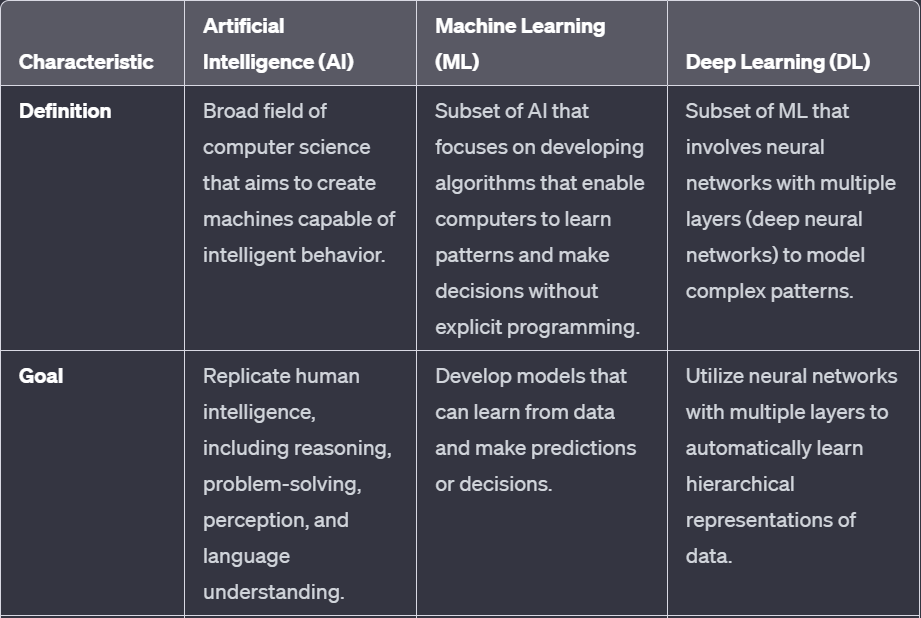
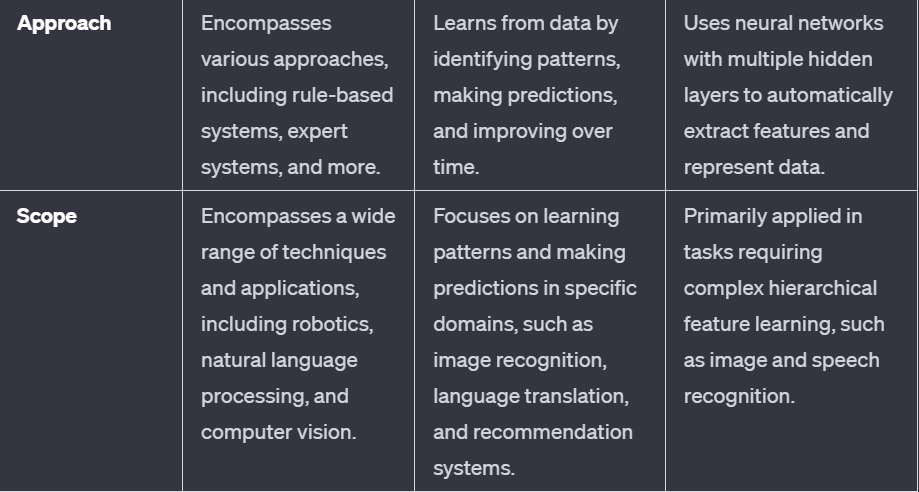
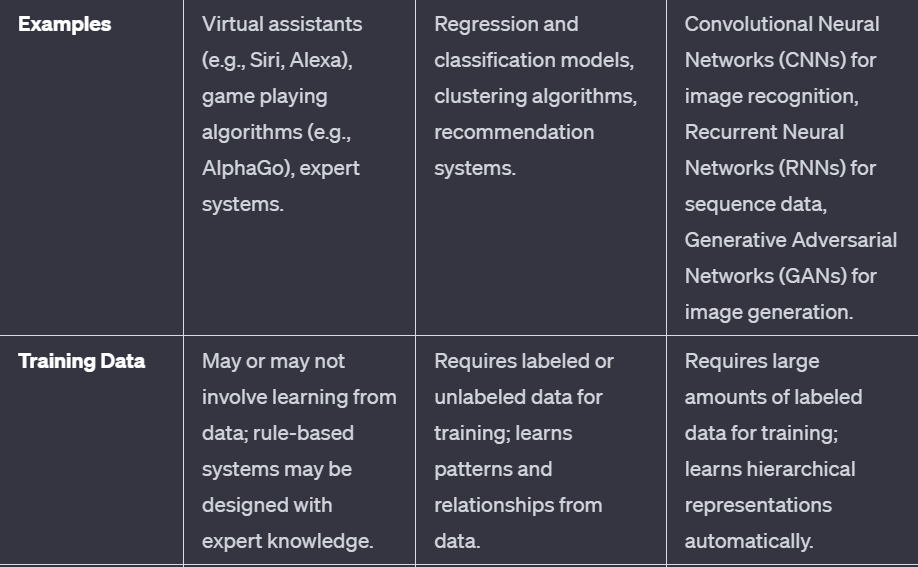
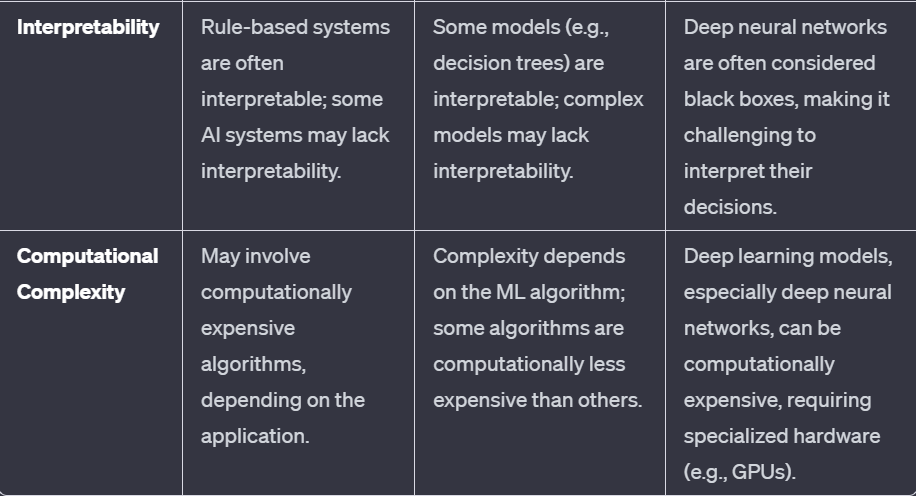
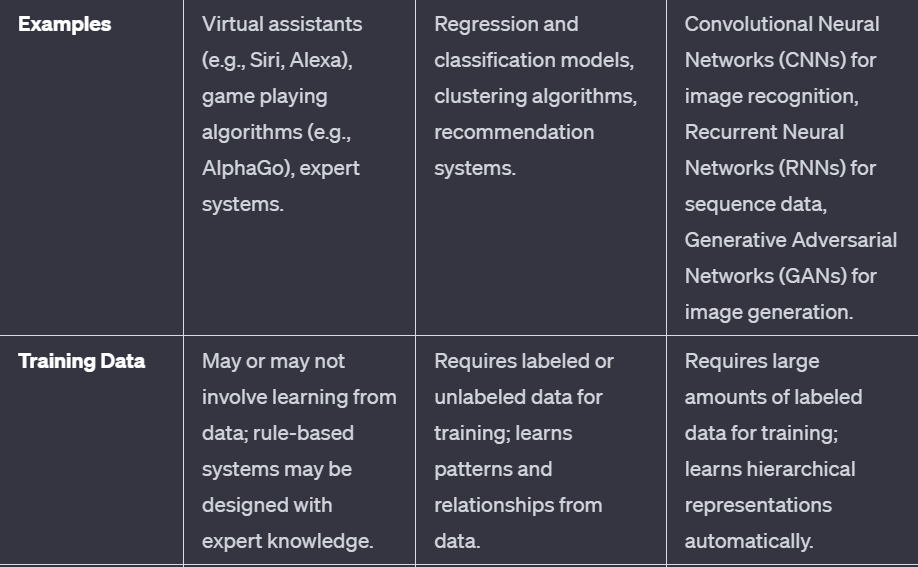
**Machine learning**

****

****

**Important Terms**

**NP** : Non polynomial whose time complexity grows exponentially is used as np problem mostly it is seen in uninformed search where we do not have knowledge of the domain and we define all the states where np problem occurs like ex 8 puzzle problem Worst case complexity of 8 puzzle problem is b^d which is 3^20 and in 15 puzzle its 10^13 this is what np is

**Heuristic :**Technique designed to solve a problem This mostly reduces the time complexity by converting np to p and thus reduces the time complexity. For example, we look at the depth first search,me We see whose cost is less and we go deeper into that side. The other side we don’t go this is what the heuristic used in informed search which gives us optimal values. cost is called as heuristic value

**UNIT-1**

* 1. **State Space Search**
* No of states in which a problem can go
* efn
* It is a problem-solving technique used in Artificial Intelligence (AI) to find the solution path from the initial state to the goal state by exploring the various states.
* The state space search approach searches through all possible states of a problem to find a solution.
* A way to mathematically represent a problem by defining all the possible states in which the problem can be.
* This is used in search algorithms to represent the initial state, goal state, and current state of the problem.
* Each state in the state space is represented using a set of variables.{States.possible Action,Result,Cost}
* The efficiency of the search algorithm greatly depends on the size of the state space,
  1. **Features**
* Exhaustiveness:State space search explores all possible states of a problem to find a solution.
* Precise: It give precise information about all the states so that machine is able to analyze it properly
* Completeness:If a solution exists, state space search will find it.
* Optimality:Searching through a state space results in an optimal solution.
  1. **Example: 8 puzzle problem**

The 8-Puzzle is a sliding puzzle that consists of a 3x3 grid with eight numbered tiles and one blank space. The goal is to rearrange the tiles from an arbitrary initial state to a goal state by sliding tiles into the blank space.

1. State Representation:

The state is represented by the arrangement of the tiles on the board. For example:

Initial State: Goal State:

1 2 3 1 2 3

4 5 6 8 4

7 8 7 6 5

1. Operators/Actions:

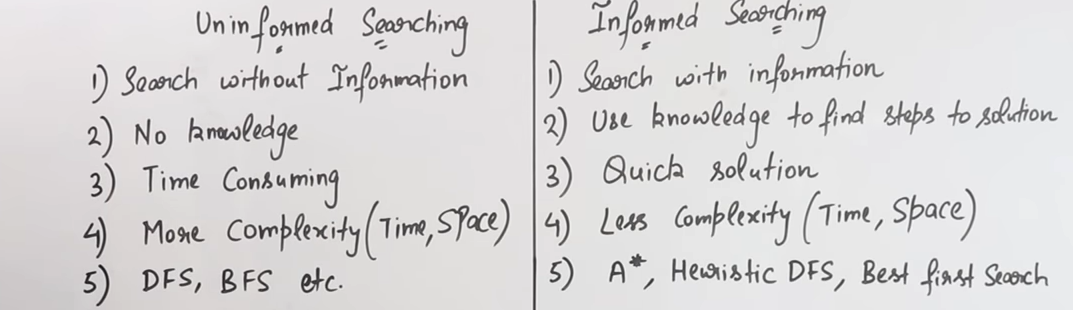
Possible actions include moving a tile adjacent to the blank space into the blank space.example Upward,downward right left

1. Goal State:

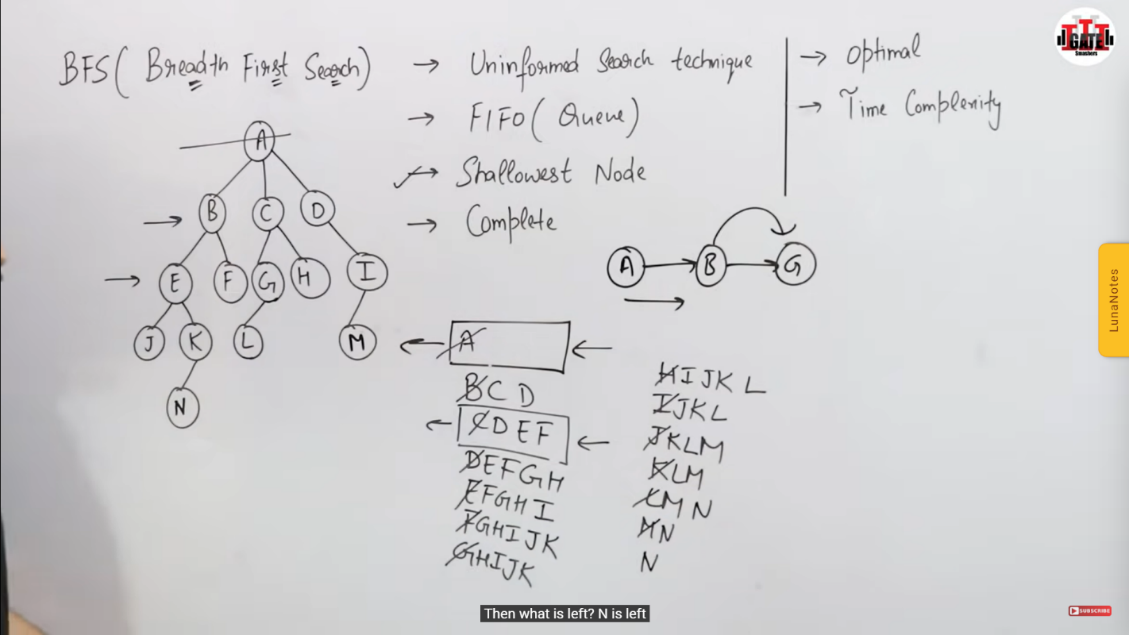
The goal state is a predefined stae and check each state that it ias a goal state or not

1. Cost: the amount of time/space/cost required to move from one state to another
   1. **Types of Space state search**

**Uninformed**: Also called as blind search don’t have additional information about the problem domain. It just have goal and initial state

**Informed Search**: uses additional information, such as heuristics, to guide the search process.

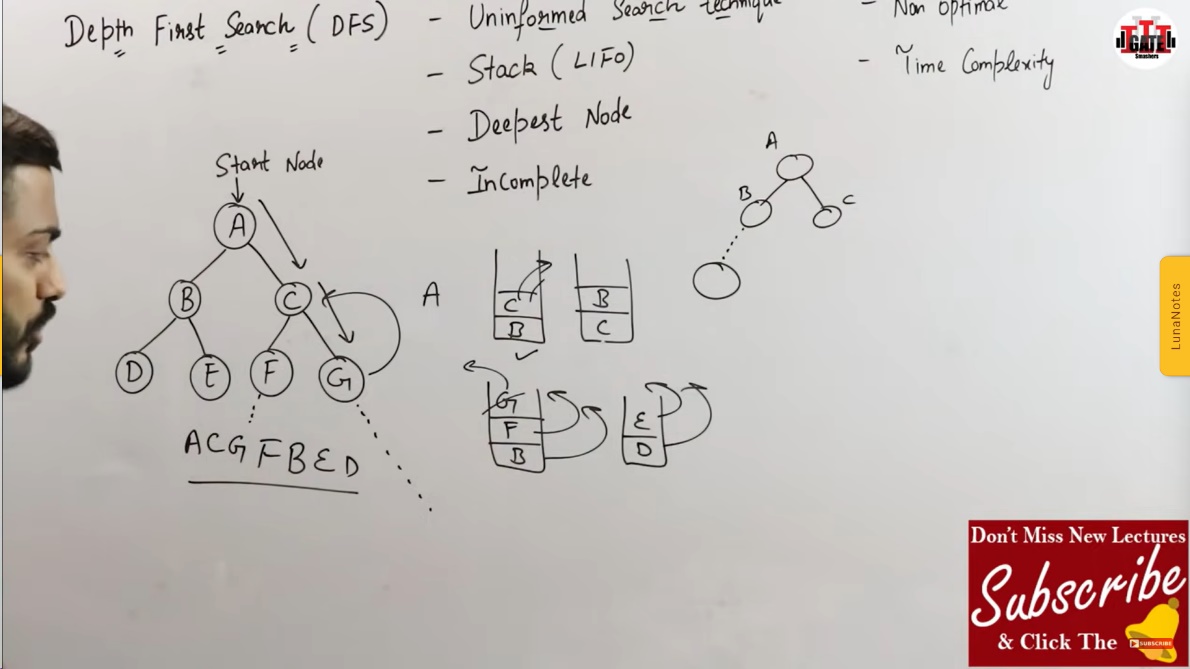
**Breadth First Search**

* It comes under uninformed search technique in which We don’t have the domain specific knowledge about the domain problem like what is the cost in reaching the goal state
* **Optimal**: Here optimal means shortest value given to reach the goal state if value is constant or not it always provide optimal or shortest path to reach at goal state
* **Complete**: it always provides a solution to the propblem
* **Shallow** **node**:unlike dfs it travels through level by level which is very important as reaches to every shallowest (neighbour) node so if there is a node at a level it first not traverse deeper
* **FIFO** : **Queue**
* ****Example:

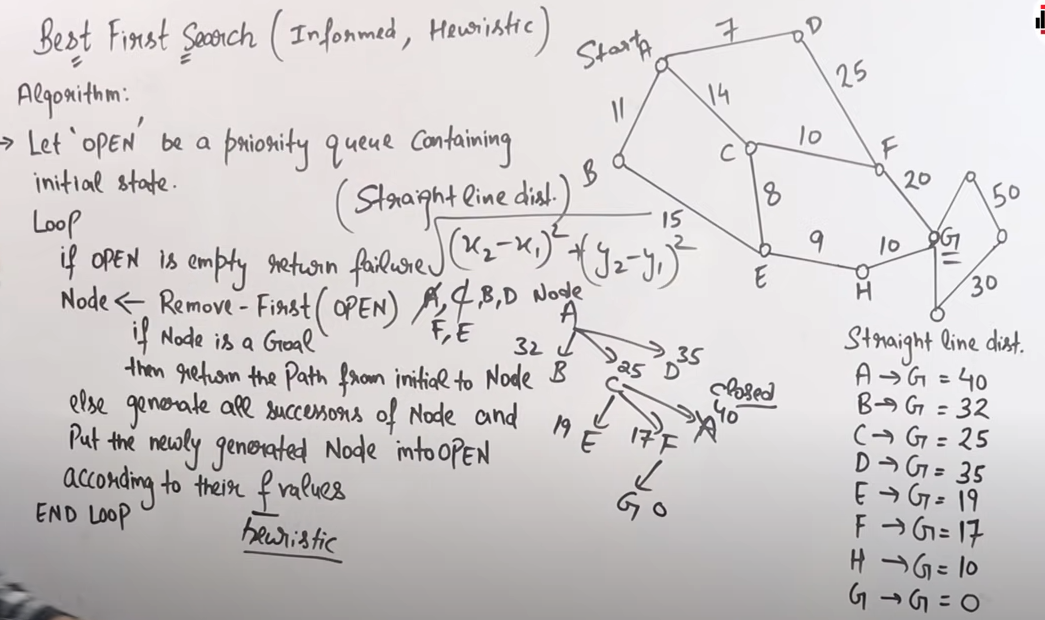
Disadvantages:

* It requires lots of memory since each level of the tree must be saved into memory to expand the next level.
* BFS needs lots of time if the solution is far away from the root node.
* Complexity : O(bd). b branches D is depth

**Depth First Search**

• Complexity : O(bd). b branches D is depth

**Best First Search**



* Informed Search technique
* Provides good solution not optimal solution
* Provides normal complexity but in worst case O(bd)
* What happens in best first search is we start from A to the goal g and we will check for each value the cost of reaching up to the goal then we will check from the starting node A there are three nodes B C and D to move now we will check from B to G what is the cost from C to G what is the cost and from D to G what is the cost then we will take the least value of reaching to the node goal then now we have seen C has the least value that is 25 then we will not approach to B and D and we will not check on B and D this is called as informed or heuristic method so we will go through C and now from we will check from C there are three nodes E F and A but we will cancel that A because it is already been taken so we will put them in a closed then we will check from E to F and now we will take the least value from E to F so F has the least value then we will check from F to G and then this is on now we have reached from F2 as goal so this is how we have get our best first search

**Hill climbing search**

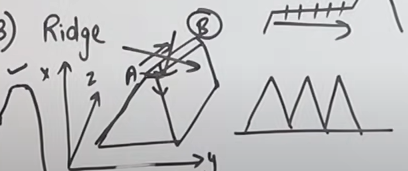
* Local search algorithm which continuously moves in the direction of increasing elevation/value to find the peak of the mountain or best solution to the problem. It terminates when it reaches a peak value where no neighbor has a higher value.
* No backtracking
* doesn’t have knowledge of global domain they have knowledge of local domain also called greedy local search as it only looks to its good immediate neighbor state and not beyond that.
* Hill Climbing is mostly used when a good heuristic is available.
* we don't need to maintain and handle the search tree or graph as it only keeps a single current state.
* **Examples** of Hill climbing algorithm is Traveling-salesman Problem in which we need to minimize the distance traveled by the salesman.
* **Problems**
* **Local Maximum**: A local maximum is a peak state in the landscape which is better than each of its neighboring states, but there is another state also present which is higher than the local maximum.

**Solution**: Backtracking ,the algorithm can backtrack the search space and explore other paths by creating alist of available paths as well.

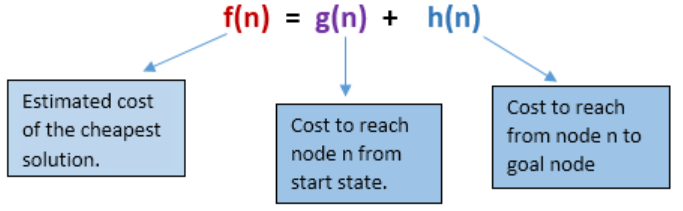
* **Plateau** is the flat area of the search space in which all the neighbor states of the current state contains the same value, because of this algorithm does not find any best direction to move. A hill-climbing search might be lost in the plateau area.

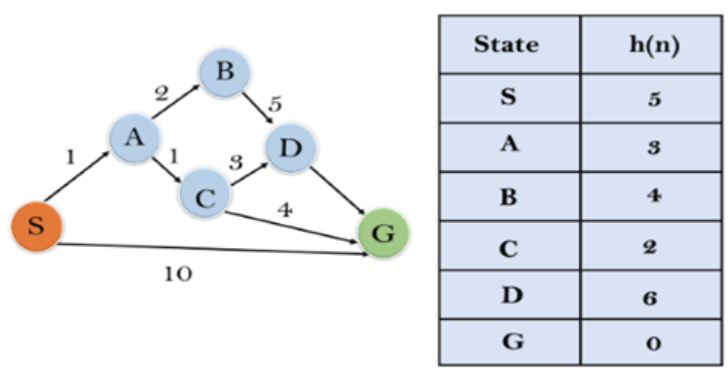
**Solution**: Randomly select a state which is far away from the current state so it is possible that the algorithm could find non-plateau region.

* **Ridge** is a special form of the local maximum. an area which is higher than its surrounding areas and here direction can’t be changed so if it is moving in A direction and gets the maximum and there exist B also but higher in different direction it will not be able to find the value

**Solution :** Moving in diferent direction is the only solution

**A\* search Technique**

* Informed Search Technique
* A\* search is the most commonly known form of best-first search.
* It uses heuristic function h(n), and cost to reach the node n from the start state g(n).
* A\* search algorithm finds the shortest path through the search space using the heuristic function.
* This search algorithm expands less search tree and provides optimal result faster.
* In A\* search algorithm, we use search heuristic as well as the cost to reach the node. Hence we can combine both costs as following, and this sum is called as a fitness number.
* ****A\* search algorithm is optimal and complete.

****

**Initialization: {(S, 5)}**

**Iteration1: {(S--> A, 4), (S-->G, 10)}**

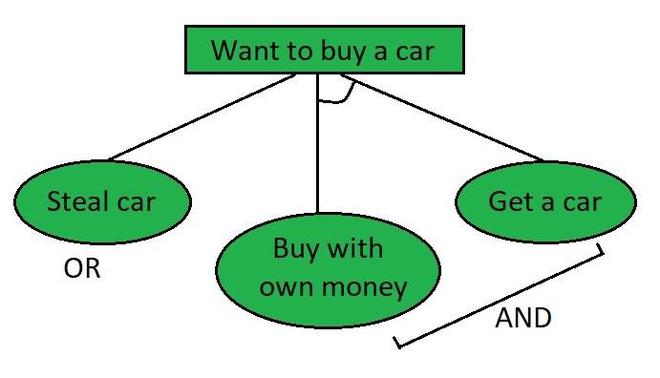
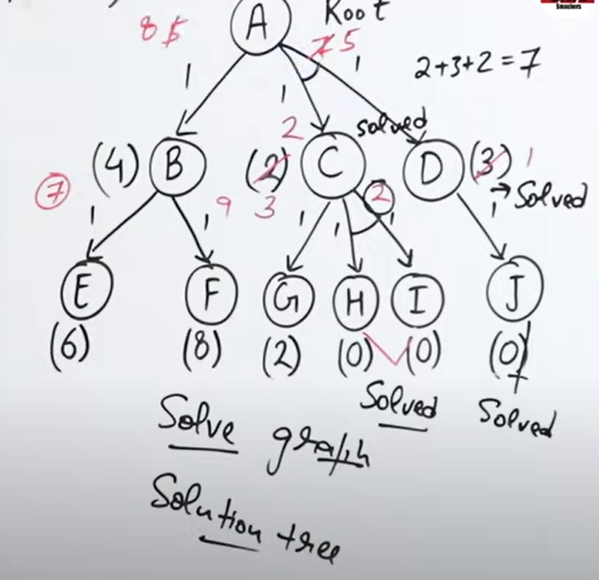
**Iteration2: {(SA--> C, 4), (SA-->B, 7)**

**Iteration3: {(SAC--->G, 6), (SAC--->D, 11),**

**Iteration 4 will give the final result, as S--->A--->C--->G it provides the optimal path with cost 6.**

**AO\* Algorithm**

* Gatesmasher ao\* algo
* Performs Best-first search.
* divides any given difficult problem into a smaller group of problems that are then resolved using the AND-OR graph concept.
* AND OR graphs are specialized graphs that are used in problems that can be divided into smaller problems.
* The AND side of the graph represents a set of tasks that must be completed to achieve the main goal,
* hile the OR side of the graph represents different methods for accomplishing the same main goal.



Example

**Means End Analysis**

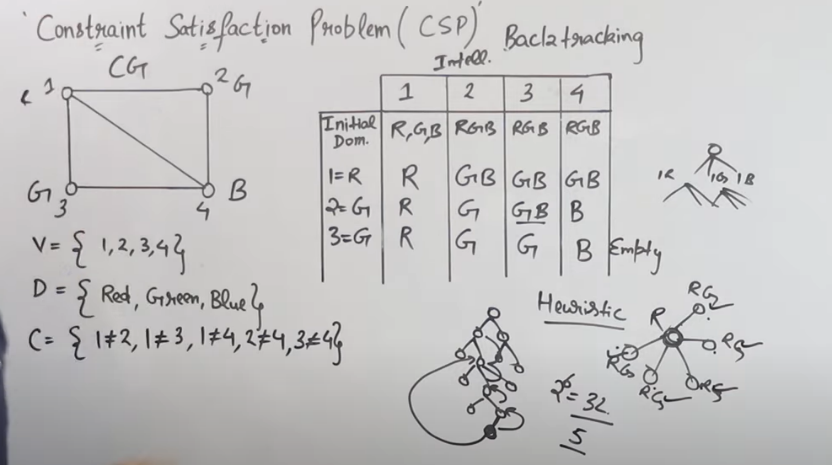
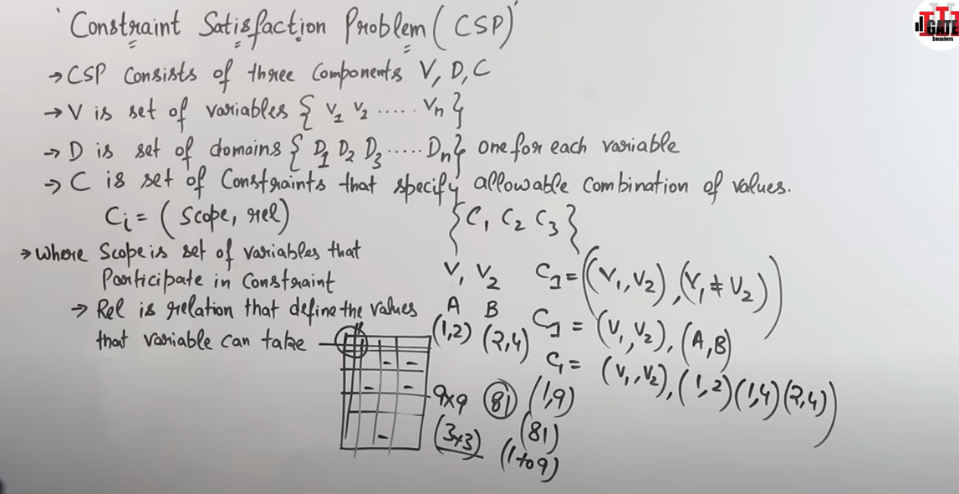
* mixture of the two directions is appropriate for solving a complex and large problem. Such a mixed strategy, make it possible that first to solve the major part of a problem and then go back and solve the small problems arise during combining the big parts of the problem. Such a technique is called Means-Ends Analysis.
* Means-Ends Analysis is problem-solving techniques used in Artificial intelligence for limiting search in AI programs.
* It is a mixture of Backward and forward search technique.
* The MEA analysis process centered on the evaluation of the difference between the current state and goal state.

**How means-ends analysis Works:**The means-ends analysis process can be applied recursively for a problem

1. Evaluate the difference between Initial State and final State.
2. Select the various operators which can be applied for each difference.
3. Apply the operator at each difference, which reduces the difference between the current state and goal state.

**Constraint Satisfaction Problem**

Basically what happens

1. There are three variables or components in CSP =>V,D,C (define in below image)
2. C is a constraint that are applied on V
3. So example is also given in an image in which
4. Constraint V1 and V2 should not be qual
5. Domain can be for v1->1,2 || V2 ->2,4
6. Variable is v1 and v2
7. So problem is we have to satisfy states or problem in given range and constraint

**Unit-2**

**Forward Reasoning:**

* It involves starting with available information (facts or premises) and moving forward to draw conclusions or make predictions.
* Goal -oriented approach where the system uses existing knowledge to derive new knowledge or reach a specific goal.
* **Knowledge Base:**
  + Forward reasoning relies on a knowledge base that contains a set of facts, rules, and relationships about the problem
  + The knowledge base serves as the foundation for drawing conclusions.
* **Initial Facts:**
  + The process begins with an initial set of known facts.
  + These facts can be directly observed or obtained from external sources.
* **Inference** 
  + The inference engine is responsible for applying the rules and logical operations defined in the knowledge base to draw new conclusions.
* **Goal or Objective:**
  + Forward reasoning often has a specific goal or objective. The system aims to derive new information that contributes to achieving the desired outcome.

**Advantages**

* Transparency: The reasoning process is transparent and easy to understand.
* Efficiency: It can be computationally efficient, especially when the goal is well-defined.
* Goal-Oriented: Well-suited where the objective is to reach a specific goal.

**Limitations**

* **May Miss Alternatives**: It may not explore alternative paths or consider other possibilities not covered by existing rules.
* **Dependency on Knowledge Base**: The effectiveness depends on the completeness and accuracy of the knowledge base.

**Backward Reasoning**

* System starts with a goal and works backward to determine what facts or conditions are needed to achieve that goal.
* Problem-solving approach that begins with the end objective and systematically reasons backward to identify the steps or conditions necessary to reach that objective**.**
* **Goal or Objective**:
  + Backward reasoning begins with a specific goal or desired outcome that the system aims to achieve.
* **Knowledge Base:**
  + The knowledge base contains a set of rules, facts, and relationships that describe the domain or problem space.
  + It serves as the foundation for determining the conditions required to achieve the goal.
* **Inference Engine:**
  + The inference engine is responsible for applying rules and logical operations in reverse order, starting from the goal and working backward to the initial conditions.
* **Working Memory:**
  + The system maintains a working memory that stores the current state of knowledge and the conditions being considered.

**Advantages**

* **Goal-Oriented**: Well-suited for tasks where the objective is to achieve a specific goal.
* **Efficient for Goal Achievement**: Can be efficient when the goal is clearly defined, and the system needs to work backward to find the required conditions.

**Limitations**

* **May Miss Alternative Paths:** It may not explore alternative paths or consider other possibilities not covered by existing rules.
* **Dependency on Knowledge Base**: The effectiveness depends on the completeness and accuracy of the knowledge base.

**Monotonic Reasoning:**

* Form of reasoning where the addition of new information never causes previously drawn conclusions to be revised or invalidated.
* It follows a "more information is always better" principle.
* **Certainty and Stability:** Conclusions reached through monotonic reasoning are considered stable and certain.
* **Inference Process:** The inference process is straightforward and based on the available information at any given time.
* **Example:**
  + "All humans are mortal" and "John is a human," then it can be concluded that "John is mortal." If later it is discovered that John is also a professor, this new information doesn't affect the previous conclusion.

**Non-monotonic Reasoning**:

* Form of reasoning where the addition of new information may lead to the revision or abandonment of previously made conclusions.
* It allows for reasoning under uncertainty and the possibility of changing beliefs.
* **Reevaluation of Conclusions**: New information may cause a reevaluation of previously drawn conclusions. Conclusions are not necessarily fixed.
* **Handling Uncertainty:** Non-monotonic reasoning is particularly useful in handling situations where information is incomplete or uncertain.
* **Default Assumptions**: It often involves default assumptions or rules that can be overridden by new information.
* Example:
  + "Birds can fly." Based on this, one might initially conclude that "Robins can fly." However, if later information is provided that "This specific robin is a mechanical toy," the previous conclusion may be revised because the default assumption about birds flying does not apply to mechanical toys.

**Reasoning with uncertainties**

Form of reasoning that deals with incomplete, imprecise, or uncertain information.

* **Uncertainty Types**:
  + Arises due to incomplete knowledge or information.
  + Arises due to inherent randomness or variability in a system.
* **Fuzzy Logic:**
  + Fuzzy logic allows for the representation of degrees of truth or membership in a set.
  + It is used when information is imprecise and boundaries between categories are not well-defined.

**UNIT-3**

**1.1 What is Machine Learning**

* Machine learning (ML) is a subdomain of artificial intelligence (AI) that focuses on developing systems that learn—or improve performance—based on the data they ingest.
* Machine learning is a field of technology that allows machine the ability to learn and improve their performance on a specific task without being explicitly programmed
* Machine learning is like teaching a computer to perform a task by showing it a large number of examples. Instead of explicitly telling the computer how to do the task, we let machines to figure out the patterns and rules on its own by the given data.

**1.2 How does Machine Learning work**

A machine learning system builds prediction models, learns from previous data, and predicts the output of new data whenever it receives it. The amount of data helps to build a better model that accurately predicts the output, which in turn affects the accuracy of the predicted output.



**1.3 Feature of Machine learning**

* Machine can learn itself from past data and automatically improve.
* From the given dataset it detects various patterns on data.
* It will become more easy to target relatable customer base.

**1.4 Why Machine learning**

* **Data Abundance**: In the digital age, vast amounts of data are generated every day. Machine learning algorithms can extract patterns from this data, which would be impossible for humans to process manually.

By which many complex problems can be resolved

* **Automation:** Machine learning enables automation of tasks and processes. It can replace manual and repetitive tasks,
* **Hidden Patterns:** Businesses and services increasingly aim to provide personalized experiences to users by Finding hidden patterns and extracting useful information from the user data.
* **Decision making** in various sector including finance

**1.5 Types of machine learning problems**

**Supervised Learning:**

* sample labelled data are provided to the machine learning system for training, and the system then predicts the output based on the training data.
* The system uses labelled data to build a model that understands the datasets and learns about each one. After the training and processing are done, we test the model with sample data to see if it can accurately predict the output.
* Types of problems: Classification (categorizing data into classes) and Regression (predicting a continuous value).
* Example: Example: Handwriting Recognition where 1,2,3,4,5 are labels and images are their corresponding output

**Unsupervised Learning:**

* Deals with unlabeled data.
* The algorithm learns patterns and structures in the data without explicit guidance.
* input data does not have corresponding labels or categories
* algorithm needs to act on that data without any supervision
* Types of problems: Clustering (grouping similar data points) and Dimensionality Reduction (reducing the number of features while retaining important information).
* Example: Customer Segmentation it group customers into segments like "high-value shoppers," "young and budget-conscious customers," and "occasional buyers. By finding patterns And this is used to target to specific customer groups based on their characteristics and behaviors.

**Reinforcement Learning:**

* feedback-based learning method,
* The agent learns automatically with these feedbacks and improves its performance.
* a learning agent gets a reward for each right action and gets a penalty for each wrong action.
* The goal of an agent is to get the most reward points, and hence, it improves its performance.
* Types of problems: Games (e.g., chess, Go), robotics, and autonomous systems.
* The robotic dog, which automatically learns the movement of his arms,

**1.6 Lifecycle of ml**

1. **Gathering of data**

* The goal of this step is to identify and obtain all data-related problems.
* We need to identify the different data sources, as data can be collected from various sources such as files, database, internet, or mobile devices
* Collect data
* The quantity and quality of the collected data will determine the efficiency of the output.
* The more will be the data, the more accurate will be the prediction.
* Integrate the data obtained from different sources

1. **Data preparation**
   * put all data together, and then randomize the ordering of data.
   * This step can be further divided into two processes:
   * **Data exploration:**
     1. It is used to understand the nature of data the characteristics, format, and quality of data.
     2. find Correlations, general trends, and outliers.
   * **Data pre-processing:**
     1. data cleaning to eliminate irregularities or blunders, standardization to scale data inside a particular reach, highlight scaling
2. **Data Wrangling**
   * Data wrangling is the process of cleaning and converting raw data into a useable format.
   * It is the process of cleaning the data, selecting the variable to use, and transforming the data in a proper format to make it more suitable for analysis in the next step.
   * It is one of the most important steps of the complete process.
   * Cleaning of data is required to address the quality issues.
   * Sometimes few of the data may not be useful. So they can have various issues, including:
     + Missing Values
     + Duplicate data
     + Invalid data
     + Noise
   * various filtering techniques is used to clean the data.
   * It is mandatory to detect and remove the above issues because it can negatively affect the quality of the outcome.
3. **Data Analysis**

* This step involves:
  + Selection of analytical techniques
  + Building models
  + Review the result
* The aim of this step is to build a machine learning model to analyse the data using various analytical techniques and review the outcome.
* It starts with the determination of the type of the problems, where we select the machine learning techniques such as Classification, Regression, Cluster analysis, Association, etc. then build the model using prepared data, and evaluate the model.
* So here we take the data and use machine learning algorithms to build the model.

1. **Train Model**

* We train our model to improve its performance for better outcome of the problem.
* We provide datasets to train the model using various machine learning algorithms.
* Training a model is required so that it can understand the various patterns, rules, and, features.

1. **Test Model**

* In this step, we check for the accuracy of our model by providing a test dataset to it.
* Testing the model determines the percentage accuracy of the model as per the requirement of project or problem.

**7. Deployment**

* Here we deploy the model in the real-world system.
* If the above-prepared model is producing an accurate result as per our requirement with acceptable speed, then we deploy the model in the real system.
* Before deploying check whether it is improving its performance using available data or not.
* The deployment phase is similar to making the final report for a project.

**1.7 Data pre-processing (handling missing values)**

<https://www.javatpoint.com/data-preprocessing-machine-learning>

**1.8 labels and features**

* Let's first build the model by training it. During training, we provide the model with features and after prediction what we get is label .
* For example, if we have a dataset containing information like age, weight, height, gender, and calories burned, these are the features.
* What we want to predict, in this case, is how many calories were burned. So, the calories burned would be our output, and we call it the label.
* If we were to predict the gender, then gender would be our label
* Features: all the columns like age salary input
* Labels: output data

**1.9 Steps Involved in Supervised Learning:**

* First Determine the type of training dataset
* Collect/Gather the labelled training data.
* Split the training dataset into training dataset, test dataset, and validation dataset.

**Example : house price :**

rows : different houses

columns: different features : area , bedroom ,sale price

**x**: columns such as area ,bedroom or independent variable

**y**= columns such as sale price that we will predict a dependent variable

**x\_train**: 80% of the x to teach the model

**y\_train:** 80% of the y to teach for the corresponding value of x

**x\_test :** 20% of the x to test

**y\_test:** predict the sale price(y) corresponding to its x\_test

while building/teaching model we provide model with x\_train. y\_train , x\_test then at the end we test the model by making it predict the x\_test and check if its producing correct y\_test

* Determine the input features of the training dataset, which should have enough knowledge so that the model can accurately predict the output.
* Determine the suitable algorithm for the model, such as support vector machine, decision tree, etc.
* Execute the algorithm on the training dataset. Sometimes we need validation sets as the control parameters, which are the subset of training datasets.
* Evaluate the accuracy of the model by providing the test set. If the model predicts the correct output, which means our model is accurate.



* 1. **Regression Analysis**
* Regression analysis is a statistical method to model the relationship between a dependent (target) and independent (predictor) variables with one or more independent variables.
* It helps us to understand how the value of the dependent variable is changing corresponding to an independent variable when other independent variables are held fixed.
* It predicts continuous/real values such as temperature, age, salary, price, etc.
* Regression is a supervised learning technique which helps in finding the correlation between variables and enables us to predict the continuous output variable based on the one or more predictor variables.
* Regression shows a line or curve that passes through all the datapoints on target-predictor graph in such a way that the vertical distance between the datapoints and the regression line is minimum
  1. **Terminologies Related to the Regression Analysis:**
     + **Dependent Variable**: The main factor in Regression analysis which we want to predict or understand is called the dependent variable. It is also called target variable.
     + **Independent Variable**: The factors which affect the dependent variables or which are used to predict the values of the dependent variables are called independent variable, also called as a predictor.
     + **Outliers:** Outlier is an observation which contains either very low value or very high value in comparison to other observed values. An outlier may hamper the result, so it should be avoided.
     + **Multicollinearity**: If the independent variables are highly correlated with each other than other variables, then such condition is called Multicollinearity. It should not be present in the dataset, because it creates problem while ranking the most affecting variable.

<https://www.javatpoint.com/overfitting-and-underfitting-in-machine-learning>

* + - **Underfitting :**
    - If our algorithm does not perform well even with training dataset, then such problem is called underfitting
    - Underfitting occurs when our machine learning model is not able to capture the underlying trend of the data.
    - he model is not able to learn enough from the training data, and hence it reduces the accuracy and produces unreliable predictions.
    - An underfitted model has high bias and low variance
    - **Overfitting:**
      * If our algorithm works well with the training dataset but not well with test dataset, then such problem is called Overfitting.
      * Overfitting occurs when our machine learning model tries to cover all the data points or more than the required data points present in the given dataset.
      * So model starts caching noise and inaccurate values present in the dataset, and all these factors reduce the efficiency and accuracy of the model.
      * The overfitted model has low bias and high variance.
      * The chances of occurrence of overfitting increase as much we provide training to our model.
      * The more we train our model, the more chances of occurring the overfitted model.
      * Overfitting is themain problem that occurs in supervised learning.
  1. **Linear Regression**
* Statistical method that is used for predictive analysis.
* It makes predictions for continuous/real or numeric variables such as sales, salary, age, product price
* Linear regression algorithm shows a linear relationship between a dependent (y) and one or more independent (y) variables.
* Since linear regression shows the linear relationship, which means it finds how the value of the dependent variable is changing according to the value of the independent variable.
* The linear regression model provides a sloped straight line representing the relationship between the variables.



* **Positive Linear Relationship**: dependent variable increases on the Y-axis and independent variable increases on X-axis
* **Negative Linear Relationship**: dependent variable decreases on the Y-axis and independent variable increases on the X-axis
* Mathematically

y=mx+b

* + Where:
  + y is the dependent variable (the variable you are trying to predict).
  + x is the independent variable (the feature used for prediction).
  + m is the slope of the line, which represents the rate of change of y with respect to x.
  + b is the y-intercept, which is the value of y when x is zero.
* A linear line showing the relationship between the dependent and independent variables is called a regression line or **best fit line**
* **The best fit line:**
* It represents the linear equation that best fits a given set of data points**.**
* It ensures the error between predicted values and actual values should be minimized.
* The best fit line will have the least error.
* The different values for weights gives a different line of regression
* So best values are calculated for best fit by using **cost function**
* **Cost Function**
* Mathematical function that measures the difference between predicted values and the actual values in dataset.
* The purpose of a cost function is to quantify how well or poorly a machine learning model is performing in terms of its ability to make predictions.
* The model aims to minimize this cost function during the training process by adjusting its internal parameters (weights and biases) to make better predictions.
* **Calculate the Cost Function (Mean Squared Error)**
* The cost function (mean squared error) measures the difference between the predicted values and the actual values.
* It is used to assess how well the best-fit line fits the data.
* The formula for mean squared error (MSE) is: 1/2n ∑ i=1to n(yi – yi)^2
  1. **Least square**
* The Least Squares Method is used to derive a generalized linear equation between two variables, one of which is independent and the other dependent on the former
* A statistical method that is used to find the equation of the line of best fit related to the given data. This method aims at reducing the sum of squares of deviations as much as possible
* **Limitation:** The least squares method assumes that the data is evenly distributed and doesn’t contain any outliers for deriving a line of best fit
* **Steps:**

Step 1: Denote the independent variable values as xi and the dependent ones as yi.

Step 2: Calculate the average values of xi and yi as X and Y.

Step 3: Presume the equation of the line of best fit as y = mx + c, where m is the slope of the line and c represents the intercept of the line on the Y-axis.

Step 4: The slope m can be calculated from the following formula:

* c = Y – mX
* **Example :**https://www.geeksforgeeks.org/least-square-method/
  1. **Residual Sum of Squares (RSS)**
* Residual sum of squares is used to calculate the variance of the data in terms of error or residuals.
* It is used to calculate the error left between regression data and regression function after running the model.
* The smaller the value of the residual sum of squares, the better the model.
* 
  + yi is the ith value of variable to be predicted,
  + f(xi) is the predicted value, and
  + n is the number of terms or variables.
  1. **Regression Sum of Squares (SSR)**
* The regression sum of squares measures how well the model is and how close is the predicted value to the expected value.
* 
  + Xi is the ith observation of the set,
  + is the mean of the dataset, and
  + n is the number of observations.
  1. **Total Sum of Squares (TSS)**
* It is used to denote the amount of variation in the dependent variable.
* It is the sum of the regression sum of squares and the residual sum of squares.
* It is calculated as:
  + TSS = RSS + SSR
  1. **Odds**
* Odds are the probability of an event occurring divided by the probability of the event not occurring
* It is calculated as the ratio of the probability of the event happening to the probability of it not happening
* Odds= P(Non-Event)/P(Event)

​

* 1. **Odd Ratio**
* The odds ratio is a measure of how the odds of an event change with a one-unit change in an independent variable.
* It provides insights into the impact of an independent variable on the likelihood of an event occurring.
* Xi=eBi

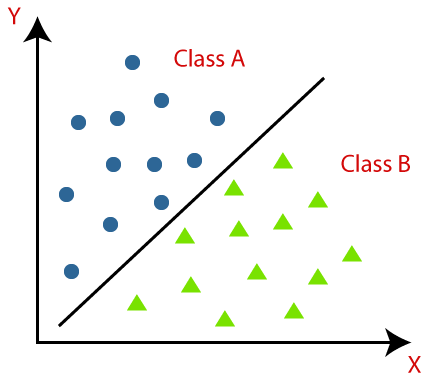
Bi is the coefficient of the independent variable xi obtained from the logistic regression model.

* 1. **In linear Regression the checks which ensures to get the best possible result from the given dataset.** 
     + Linear relationship between the features and target:
     + Linear relationship between the dependent and independent variables.
     + Small or no multicollinearity between the features:
       - Multicollinearity means high-correlation between the independent variables.
       - Due to multicollinearity, it may difficult to find the true relationship between the predictors and target variables
       - it is difficult to determine which predictor variable is affecting the target variable and which is not.
  2. **Types of Linear Regression**
* **Simple Linear Regression:**
  + If a single independent variable is used to predict the value of a numerical dependent variable
* **Multiple Linear regression:**
  + If more than one independent variable is used to predict the value of a numerical dependent variable
  1. **Simple Linear Regression**
* Relationship between a dependent variable and a single independent variable.
* The relationship shown by a Simple Linear Regression model is linear or a sloped straight line, hence it is called Simple Linear Regression.
* Dependent variable must be a continuous/real value.
* The independent variable can be measured on continuous or categorical values.
* **Objective**
  + - * Model the relationship between the two variables. Such as the relationship between Income and expenditure, experience and Salary, etc.
      * Forecasting new observations. Such as Weather forecasting according to temperature,
      * can be represented using the below equation:
      * y= a1x+c
  1. **Multiple Linear Regression**
* Algorithms which models the linear relationship between a single dependent continuous variable and more than one independent variable.
* Uses Two or more independent variables to predict the same dependent variable
* The goal is to model a linear relationship between the dependent variable and multiple independent variables.
* The equation for multiple linear regression can be written as:
  + - * y=b0+b1x1+b2x2+b3x3+…+bnxn

​

**3.1 Classification**

* Supervised Learning technique that is used to identify the category of new observations on the basis of training data.
* Program learns from the given dataset and then classifies new observation into a number of classes or groups. Such as, Yes or No, 0 or 1, Spam or Not Spam, cat or dog, etc.
* Classes can be called as targets/labels or categories
* The output variable is a category, not a value, such as "Green or Blue", "fruit or animal", etc.
* It takes labelled input data, which means it contains input with the corresponding output.
* Discrete output function(y) is mapped to input variable(x).
* y=f(x), where y = categorical output
* The main goal is to identify the category of a given dataset



**3.2 Learners in Classification Problems:**

**Lazy Learners**:

* It firstly stores the training dataset and wait until it receives the test dataset.
* classification is done on the basis of the most related data stored in the training dataset.
* It takes less time in training but more time for predictions. Ex: K-NN algorithm

**Eager Learners** :

* Eager Learners develop a classification model based on a training dataset before receiving a test dataset.
* Opposite to Lazy learners, Eager Learner takes more time in learning, and less time in prediction.
* Example: Decision Trees, Naïve Bayes, ANN.

**3.3 Logistic Regression**

* Logistic regression predicts the output of a categorical dependent variable.
* It doesn’t give exact value as 0 and 1, it gives the probabilistic values which lie between 0 and 1.
* instead of fitting a regression line, we fit an "S" shaped logistic function, which predicts two maximum values (0 or 1).
* It has the ability to provide probabilities and classify new data using continuous and discrete datasets.



**3.4 Sigmoid Function**

* The sigmoid function is a mathematical function used to map the predicted values to probabilities.
* It maps any real value into another value within a range of 0 and 1.
* The value of the logistic regression must be between 0 and 1, which cannot go beyond this limit, so it forms a curve like the "S" form. The S-form curve is called the Sigmoid function or the logistic function.
* In logistic regression, we use the concept of the threshold value, which defines the probability of either 0 or 1. Such as values above the threshold value tends to 1, and a value below the threshold values tends to 0
* equation of the straight line



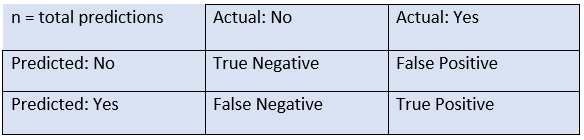
* y can be between 0 and 1 only, so for this let's divide the above equation by (1-y):

Logistic Regression in Machine Learning

* But we need range between -[infinity] to +[infinity], then take logarithm of the equation it will become:

Logistic Regression in Machine Learning

**4.1 Confusion Matrix**

* The confusion matrix is a matrix used to determine the performance of the classification models for a given set of test data.
* It can only be determined if the true values for test data are known.
* For 2 prediction class the matrix is of 2\*2 table, for 3 classes, it is 3\*3 table
* The matrix is divided into two dimensions, that are predicted values and actual values along with the total number of predictions.
* Predicted values are values, predicted by the model, and actual values are the true values for the given observations.
* **Need for Confusion Matrix**
* It evaluates the performance of the classification models, when they make predictions on test data, and tells how good our classification model is.
* It not only tells the error made by the classifiers but also the type of errors such as it is either type-I or type-II error.
* With the help of the confusion matrix, we can calculate the different parameters for the model, such as accuracy, precision
* **True Negative**: Model has given prediction No, and the real or actual value was also No.
* **True Positive**: The model has predicted yes, and the actual value was also true.
* **False Negative**: The model has predicted no, but the actual value was Yes, also called a Type-II error.
* **False Positive**: The model has predicted Yes, but the actual value was No. also called a Type-I error.
* **Accuracy:** Accuracy is a measure of how many correctly predicted instances there are out of the total instances. It's the most basic measure of a model's performance and is calculated as
  + **Accuracy**= Total Number of Predictions/Number of Correct Predictions
  + 
* **Precision** measures the fraction of true positive predictions among all positive predictions.
  + It's used when you want to minimize false positives.
  + **Precision**=True Positives/True Positives + False Positives
* **Recall** (or Sensitivity) measures the fraction of true positive predictions among all actual positives. It's used when you want to minimize false negatives.
  + **Recall** = True Positives/True Positives +False Negatives
* **F1-Score**: The F1-Score is the harmonic mean of precision and recall.
  + It provides a balance between the two metrics and is useful when you want to find a good compromise between precision and recall.
  + **F1-Score**= 2⋅Precision⋅Recall/ Precision+Recall
* <https://www.javatpoint.com/bias-and-variance-in-machine-learning>
* **Bias**:
* Bias is a prediction error that is introduced in the model due to oversimplifying the machine learning algorithms.
* the difference between the predicted values and the actual values.
* While making predictions, a difference occurs between prediction values made by the model and actual values/expected values, and this difference is known as bias errors
* Low Bias Models: Low bias models are flexible and make fewer assumptions about the data.
* They have the capacity to capture complex patterns.
* Overfitting Risk: When a model is very flexible (low bias), it can fit the training data extremely closely, even to the extent of capturing random fluctuations or noise in the data.
* Capturing Noise: This means that the model doesn't just learn the meaningful patterns but also learns the irregularities and randomness present in the training data.
* It tries to explain every data point, even if some of them are just due to chance.
* **Variance**:
* If the machine learning model performs well with the training dataset, but does not perform well with the test dataset, then variance occurs.
* variance tells that how much a random variable is different from its expected value
* Low variance means there is a small variation in the prediction of the target function with changes in the training data set.
* High variance shows a large variation in the prediction of the target function with changes in the training dataset.
* A model that shows high variance learns a lot and perform well with the training dataset, and does not generalize well with the unseen dataset. As a result, such a model gives good results with the training dataset but shows high error rates on the test dataset.

**UNIT -4**

**What is Dimensionality Reduction?**

* The number of input features, variables, or columns present in a given dataset is known as dimensionality, and the process to reduce these features is called dimensionality reduction.
* it is very difficult to visualize or make predictions for the training dataset with a high number of features, for such cases, dimensionality reduction techniques are required to use.
* It is a way of converting the higher dimensions dataset into lesser dimensions dataset ensuring that it provides similar information." Used for a better fit predictive model while solving the classification and regression problems.
* Ex:speech recognition, signal processing, bioinformatics data visualization, noise reduction

**Dimensionality Reduction Technique**

**The Curse of Dimensionality**

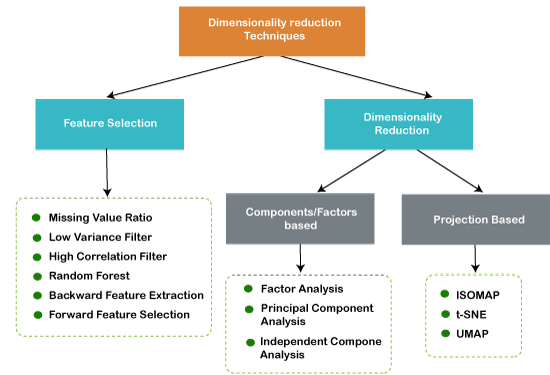
Handling the high-dimensional data is very difficult in practice As the number of features increases, the number of samples also gets increased proportionally, and the chance of overfitting also increases. This leads to model inefficiency called as curse of dimensionality

**Advantages**

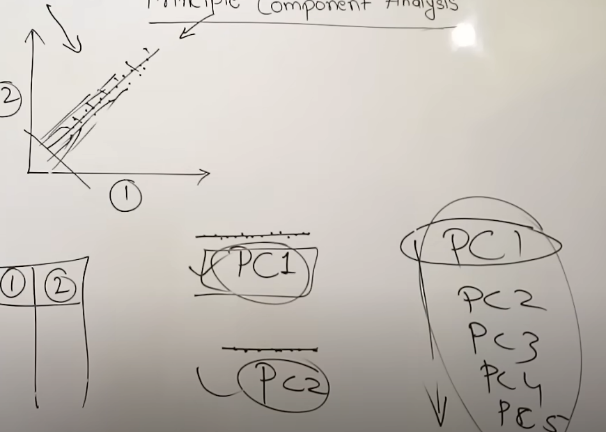
* By reducing the dimensions of the features, the space required to store the dataset also gets reduced.
* Less Computation training time is required for reduced dimensions of features.
* Reduced dimensions of features of the dataset help in visualizing the data quickly.
* It removes the redundant features (if present) by taking care of multicollinearity.

**Disadvantages**

* Some data may be lost due to dimensionality reduction.
* In the PCA dimensionality reduction technique, sometimes the principal components required to consider are unknown.

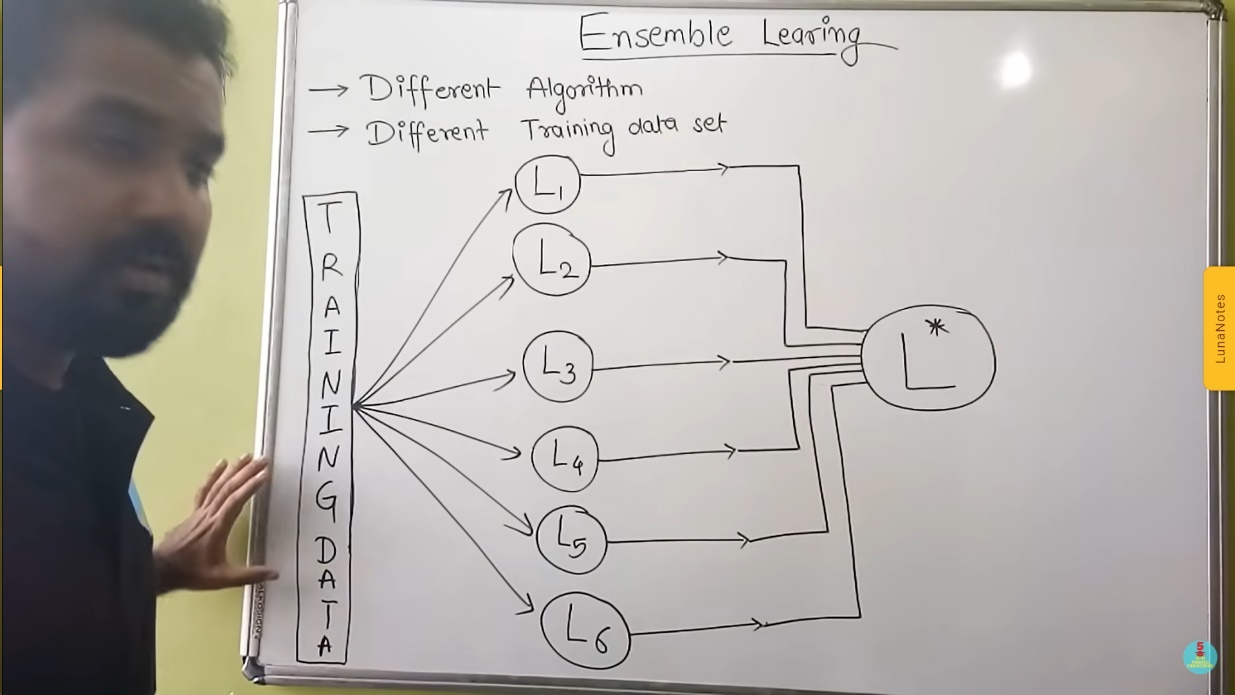
**Approaches of Dimension Reduction**

**Principal Component Analysis**

* Principal Component Analysis is an unsupervised learning algorithm that is used for the dimensionality reduction
* Due to overfitting model gets trained on each and every point which is not effective model so pca comes in pca tries to reduce the dimensions or attributes
* The number of these PCs are either equal to or less than the original features present in the dataset. If there are 4 attributes then there can be only less that or equal to 4 attributes in pca so PC1,PC2,PC3,PC4 can be created
* Converts the observations of correlated features into a set of linearly uncorrelated features with the help of orthogonal transformation.
* All these PCs should be orthoganl means independent of each other
* PCA generally tries to find the lower-dimensional surface to project the high-dimensional data.
* Example: image processing, movie recommendation system,

**Important Terms**

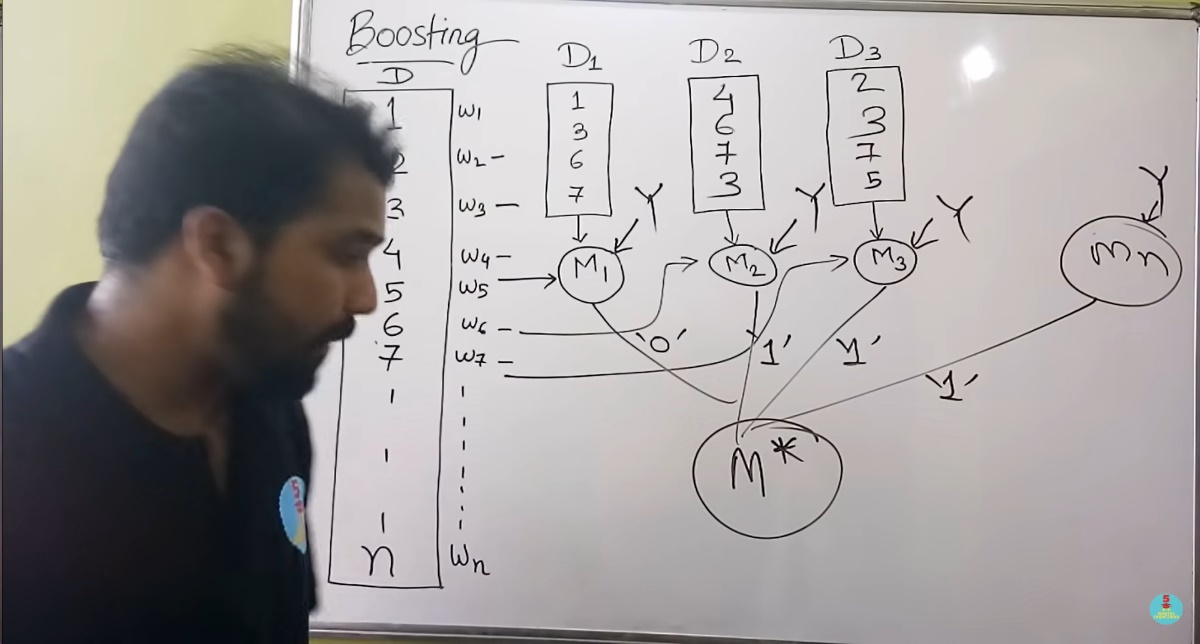
* + **Dimensionality**: It is the number of features or variables present in the given dataset. More easily, it is the number of columns present in the dataset.
  + **Correlation**: It signifies that how strongly two variables are related to each other. Such as if one changes, the other variable also gets changed. The correlation value ranges from -1 to +1. Here, -1 occurs if variables are inversely proportional to each other, and +1 indicates that variables are directly proportional to each other.
  + **Orthogonal:** It defines that variables are not correlated to each other, and hence the correlation between the pair of variables is zero.
  + **Eigenvectors:** If there is a square matrix M, and a non-zero vector v is given. Then v will be eigenvector if Av is the scalar multiple of v.
  + **Covariance Matrix**: A matrix containing the covariance between the pair of variables is called the Covariance Matrix.
* The principal component must be the linear combination of the original features.
* These components are orthogonal, i.e., the correlation between a pair of variables is zero.
* The importance of each component decreases when going to 1 to n, it means the 1 PC has the most importance, and n PC will have the least importance.



**Ensemble Learning**

* Ensemble learning helps improve machine learning results by combining several models
* Here L1,l2,l3,l4 are several different models working on differente algo like random forest svm decision tree etc these are called base learners or weak classifiers and L\* called as strong classifier
* It is based on decision of multimple models that’s why its accuracy is greater than all base learners

**Bagging**

* It describes the way How strong classifier generates its output
* Firstly it used Random samples of dataset for each model so if there are 100 records so 15 records to l1 20 records to L2 and that can be in random orders of record and also with replacement means record1 can be in record2 also or record3 laso
* Secondly after modelling on the basis of diff algo models are created then strong classifier need to decide which is the correct output as every model has their own output and can be dfferent
* So here strong classifier uses voting mechanism and generates output ex if out of 10, 6 says yes then output is yes

**Boosting**

* In boosting with all the records they have their corresponding weights which are used to decide to be selected in the dataset
* Secondly it uses step by step procedure or sequential flow of training models
* So initially every record has same weight means every record has equal probabilty to get selected in the dataset(so this dataset contains only few records not all) so it gives training data and after training it gives all the data to test where as it is is a weak classifier it gives some of the output wrong
* Those values which model1 has predicted wrong increase their weight so that they can be selcted in model2 training again process is repeated who have wrong output ,their weights are increased and train to third model this is done n times
* after that on the basis of voting strong classifier choose the output which have highest vote

**Neural network :**

* + It is a branch of artificial intelligence which is inspired by human brains
  + It consist of interconnected nodes called as neurons used for communication
  + It consist of a layered structure just like human brain
  + The idea behind the neural network was to learn from its mistake
  + It creates an adaptive system that computers use to learn from their mistakes and improve continuously.
  + Thus, artificial neural networks attempt to solve complicated problems, like summarizing documents or recognizing faces, with greater accuracy.

**ANN**

* Mechanism for information processing that draws inspiration from the brain is called an artificial neural network (ANN). ANNs learn via imitation just like people do.
* ANN is tailored for a particular purpose, including such pattern classification or data classification

**Components**

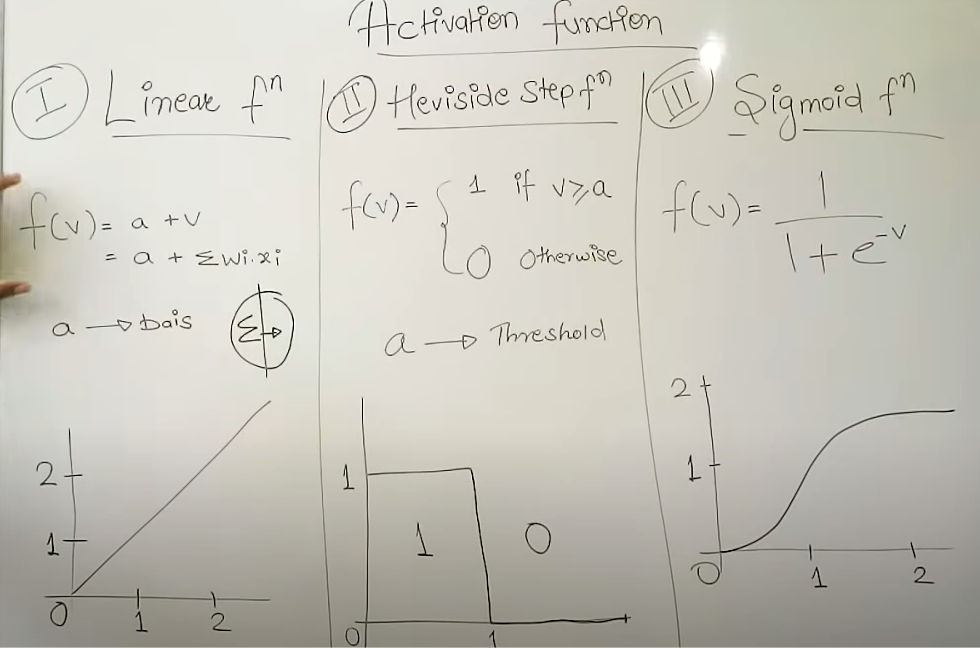
**Input Layer**

* The data will be accepted by this layer and forwarded to the remainder of the network.
* This layer allows feature input. It feeds the network with data from the outside world;
* nodes simply transmit the information (features) to the hidden units.

**Hidden Layer**

* The nodes in this layer are not visible to the outside world.
* Any features entered through to the input layer are processed by the hidden layer in any way, with the results being sent to the output layer.
* Also called as concealed layer
* either there are one or many hidden layers.
* Hidden layers are what give neural networks their exceptional performance and intricacy.
* They carry out several tasks concurrently, including data transformation and automatic feature generation.

**Output Layer**

* This layer raises the knowledge that the network has acquired to the outside world.
* The output layer is the final layer
* The output layer contains the answer to the issue.
* We receive output from the output layer after passing raw photos to the input layer.

**Activation Function**

* Multiple outputs are combined together and give to activation function and it generates the output so it is reponsible for generating output
* It is one that outputs a smaller value for tiny inputs and a higher value if its inputs are greater than a threshold.
* It "fires" if the inputs are big enough; otherwise, nothing happens.
* It is a gate that verifies how an incoming value is higher than a threshold value.
* they introduce non-linearities in neural networks and enable the neural networks can learn powerful operations, activation functions are helpful.
* A feedforward neural network might be refactored into a straightforward linear function or matrix transformation on to its input if indeed the activation functions were taken out.
* Explanation: As we are aware, neurons in neural networks operate in accordance with weight, bias, and their corresponding activation functions. Based on the mistake, the values of the neurons inside a neural network would be modified. This process is known as back-propagation. Back-propagation is made possible by activation functions since they provide the gradients and error required to change the biases and weights.
* **Types of neural network: 4 types**
  + **Feedforward Neural Networks (FNN):** Feedforward neural networks process data in one direction, from the input node to the output node. Every node in one layer is connected to every node in the next layer. A feedforward network uses a feedback process to improve predictions over time.
  + **Convolutional Neural Networks (CNN):** A neural network that is designed to process input data that has a grid-like structure, such as an image. It uses convolutional layers and to extract features from the input data.
  + **Recurrent Neural Networks (RNN):** A neural network that can operate on input sequences of variable length, such as text. It uses weights to make structured predictions.
  + **Long Short-Term Memory (LSTM) :** Specialized RNN architectures designed to address the vanishing gradient problem .It uses memory cells and gates to selectively read, write, and erase information.

**Feed Forward**

* Also known as a multilayer perceptron (MLP), is a type of artificial neural network where the information moves in only one direction—forward—from the input layer, through the hidden layers and finally to the output layer.
* The term "feedforward" signifies the flow of data through the network without any feedback loops.
* Feedforward Neural Networks provide a foundation for more complex neural network architectures
* **Neurons and Connections**:
  + Each neuron in a layer is connected to every neuron in the next layer.
  + Neurons in a layer do not have connections with neurons in the same layer or in previous layers.
* **Activation Functions**:
  + Each neuron typically has an activation function that determines its output based on the weighted sum of its inputs.
* **Training:**
  + Feedforward Neural Networks are trained using supervised learning. The weights and biases of the neurons are adjusted during training to minimize the difference between the predicted output and the actual output.
  + Backpropagation is a common algorithm used for training feedforward neural networks.

**Back Propogation Algorithm**

* It is a supervised learning algorithm commonly used for training artificial neural networks, including feedforward neural networks.
* The goal of backpropagation is to minimize the error between the predicted output of the neural network and the actual target output by adjusting the weights and biases of the network.

**step-by-step overview**

* **Initialization:**
  + Initialize the weights and biases of the neural network randomly. These weights and biases determine how much each input contributes to the neuron's output.
* **Forward Pass:**
  + Perform a forward pass to calculate the predicted output. This involves applying the weighted sum of inputs and the activation function for each neuron in each layer, progressing from the input layer to the output layer.
* **Compute Loss:**
  + Calculate the loss, which represents the difference between the predicted output and the actual target output.
* **Backward Pass (Backpropagation)**
  + Compute the gradient which involves calculating how much the loss would change with a small change in each weight and bias.
  + Update the weights and biases in the opposite direction of the gradient to minimize the loss. This step involves adjusting the parameters using a learning rate, which determines the size of the update.
* **Repeat:**
  + Repeat the forward pass, loss computation, and backward pass steps for multiple iterations

**Recommendation System**

* It is an application of (AI) that provides **personalized suggestions** or recommendations to users.
* These systems are widely used in various **online platforms** to help users discover new products, services, or content that align with their preferences.
* rely **on user data**, including explicit feedback (ratings, reviews) and implicit feedback (clicks, views), to learn user preferences.
* It **identify patterns and similarities** among users or items to make recommendations. This can be user-based or item-based.
* It recommends **items similar to those the user** has liked or interacted with, based on the features or content of the items.
* **Deep learning** models, including neural networks, are increasingly used for recommendation systems to capture complex patterns and representations.

**Types**:

* **Collaborative Filtering**:
  + **User-Based Collaborative Filtering**: Recommends items based on the preferences of users with similar tastes.
  + **Item-Based Collaborative Filtering:** Recommends items similar to those the user has liked or interacted with.
* **Content-Based Filtering:**
  + Recommends items based on the features or content of the items and the user's preferences.

**Content Based Recommendation**

* **Item Representation**: Items are represented by a set of features. These features could be keywords, tags, or other descriptors that characterize the content.
* **User Profile:** The system builds a user profile based on the features of items the user has interacted with in the past. This profile represents the user's preferences.
* **Similarity Calculation**: Similarity measures (e.g., cosine similarity) are used to determine the likeness between the user profile and the features of other items in the system.
* **Recommendation**: Items with high similarity scores to the user profile are recommended to the user.

**Example: Movie Recommendation System:**

* **Item Representation**:Each movie is represented by features such as genre, director, actors, and keywords
* **User Profile**:The user profile is constructed based on movies the user has rated or liked. For example, if a user enjoys action movies with specific actors, those preferences become part of their profile.
* **Similarity Calculation**:Similarity is calculated based on the overlap of features between the user profile and other movies.
* **Recommendation**:Movies with high cosine similarity to the user profile are recommended. If the user likes action movies with certain actors, the system might recommend other action movies featuring those actors.

**Limitations of recommendation systems:**

* **Cold Start Problem**:
  + **Definition**: It refers to the difficulty in making accurate recommendations for new users or items that have limited or no interaction history.
  + **Challenge**: Without sufficient data, the system may struggle to understand the preferences of new users or to recommend new items accurately.
* **Data Sparsity:**
  + **Definition**: Many recommendation systems rely on user-item interaction data. In many cases, this data can be sparse, meaning that users may not interact with a large portion of available items.
  + **Challenge**: Sparse data can lead to challenges in accurately capturing user preferences and finding meaningful patterns.
* **Popularity Bias:**
  + **Definition**: Popular items tend to be recommended more frequently, leading to a bias toward already popular or mainstream items.
  + **Challenge**: Less popular or niche items may not receive enough exposure, hindering diversity in recommendations.
* **Limited Serendipity:**
  + **Definition**: Recommendation systems often focus on providing users with items similar to what they have already interacted with, reducing the element of surprise or serendipity.
  + **Challenge**: Users may miss out on discovering new and diverse content that does not align with their existing preferences.
* **Profile Drift:**
  + **Definition**: User preferences can change over time, but some recommendation systems may not adapt quickly to these changes, leading to a form of profile drift.
  + **Challenge**: Stale recommendations that do not reflect the user's current interests may result in a decline in user satisfaction.
* **Privacy Concerns:**
  + **Definition**: Recommendation systems often require access to user data to make personalized suggestions, raising privacy concerns.
  + **Challenge**: Balancing personalization with user privacy is crucial, and users may be hesitant to share sensitive information.
* **Cross-Domain Challenges:**
  + **Definition**: Extending recommendations across different domains (e.g., from movies to books) can be challenging due to the distinct nature of each domain.
  + **Challenge**: Models trained on one domain may not generalize well to others, limiting the effectiveness of cross-domain recommendations.