
  - ✓ **International Data Corporation** (IDC) forecasts that 180 zettabytes of data will be generated by 2025.



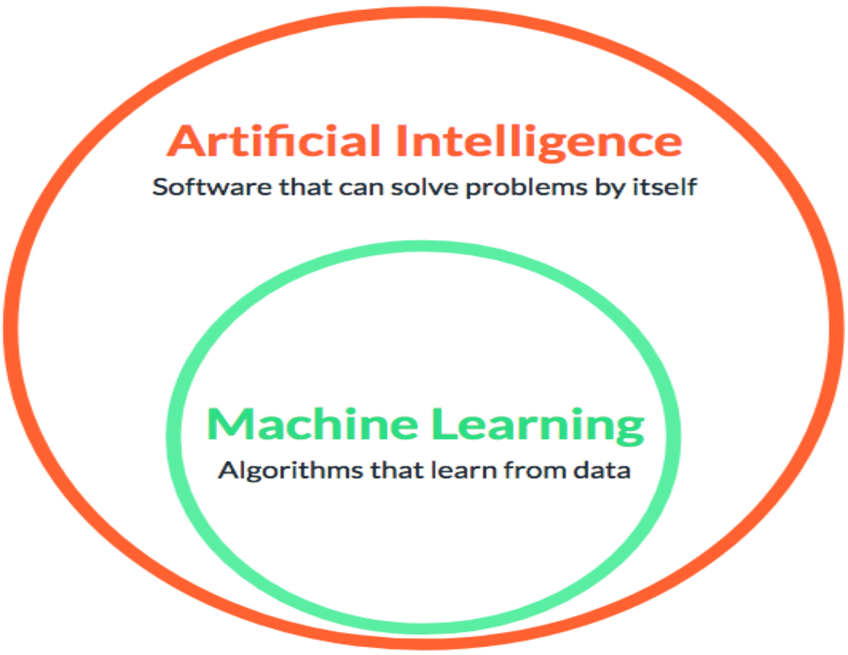
- ✓ This explosion of data has given rise to a new economy known as the **Data Economy.**
- ✓ Data is the new oil that is precious but useful only when cleaned and processed.
- ✓ There is a constant battle for ownership of data between enterprises to derive benefits from it.

❖ **Define Artificial Intelligence**

Artificial intelligence refers to the simulation of human intelligence in machines that are programmed to think like humans and mimic their actions. The term may also be applied to any machine that exhibits traits associated with a human mind such as learning and problem-solving.

2.1.2. Definition of Machine Learning

❖ **Relationship between AI and ML**



Machine Learning is an approach or subset of Artificial Intelligence that is based on the idea that machines can be given access to data along with the ability to learn from it.

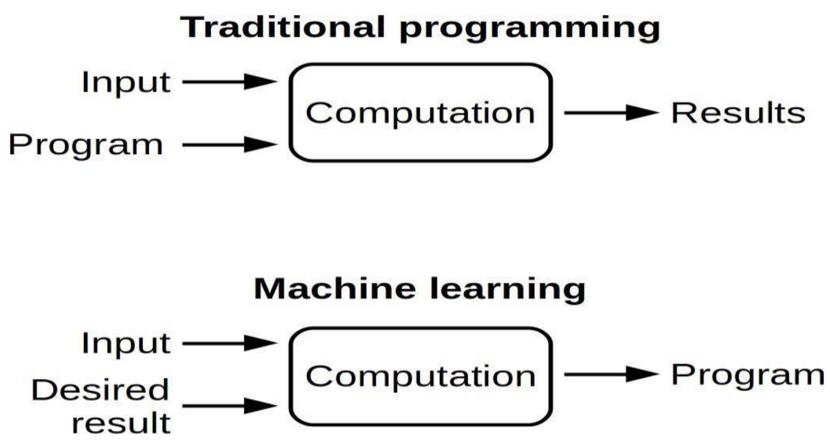
- **Define Machine Learning**

Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. Machine learning focuses on the development of computer programs that can access data and use it learn for themselves.

- **Features of Machine Learning**
  - ✓ Machine Learning is computing-intensive and generally requires a large amount of training data.
  - ✓ It involves repetitive training to improve the learning and decision making of algorithms.
  - ✓ As more data gets added, Machine Learning training can be automated for learning new data patterns and adapting its algorithm.

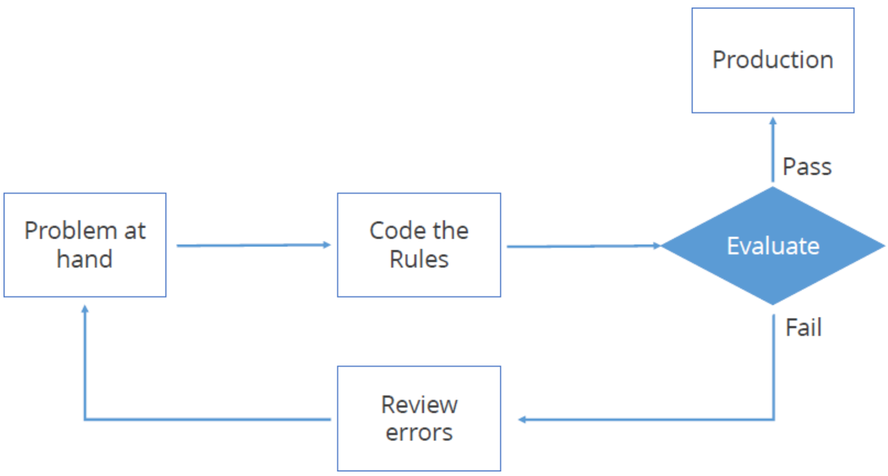
2.1.3. Machine Learning Algorithms

- **Traditional Programming vs. Machine Learning Approach**



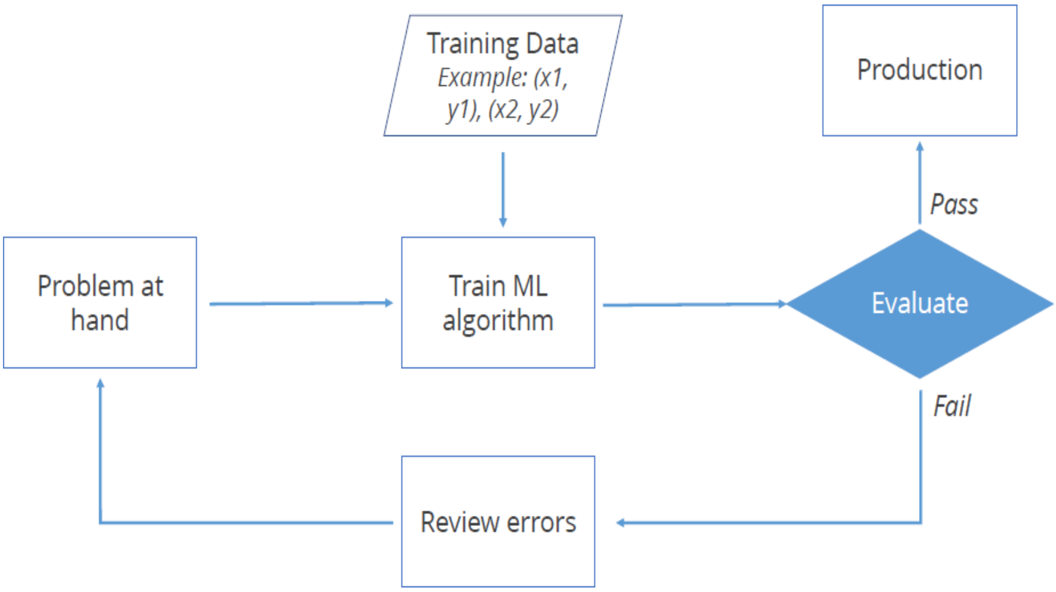
- **Traditional Approach**

Traditional programming relies on **hard-coded rules.**



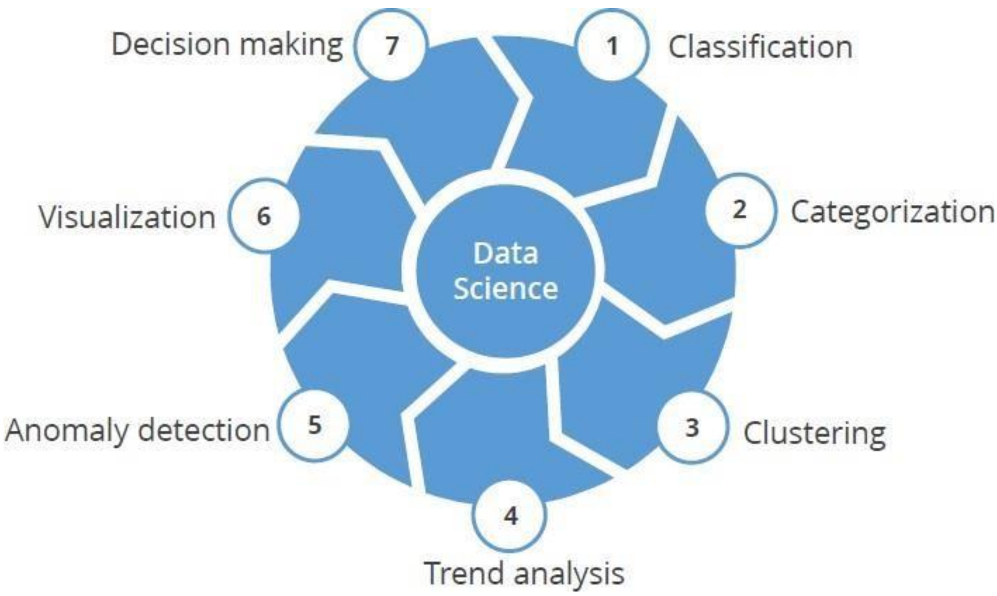
❖ Machine Learning Approach

Machine Learning relies on learning patterns based on sample data.



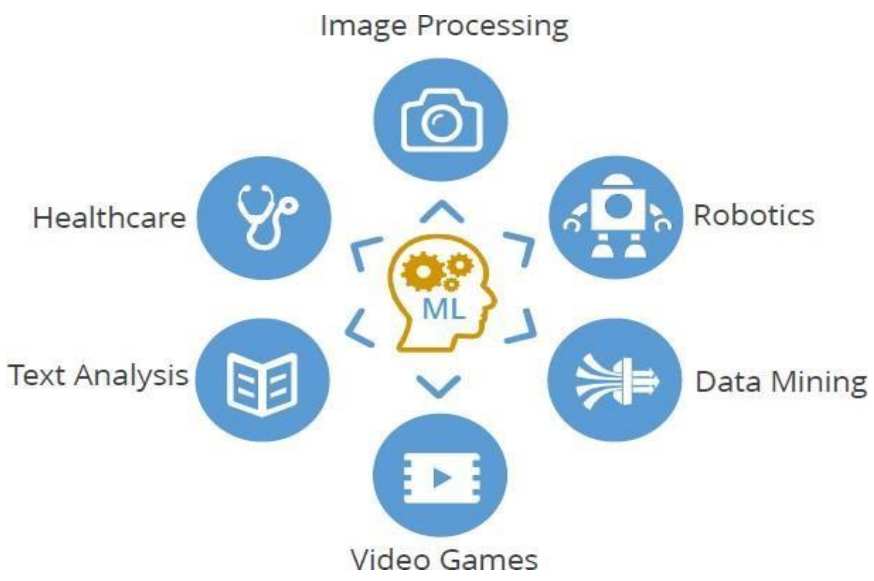
❖ Machine Learning Techniques

- ✓ Machine Learning uses a number of theories and techniques from Data Science.



- ✓ Machine Learning can learn from **labelled data** (known as supervised learning) or **unlabeled data** (known as unsupervised learning).

2.1.4. Applications of Machine Learning



- ❖ **Image Processing**
  - ✓ Optical Character Recognition (OCR)
  - ✓ Self-driving cars
  - ✓ Image tagging and recognition
- ❖ **Robotics**
  - ✓ Industrial robotics
  - ✓ Human simulation
- ❖ **Data Mining**
  - ✓ Association rules
  - ✓ Anomaly detection
  - ✓ Grouping and Predictions
- ❖ **Video games**
  - ✓ Pokémon
  - ✓ PUBG
- ❖ **Text Analysis**
  - ✓ Spam Filtering
  - ✓ Information Extraction
  - ✓ Sentiment Analysis
- ❖ **Healthcare**
  - ✓ Emergency Room & Surgery
  - ✓ Research
  - ✓ Medical Imaging & Diagnostics

Daily to-do tasks

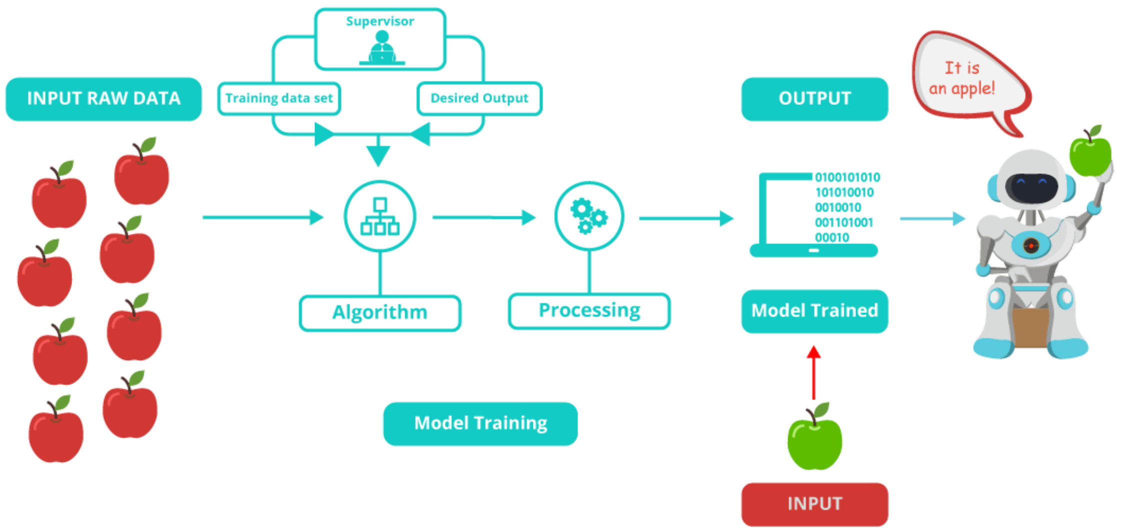
1	Topic	Dataset	Days	Description	notebook name
2	Simple Linear Regression	20311200375 21 (1)	1	simple linear regression apply with analog1, dt, spd, ign	sl.py
3	Multiple Linear Regression	20311200375 21 (1)	1	independent columns = analog1,dt	ml.py
4	Polynomial Regression	20311200375 21 (1)	1	independent columns = analog1,dt dependent = fuel_data	pl.py
5	Logistic regression	20311200375 21 (1)	1	independent columns = analog1,dt dependent = fuel_data	logistic.py
6	Decision Tree	20311200375 21 (1)	1	independent columns = analog1,dt dependent = fuel_data	decisiontree.py
7	Gradientboosting	20311200375 21 (1)	2	independent columns = analog1,dt best result compare to others model	gradient.py
8	random forest	20311200375 21 (1)	2	independent columns = analog1,dt,llv,rtv 2nd best	random.py
9					
10	Simple Linear Regression	19481200066 70.csv	2	simple linear regression apply with dt, llv, rtv, tilt	sl.py
11	Multiple Linear Regression	19481200066 70.csv		independent columns = dt, llv, rtv, tilt	ml.py
12	Polynomial Regression	19481200066 70.csv	1	independent columns = dt, llv, rtv, tilt dependent = fuel_data	pl.py
13	Logistic regression	19481200066 70.csv	1	independent columns = dt, llv, rtv, tilt dependent = fuel_data	logistic.py
14	Decision Tree	19481200066 70.csv	1	independent columns = dt, llv, rtv, tilt dependent = fuel_data	decisiontree.py
15	Gradientboosting	19481200066 70.csv	1	independent columns = dt, llv, rtv, tilt best result compare to others model	gradient.py
16	random forest	19481200066 70.csv	2	independent columns = dt, llv, rtv, tilt 2nd best	random.py
17					
18	Polynomial Regression	Dataset_3.csv	2	independent columns = analog1, dt, llv, rtv, spd, ign dependent = fuel_data	pr.py
19	Gradientboosting	Dataset_3.csv	1	independent columns = analog1, dt, llv, rtv, spd, ign dependent = fuel_data	gb.py
20	random forest	Dataset_3.csv	1	independent columns = analog1, dt, llv, rtv, spd, ign dependent = fuel_data	rf.py
21					
22					
23	Gradientboosting	MP-09-HH-8206.csv	1	independent columns = analog1, dt, llv, rtv, spd, ign dependent = fuel_data	Gb.py
24	random forest	MP-09-HH-8206.csv	1	independent columns = analog1, dt, llv, rtv, spd, ign dependent = fuel_data	rf.py
25	clustering	MP-09-HH-8206.csv	2	Hierarchical clustering , K-means Clustering	hiera.py , kmeans.py
26	task of 5-11	MP-09-HH-8206.csv	1	task of predicting result with manually puting test data like first 10,20	Task_5-11.py
27	task of 6-11	MP-09-HH-8206.csv	2	try to add new column and apply decision tree	Task_6-11.py
28	task of 13-11	any dataset	1	Feature_selection with features like co-relation-map, heat-map with seaborn	Task_13-11.py
29	task of 20-11	MP-09-HH-8206.csv	2	learn supervised learning like association	Task_20-11.py
30	task of 30-11	MP-09-HH-8206.csv	1	clustering with new dataset	Task_23-11.py
31	time-series	try with all datasets if possible	3	Learn basic and try to apply time series with arima model, in these types of datasets time-series is	timeseries.py
32	stream-line	learn streamline	2	Learn stream line of basics.	streamline.py
33	EDA	try to learn	2	Learn basics of Exploratry data analysis	Eda.py

2.2. Techniques of Machine Learning

2.2.1. Supervised Learning

❖ **Define Supervised Learning**

Supervised learning is the machine learning task of learning a function that maps an input to an output based on example input-output pairs. It infers a function from labeled training data consisting of a set of training examples.



In supervised learning, each example is a pair consisting of an input object (typically a vector) and a desired output value (also called the supervisory signal).

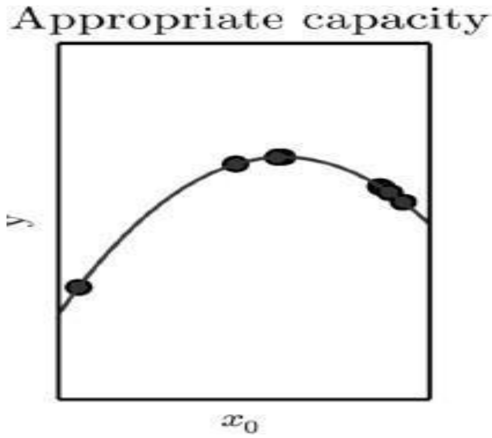
❖ Supervised Learning Flow

- ✓ **Data Preparation**
  - Clean data
  - Label data (x, y)
  - Feature Engineering
  - Reserve 80% of data for Training (Train_X) and 20% for Evaluation (Train_E)
- ✓ **Training Step**
  - Design algorithmic logic
  - Train the model with Train X
  - Derive the relationship between x and y, that is, y = f(x)
- ✓ **Evaluation or Test Step**
  - Evaluate or test with Train E
  - If accuracy score is high, you have the final learned algorithm y = f(x)
  - If accuracy score is low, go back to training step
- ✓ **Production Deployment**
  - Use the learned algorithm y = f(x) to predict production data.

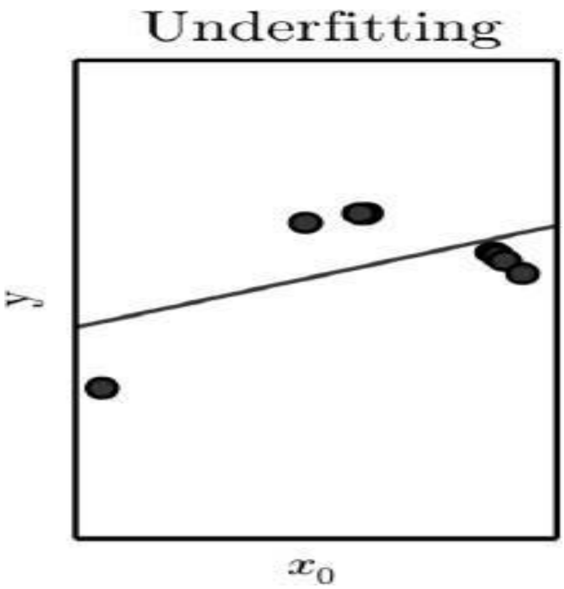
The algorithm can be improved by more training data, capacity, or algo redesign.

❖ Testing the Algorithms

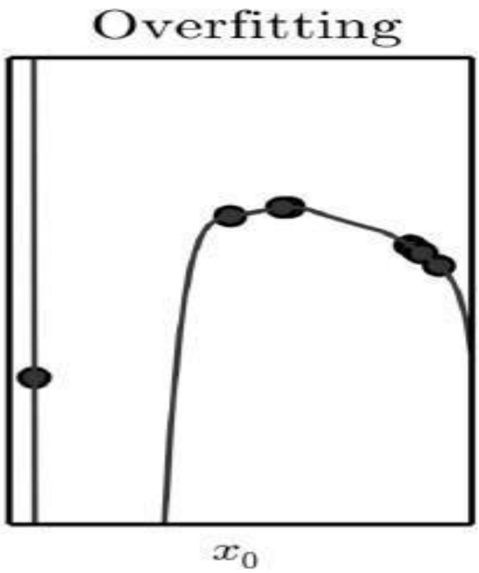
- ✓ Once the algorithm is trained, test it with test data (a set of data instances that do not appear in the training set).
- ✓ A well-trained algorithm can predict well for new test data.



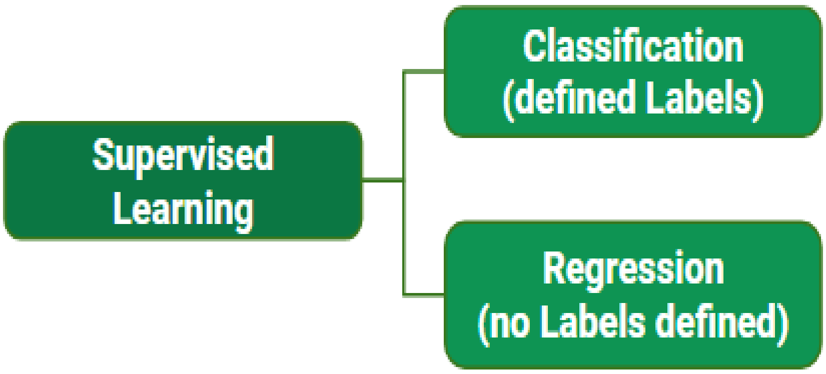
- ✓ If the learning is poor, we have an underfitted situation. The algorithm will not work well on test data. Retraining may be needed to find a better fit.



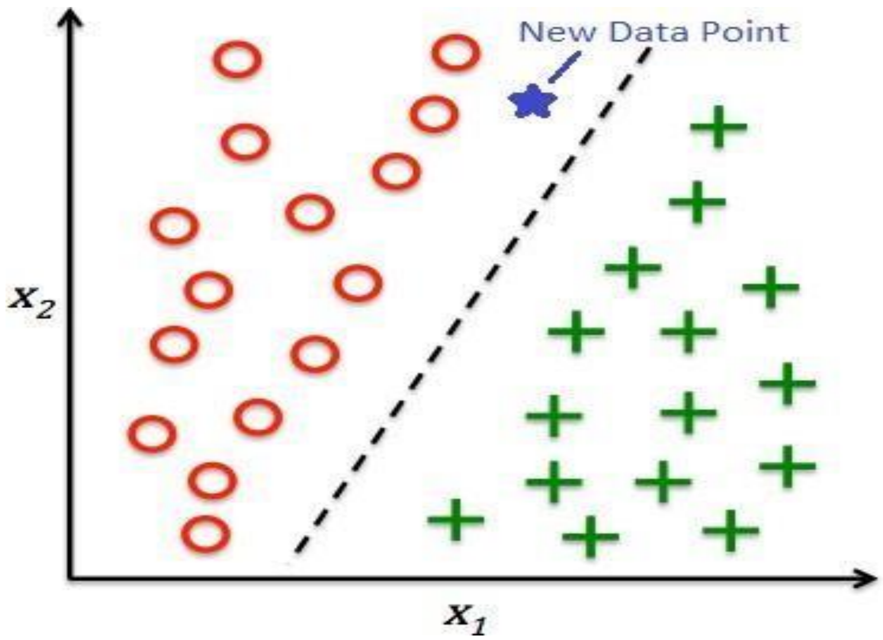
- ✓ If learning on training data is too intensive, it may lead to overfitting–a situation where the algorithm is not able to handle new testing data that it has not seen before. The technique to keep data generic is called **regularization.**



- **Examples of Supervised Learning**
  - Voice Assistants
  - Gmail Filters
  - Weather Apps
- **Types of Supervised Learning**



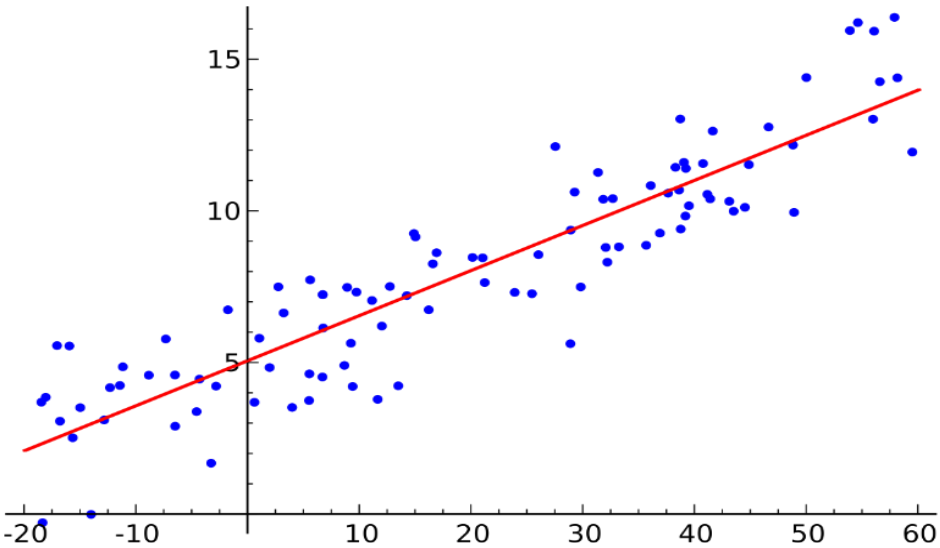
  - **Classification**
    - Answers "**What class**?"



- Applied when the output has finite and discreet values Example: Social media sentiment analysis has three potential outcomes, positive, negative, or neutral

✓ **Regression**

- Answers “**How much**?”

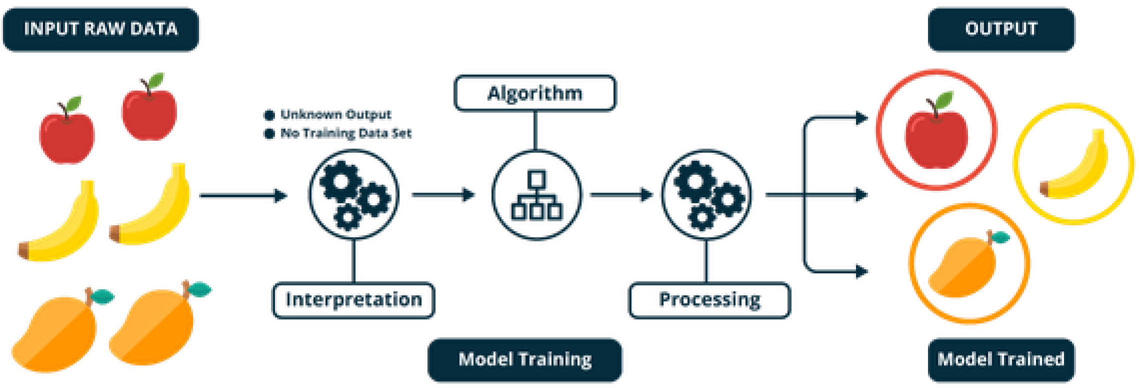


- Applied when the output is a continuous number
- A simple regression algorithm: **y = wx + b**. Example: relationship between environmental temperature (y) and humidity levels (x)

2.2.2. Unsupervised Learning

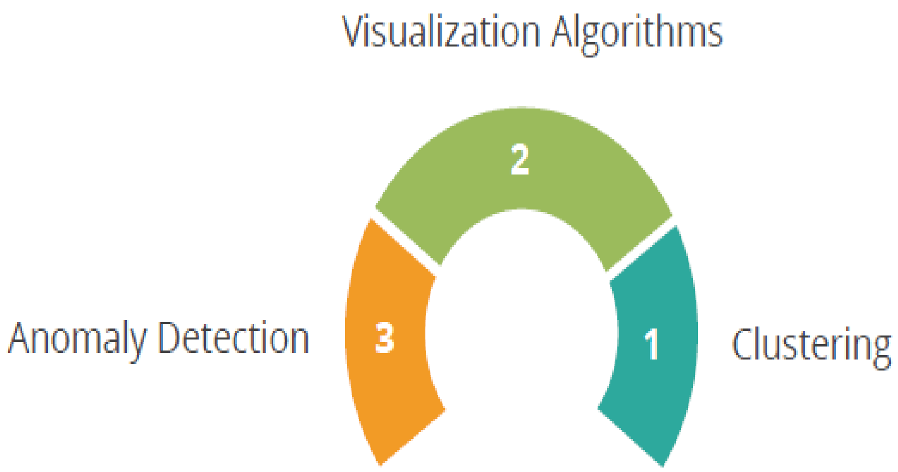
❖ **Define Unsupervised Learning**

Unsupervised learning is the training of machine using information that is neither classified nor labeled and allowing the algorithm to act on that information without guidance.



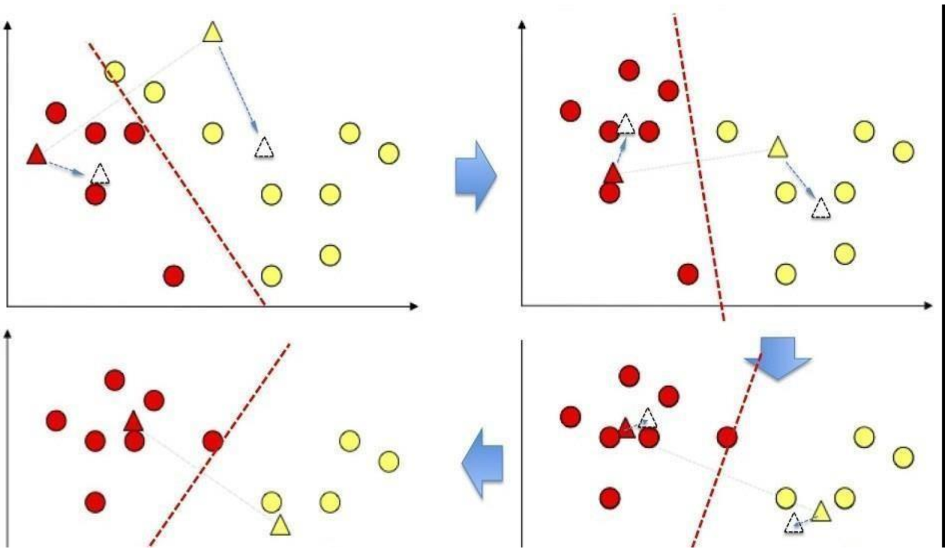
Here the task of machine is to group unsorted information according to similarities, patterns and differences without any prior training of data.

- **Types of Unsupervised Learning**



- **Clustering**

The most common unsupervised learning method is cluster analysis. It is used to find data clusters so that each cluster has the most closely matched data.



✓ **Visualization Algorithms**

Visualization algorithms are unsupervised learning algorithms that accept unlabeled data and display this data in an intuitive 2D or 3D format. The data is separated into somewhat clear clusters to aid understanding.

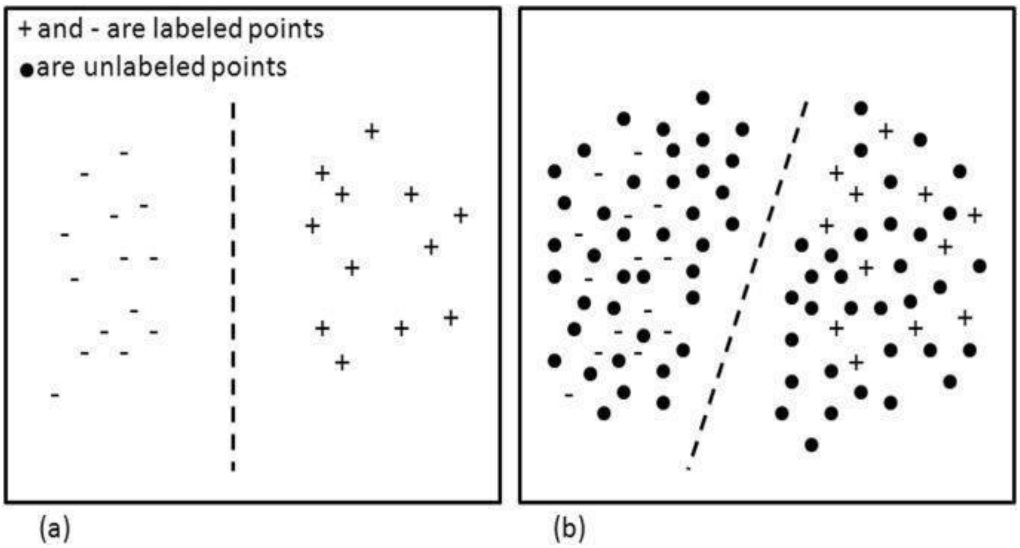
✓ **Anomaly Detection**

This algorithm detects anomalies in data without any prior training.

2.2.3. Semi- supervised Learning

❖ **Define Semi-supervised Learning**

Semi-supervised learning is a class of machine learning tasks and techniques that also make use of unlabeled data for training – typically a small amount of labeled data with a large amount of unlabeled data.



Semi-supervised learning falls between unsupervised learning (without any labeled training data) and supervised learning (with completely labeled training data).

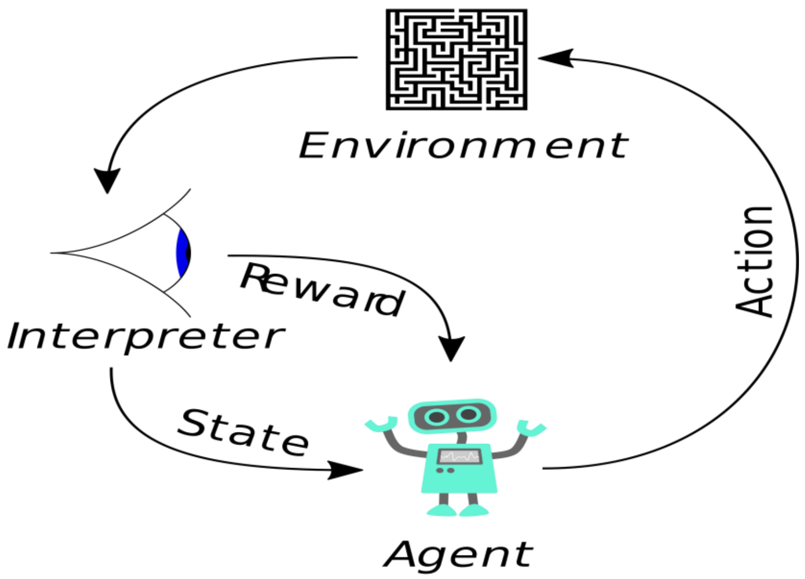
❖ **Example of Semi-supervised Learning**

- Google Photos automatically detects the same person in multiple photos from a vacation trip (clustering –unsupervised).
- One has to just name the person once (supervised), and the name tag gets attached to that person in all the photos.

2.2.4. Reinforcement Learning

❖ **Define Reinforcement Learning**

Reinforcement Learning is a type of Machine Learning that allows the learning system to observe the environment and learn the ideal behavior based on trying to maximize some notion of cumulative reward.



It differs from supervised learning in that labelled input/output pairs need not be presented, and sub-optimal actions need not be explicitly corrected. Instead the focus is finding a balance between exploration (of uncharted territory) and exploitation (of current knowledge)

❖ **Features of Reinforcement Learning**

- The learning system (agent) observes the environment, selects and takes certain actions, and gets rewards in return (or penalties in certain cases).
- The agent learns the strategy or policy (choice of actions) that maximizes its rewards over time.

❖ **Example of Reinforcement Learning**

- In a manufacturing unit, a robot uses deep reinforcement learning to identify a device from one box and put it in a container.
- The robot learns this by means of a rewards-based learning system, which incentivizes it for the right action.

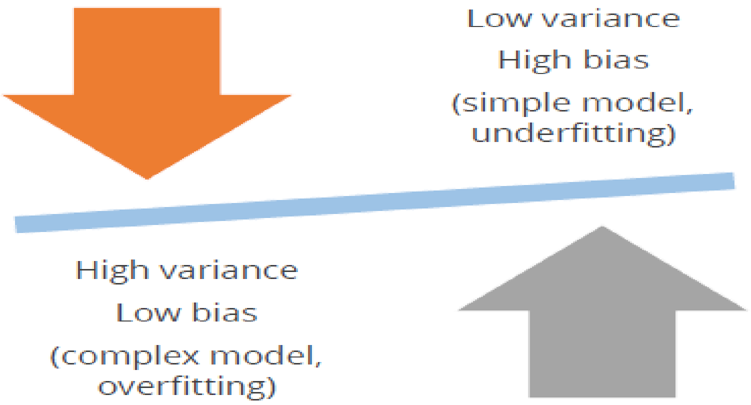
2.2.5. Some Important Considerations in Machine Learning

❖ **Bias & Variance Tradeoff**

- ➢ Bias refers to error in the machine learning model due to wrong assumptions. A high-bias model will underfit the training data.
- ➢ Variance refers to problems caused due to overfitting. This is a result of over-sensitivity of the model to small variations in the training data. A model with

many degrees of freedom (such as a high-degree polynomial model) is likely to have high variance and thus overfit the training data.

- **Bias & Variance Dependencies**
  - Increasing a model's complexity will reduce its bias and increase its variance.



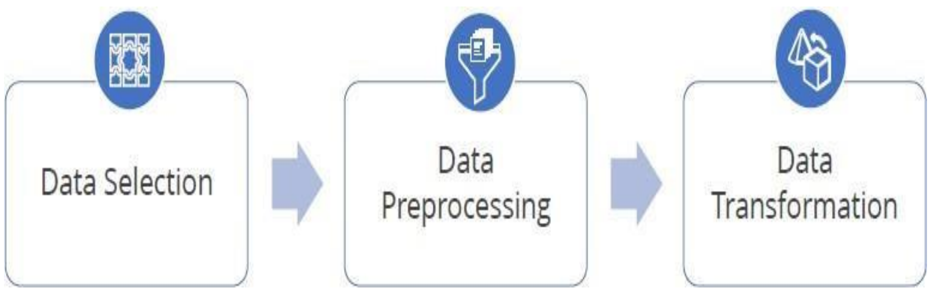
  - Conversely, reducing a model's complexity will increase its bias and reduce its variance. This is why it is called a tradeoff.
- **What is Representational Learning**

In Machine Learning, Representation refers to the way the data is presented. This often make a huge difference in understanding.

2.3. Data Preprocessing

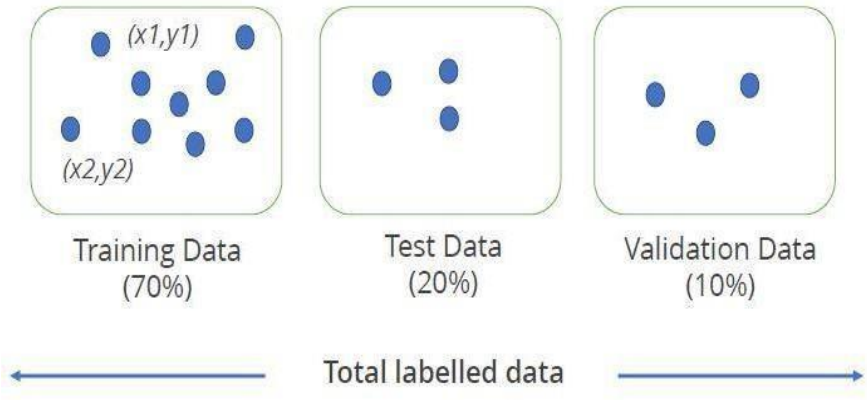
2.3.1. Data Preparation

- **Data Preparation Process**
  - ✓ Machine Learning depends largely on test data.
  - ✓ Data preparation involves data selection, filtering, transformation, etc.



  - ✓ Data preparation is a crucial step to make it suitable for ML.
  - ✓ A large amount of data is generally required for the most common forms of ML.

- ❖ **Types of Data**
  - ✓ Labelled Data or Training Data

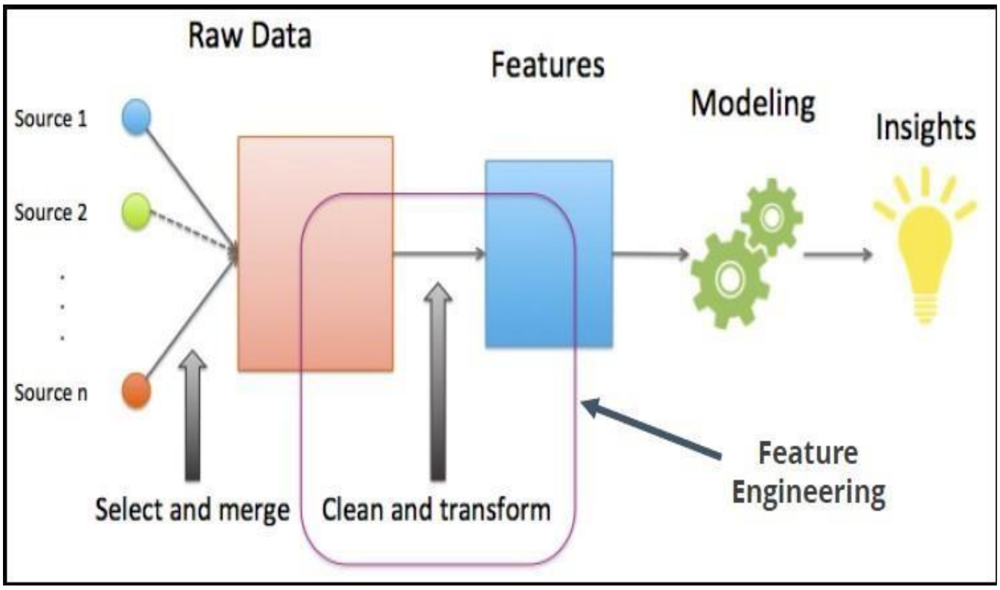


  - ✓ Unlabeled Data
  - ✓ Test Data
  - ✓ Validation Data

2.3.2. Feature Engineering

- ❖ **Define Feature Engineering**

The transformation stage in the data preparation process includes an important step known as Feature Engineering.



Feature Engineering refers to selecting and extracting right features from the data that are relevant to the task and model in consideration.

- ❖ **Aspects of Feature Engineering**
  - ✓ Feature Selection

Most useful and relevant features are selected from the available data

  - ✓ Feature Addition

New features are created by gathering new data

- ✓ Feature Extraction

Existing features are combined to develop more useful ones

- ✓ Feature Filtering

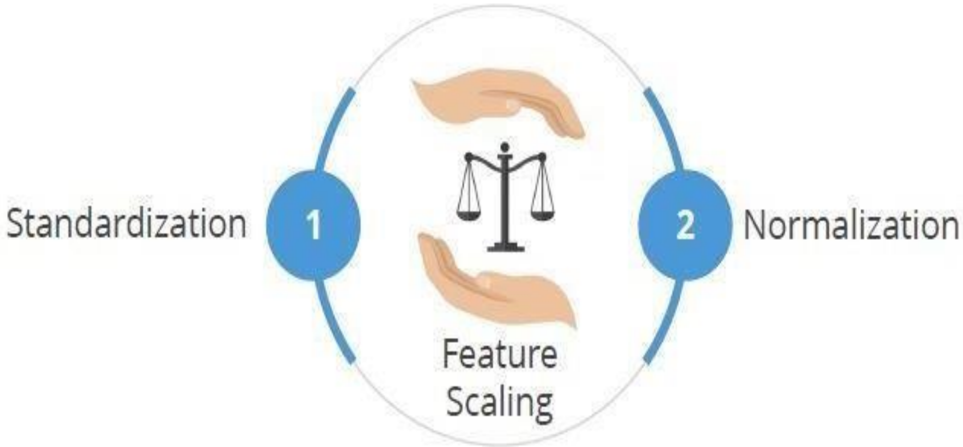
Filter out irrelevant features to make the modelling step easy

2.3.3. Feature Scaling

❖ **Define Feature Scaling**

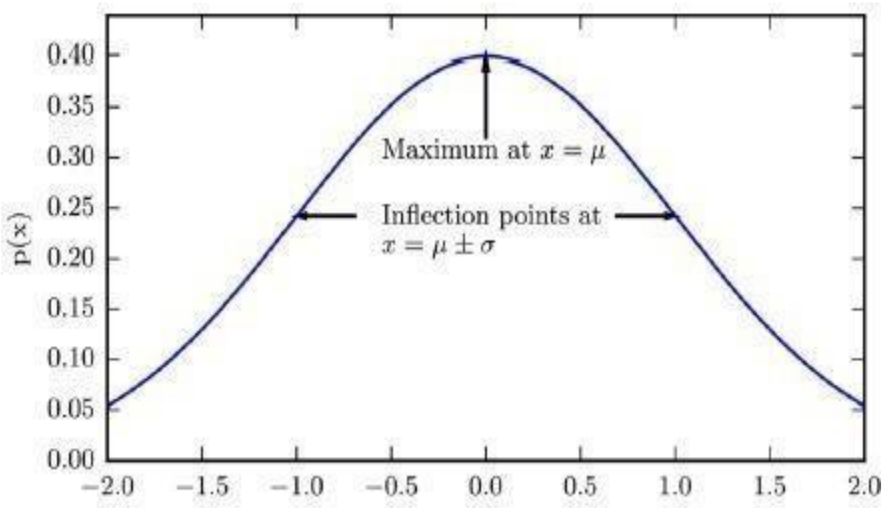
- ✓ Feature scaling is an important step in the data transformation stage of data preparation process.
- ✓ Feature Scaling is a method used in Machine Learning for standardization of independent variables of data features.

❖ **Techniques of Feature Scaling**



- ✓ **Standardization**
  - Standardization is a popular feature scaling method, which gives data the property of a standard normal distribution (also known as Gaussian distribution).
  - All features are standardized on the normal distribution (a mathematical model).
  - The mean of each feature is centered at zero, and the feature column has a standard deviation of one.

$$x'_j = \frac{x_j - \mu_j}{\sigma_j}$$



The ML library scikit-learn implements a class for standardization called StandardScaler, as demonstrated here:

```
>>> from sklearn.preprocessing import StandardScaler
>>> stdsc = StandardScaler()
>>> X_train_std = stdsc.fit_transform(X_train)
>>> X_test_std = stdsc.transform(X_test)
```

✓ **Normalization**

- In most cases, normalization refers to rescaling of data features between 0 and 1, which is a special case of Min-Max scaling.

$$x_{norm}^{(i)} = \frac{x^{(i)} - x_{\min}}{x_{\max} - x_{\min}}$$

- In the given equation, subtract the min value for each feature from each feature instance and divide by the spread between max and min.
- In effect, it measures the relative percentage of distance of each instance from the min value for that feature.

The ML library scikit-learn has a MinMaxScaler class for normalization.

It implements normalization as explained on the previous slide.

```
>>> from sklearn.preprocessing import MinMaxScaler
>>> mms = MinMaxScaler()
>>> X_train_norm = mms.fit_transform(X_train)
>>> X_test_norm = mms.transform(X_test)
```

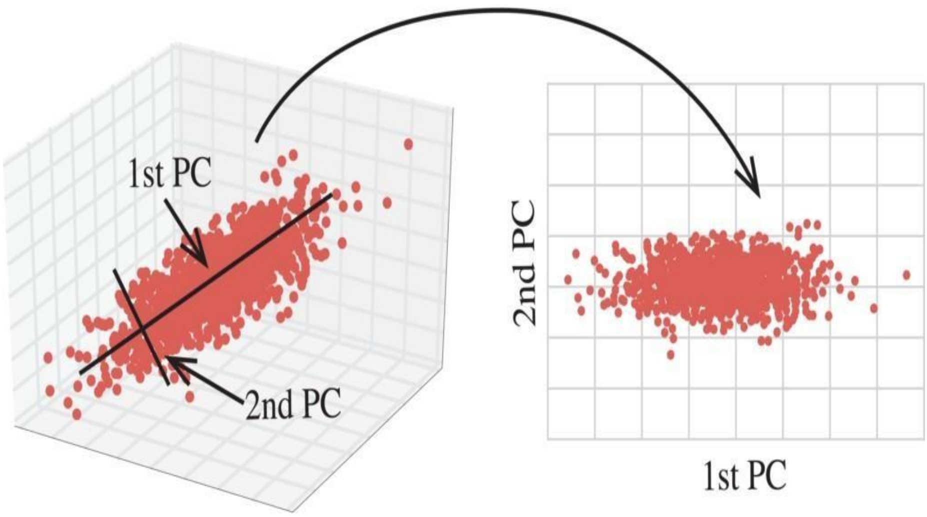
2.3.4. Datasets

- Machine Learning problems often need training or testing datasets.
- A dataset is a large repository of structured data.
- In many cases, it has input and output labels that assist in Supervised Learning.

2.3.5. Dimensionality Reduction with Principal Component Analysis

❖ **Define Dimensionality Reduction**

- ✓ Dimensionality reduction involves transformation of data to new dimensions in a way that facilitates discarding of some dimensions without losing any key information.



❖ **Define Principal Component Analysis (PCA)**

- ✓ Principal component analysis (PCA) is a technique for dimensionality reduction that helps in arriving at better visualization models.

1. Let $\mu = \frac{1}{m}\sum_{i=1}^{m} x^{(i)}$.
2. Replace each $x^{(i)}$ with $x^{(i)} - \mu$.
3. Let $\sigma_j^2 = \frac{1}{m}\sum_i (x_j^{(i)})^2$
4. Replace each $x_j^{(i)}$ with $x_j^{(i)}/\sigma_j$.

❖ **Applications of PCA**
  - ✓ Noise reduction
  - ✓ Compression
  - ✓ Preprocess

2.4. Math Refresher

2.4.1. Concept of Linear Algebra

❖ **Linear Equation**
  - Linear algebra is a branch of mathematics that deals with the study of vectors and linear functions and equations.

A linear equation of n variables is of the form:

$a_1x_1 + a_2x_2 + \ldots + a_nx_n = b$

where $x_1, x_2, \ldots, x_n$ are the unknown quantities to be found, $a_1, \cdots, a_n$ are the coefficients (given numbers), and b is the constant term.

  - A linear equation does not involve any products, inverses, or roots of variables. All variables occur only to the first power and not as arguments for trigonometric, logarithmic, or exponential functions.

❖ **System of Linear Equations**
  - A system of linear equations is a finite collection of linear equations.

A linear system of m equations in n variables has the form:

$a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = b_1$
$a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = b_2$
....
$a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n = b_m$

In the case of a single linear equation, a linear system can have infinitely many solutions, one solution, or no solutions at all.

- A linear system that has a solution is called consistent, and the one with no solution is termed inconsistent.

- **Matrix**

$$
\begin{array}{c}
\\ 1 \\ 2 \\ 3 \\ \vdots \\ m
\end{array}
\begin{array}{c}
\begin{array}{ccccc} 1 & 2 & \cdots & n \end{array} \\
\begin{bmatrix}
a_{11} & a_{12} & \cdots & a_{1n} \\
a_{21} & a_{22} & \cdots & a_{2n} \\
a_{31} & a_{32} & \cdots & a_{3n} \\
\vdots & \vdots & \vdots & \vdots \\
a_{m1} & a_{m2} & \cdots & a_{mn}
\end{bmatrix}
\end{array}
$$

An m × n matrix: the m rows are horizontal and the n columns are vertical. Each element of a matrix is often denoted by a variable with two subscripts. For example, a2,1 represents the element at the second row and first column of the matrix.

- ✓ **Addition**

Two matrices can be added only if they have the same number of rows and columns. Also, during addition, A + B = B + A

Consider the following two matrices:

$$A = \begin{pmatrix} 22 & 32 \\ 11 & 16 \end{pmatrix} \qquad B = \begin{pmatrix} 13 & 8 \\ 13 & 16 \end{pmatrix}$$

$$A + B = \begin{pmatrix} 22 + 13 & 32 + 8 \\ 11 + 13 & 16 + 16 \end{pmatrix}$$

The corresponding elements in the rows are added.

$$A + B = \begin{pmatrix} 35 & 40 \\ 24 & 32 \end{pmatrix}$$

- ✓ **Subtraction**

Two matrices can be subtracted only if they have the same number of rows and columns. Also, during subtraction, A -B not equal to B -A

Now consider the same matrices again:

$$A = \begin{pmatrix} 22 & 32 \\ 11 & 16 \end{pmatrix} \qquad B = \begin{pmatrix} 13 & 8 \\ 13 & 16 \end{pmatrix}$$

$$A - B = \begin{pmatrix} 22 - 13 & 32 - 8 \\ 11 - 13 & 16 - 16 \end{pmatrix}$$

The corresponding elements in the rows are subtracted.

$$A - B = \begin{pmatrix} 9 & 24 \\ -2 & 0 \end{pmatrix}$$

✓ Multiplication

The matrix product AB is defined only when the number of columns in A is equal to the number of rows in B. BA is defined only when the number of columns in B is equal to the number of rows in A. AB is not always equal to BA.

Consider the same matrices again:

$$A = \begin{pmatrix} 22 & 32 \\ 11 & 16 \end{pmatrix} \qquad B = \begin{pmatrix} 13 & 8 \\ 13 & 16 \end{pmatrix}$$

Let AB = C. To compute the value of every element in the 2 x 2 matrix C, use the formula $C_{ik} = \Sigma_j A_{ij}B_{jk}$,

Dot product

$$A.B = \begin{pmatrix} (22 \times 13) + (32 \times 13) & (22 \times 8) + (32 \times 16) \\ (11 \times 13) + (16 \times 13) & (11 \times 8) + (16 \times 16) \end{pmatrix}$$

The 1st and 2nd rows of A are multiplied with the 1st and 2nd columns of B and added.

$$C = \begin{pmatrix} 702 & 688 \\ 351 & 344 \end{pmatrix}$$

✓ Transpose

A transpose is a matrix formed by turning all the rows of a given matrix into columns and vice versa. The transpose of matrix A is denoted as A^T

From the previous examples:

$$A = \begin{pmatrix} 22 & 32 \\ 11 & 16 \end{pmatrix}$$

$$A^T = \begin{pmatrix} 22 & 11 \\ 32 & 16 \end{pmatrix}$$

The rows become columns and vice versa.

✓ Inverse

An n-by-n square matrix A is called invertible (also nonsingular or nondegenerate) if there exists an n-by-n square matrix B such that

$AB = BA = I_n$

where I_n denotes the n-by-n **identity matrix** and the multiplication used is ordinary matrix multiplication.

When the matrix B is uniquely determined by A, it is called the inverse of A, denoted by A^{-1}

❖ Special Types of Matrix

➢ Diagonal Matrix

DIAGONAL MATRICES

A square matrix which consists of all zeros off the main diagonal is called a **diagonal matrix**.

Example:
$$\begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & \frac{1}{3} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 5 \end{bmatrix}$$

- **Symmetric Matrix**

Symmetric matrix & Skew Symmetric Matrix

- Symmetric: $A^T = A$.
- Skew-symmetric: $A^T = -A$.
- Examples:

$$\begin{bmatrix} 1 & 1 & -1 \\ 1 & 2 & 0 \\ -1 & 0 & 5 \end{bmatrix} \qquad \begin{bmatrix} 0 & 1 & -2 \\ -1 & 0 & 3 \\ 2 & -3 & 0 \end{bmatrix}$$

symmetric   skew-symmetric

- **Identity Matrix**

The identity matrix I_n is a n x n square matrix with the main diagonal of 1's and all other elements are 0's.

$$I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad I_3 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

If A is a m x n matrix, then

$$I_m A = A \text{ and } AI_n = A$$

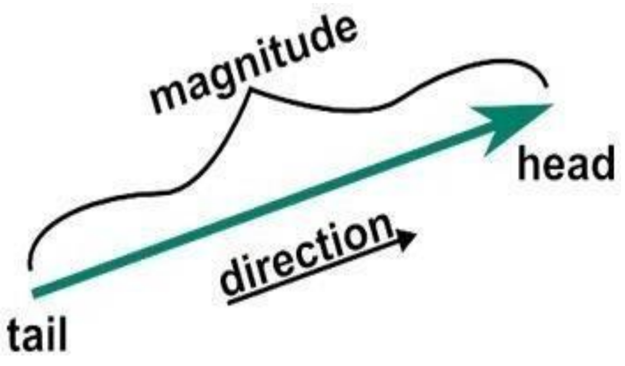
If A is a n x n matrix, then

$$AI_n = I_n A = A$$

❖ Vector

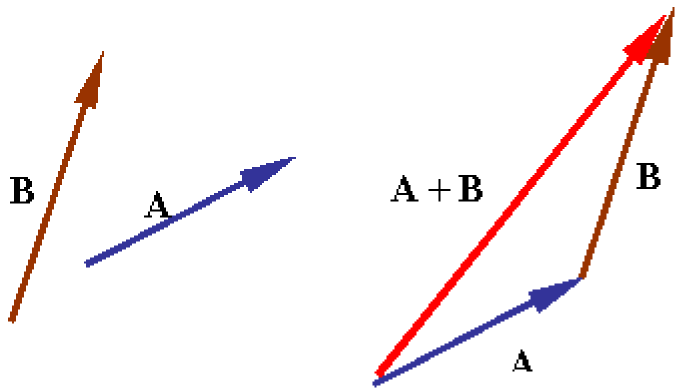
A vector (v) is an object with both magnitude (length) and direction.

It starts from origin (0,0), and its length is denoted by ||v||.



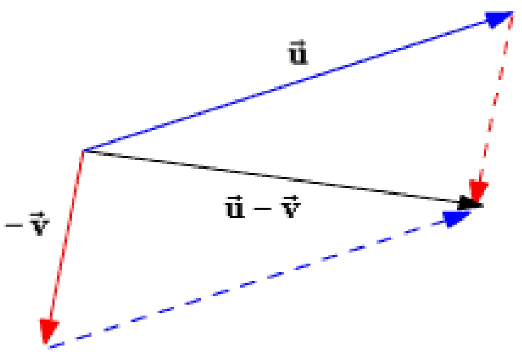
➢ Addition

- The operation of adding two or more vectors together into a vector sum is referred to as vector addition.



➢ Subtraction

- Vector subtraction is the process of subtracting two or more vectors to get a vector difference.



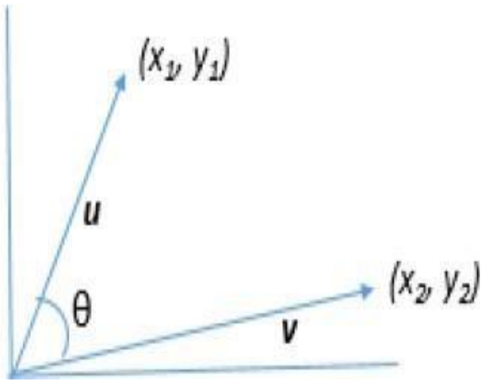
- **Multiplication**
  - Vector multiplication refers to a technique for the multiplication of two (or more) vectors with themselves.

$u*v = z$

$u*v = (x_1x_2 + y_1y_2) = \sum(x_iy_i)$

This can be shown to equal :

$u*v = ||x||\ ||y|| \cos\theta$



2.4.2. Eigenvalues, Eigenvectors, and Eigen decomposition

- **Eigenvalue & Eigenvector**
  - An eigenvector of a square matrix A is a non-zero vector such that multiplication by A alters only the scale of v.

$$Av = \lambda v$$

where λ is eigenvalue corresponding to this eigenvector.

- **Eigen decomposition**
  - Integers can be broken into their prime factors to understand them, example: 12 = 2 x 2 x 3. From this, useful properties can be derived, for example, the number is not divisible by 5 and is divisible by 2 and 3.
  - Similarly, matrices can be decomposed. This will help you discover information about the matrix.

If A has a set of eigenvectors $v_1, v_2, \ldots$ represented by matrix V, and the corresponding eigenvalues $\lambda_1, \lambda_2, \ldots$ represented by vector λ, then eigendecomposition of A is given by:

$$A = V\, diag(\lambda)\, V^{-1}$$

2.4.3. Introduction to Calculus

Calculus is the study of change. It provides a framework for modelling systems in which there is change and ways to make predictions of such models.

- **Differential Calculus**
  - ✓ Differential calculus is a part of calculus that deals with the study of the rates at which quantities change.
  - ✓ Let x and y be two real numbers such that y is a function of x, that is, y = f(x).

- ✓ If f(x) is the equation of a straight line (linear equation), then the equation is represented as y = mx + b.
- ✓ Where m is the slope determined by the following equation:

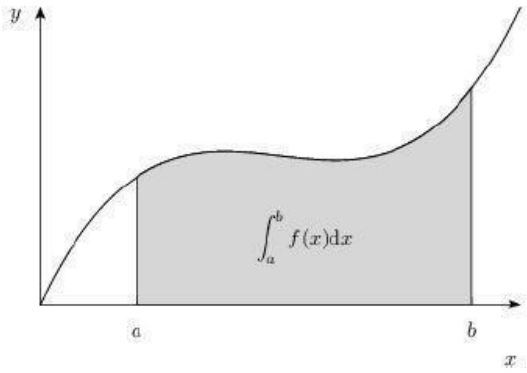
$$m = \frac{\text{change in } y}{\text{change in } x} = \frac{\Delta y}{\Delta x}$$

Δy/Δx or dy/dx is the derivative of y with respect to x and is also the rate of change of y per unit change in x.

❖ **Integral Calculus**

- ✓ Integral Calculus assigns numbers to functions to describe displacement, area, volume, and other concepts that arise by combining infinitesimal data.
- ✓ Given a function f of a real variable x and an interval [a, b] of the real line, the definite integral is defined informally as the signed area of the region in the xy-plane that is bounded by the graph of f, the x -axis, and the vertical lines x=a and x=b.

$$\int_a^b f(x)\,dx$$



2.4.4. Probability and Statistics

❖ **Probability Theory**

- ➢ Probability is the measure of the likelihood of an event's occurrence.
- ➢ Example: The chances of getting heads on a coin toss is ½ or 50%

$$P\,(\text{heads}) = \frac{1}{2} = 0.5$$

- Probability of any specific event is between 0 and 1 (inclusive). The sum of total probabilities of an event cannot exceed 1, that is, 0 <= p(x) <= 1. This implies that. p(x)dx =1 (integral of p for a distribution over x)

- **Conditional Probability**
  - Conditional Probability is a measure of the probability of an event occurring given that another event has occurred.

Probability of event A occured and event B occured

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

Probability of event A given B has occured

Probability of event B

- **Chain Rule of Probability**
  - Joint probability distribution over many random variables may be decomposed into conditional distributions over only one variable.
  - It can be represented as:

$$P(\mathrm{x}^{(1)}, \ldots, \mathrm{x}^{(n)}) = P(\mathrm{x}^{(1)}) \Pi_{i=2}^{n} P(\mathrm{x}^{(i)} \mid \mathrm{x}^{(1)}, \ldots, \mathrm{x}^{(i-1)})$$

For example: P(a, b, c) = P(a | b, c) * P(b | c) * P(c)

- **Standard Deviance**
  - Standard deviation is a quantity that expresses the value by which the members of a group differ from the mean value for the group.

$$\mathrm{SD} = \sqrt{\frac{\sum |x - \bar{x}|^2}{n}}$$

  - Standard deviation is used more often than variance because the unit in which it is measured is the same as that of mean, a measure of central tendency.

- **Variance**
  - Variance refers to the spread of the data set, for example, how far the numbers are in relation to the mean.
  - Variance is particularly useful when calculating the probability of future events or performance.

$$\sigma^2 = \frac{\Sigma(x-\mu)^2}{N}$$

- Notice that variance is just the square of standard deviation.

- **Covariance**
  - Covariance is the measure of how two random variables change together. It is used to calculate the correlation between variables.

Formula to find the mean for X

$$\mu_x = \frac{\sum_{i=1}^{n} x_i}{n}$$

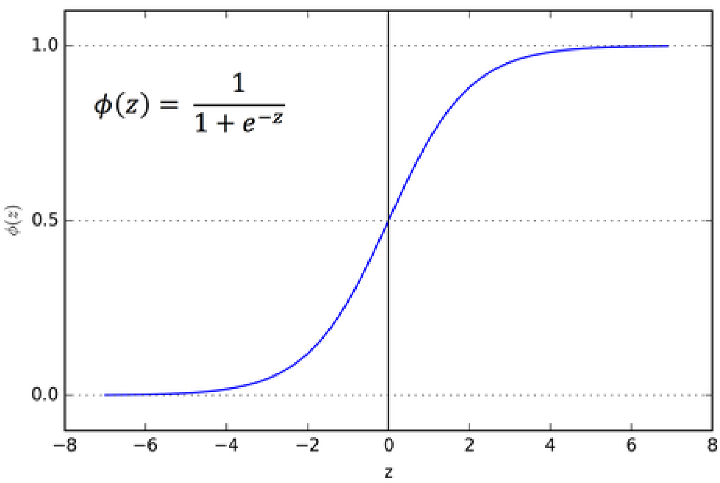
Formula to find the mean for Y

$$\mu_y = \frac{\sum_{i=1}^{n} y_i}{n}$$

Formula to find covariance of X & Y

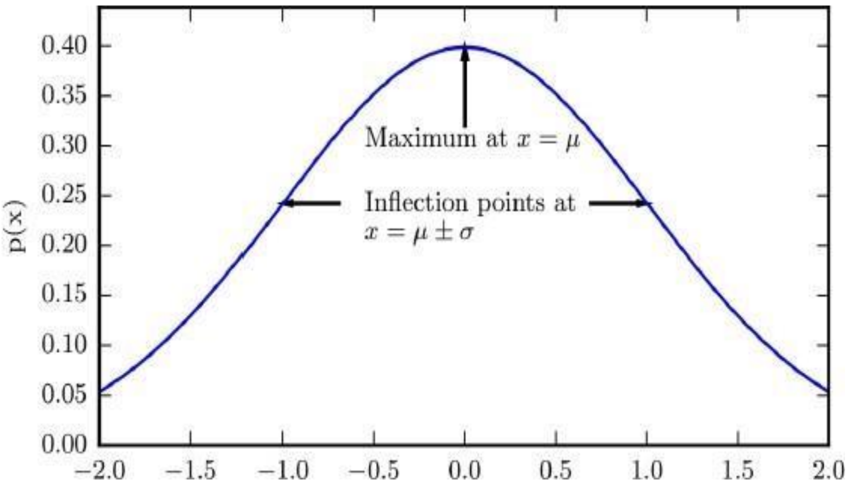
$$cov(X,Y) = \frac{\sum_{i=1}^{n}(x_i - \mu_x)(y_i - \mu_y)}{(n-1)}$$

- **Logistic Sigmoid**
  - The Logistic Sigmoid is a useful function that follows the S curve. It saturates when input is very large or very small.



❖ **Gaussian Distribution**

➢ The distribution where the data tends to be around a central value with lack of bias or minimal bias toward the left or right is called Gaussian distribution, also known as normal distribution.



$$\mathcal{N}(x; \mu, \sigma^2) = \sqrt{\frac{1}{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)$$

$$\mathcal{N}(x; \mu, \beta^{-1}) = \sqrt{\frac{\beta}{2\pi}} \exp\left(-\frac{1}{2}\beta(x-\mu)^2\right)$$

- μ = mean or peak value, which also means E[x] = μ
- σ = standard deviation and σ^2 = variance
- A "standard normal distribution" has μ = 0 and σ = 1
- For efficient handling, invert σ and use precision β

2.5. Supervised learning

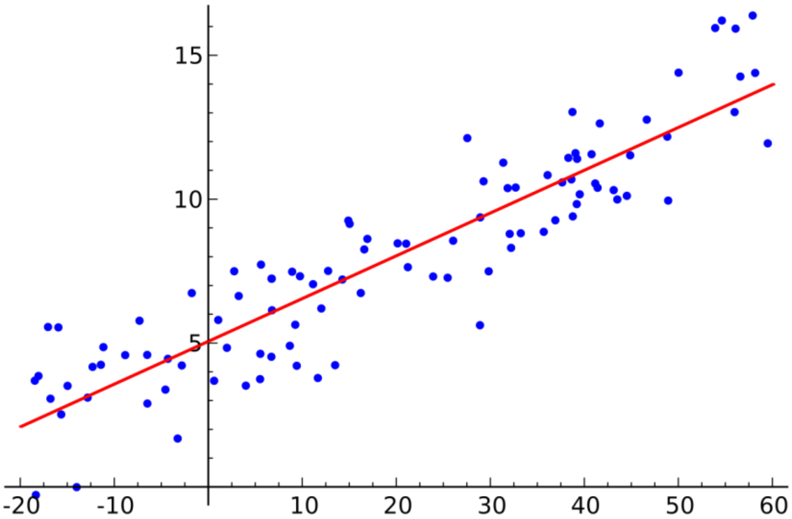
2.5.1. Regression

✓ In statistical modeling, regression analysis is a set of statistical processes for estimating the relationships among variables.

✓ It includes many techniques for modeling and analyzing several variables, when the focus is on the relationship between a dependent variable and one or more independent variables (or 'predictors').

- ✓ More specifically, regression analysis helps one understand how the typical value of the dependent variable (or 'criterion variable') changes when any one of the independent variables is varied, while the other independent variables are held fixed.

2.5.1.1. Linear Regression



- ➢ Linear regression is a linear approach for modeling the relationship between a scalar dependent variable y and an independent variable x.

$$\hat{y} = \boldsymbol{w}^{\top}\boldsymbol{x}$$

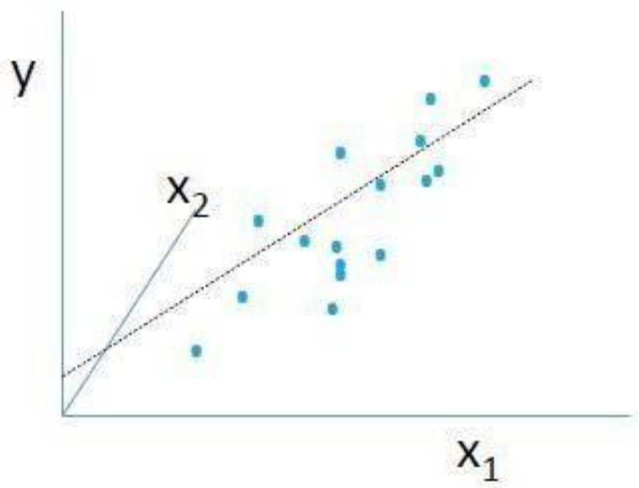
- ➢ where x, y, w are vectors of real numbers and w is a vector of weight parameters.
- ➢ The equation is also written as:

$$\mathbf{y = wx + b}$$

- ➢ where b is the bias or the value of output for zero input

2.5.1.2. Multiple Linear Regression

- It is a statistical technique used to predict the outcome of a response variable through several explanatory variables and model the relationships between them.

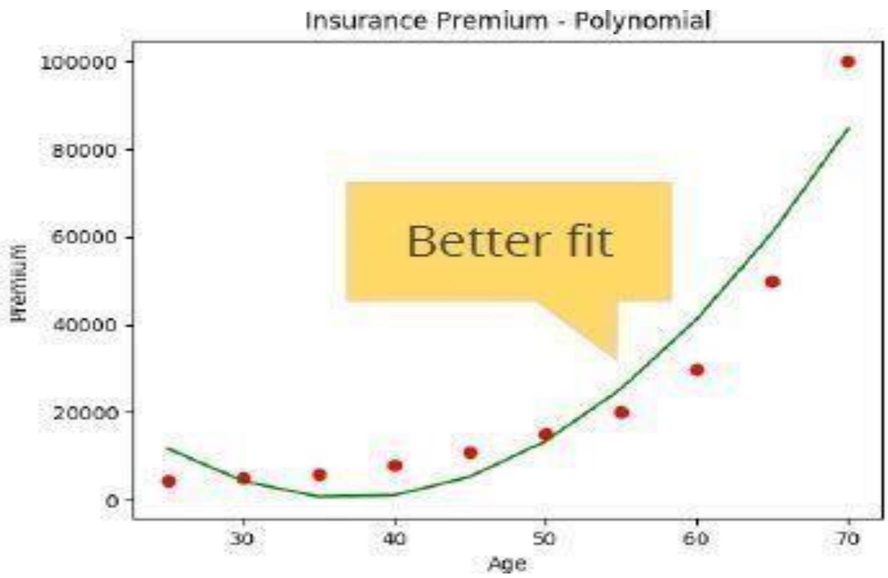


The graph shows dependent variable y plotted against two independent variables x_1 and x_2. It is shown in 3D. More independent variables (if involved) will increase the dimensions further.

- It represents line fitment between multiple inputs and one output, typically:

$$y = w_1x_1 + w_2x_2 + b$$

2.5.1.3. Polynomial Regression



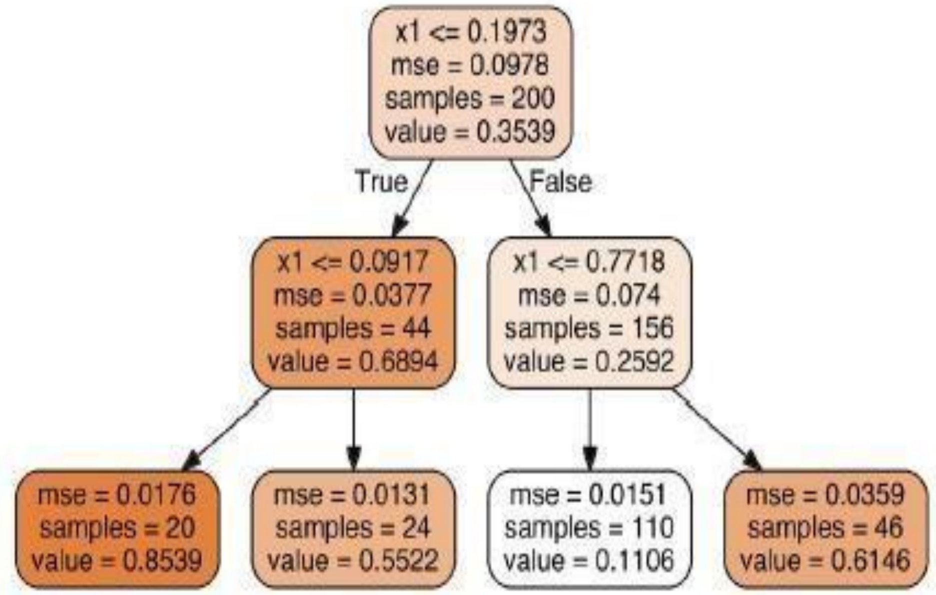
- Polynomial regression is applied when data is not formed in a straight line.
- It is used to fit a linear model to non-linear data by creating new features from powers of non-linear features.

Example: Quadratic features

$$x_2' = x_2^2$$
$$y = w_1x_1 + w_2x_2^2 + 6$$
$$= w_1x_1 + w_2x_2' + 6$$

2.5.1.4. Decision Tree Regression

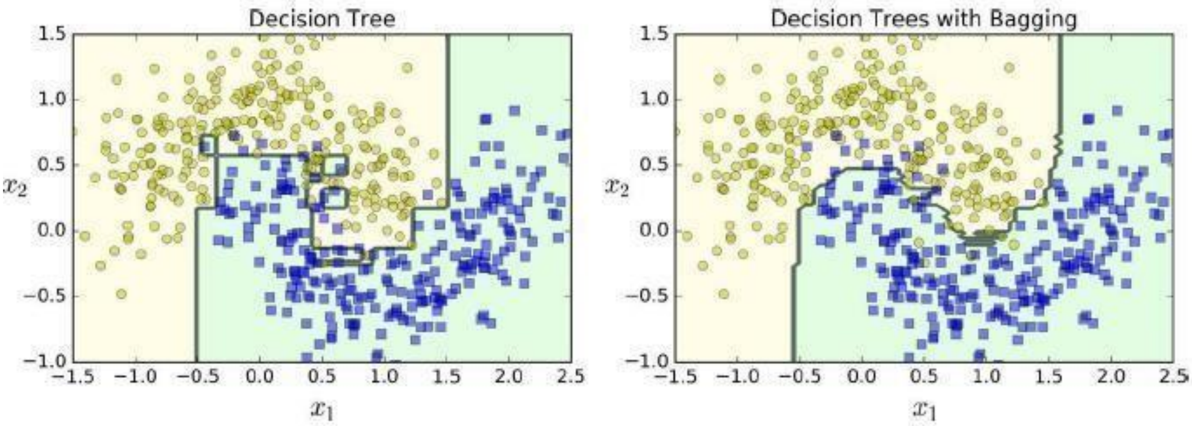
- A decision tree is a graphical representation of all the possible solutions to a decision based on a few conditions.
- Decision Trees are non-parametric models, which means that the number of parameters is not determined prior to training. Such models will normally overfit data.
- In contrast, a parametric model (such as a linear model) has a predetermined number of parameters, thereby reducing its degrees of freedom. This in turn prevents overfitting.



- **max_depth** –limit the maximum depth of the tree
- **min_samples_split** –the minimum number of samples a node must have before it can be split
- **min_samples_leaf** –the minimum number of samples a leaf node must have
- **min_weight_fraction_leaf** –same as min_samples_leaf but expressed as a fraction of total instances
- **max_leaf_nodes** –maximum number of leaf nodes
- **max_features** –maximum number of features that are evaluated for splitting at each node

2.5.1.5. Random Forest Regression

- Ensemble Learning uses the same algorithm multiple times or a group of different algorithms together to improve the prediction of a model.

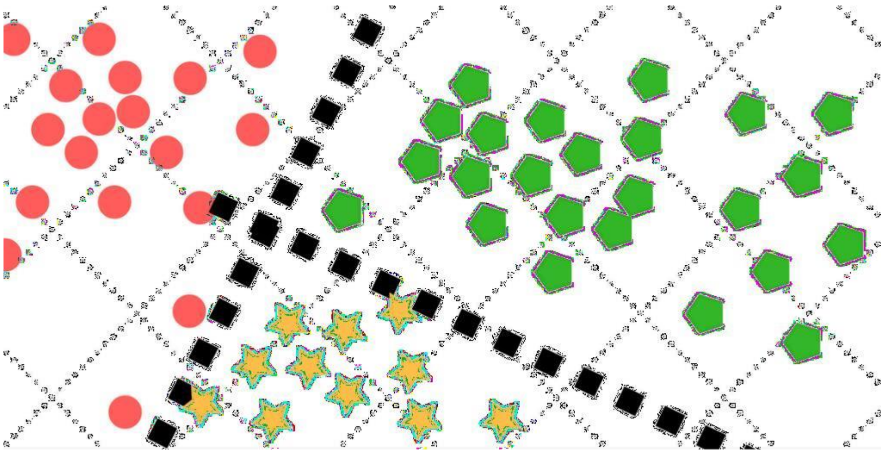


A single decision tree vs. a bagging ensemble of 500 trees

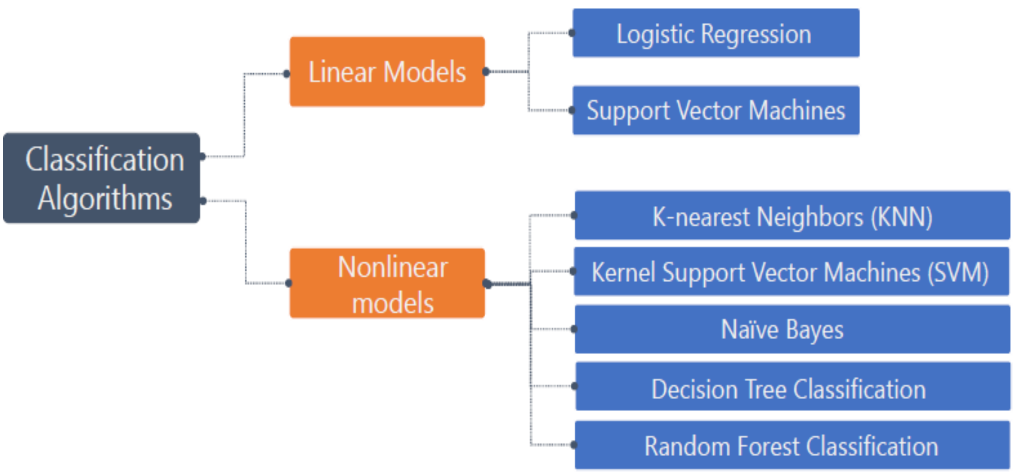
- Random Forests use an ensemble of decision trees to perform regression tasks.

2.5.2. Classification

- It specifies the class to which data elements belong to.
- It predicts a class for an input variable.
- It is best used when the output has finite and discreet values.



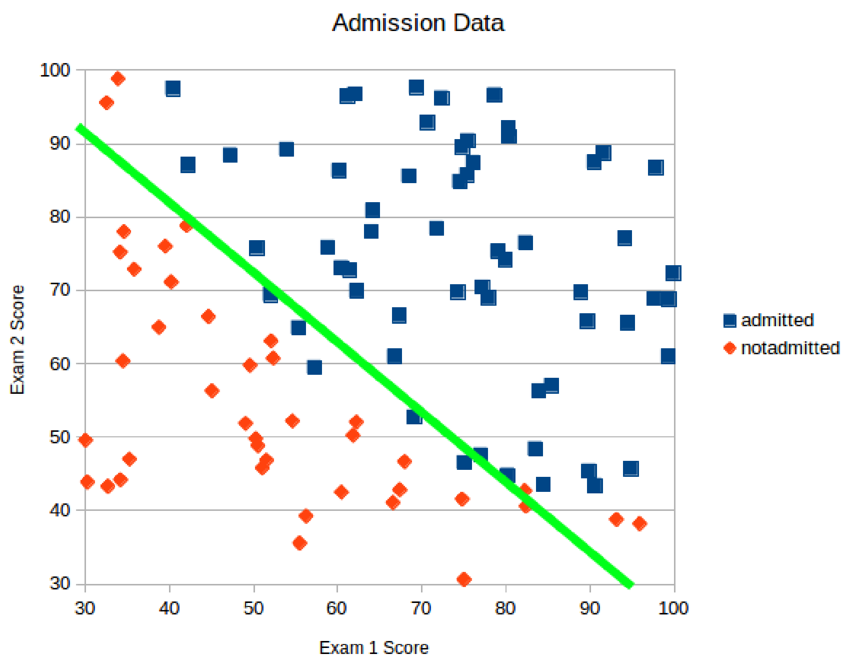
- There are 2 types of classification, **binomial and multi-class**.



2.5.2.1. Linear Models

2.5.2.1.1. Logistic Regression

- ✓ This method is widely used for binary classification problems. It can also be extended to multi-class classification problems.
- ✓ A binary dependent variable can have only **two values**, like 0 or 1, win or lose, pass or fail, healthy or sick, etc.

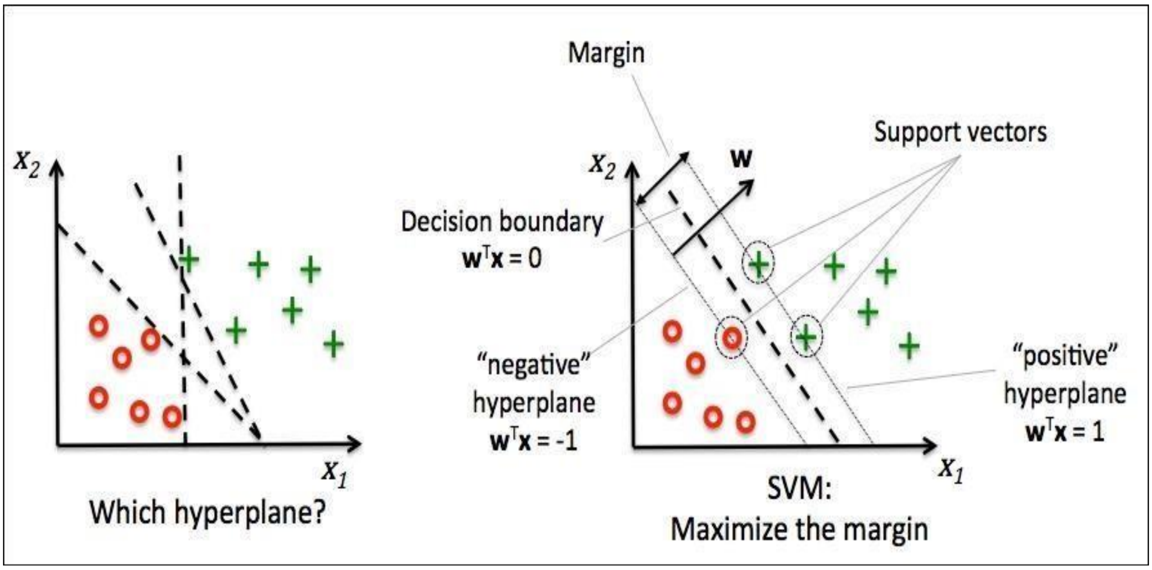


- ✓ The probability in the logistic regression is often represented by the **Sigmoid function** (also called the **logistic function** or the **S-curve**)

$$S(t) = \frac{1}{1 + e^{-t}}$$

2.5.2.1.2. Support Vector machines

- ➢ SVMs are very versatile and are also capable of performing linear or nonlinear classification, regression, and outlier detection.
- ➢ They involve detecting hyperplanes which segregate data into classes.

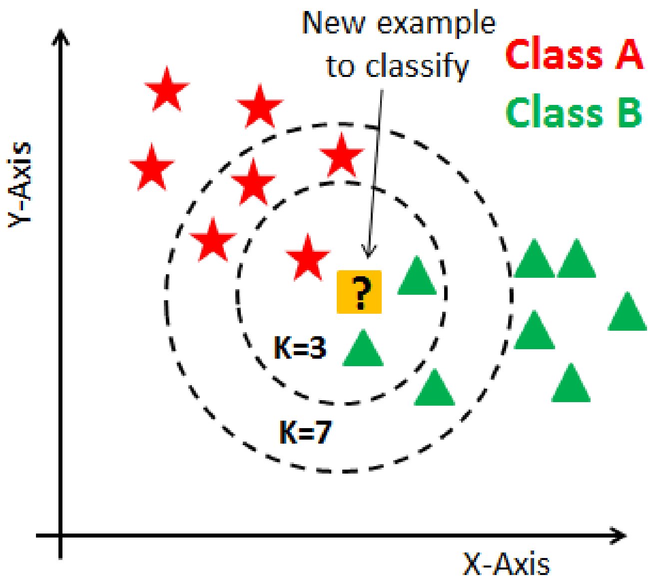


- The optimization objective is to find "**maximum margin hyperplane**" that is farthest from the closest points in the two classes (these points are called support vectors).

2.5.2.2. Nonlinear Models

2.5.2.2.1. K-Nearest Neighbors (KNN)

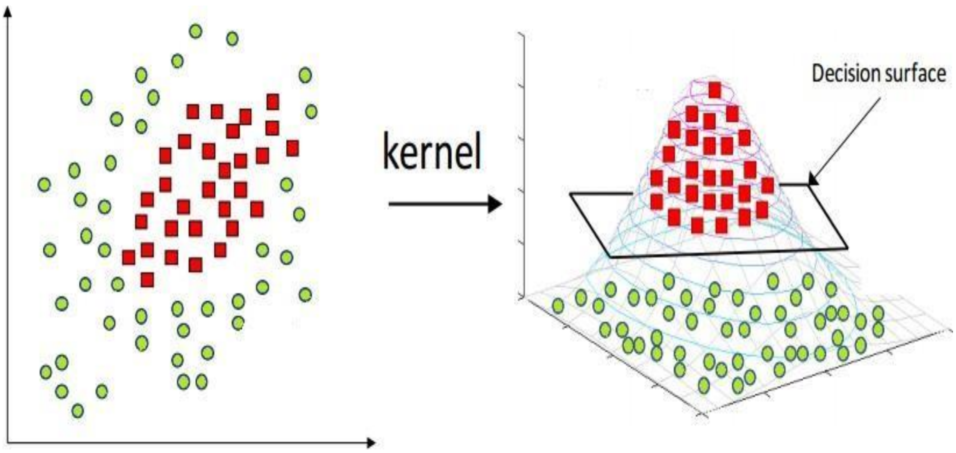
- K-nearest Neighbors algorithm is used to assign a data point to clusters based on similarity measurement.



- A new input point is classified in the category such that it has the **greatest number of neighbors** from that category.

2.5.2.2.2. Kernel Support Vector Machines (SVM)

- Kernel SVMs are used for classification of nonlinear data.
- In the chart, nonlinear data is projected into a higher dimensional space via a mapping function where it becomes linearly separable.



- A reverse projection of the higher dimension back to original feature space takes it back to nonlinear shape.

2.5.2.2.3. Naïve Bayes

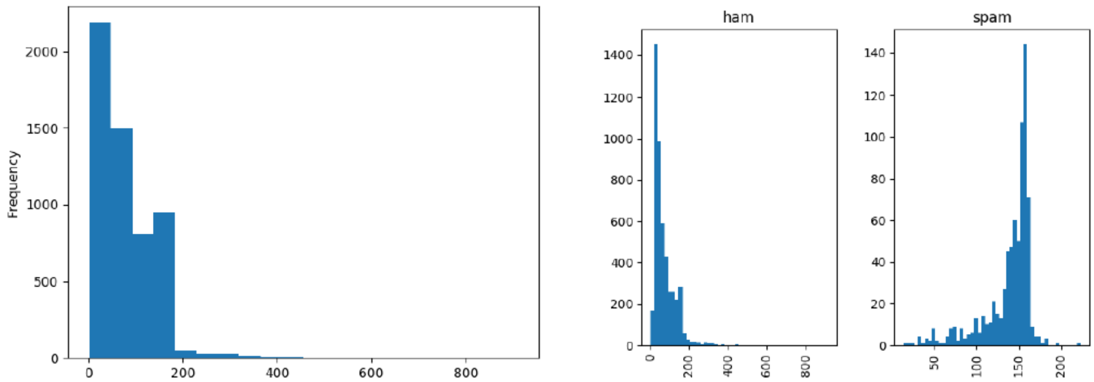
- According to Bayes model, the conditional probability P(Y|X) can be calculated as:

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)}$$

- This means you have to estimate a very large number of P(X|Y) probabilities for a relatively small vector space X.

Naïve Bayes Classifier for SMS Spam Detection

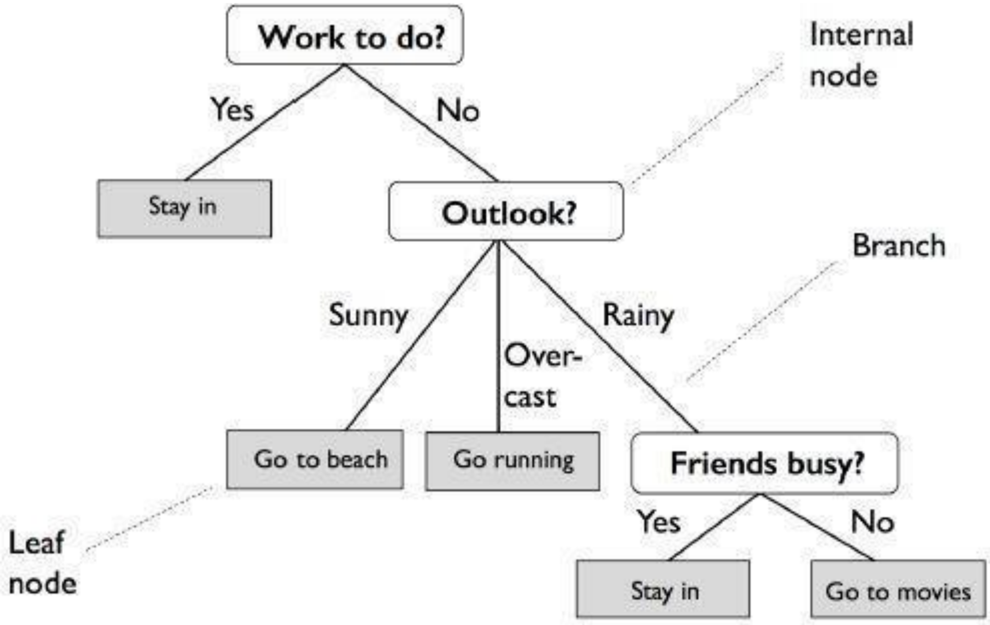
The message lengths and their frequency (in the training data set) are as shown below:



2.5.2.2.4. Decision Tree Classification

- ✓ The advantage of decision trees is that they require very little data preparation.
- ✓ They do not require feature scaling or centering at all.

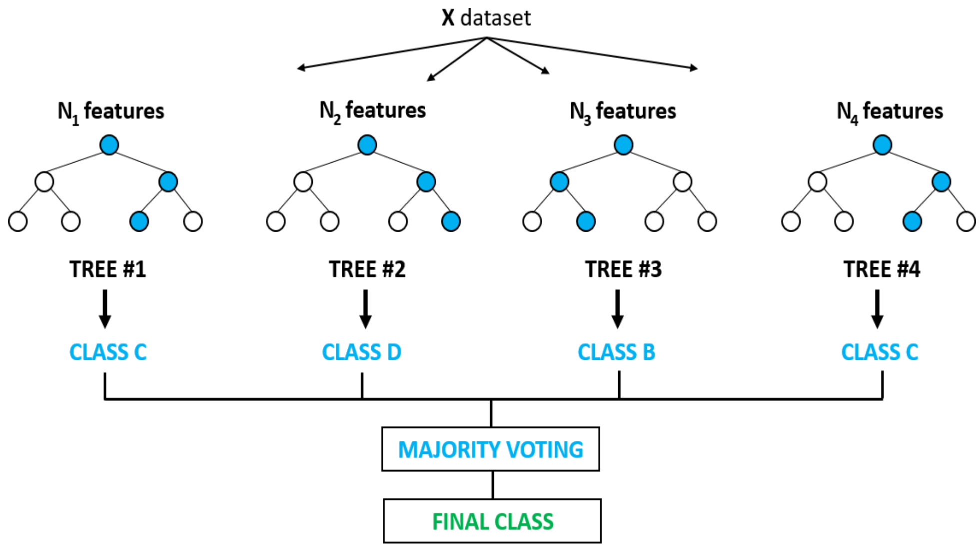
- ✓ They are also the fundamental components of Random Forests, one of the most powerful ML algorithms.



- ✓ Start at the tree root and split the data on the feature using the decision algorithm, resulting in the **largest information gain** (IG).

2.5.2.2.5. Random Forest Classification

- ➢ Random decision forests correct for decision trees' habit of overfitting to their training set.

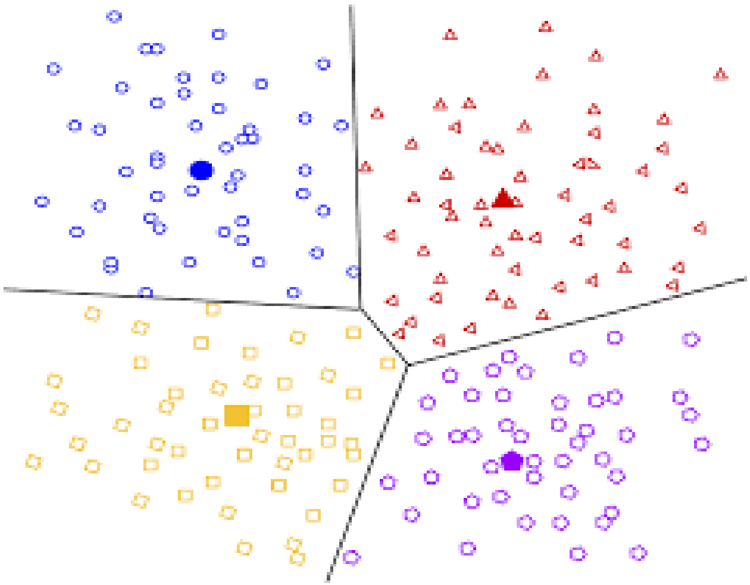


- ➢ Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks that operates by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees.

2.6. Unsupervised learning

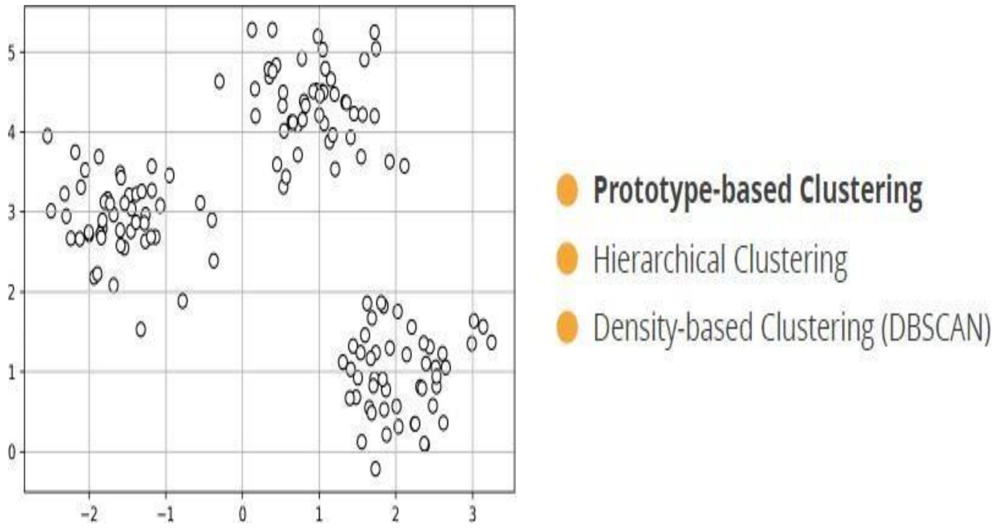
2.6.1. Clustering

2.6.1.1. Clustering Algorithms



❖ **Clustering means**

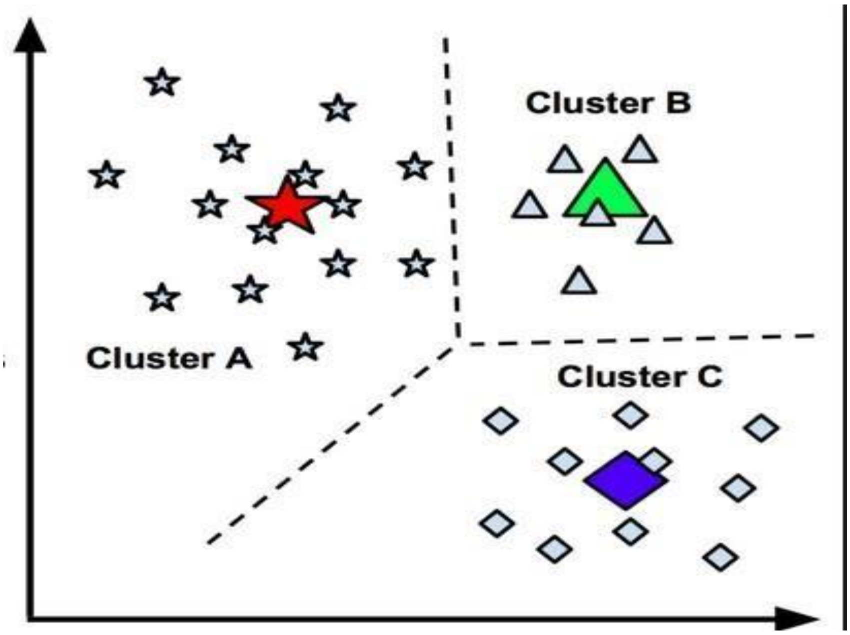
✓ Clustering is a Machine Learning technique that involves the grouping of data points.



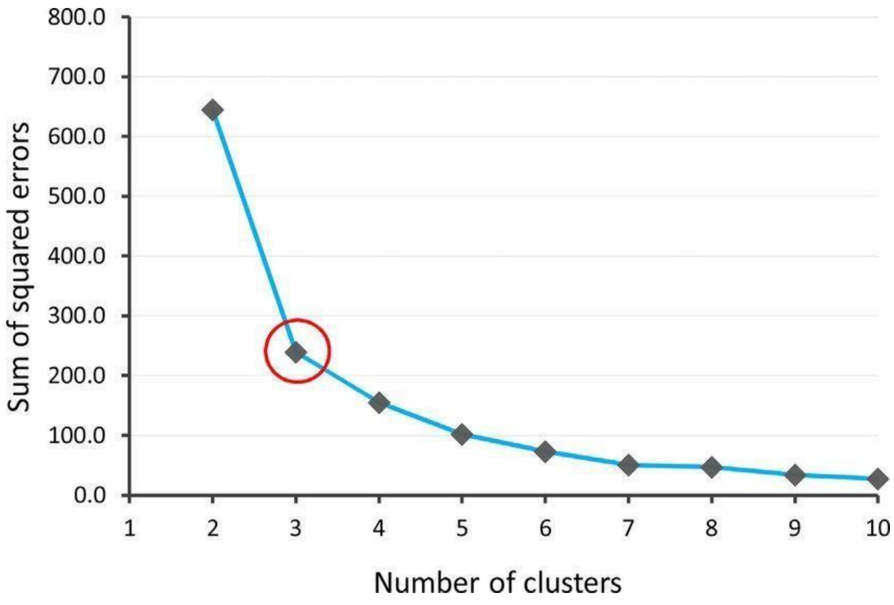
❖ **Prototype Based Clustering**

- Prototype-based clustering assumes that most data is located near prototypes; example: centroids (average) or medoid (most frequently occurring point)
- K-means, a Prototype-based method, is the most popular method for clustering that involves:
    - Training data that gets assigned to matching cluster based on similarity
    - Iterative process to get data points in the best clusters possible

2.6.1.2. K-means Clustering



- **K-means Clustering Algorithm**
  - Step 1: randomly pick k centroids
  - Step 2: assign each point to the nearest centroid
  - Step 3: move each centroid to the center of the respective cluster
  - Step 4: calculate the distance of the centroids from each point again
  - Step 5: move points across clusters and re-calculate the distance from the centroid
  - Step 6: keep moving the points across clusters until the Euclidean distance is minimized
- **Elbow Method**
  - One could plot the Distortion against the number of clusters K. Intuitively, if K increases, distortion should decrease. This is because the samples will be close to their assigned centroids. This plot is called the Elbow method.



- ➢ It indicates the optimum number of clusters at the position of the elbow, the point where distortion begins to increase most rapidly.

❖ **Euclidian Distance**

- ✓ K-means is based on finding points close to cluster centroids. The distance between two points x and y can be measured by the squared Euclidean distance between them in an m-dimensional space.

- Here, *j* refers to *j*-th dimension (or *j*-th feature) of the data point.

$$d(x,y)^2 = \sum_{j=1}^{m}(x_j - y_j)^2 = \|x-y\|_2^2$$

❖ **Examples of K-means Clustering**

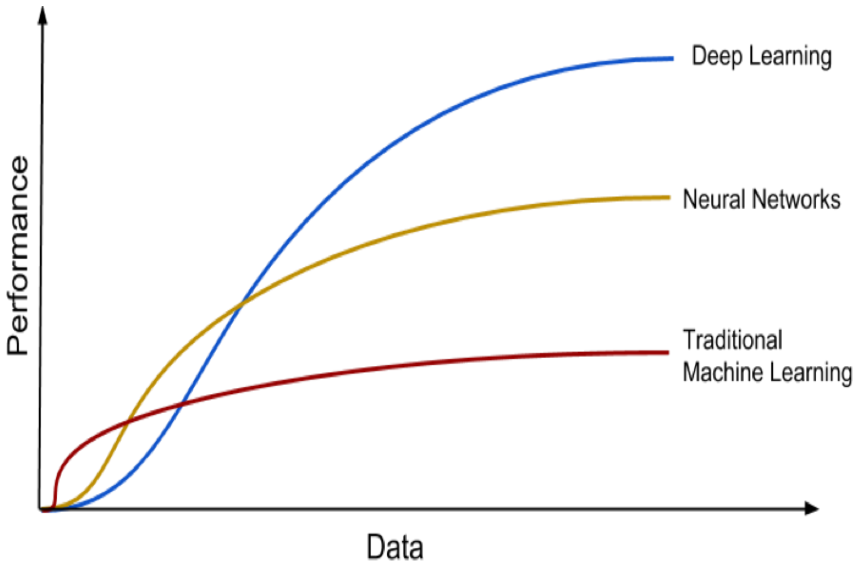
- ➢ Grouping articles (example: Google news)
- ➢ Grouping customers who share similar interests
- ➢ Classifying high risk and low risk patients from a patient pool

2.7. Introduction to Deep Learning

2.7.1. Meaning and Importance of Deep Learning

❖ **Define Deep Learning**

Deep Learning is a specialized form of Machine Learning that uses supervised, unsupervised, or semi-supervised learning to learn data representations.



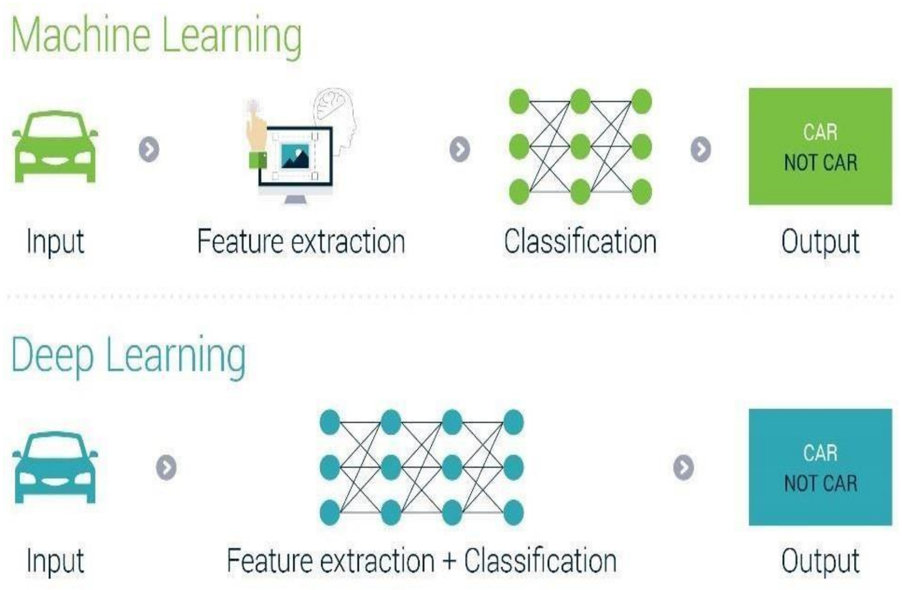
It is similar to the structure and function of the human nervous system.

❖ **Why Deep Learning**

- The vast availability of Big Data enables machines to be trained.
- Experts have discovered multi-layered learning networks that can be leveraged for deep learning as they learn in layers.

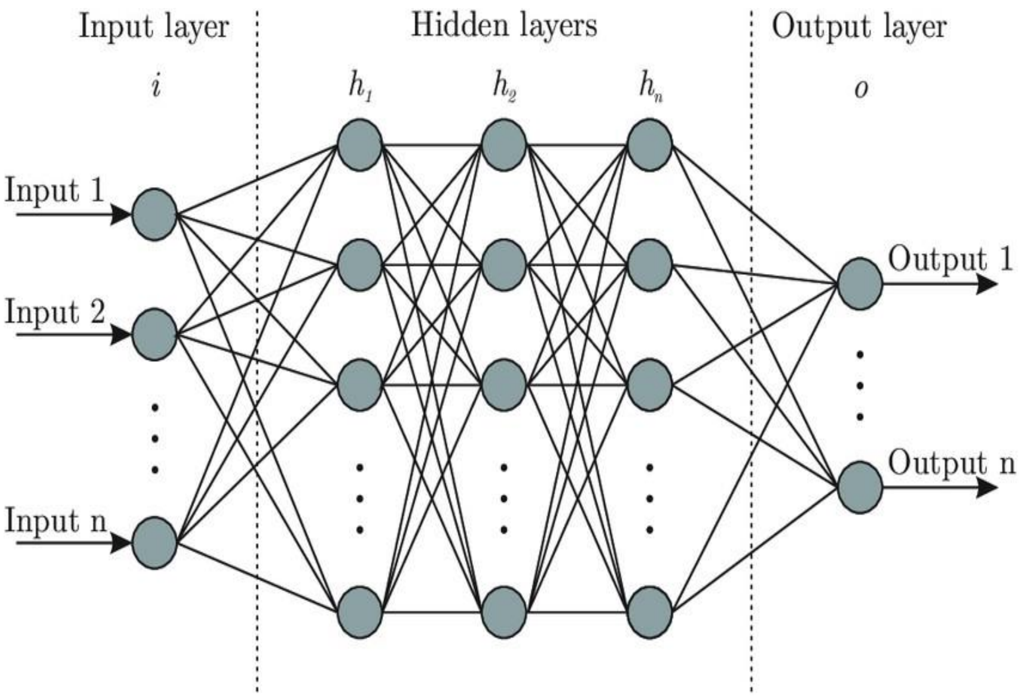
- Scientists have figured out that high-performing graphics processing units (GPU) can be used for deep learning.

- **ML Vs Deep Learning**



2.7.2. Artificial Neural Networks

- ✓ Deep learning relies on multiple layers of training.
- ✓ Artificial Neural Network is a computing system made up of a number of simple, highly interconnected processing elements which process information by their dynamic state response to external inputs.



- ✓ It is an interconnected group of nodes akin to the vast network of layers of neurons in a brain.

2.7.3. TensorFlow

- TensorFlow is the open source Deep Learning library provided by Google.
- It allows development of a variety of neural network applications such as computer vision, speech processing, or text recognition.
- It uses data flow graphs for numerical computations.

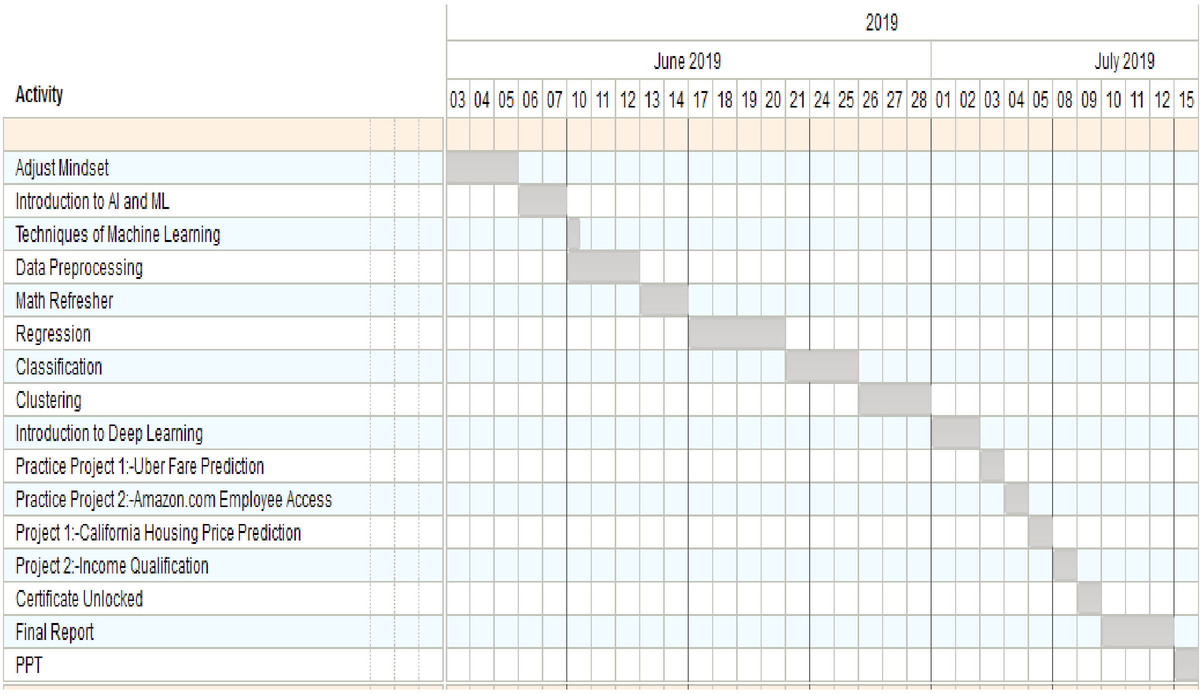
3. Reason for choosing Machine Learning

- **Learning machine learning brings in better career opportunities**
  - ✓ Machine learning is the shining star of the moment.
  - ✓ Every industry looking to apply AI in their domain, studying machine learning opens world of opportunities to develop cutting edge machine learning applications in various verticals – such as cyber security, image recognition, medicine, or face recognition.
  - ✓ Several machine learning companies on the verge of hiring skilled ML engineers, it is becoming the brain behind business intelligence.
- **Machine Learning Jobs on the rise**
  - ✓ The major hiring is happening in all top tech companies in search of those special kind of people (machine learning engineers) who can build a hammer (machine learning algorithms).
  - ✓ The job market for machine learning engineers is not just hot but it's sizzling.
  - ✓ Machine Learning Jobs on Indeed.com - 2,500+(India) & 12,000+(US)

4. Learning Outcome

- Have a good understanding of the fundamental issues and challenges of machine learning: data, model selection, model complexity, etc.
- Have an understanding of the strengths and weaknesses of many popular machine learning approaches.
- Appreciate the underlying mathematical relationships within and across Machine Learning algorithms and the paradigms of supervised and un-supervised learning.
- Be able to design and implement various machine learning algorithms in a range of real-world applications.
- Ability to integrate machine learning libraries and mathematical and statistical tools with modern technologies
- Ability to understand and apply scaling up machine learning techniques and associated computing techniques and technologies.

5. Gantt Chart



6. Bibliography

6.1. All Content used in this report is from

- ✓ https://www.simplilearn.com/
- ✓ https://www.wikipedia.org/
- ✓ https://towardsdatascience.com/
- ✓ https://www.expertsystem.com/
- ✓ https://www.coursera.org/
- ✓ https://www.edureka.co/
- ✓ https://subhadipml.tech/
- ✓ https://www.forbes.com/
- ✓ https://medium.com/
- ✓ https://www.google.com/

6.2. All Pictures are from

- ✓ https://www.simplilearn.com/
- ✓ https://www.google.com/
- ✓ https://www.wikipedia.org/
- ✓ https://www.youtube.com/
- ✓ https://www.edureka.co/

6.3. Book I referred are

- ✓ Hands-on Machine Learning with Scikit-learn & Tensorflow By Aurelien Geron
- ✓ Python Machine Learning by Sebastian Raschka