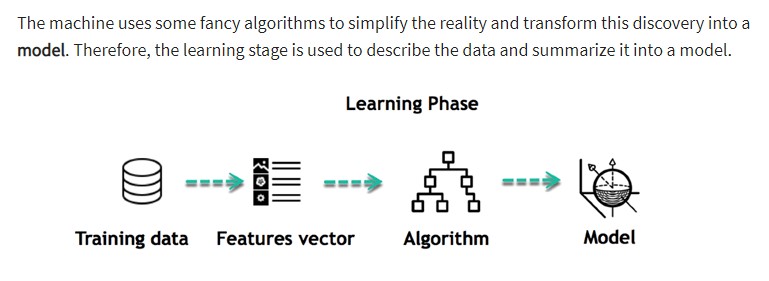
**UNIT-1**

**Chapter-01(Introduction to Machine Learning)**

1. **What is machine learning? Explain working of Machine Learning?** 
   * Machine Learning is a system that can learn from example through self-improvement and without being explicitly coded by programmer.
   * The breakthrough comes with the idea that a machine can singularly learn from the data (i.e., example) to produce accurate results.
   * Machine learning combines data with statistical tools to predict an output.
   * This output is then used by corporate to makes actionable insights.
   * Machine learning is closely related to data mining and Bayesian predictive modeling.
   * The machine receives data as input, use an algorithm to formulate answers.
   * Machine learning is also used for a variety of task like fraud detection, predictive maintenance, portfolio optimization, automatize task and so on.

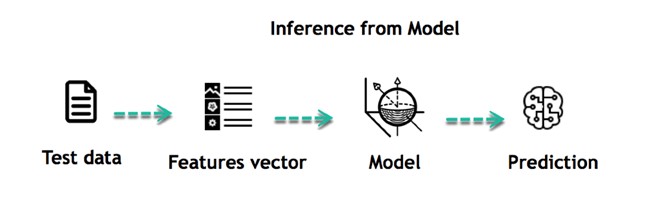
**Working Of Machine Learning:**

* + Machine learning is the brain where all the learning takes place. The way the machine learns is similar to the human being. Humans learn from experience. The more we know, the more easily we can predict.
  + By analogy, when we face an unknown situation, the likelihood of success is lower than the known situation.
  + Machines are trained the same. To make an accurate prediction, the machine sees an example. When we give the machine a similar example, it can figure out the outcome. However, like a human, if its feed a previously unseen example, the machine has difficulties to predict.
  + The core objective of machine learning is the **learning** and **inference**.
  + First of all, the machine learns through the discovery of patterns. This discovery is made thanks to the **data**. One crucial part of the data scientist is to choose carefully which data to provide to the machine.
  + The list of attributes used to solve a problem is called a **feature vector.** You can think of a feature vector as a subset of data that is used to tackle a problem.



* + **Inferring**

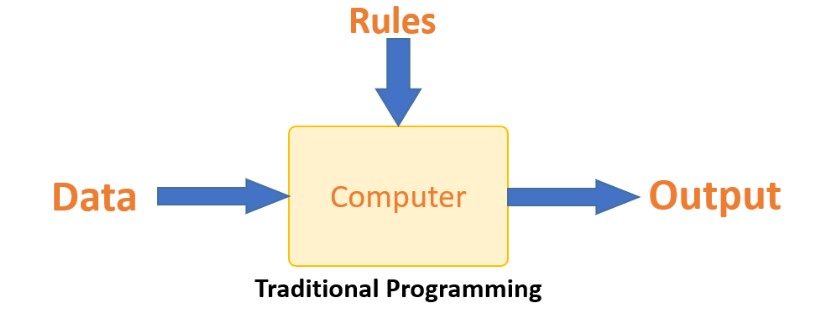
When the model is built, it is possible to test how powerful it is on never-seen-before data. The new data are transformed into a features vector, go through the model and give a prediction. This is all the beautiful part of machine learning. There is no need to update the rules or train again the model. You can use the model previously trained to make inference on new data.



1. **Differentiate between machine learning and traditional learning?**

Traditional programming differs significantly from machine learning.

* + - In traditional programming, a programmer code all the rules in consultation with an expert in the industry for which software is being developed. Each rule is based on a logical foundation; the machine will execute an output following the logical statement. When the system grows complex, more rules need to be written. It can quickly become unsustainable to maintain.



* + - Machine learning is supposed to overcome this issue. The machine learns how the input and output data are correlated and it writes a rule. The programmers do not need to write new rules each time there is new data. The algorithms adapt in response to new data and experiences to improve efficacy over time.



**List and explain applications of Machine Learning?**

* + - **Virtual Personal Assistants:** Siri, Alexa, Google Assistant help with tasks, from scheduling to setting alarms, using machine learning and voice interaction.
    - **Traffic Predictions:** Google Maps uses machine learning for real-time traffic updates and estimated arrival times.
    - **Online Transportation Networks:** Apps like Uber and Ola employ ML to estimate ride prices and predict demand surge.
    - **Video Surveillance:** ML-based systems monitor CCTV footage, detecting unusual behavior and potential crime.
    - **Social Media Services:** Facebook uses ML to personalize news feeds, suggest friends, and enable face recognition. Pinterest uses ML for image recognition.
    - **Email Spam and Malware Filtering:** ML aids in identifying spam and detecting malware patterns, enhancing security.
    - **Online Customer Support:** Chatbots on websites extract data and offer assistance using machine learning.
    - **Search Engine Result Refining:** Google refines search results using ML, learning from user behavior to improve accuracy.
    - **Product Recommendations:** Amazon, Flipkart use ML to suggest items based on user behavior and preferences.
    - **Online Fraud Detection:** PayPal employs ML to distinguish legitimate transactions from fraudulent ones.
    - **Medicine:** ML analyzes medical records, aiding doctors in understanding diseases and treatments better.
    - **Computational Biology**: ML helps biologists in understanding relationships between genes and human features.
    - **Handwriting Recognition:** ML recognizes and reads handwriting for various applications, like routing postal mail.
    - **Machine Translation:** Google Translate uses ML to instantly translate text between different languages.
    - **Driverless Cars and Autonomous Vehicles:** ML enables programs to learn and drive cars or helicopters intuitively, bypassing the need for defining strict rules.

4. **List all the types of machine learning methods and explain supervised learning in details?**

Type of Machine Learning methods:

* + - Supervised Learning.
    - Unsupervised Learning.
    - Reinforcement Learning.

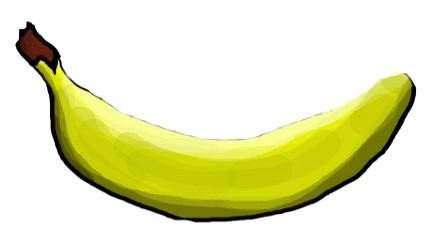
**Supervised Learning Method:**

* + - Supervised learning as the name indicates the presence of a supervisor as a teacher.
    - Basically supervised learning is a learning in which we teach or train the machine using data which is well labeled that means some data is already tagged with the correct answer.
    - After that, the machine is provided with a new set of examples(data) so that supervised learning algorithm analyses the training data(set of training examples) and produces a correct outcome from labeled data.

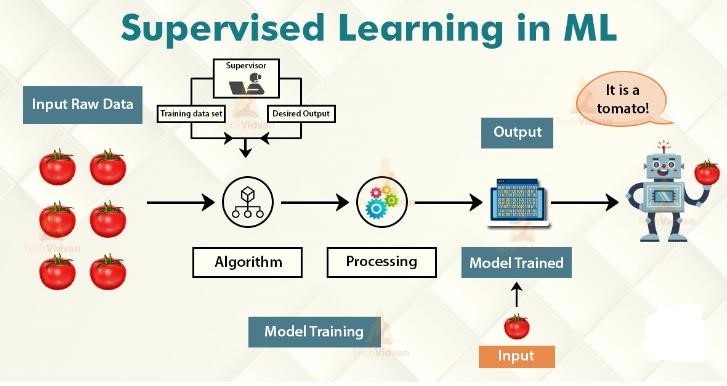
* + - For instance, suppose you are given an basket filled with different kinds of fruits. Now the first step is to train the machine with all different fruits one by one like this:



* + - If shape of object is rounded and depression at top having color Red then it will be labelled as – Apple.
    - If shape of object is long curving cylinder having color Green-Yellow then it will be labelled as – Banana.
    - Now suppose after training the data, you have given a new separate fruit say Banana from basket and asked to identify it.



* + - Since the machine has already learned the things from previous data and this time have to use it wisely.
    - It will first classify the fruit with its shape and color and would confirm the fruit name as BANANA and put it in Banana category.
    - Thus the machine learns the things from training data(basket containing fruits) and then apply the knowledge to test data(new fruit).

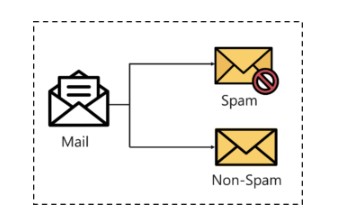


* + - There are **two** **categories** of supervised learning:
    - **Classification task**
    - **Regression task**

**Write a short note on classification and regression with example?**

**Classification**

* + - In **machine learning**, **classification** refers to a predictive modeling problem where a class label is predicted for a given example of input data.
    - Classification in [machine learning](https://www.edureka.co/blog/machine-learning-tutorial/) and statistics is a supervised learning approach in which the computer program learns from the data given to it and make new observations or classifications.
    - Classification is a process of categorizing a given set of data into classes, It can be performed on both structured or unstructured data. The process starts with predicting the class of given data points. The classes are often referred to as target, label or categories.
    - The classification predictive modeling is the task of approximating the mapping function from input variables to discrete output variables. The main goal is to identify which class/category the new data will fall into.



**Regression**

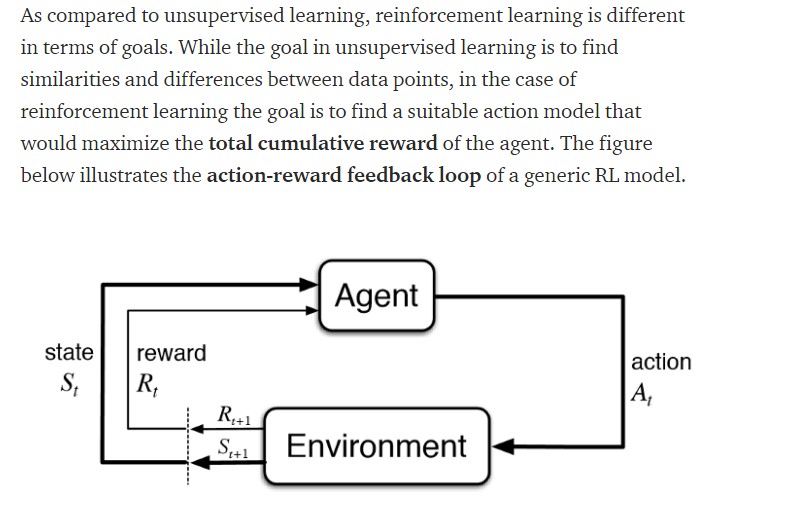
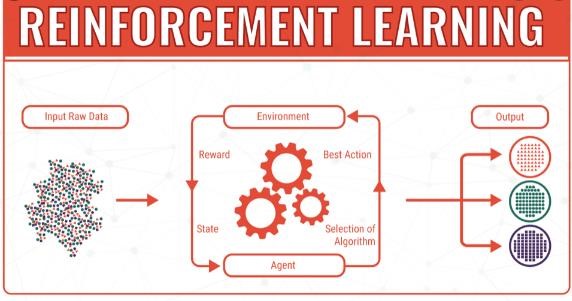
* + - **Regression** analysis consists of a set of **machine learning** methods that allow us to predict a continuous outcome variable (y) based on the value of one or multiple predictor variables (x). Briefly, the goal of **regression** model is to build a mathematical equation that defines y as a function of the x variables.
    - Regression models are used to predict a continuous value. Predicting prices of a house given the features of house like size, price etc is one of the common examples of Regression. It is a supervised technique.

* + - **Types of Regression**
      1. Simple Linear Regression
      2. Polynomial Regression
      3. Support Vector Regression 4. Decision Tree Regression

5. Random Forest Regression

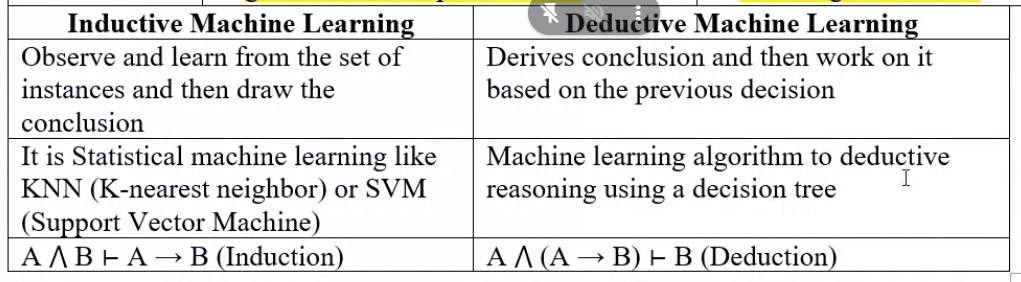
**Explain reinforcement learning with example?**

* Reinforcement Learning is a feedback-based Machine learning technique in which an agent learns to behave in an environment by performing the actions and seeing the results of actions. For each good action, the agent gets positive feedback, and for each bad action, the agent gets negative feedback or penalty.
* In Reinforcement Learning, the agent learns automatically using feedbacks without any labeled data, unlike supervised learning.
* Since there is no labeled data, so the agent is bound to learn by its experience only.
* The agent interacts with the environment and explores it by itself. The primary goal of an agent in reinforcement learning is to improve the performance by getting the maximum positive rewards.
* The agent learns with the process of hit and trial, and based on the experience, it learns to perform the task in a better way. Hence, we can say that "Reinforcement learning is a type of machine learning method where an intelligent agent (computer program) interacts with the environment and learns to act within that." How a Robotic dog learns the movement of his arms is an example of Reinforcement learning.



**Differentiate between Inductive and deductive learning?**

|  |  |  |
| --- | --- | --- |
| **Basis of**  **Comparison** | **Deductive Reasoning** | **Inductive Reasoning** |
| **Definition** | Deductive reasoning is the form of valid reasoning, to deduce new information or conclusion from knowledge related facts and information. | Inductive Reasoning arries at a conclusion by the process of generalization using specific facts or data. |
| **Approach** | Deductive reasoning follows a top-down approach. | Inductive reasoning follows a bottom-up approach. |
| **Starts from** | Deductive reasoning starts from Premises. | Inductive reasoning starts from Conclusion. |
| **Validity** | In Deductive reasoning, the conclusion must be true if the premises are true. | In Inductive reasoning, the truth of premises does not guarantee the truth of conclusions. |
| **Usage** | Use of Deductive reasoning is difficult, as we need facts which must be true. | |  |  | | --- | --- | | Use of inductive reasoning is fast | | | and easy | , as we need evidence |   instead of true facts.  We often use it in our daily life. |
| **Process** | Theory → Hypothesis → Patterns → Confirmation. | Observation → Patterns → Hypothesis → Theory. |
| **Arguments** | In Deductive reasoning, arguments may be valid or invalid. | In Inductive reasoning, arguments may be weak or strong. |
| **Structure** | Deductive reasoning reaches from general facts to specific facts. | Inductive reasoning ranges from specific facts to general facts. |



**With diagram Explain Case based learning?**

* Case-based reasoning (CBR), broadly construed, is the process of solving new problems based on the solutions of similar past problems.
* An auto mechanic who fixes an engine by recalling another car that exhibited similar symptoms is using case-based reasoning.
* Case-based reasoning has been formalized for purposes of computer reasoning as a **four-step process**:
  1. **Retrieve:**

Given a target problem, retrieve from memory cases relevant to solving it. A case consists of a problem, its solution, and, typically, annotations about how the solution was derived.

* 1. **Reuse:**

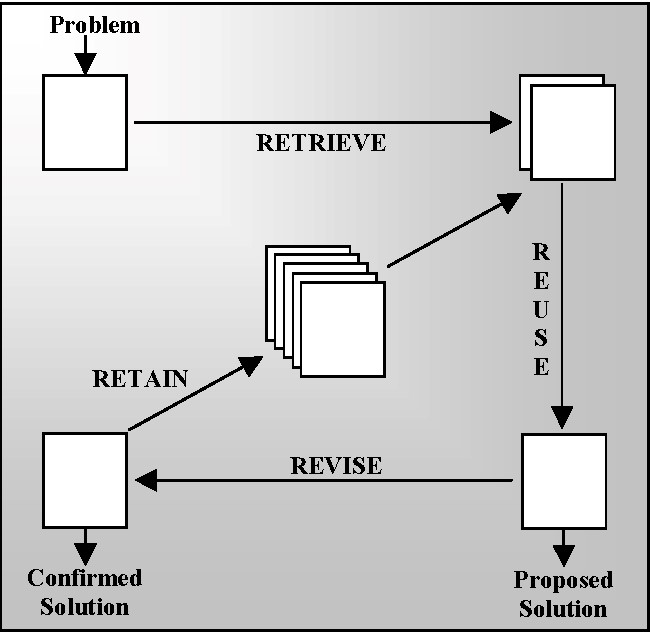
Map the solution from the previous case to the target problem. This may involve adapting the solution as needed to fit the new situation.

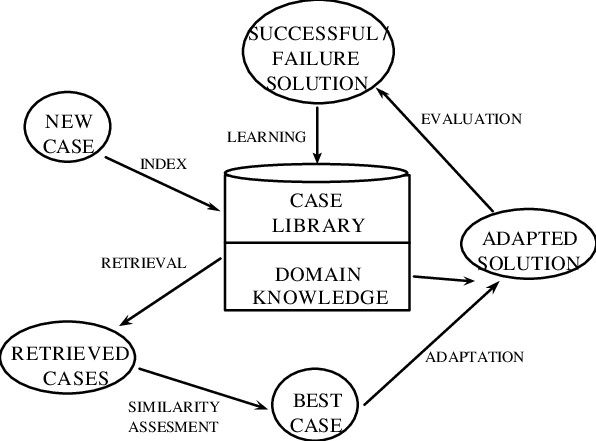
* 1. **Revise:**

Having mapped the previous solution to the target situation, test the new solution in the real world (or a simulation) and, if necessary, revise.

* 1. **Retain:**

After the solution has been successfully adapted to the target problem, store the resulting experience as a new case in memory.



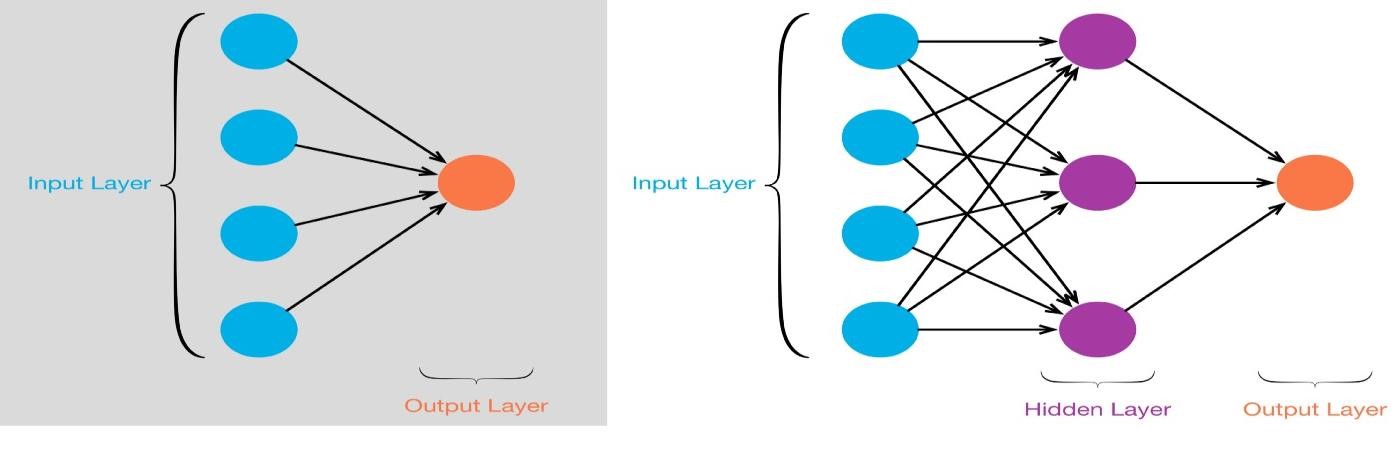


**Chapter-02(Artificial Neural Networks)**

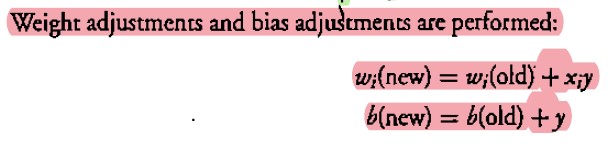
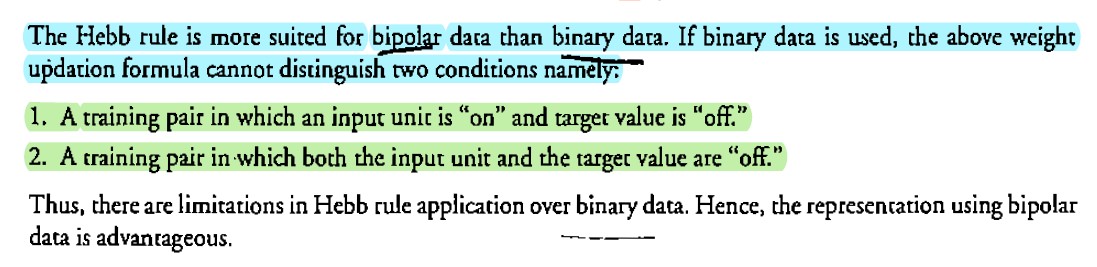
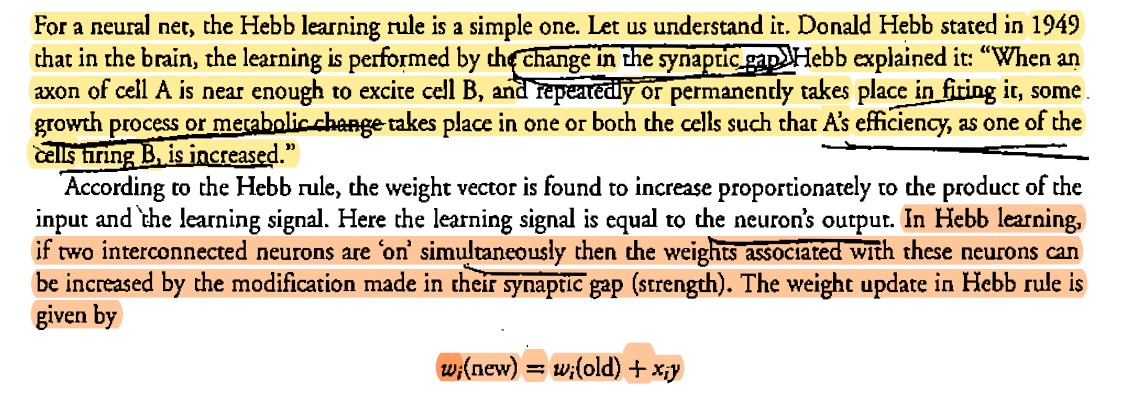
1. **Draw and explain artificial neural network model?**

* A neural network is a processing device, either an algorithm or an actual hardware, whose design was inspired by the design and functioning of animal brains and components.
* An artificial neural network (ANN) may be defined as an information processing model that is inspired by the way biological nervous systems, such as the brain, process information. This model is to replicate only the most basic functions of the brain.
* The key element of ANN is the novel structure of is information processing system. An ANN is composed of a large number of highly interconnected processing elements (neurons) working in union to solve specific problems.

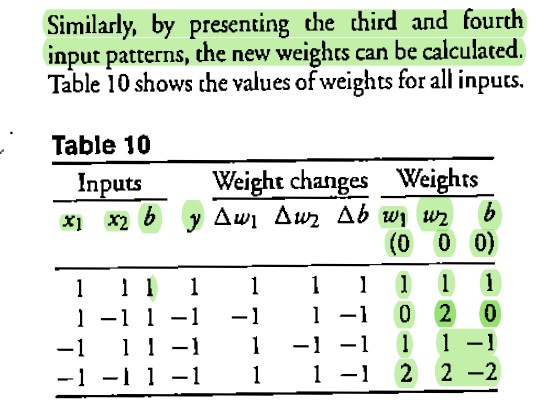
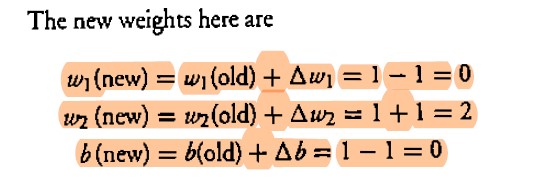
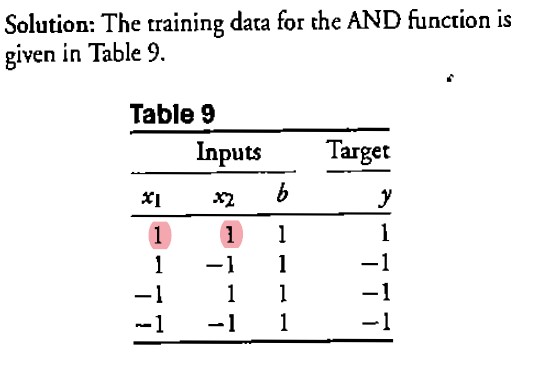
* **Connections**
* The neurons should be visualized for their arrangements in layers.
* A neural network consists of a set of highly interconnected processing elements (neurons) such that each processing element output is found to be connected through weight to the other processing elements or to itself.
* Hence, the arrangement of these processing elements and geometry are essential for neural network



1. **Explain Hebb Rule with training algorithm with its flow chart.**



1. **Desing Hebb network to train AND/NOT function use bipolar inputs and targets.**



1. **Explain M-P model with diagram.** 
   * The Mculloh-pitts neuron model was the earliest neural network discovered in 1943.
   * It should be noted that the activation of a M-P neuron is binary, that is, at any time the the neuron may fire or may not fire.
   * The weight associated with the communication links may be excitatory(weight is positive) inhibitory (weight is negative).
   * The threshold plays a major role in M-P neuron there is a fixed threshold for each neuron, and if the net input to the neuron is greater then the threshold then the neuron fires. • The activation function use for mp model is

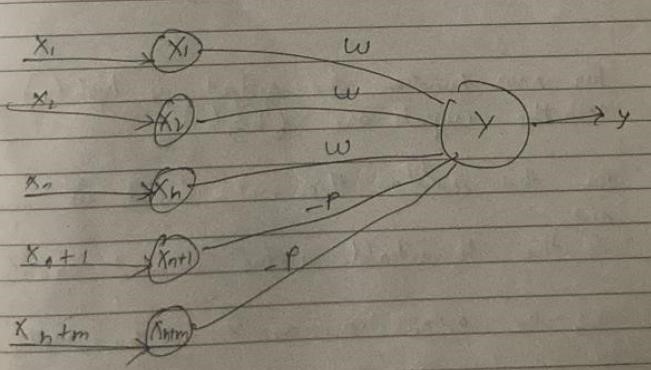
1 𝑖𝑓 𝑦𝑖𝑛 ≥ 𝜃

F (yin) = {

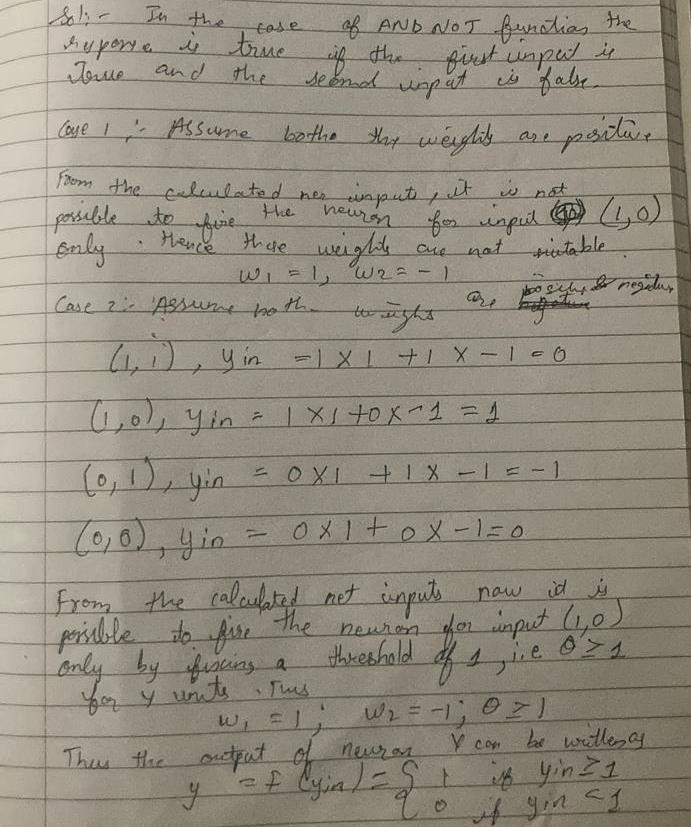
0 𝑖𝑓 𝑦𝑖𝑛 < 𝜃

* + For inhibitior to be absolete, the threshold with the activation function should satisfy the following condition

𝜃 > n w - p



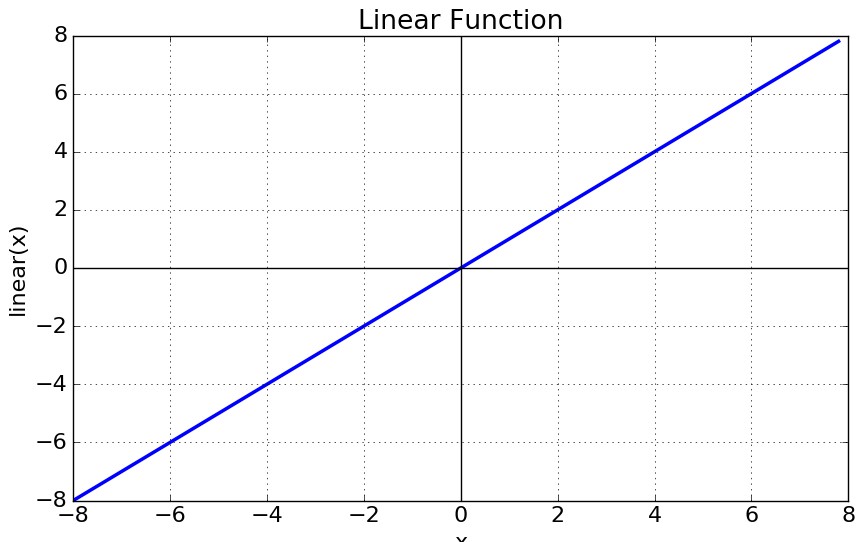
1. **Implement AND function using M-P model use bipolar inputs and targets.**



1. **What is activation function? List and explain various types of activation?** 
   * It’s just a thing function that you use to get the output of node. It is also known as **Transfer Function**.
   * **Why we use Activation functions with Neural Networks?**
   * It is used to determine the output of neural network like yes or no. It maps the resulting values in between 0 to 1 or -1 to 1 etc. (depending upon the function).
   * The Activation Functions can be basically divided into 2 types-
     1. Linear Activation Function
     2. Non-linear Activation Functions

**Linear or Identity Activation Function**

* As you can see the function is a line or linear. Therefore, the output of the functions will not be confined between any range.



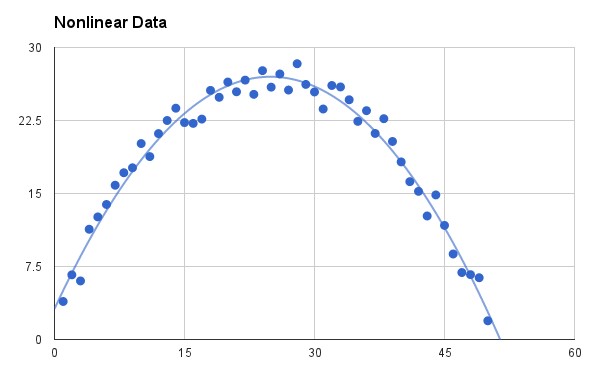
**Equation :** f(x) = x

**Range :** (-infinity to infinity)

It doesn’t help with the complexity or various parameters of usual data that is fed to the neural networks.

**Non-linear Activation Function**

* The Nonlinear Activation Functions are the most used activation functions. Nonlinearity helps to makes the graph look something like this
* It makes it easy for the model to generalize or adapt with variety of data and to differentiate between the output.



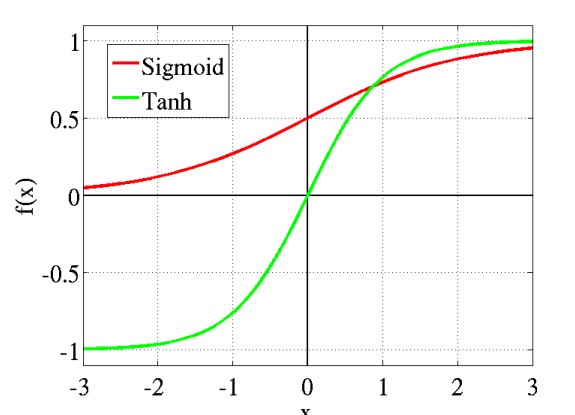
**Sigmoid or Logistic Activation Function** The Sigmoid Function curve looks like a S-shape.



* The main reason why we use sigmoid function is because it exists between **(0 to 1).**
* Therefore, it is especially used for models where we have to **predict the probability** as an output. • Since probability of anything exists only between the range of **0 and 1,** sigmoid is the right choice.
* The function is **differentiable**.That means, we can find the slope of the sigmoid curve at any two points.

**Tanh or hyperbolic tangent Activation Function**

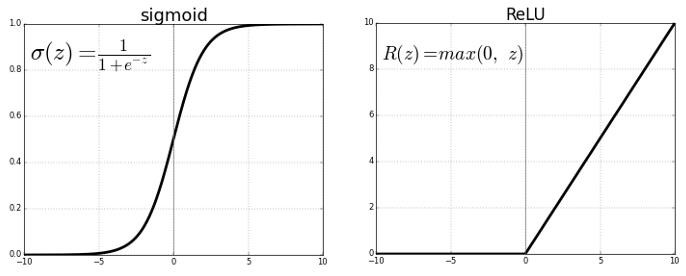
* tanh is also like logistic sigmoid but better. The range of the tanh function is from (-1 to 1). tanh is also sigmoidal (s - shaped)



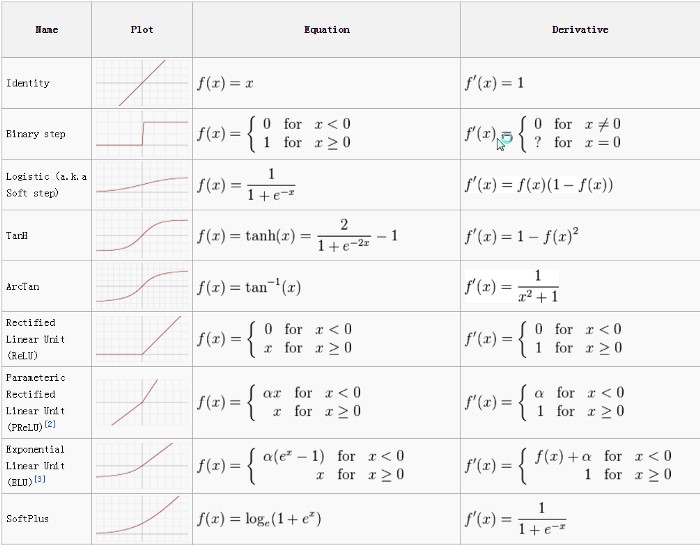
* The advantage is that the negative inputs will be mapped strongly negative and the zero inputs will be mapped near zero in the tanh graph.
* The function is **differentiable**.
* The function is **monotonic** while its **derivative is not monotonic**.
* The tanh function is mainly used classification between two classes.

**ReLU (Rectified Linear Unit) Activation Function**

* The ReLU is the most used activation function in the world right now.Since, it is used in almost all the convolutional neural networks or deep learning.

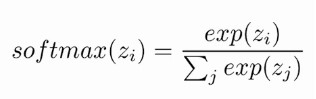


* As you can see, the ReLU is half rectified (from bottom). f(z) is zero when z is less than zero and f(z) is equal to z when z is above or equal to zero.
* **Range:** [ 0 to infinity)
* The function and its derivative **both are** **monotonic**.



**SoftMax Activation Function**

* The softmax function is used as **the activation function in the output layer of neural** network models that predict a multinomial probability distribution.
* That is, softmax is used as the activation function for multi-class classification problems where class membership is required on more than two class labels.



7. **Write a short note on a)Weights b)Bias c)Learning Rate**

1. **Weights:**

Definition: Weights in a neural network are parameters that determine the strength of connections between neurons.

Importance: They control the influence of one neuron on another, influencing the output.

1. **Bias:**

Definition: Bias is an additional parameter in a neuron that allows shifting the activation function. Importance: It provides flexibility to the model, enabling better fit to data.

1. **Learning Rate:**

Definition: Learning rate is a hyperparameter that controls the size of the steps taken during the optimization process.

Importance: It affects the convergence and stability of the model during training. A too high learning rate may cause overshooting, and a too low learning rate may result in slow convergence or getting stuck in local minima. Adjustments are made through experimentation.

# Chapter-03(Classification)

1. **Define classification? What are its applications?** 
   * + Classification models predict discrete value or class, while Regression models predict a continuous value.
     + Classification is a classical method which is used by machine learning researchers and statisticians for predicting the outcome of unknown samples.
     + It is used for categorization of objects (or things) into a given discrete number of classes. • Classification problems can be of two types, either binary or multiclass
     + Classification is defined as two types:
     + Posteriori classification
     + Priori classification

**Applications of Classification:**

* + - To analyze the credit history of bank customers to identify if it would be risky or safe to grant them loans.
    - To analyze the purchase history of a shopping mall’s customers to predict whether they will buy a certain product or not.
    - To predict how much a given customer will spend during a sale.
    - To predict the salary-package of a student that he/she may get during his/her placement.
    - Other Applications of Classifications:
    - Email spam filtering
    - Credit risk assessment
    - Medical diagnosis
    - Image classification
    - Sentiment analysis.
    - Fraud detection
    - Quality control
    - Recommendation systems

1. **Write a short note on Binary and Multiclass classification with example?**

* + Classification is a classical method which is used by machine learning researchers and statisticians for predicting the outcome of unknown samples. It is used for categorization of objects (or things) into given discrete number of classes. Classification problems can be of two types, either binary or multiclass.

* + In **binary classification** the target attribute can only have two possible values. For example, a tumor is either cancerous or not, a team will either win or lose, a sentiment of a sentence is either positive or negative and so on.

* + In **multiclass classification**, the target attribute can have more than two values. For example, a tumor can be of type 1, type 2 or type 3 cancer; the sentiment of a sentence can be happy, sad, angry or of love; news stories can be classified as weather, finance, entertainment or sports news.

1. **Explain guidelines for size and quality of training data in classification.**

* + There should be a balance between the number of training samples and independent attributes.

* + It has been observed that generally, the number of training samples required is likely to be relatively small if the number of independent or input attributes is small and similarly, number of training samples required is likely to be relatively large if the number of independent or input attributes is large.

* + The quality of the classifier depends upon the quality of the training data. If there are two or more than two classes, then sufficient training data should be available belonging to each of these classes.

* + Researchers have developed a number of classifiers that include: Decision Tree, Naïve Bayes, Support Vector Machine and Neural Networks.

1. **Differentiate between binary and multiclass classification?**

|  |  |  |
| --- | --- | --- |
| **Parameters** | **Binary classification** | **Multi-class classification** |
| No. of classes | It is a classification of two groups, i.e. classifies objects in at most two classes. | There can be any number of  classes in it, i.e., classifies the object into more than two classes. |
| Algorithms used | Logistic Regression k-Nearest Neighbors  Decision Trees  Support Vector Machine  Naive Bayes | k-Nearest Neighbors  Decision Trees Naive Bayes Random Forest.  Gradient Boosting |
| Examples | Email spam detection (spam or not).  Churn prediction (churn or not). Conversion prediction (buy or not). | Face classification.  Plant species classification.  Optical character recognition. |

# UNIT-II Chapter-01(Regression)

1. **What is regression explain with examples?** 
   * Regression is a supervised machine learning technique which is used to predict continuous values.
   * The ultimate goal of the regression algorithm is to plot a best-fit line or a curve between the data.
   * Linear regression allows us to plot a linear equation, i.e., a straight line.
   * Predicting prices of a house given the features of house like size, price etc is one of the common examples of Regression. It is a supervised technique.
   * The three main metrics that are used for evaluating the trained regression model are variance, bias and error. If the variance is high, it leads to overfitting and when the bias is high, it leads to underfitting.
   * Based on the number of input features and output labels, regression is classified as linear (one input and one output), multiple (many inputs and one output) and multivariate (many outputs).
   * Linear regression allows us to plot a linear equation, i.e., a straight line. We need to tune the coefficient and bias of the linear equation over the training data for accurate predictions.

1. **Explain Linear Regression with example?**

* Linear regression is used to predict the relationship between two variables by applying a linear equation to observed data. There are two types of variables, one variable is called an independent variable, and the other is a dependent variable.
* **Use of Linear Regression**

o Linear regression is commonly used for predictive analysis

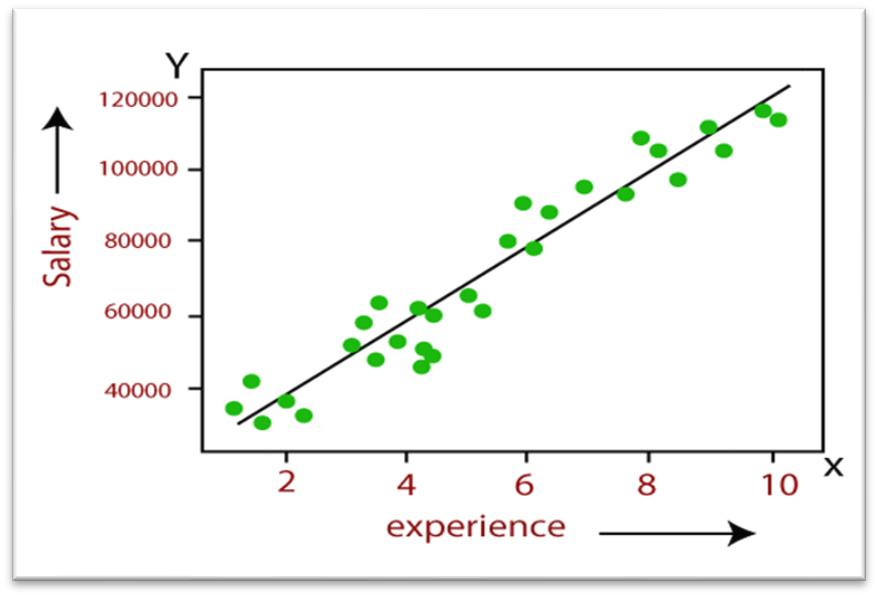
* **Examples of Linear Regression** o The weight of the person is linearly related to their height. So, this shows a linear relationship between the height and weight of the person. According to this, as we increase the height, the weight of the person will also increase.

**Linear Regression Equation is given below:**

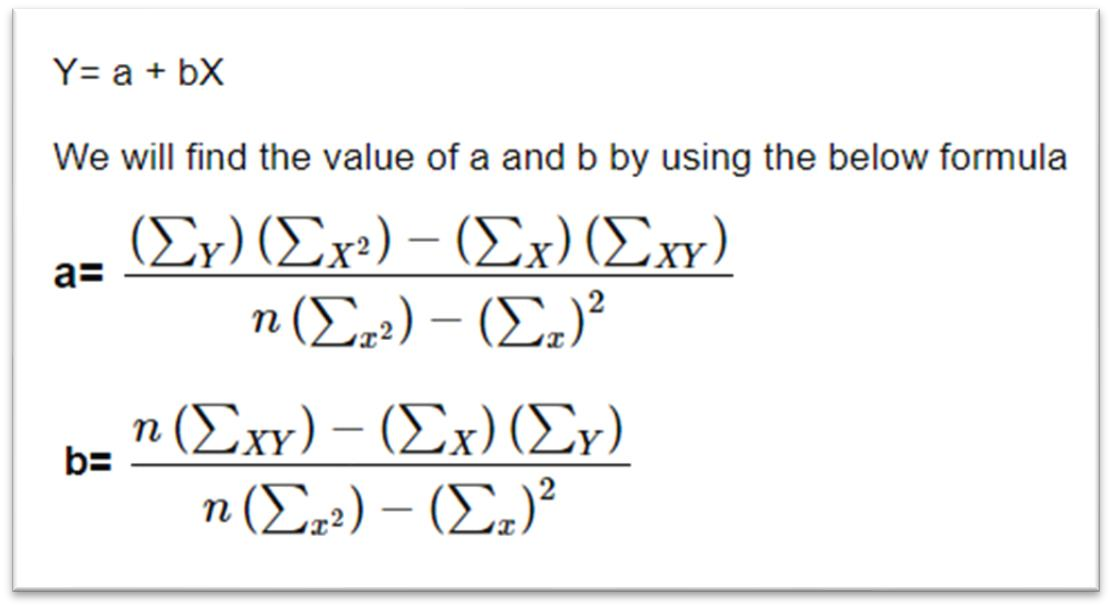
**Y=ax + b**

**where X is the independent variable and it is plotted along the x-axis Y is the dependent variable and it is plotted along the y-axis Here, the slope of the line is b, and a is the intercept (the value of y when x = 0)**

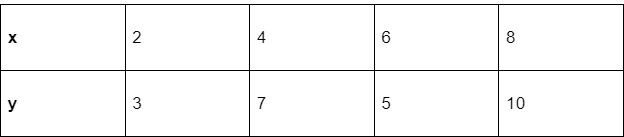
* Linear regression shows the linear relationship between the independent variable (X-axis) and the dependent variable (Y-axis), hence called linear regression.
* If there is only one input variable (x), then such linear regression is called **simple linear regression**. And if there is more than one input variable, then such linear regression is called **multiple linear regression**.
* The relationship between variables in the linear regression model can be explained using the below image. Here we are predicting the salary of an employee on the basis of **the year of experience**.



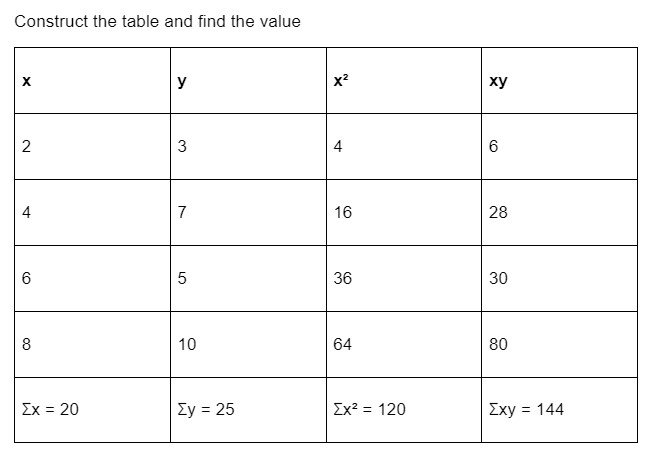
**Formula**



1. **Find a linear regression equation for the following two sets of data:**



**Solution:** To find the linear regression equation we need to find the value of Σx, Σy, Σx2(x Square) and Σxy



The formula

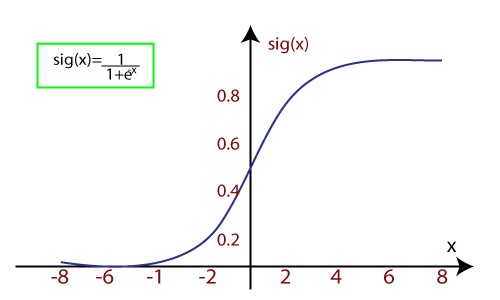
of the linear equation is y=a+bx. Using the formula we will find the value of a and b

|  |  |
| --- | --- |
|  |  |

1. **What is Logistic Regression? Explain its applications?** 
   * Logistic regression is a popular and fundamental statistical method used in machine learning for binary classification tasks, where the goal is to predict one of two possible outcomes (e.g., yes/no, spam/ham, 1/0).
   * Logistic regression is another supervised learning algorithm which is used to solve the classification problems. In **classification problems**, we have dependent variables in a binary or discrete format such as 0 or 1.
   * Logistic regression algorithm works with the categorical variable such as 0 or 1, Yes or No, True or False, Spam or not spam, etc.
   * It is a predictive analysis algorithm which works on the concept of probability.
   * Logistic regression uses **sigmoid function** or logistic function which is a complex cost function. This sigmoid function is used to model the data in logistic regression. The function can be represented as:

 o f(x)= Output between the 0 and 1 value. o x= input to the function o e= base of natural logarithm.

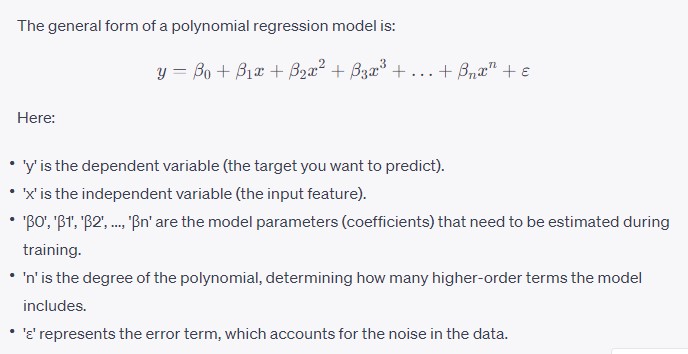
When we provide the input values (data) to the function, it gives the S-curve as follows

 o It uses the concept of threshold levels, values above the threshold level are rounded up to 1, and values below the threshold level are rounded up to 0.

There are three types of logistic regression:

* + **Binary(0/1, pass/fail)** o **Multi(cats, dogs, lions)** o **Ordinal(low, medium, high)**

1. **Explain polynomial regression with example?** 
   * Polynomial regression is a type of regression analysis used in machine learning and statistics to model the relationship between a dependent variable (target) and one or more independent variables (features) when the relationship is not linear but can be approximated by a polynomial function.
   * In simple linear regression, the relationship between the independent variable 'x' and the dependent variable 'y' is represented as a straight line. However, in many real-world scenarios, the relationship may be more complex. Polynomial regression addresses this by allowing the model to fit a polynomial equation to the data.



# UNIT-II Chapter-02(Supervised Learning)

1. **Explain Bayes theorem with its formula and example?**

* It is based on Bayes theorem given by Thomas Bayes in the middle of the eighteenth century.
* It is based on a hypothesis that the given data belongs to a particular class.
* In this theorem probability is calculated for the hypothesis to be true.
* P(A) refers to the probability of occurrence of event A, while P(A|B) refers to the conditional probability of event A given that event B has already occurred.

**Bayes theorem** is defined as follows

**P(A|B) = P(B|A) P(A)/P(B)** …(1)

Let us first prove this theorem.

We already know that

**P(A|B) = P(A & B)/P(B)**  …(2)

[It is the probability that next event will be A when B has already happened] Similarly,

**P(B|A) = P (B & A)/P(A)** …(3)

From equation (3):

**P (B & A) = P(B|A) \* P(A)**

By putting this value of P (B & A) in equation (2), we get ….

**P(A|B) = P(B|A) P(A)/P(B)**

This proves the Bayes theorem.

According to Bayes theorem: **P(Ci |X) = P(X|Ci ) P(Ci )/P(X)** Where:

**P(Ci|X)** is the conditional probability of class being Ci.

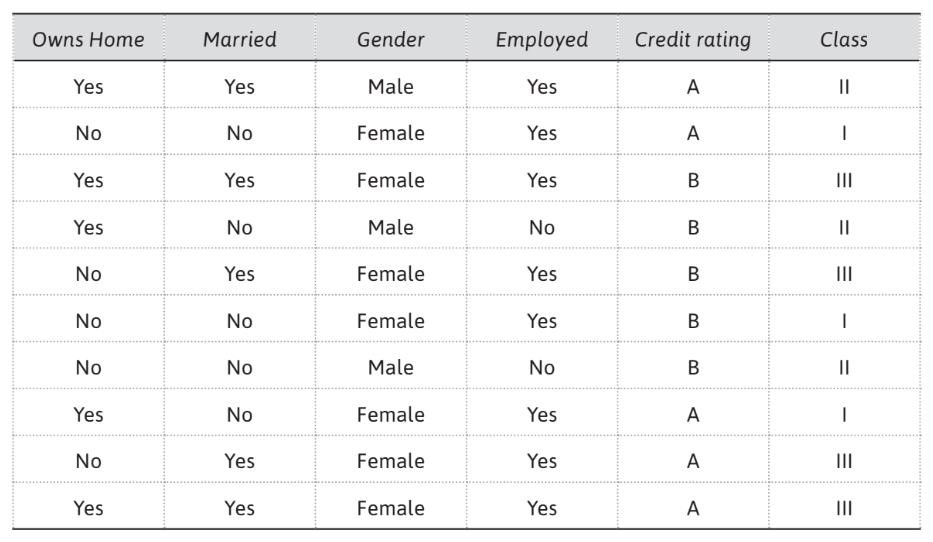
**P(X|Ci)** is the conditional probability of the record being X.

**P(Ci)** is the probability that an object belongs to class Ci. **P(X)** is the probability of occurrence of record X.

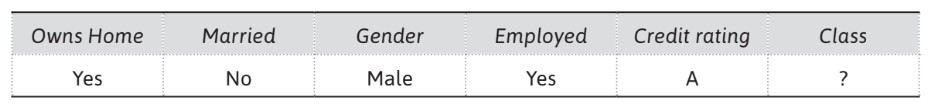
Here, we have already made the hypothesis that X has already occurred so P(X) is 1, so we have to calculate P(X|Ci) and P(Ci) in order to find required value.

**Example:**

To further understand this concept let us consider the following database, where the class of customer is defined based on his/her marital status, gender, employment status and credit rating as shown below.



Here, we have to predict the class of occurrence of X and let us suppose X is as shown below.



Then we have to calculate the probability for class Ci, when X has already occurred. Thus it is P(Ci | Yes, No, Male, Yes, A) = P(Yes, No, Male, Yes, A | Ci) \* P (Ci)

Let us first calculate probability of each class, i.e., P(Ci). Here, we have three classes, i.e., I, II and III. There are a total 10 instances in the given dataset and there are three instances for class I, three for class II and four for class III. Thus, Probability of class I, i.e., P(I) = 3/10 = 0.3

Probability of class II, i.e., P(II) = 3/10 = 0.3

Probability of class III, i.e., P(III) = 4/10 = 0.4

Now, let us calculate P(X|Ci), i.e., P(Yes, No, Male, Yes, A | Ci)P(Yes, No, Male, Yes, A | Ci) =

P(Owns Home = Yes| Ci) \* P(Married = No| Ci) \* P(Gender = Male|Ci) \* P(Employed = Yes| Ci) \* P (Credit rating = A| Ci)

Thus, we need to calculate

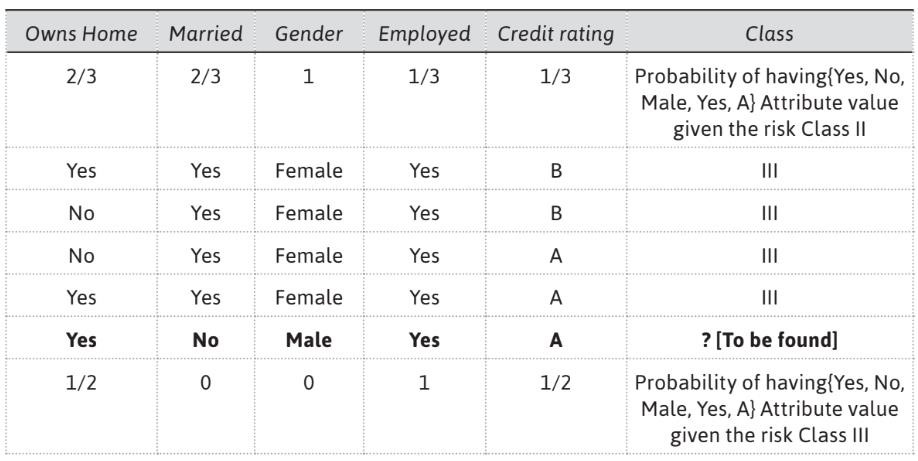
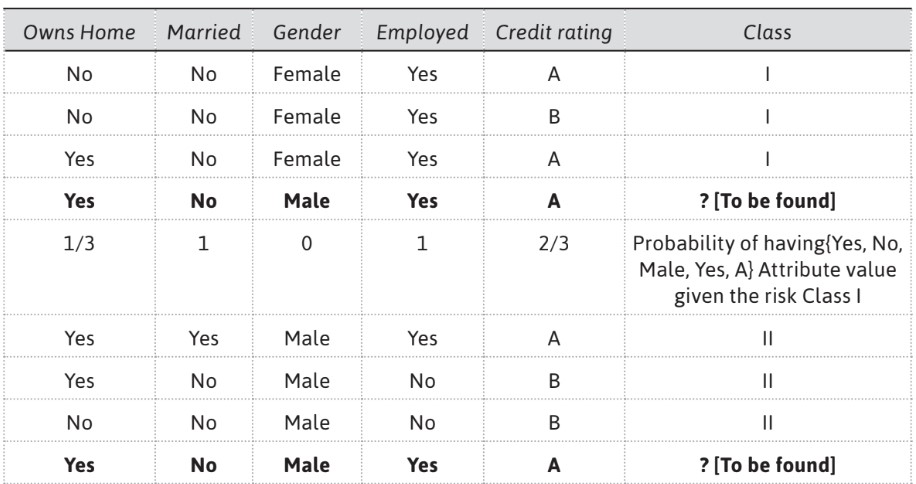
P(Owns Home = Yes|Ci),

P(Married = No| Ci),

P(Gender = Male|Ci),

P(Employed = Yes|Ci),

P (Credit rating = A|Ci).



Thus, P(X|I) = 1/3 \* 1 \* 0 \* 1 \* 2/3 = 0

P(X|II) = 2/3 \*2/3 \* 1 \* 1/3 \* 1/3 = 4/81

P(X|III) = 1/2 \* 0 \* 0 \* 1 \* 1/2 = 0

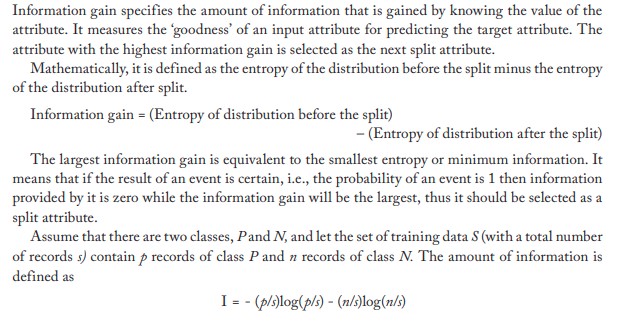
P(I | Yes, No, Male, Yes, A) = P(Yes, No, Male, Yes, A | I) \* P (I) = 0\*0.3 = 0

P(II | Yes, No, Male, Yes, A) = P(Yes, No, Male, Yes, A | II) \* P (II) = 4/81\*0.3 = 0.0148

P(III | Yes, No, Male, Yes, A) = P(Yes, No, Male, Yes, A | III) \* P (III) = 0\*0.4 = 0

Therefore, X is assigned to Class II. It is very unlikely in real life datasets that the probability of class comes out to be 0 as in this example.

1. **Define Information gain with examples?**



1. **What are the advantages and disadvantages of decision tree classification?**

**Advantages of the decision tree method**

The main advantages of a decision tree classifier are as follows:

* + The rules generated by a decision tree classifier are easy to understand and use.
  + Domain knowledge is not required by the decision tree classifier.
  + Learning and classification steps of the decision tree are simple and quick

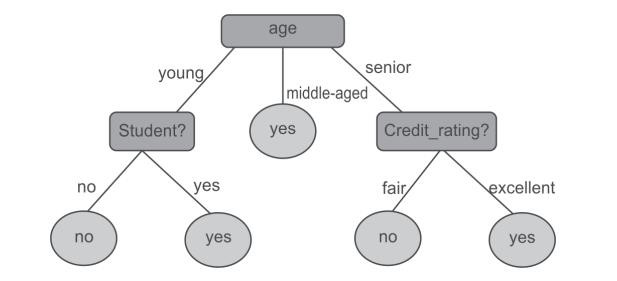
**Disadvantages of the decision tree**

The disadvantages of a decision tree classifier can be:

* + Decision trees are easy to use compared to other decision-making models, but preparing decision trees, especially large ones with many branches, are complex and time-consuming affairs.
  + They are unstable, meaning that a small change in the data can lead to a large change in the structure of the optimal decision tree.
  + They are often relatively inaccurate.

1. **What is decision tree? Explain the concept of Gini index with example?**

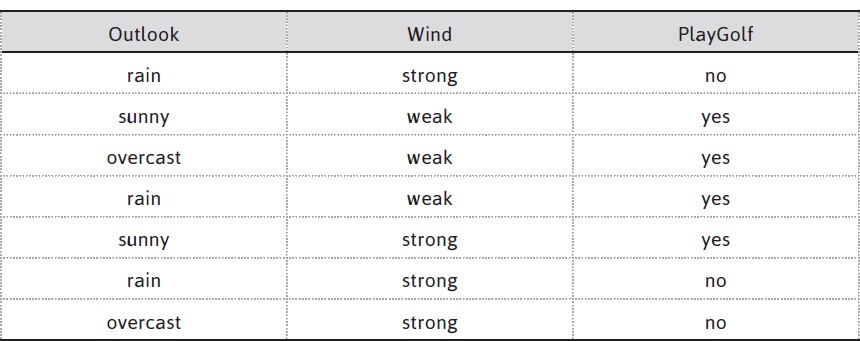
* In the decision tree classifier, predictions are made by using multiple ‘if…then...’ conditions which are similar to the control statements in different programming languages.
* The decision tree structure consists of a root node, branches and leaf nodes. Each internal node represents a condition on some input attribute, each branch specifies the outcome of the condition and each leaf node holds a class label. The root node is the topmost node in the tree.
* The decision tree shown below represents a classifier tasked for predicting whether a customer will buy a laptop or not. Here, each internal node denotes a condition on the input attributes and each leaf node denotes the predicted outcome (class). By traversing the decision tree, one can analyze that if a customer is middle aged then he will probably buy a laptop, if a customer is young and a student then he will probably not buy a laptop. If a customer is a senior citizen and has an excellent credit rating then he can probably buy a laptop. The system makes these predictions with a certain level of probability.



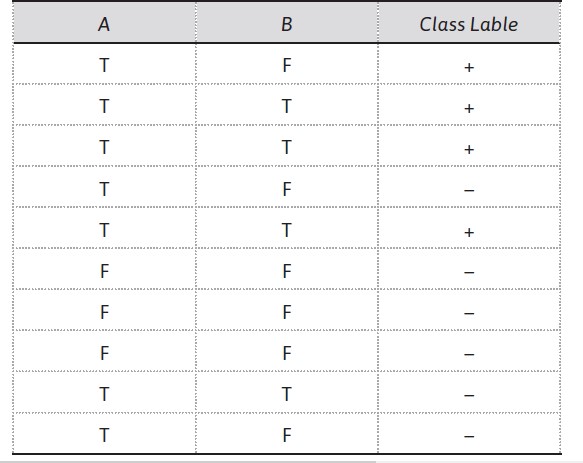
**Gini Index:**

* The Gini Index is used to represent level of equality or inequality among objects. It can also be used to make decision trees.
* The Gini Index always ranges between 0 and 1
* It was designed to define the gap between the rich and the poor, with 0 signifying perfect equality where all people have the same income while 1 demonstrating perfect inequality where only one person gets everything in terms of income and rest of the others get nothing.
* From this, it is evident that if Gini Index is very high, there will be huge inequality in income distribution.

1. **Identify the attribute that will act as the root node of a decision tree to predict golf for following database with Gini Index. Indicate all the intermediate steps.**



1. **Calculate the information gain when splitting on A and B. Which attribute would the decision tree induction algorithm choose?**



# UNIT-II

**Chapter-03(Rule Based System)**

1. **Explain rule-based system with example?**
2. **List and explain applications of Rule-Based System.**
3. **Define Rule Based? Write down syntax to create rules in rule based?**
4. **Explain with example any one Rule-Based Model?**

# UNIT-III Chapter-01(Unsupervised Learning) Chapter-02(Applications of Cluster Analysis)

1. **Briefly explain unsupervised learning? Explain Clustering analysis with example?**
2. **List all the applications of clustering? Explain any 5 in details?** 
   * **Marketing**: It helps marketers find out distinctive groups among their customer bases, and this knowledge helps them improve their targeted marketing programs.
   * **Land use:** Clustering is used for identifying areas of similar land use from the databases of earth observations.
   * **Insurance**: Clustering is helpful for recognizing clusters of insurance policyholders with a high regular claim cost.
   * **City-planning**: It also helps in identifying clusters of houses based on house type, geographical location, and value.
   * **Earthquake studies**: Clustering is helpful for analysis of earthquakes as it has been noticed that earthquake epicenters are clustered along continent faults.
   * **Biology studies**: Clustering helps in defining plant and animal classifications, identifying genes with similar functionalities, and in gaining insights into structures inherent to populations.
   * **Web discovery**: Clustering is helpful in categorizing documents on the web for information discovery.
   * **Fraud detection**: Clustering is also helpful in outlier detection applications such as credit card fraud detection.

1. **Explain desired features of clustering analysis?**

The desired feature of an ideal clustering technique is that intra-cluster distances should be minimized and inter-cluster distances should be maximized.

* + - **Scalability**: Clustering algorithms should be capable of handling small as well as large datasets smoothly.
    - **Ability to handle different types of attributes**: Clustering algorithms should be able to handle different kinds of data such as binary, categorical and interval-based (numerical) data.
    - **Independent of data input order**: The clustering results should not be dependent on the ordering of input data.
    - **Identification of clusters with different shapes**: The clustering algorithm should be capable of identifying clusters of any shape.
    - **Ability to handle noisy data**: Usually, databases consist of noisy, erroneous or missing data, and algorithm must be able to handle these.
    - **High performance**: To have a high performance algorithm, it is desirable that the algorithm should need to perform only one scan of the dataset. This capability would reduce the cost of input-output operations.
    - **Interpretability**: The results of clustering algorithms should be interpretable, logical and usable.
    - **Ability to stop and resume**: For a large dataset, it is desirable to stop and resume the task as it can take a huge amount of time to accomplish the full task and breaks may be necessary.
    - **Minimal user guidance**: The clustering algorithm should not expect too much supervision from the analyst, because commonly the analyst has a limited knowledge of the dataset.

In clustering, distance metrics play a vital role in comprehending the similarity between the objects.

1. **List and explain with example various metrics used in clustering analysis?**

A distance metric is a function d(x, y) that specifies the distance between elements of a set as a nonnegative real number. Two elements are equal under a particular metric if the distance between them is zero. Distance functions present a method to measure the closeness of two elements.

important distance metrics used in the measuring similarity among objects:

* + Euclidean distance
  + Manhattan distance

**Euclidean distance:**

Euclidean distance is mainly used to calculate distances. The distance between two points in the plane with coordinates (x, y) and (a, b) according to the Euclidean distance formula is given by:

**Euclidean dist((x, y), (a, b)) = √(x - a)² + (y - b)²**

For example,

the (Euclidean) distance between points (-2, 2) and (2, -1) is calculated as

Euclidean dist((-2, 2), (2, - 1)) = √(-2 - (2))² + (2 - (-1))²

= √(-4)² + (3)²

= √16 + 9

= √25

= 5

**Manhattan distance:**

Manhattan distance is also called L1-distance. It is defined as the sum of the lengths of the projections of the line segment between the two points on the coordinate axes. For example, the distance between two points in the plane with coordinates (x, y) and (a, b) according to the Manhattan distance formula, is given by:

**Manhattan dist((x, y), (a, b)) = | x – a | + | y – b |**

Using the formula of Manhattan distance, we can calculate the similarity distance among persons. The calculation for the distance between person 1 and 2 is: Manhattan dist((30, 70), (40, 54)) = |30 - 40| + |70 – 54|

= |- 10 | + | 16|

= 10+16

= 26

The calculation for the distance between person 1 and 3 is:

Manhattan dist ((30, 70), (80, 50)) = |30 – 80| + |70 – 50|

= 50+20

= 70

The calculation for the distance between person 2 and 3 is:

Manhattan dist((40, 54), (80, 50)) = |40 –80| + |54 – 50|

= 40+4

= 44

This indicates that the persons 1 and 2 are most similar while person 1 and person 3 are most dissimilar and it produces the same conclusion as Euclidean distance.

**Chebyshev distance**

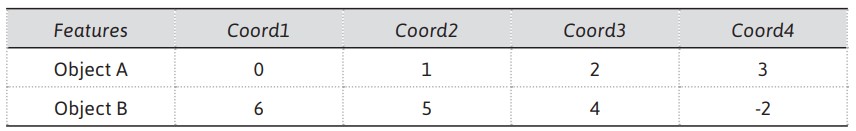
It is also called as chessboard distance because, in a game of chess, the minimum number of moves required by a king to go from one square to another on a chessboard equals Chebyshev distance between the centers of the squares.

Chebyshev distance is defined on a vector space, where the distance between two vectors is the maximum value of their differences along any coordinate dimension. Formula of Chebyshev distance is given by:

**Chebyshev dist((r1, f1), (r2, f2)) = max(|r2−r1|,|f2-f1|)**

**Example:**

Using the formula of Chebyshev distance, let us find the distance between object A and object B.



Object A coordinate = {0,1,2,3}

Object B coordinate = {6,5,4,-2}

According to Chebyshev distance formula

D = max(|r2−r1|,|f2−f1|)

= max(|6-0|,|5-1|,|4-2|,|-2-3)|)

= max(6,4,2,5)

= 6

distance between person 1 and 2 is:

Chebyshev dist((30, 70), (40, 54)) = max( |30 - 40|, |70 – 54|)

= max(|- 10 |, | 16|)

= 16

The calculation for the distance between person 1 and 3 is:

Chebyshev dist((30, 70), (80, 50)) = max(|30 – 80| , |70 – 50|)

= max(50,20)

= 50

The calculation for the distance between person 2 and 3 is:

= max(|40 – 80|, |54 – 50|)

Chebyshev dist((40, 54), (80, 50)) = max(40,4)

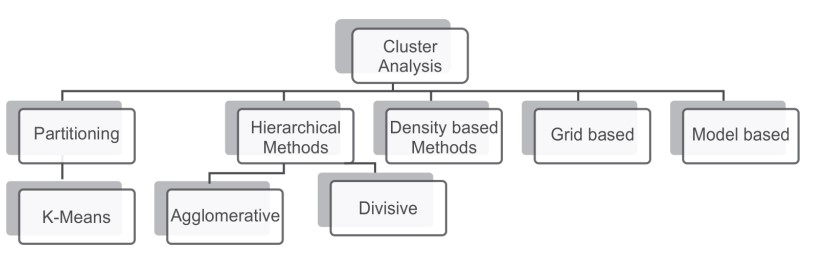
= 40

This indicates that the persons 1 and 2 are most similar while person 1 and person 3 are most dissimilar and it produces the same conclusion as Euclidean and Manhattan distance.

1. **List all the clustering algorithm and explain partition clustering?**

* + Clustering methods/algorithms can be categorized into five categories which are given as follows: Partitioning method: It constructs random partitions and then iteratively refines them by some criterion.
  + Hierarchical method: It creates a hierarchical decomposition of the set of data (or objects) using some criterion.
  + Density-based method: It is based on connectivity and density functions.
  + Grid based method: It is based on a multiple-level granularity structure.
  + Model based method: A model is considered for each of the clusters and the idea is to identify the best fit for that model.

The categorization of clustering algorithms is shown below:



**Partitioning Clustering:**

Clustering is the task of splitting a group of data or dataset into a small number of clusters.

For example, the items in a grocery store are grouped into different categories (butter, milk, and cheese are clustered in dairy products).

In the partitioning method, we cluster objects based on attributes into a number of partitions.

The k-means clustering is an important technique which falls under partitioning clustering.

1. **Explain k-means clustering algorithm with its flow chart?**

In the k-means clustering algorithm, n objects are clustered into k clusters or partitions based on attributes, where k < n and k is a positive integer number.

In simple words, in k-means clustering algorithm, the objects are grouped into ‘k’ number of clusters based on attributes or features.

The grouping of objects is done by minimizing the sum of squares of distances, i.e., a Euclidean distance between data and the corresponding cluster centroid.

**Working of the k-means algorithm:**

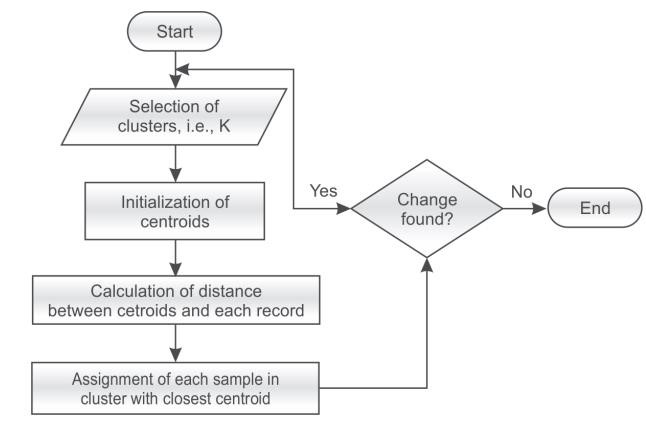
k-means clustering algorithm can be illustrated in five simple steps **Step 1:** Start with a selection of the value of k where k = number of clusters

**Step 2:** Creation of distance matrix between the centroids and each pattern

**Step 3:** Assign each sample in the cluster with the closest centroid (minimal distance)

**Step 4:** Update the new centroids for each cluster

**Step 5:** Repeat until no further change occurs



7. **Explain Hierarchical Clustering Algorithms (HCA) with its types?**

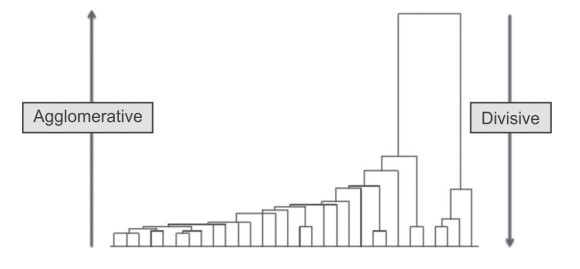
* Hierarchical clustering is a type of cluster analysis which seeks to generate a hierarchy of clusters.
* It is also called Hierarchical Cluster Analysis (HCA).
* Hierarchical clustering methods build a nested series of clusters in comparison to partitioned methods that generate only a flat set of clusters.
* There are two types of Hierarchical clustering: o agglomerative o divisive.

**Agglomerative:** This is a ‘bottom-up’ approach. In this approach, each object is a cluster by itself at the start and its nearby clusters are repetitively combined resulting in larger and larger clusters until some stopping criterion is met.

The stopping criterion may be the specified number of clusters or a stage at which all the objects are combined into a single large cluster that is the highest level of hierarchy as shown in Figure.

**Divisive:** This is a ‘top-down’ approach. In this approach, all objects start from one cluster, and partitions are performed repeatedly resulting in smaller and smaller clusters until some stopping criterion is met or each cluster consists of the only object in it as shown in Figure below.

Generally, the mergers and partitions are decided in a greedy manner.



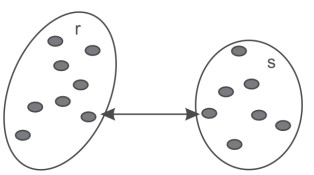
8. **List and explain role of linkage metrics used in HCA?**

* It is required to compute the distance between clusters by using some metrics before any clustering is performed.
* These metrics are called linkage metrics.
* It will be beneficial to identify the proximity matrix consisting of the distance between each point using a distance function.
* The important linkage metrics to measure the distance between each cluster are:
  + Single linkage o Complete linkage
  + Average linkage

**Single linkage**

* In single linkage hierarchical clustering, the shortest distance between two points in each cluster is considered as the distance between two clusters.
* For example, the distance between two clusters ‘r’ and ‘s’ to the left is equal to the length of the arrow between their two closest points as shown below and the formula of calculation of distance using single-linkage hierarchical clustering is given by:

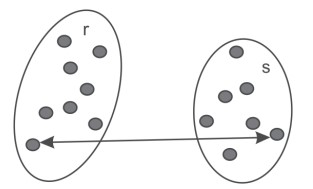
L(r,s) = min(D(xri, xsj))



**Complete linkage**

* In complete linkage hierarchical clustering, the longest distance between two points in each cluster is considered as the distance between two clusters.
* For example, the distance between clusters ‘r’ and ‘s’ to the left is equal to the length of the arrow between their two farthest points as shown in Figure and the formula of calculation of distance using complete linkage hierarchical clustering is given by:

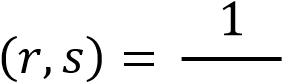
L(r,s) = max(D(xri, xsj))



**Average linkage**

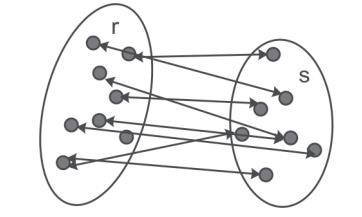
* In average linkage hierarchical clustering, the average distance between each point in one cluster to every point in the other cluster is considered as the distance between two clusters.
* For example, the distance between clusters ‘s’ and ‘r’ to the left is equal to the average length of each arrow connecting the points of one cluster to the other as shown in Figure and the formula of calculation of distance using average linkage hierarchical clustering is given by:

𝑛𝑌 𝑛𝑆

𝐿 ∑ ∑ 𝐷(𝑥𝑟𝑖, 𝑥𝑠𝑗)

𝑛𝑟𝑛𝑠

𝑖=1 𝑗=1



9. **Explain density-based clustering with its algorithm?**

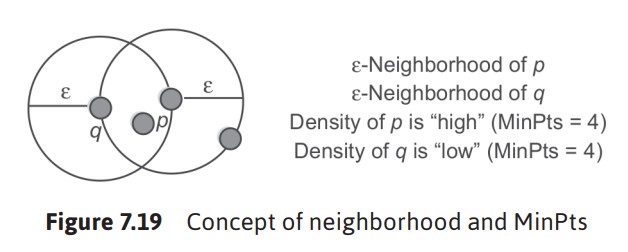
* Density-based clustering algorithms perform the clustering of data by forming the cluster of nodes based on the estimated density distribution of corresponding nodes in the region.
* These clusters are separated by regions of lower object density.
* A cluster is defined as a maximal set of density-connected points.
* A Density-based clustering method is known as DBSCAN (Density-Based Spatial Clustering of Applications with Noise).
* Familiarity with the following terms is a must to understand the workings of the DBSCAN algorithm.

**Neighborhood (e)**

Neighborhood is an important term used in DBSCAN.

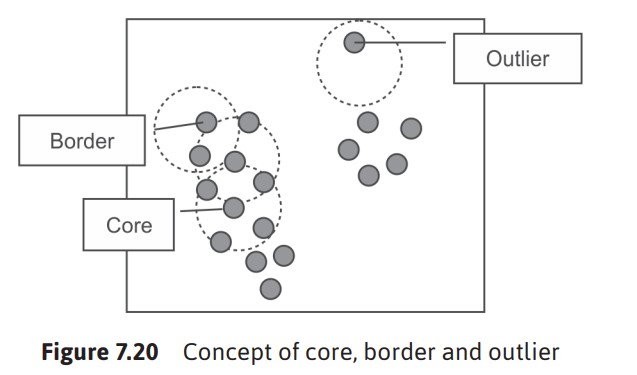
It represents objects within a certain radius of from a centroid type object.

The high-density neighborhood results if an object contains at least MinPts (minimum points) of objects in its neighborhood.



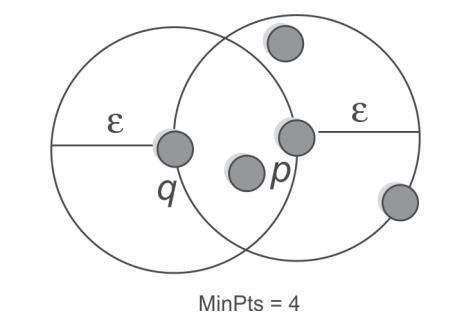
**Core, Border, and Outlier:**

A point is known as a core point if it has more than a specified number of points (MinPts) within neighborhood (e). These points must lie at the interior of a cluster. A border point is a point if it has fewer than MinPts within neighborhood (e), but it is in the neighborhood of a core point. A point is a noise or outlier point if it is neither a core point nor a border point. The concept of Core, Border, and Outlier is illustrated by considering Minpts = 4



**Density reachability; Directly and Indirectly density-reachable**

DBSCAN’s definition of a cluster is defined based on the concept of density reachability. Generally, a point q is directly density-reachable from a point p if it is not more distant than a given distance e (i.e., is part of its e-neighborhood). However, one may consider that p and q belong to the same cluster if point p is surrounded by necessarily many points.



**DBSCAN algorithm**

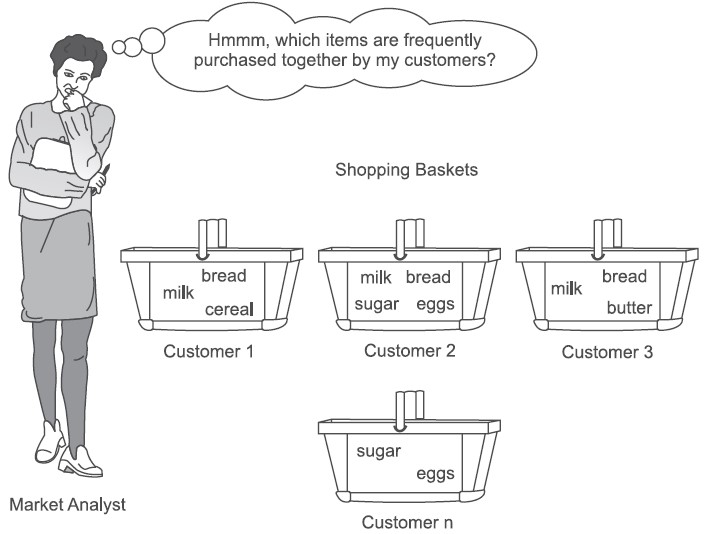
After the understanding of concepts of neighborhood, core, border, outlier points and density reachability, DBSCAN algorithm can be summarized as given below.

Algorithm for DBSCAN

1. Find the e (epsilon) neighbors of every point, and identify the core points with more than minPts neighbors.
2. Find the connected components of core points on the neighbourhood graph, ignoring all non-core points. 3. Assign each non-core point to a nearby cluster if the cluster is an e (eps) neighbor, otherwise assign it to noise.

# UNIT-III Chapter-03(Association Rules)

1. **Define Association Rule? Explain market basket analysis with example?** 
   * Association rule mining can be defined as identification of frequent patterns, correlations, associations, or causal structures among sets of objects or items in transactional databases
   * Association rule mining often known as ‘market basket’ analysis is very effective technique to find the association of sale of item X with item Y.
   * In simple words, market basket analysis consists of examining the items in the baskets of shoppers checking out at a market to see what types of items ‘go together’



* + This is commonly known as market basket data analysis.
  + Association rule mining can also be used in applications like marketing, customer segmentation, web mining, medicine, adaptive learning, finance and bioinformatics, etc

1. **Explain the concept of support and confidence with example?**

The metrics to judge the strength and accuracy of the rule are as follows:

* + Support
  + Confidence
  + Lift

**Support**

Let N is the total number of transactions. Support of X is represented as the number of times X

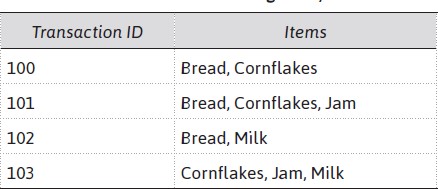
appears in the database divided by N, while the support for X and Y together is represented as the number of times they appear together divided by N as given below.

Support(X) = (Number of times X appears) / N = P(X)

Support(XY) = (Number of times X and Y appear together) / N = P(X ∩ Y)

1. **Explain the concept of lift with example?**

1. **Explain Naïve algorithm used for association rule generation?**
2. **Explain Apriori algorithm?**
3. **Explain The Naïve Algorithm for Finding Association Rules with example?**
4. **Used the following data set for example.**



1. **Apply the Naïve algorithm for the above database to identify frequent item sets with minimum support of 50%. Also find association rules having confidence of more than 75%.**

1. **Apply the Apriori algorithm for the following database to identify frequent item sets with minimum support of 50%. Also find association rules having confidence of more than 75%.**

