University of Technology
الجامعة التكنولوجية

Computer Science Department
قسم علوم الحاسوب

Principles of AI
أساسيات الذكاء الاصطناعي

Dr. Raghad Nazar
د. رغد نزار

cs.uotechnology.edu.iq

Lecturer 1 - **introduction** TO AI

By Dr.Raghad Alshabandar

1.1 What is Artificial intelligence (AI)

It is an area of computer science which aims to develop machines to perform tasks like how a human would perform.

- It can be defined as "the automation of activities" it attributes to preform task without human intervention such as, thinking and rationality, problem-solving, decision-making, and learning.
- The AI involve set Predefined algorithms to execute a series of instructions and rules to guide AI system to train data and generate output.
- Many AI algorithms learn directly from data by identifying patterns and relationships, eliminating the need for rule-based instructions typically required in traditional programming.

1.2 What are Artificial Intelligence Types

- **Weak AI (Narrow AI)**: It is a type of Artificial intelligence that has limited capacity. It can only simulate the limited range of human cognitive functions. Currently, it has been implemented to solve complex real-world problems.
  - ✓ Example (website suggested purchases, Siri, Alexa, Google Assistant. chatbot).

- **Strong AI**: It is a form of machine intelligence that has intellectual capacity equal to human intelligence. With strong AI, the computer can be trained to have perception, feeling, and self-awareness. It can provide smarter business decisions and advice.
  - ✓Example (GPT-4,Human-LikRobots (SOPHIA,SURENA-university teacher) , Tesla Robot.

1.3 Artificial Intelligence Pioneer

The first Pioneer of AI is Alan Turing, he is British mathematician, from the University of Manchester, who was invented the Turing machine in 1936.

1.3.1 Turing Contribution

- Invent Turing machines
- Breaking Enigma code during the second war that was used for military purposes, he developed Bombe machine which was deciphered the Enigma-encrypted messages.
- proposed first program allow users to play a complete chess game.

1.3.2 Turing machines Description

- Turing machines are a computational device that can simulate any program to solve problems in computer science.
- It is much like an algorithm; a Turing machine operates on an input string of bits. The machine's head reads the input and writes the output.
- The output is derived by a set of transition functions (essentially lines of code) that dictate how the machine responds to each input.

1.4 History Of Artificial Intelligence

Period	Age Name	Author	Key Events and Contributions
1943-1956	Dark Ages	• Alan Turing • Warren McCulloch & Pitts • Claude Shannon	Turing machine, breaking Enigma code Neural networks. Chess-playing algorithms(min-max search algorithm. heuri and 'strategic' AI).
1956-1960	Great Expectations	• Lotfi Zadeh	-knowledge representation, learning algorithms, -Neural computing and computing with words.
1970 -1980	Golden age	• Stanford University	-Expert systems has three (components: a knowledge base , inference engine , input/output interface) -MYCIN (medical diagnostics Expert systems was used fo diagnosing blood diseases).
1980-1990		• Hopfield • Kohonen • Barto, Sutton and Anderson • Bryson and Ho • Judea Pearl Christopher Manning and Hinrich Schütze	Introducing Neural Networks with feedback Introduce self-organizing maps neural networks Publish paper on reinforcement learning Introduce back-propagation learning algorithm. - Introduce Bayesian networks. - Develop statistics
1990-2000	**Internet Era**	• Garry Kasparov • Dragon Systems • **Geoffrey Hinton**, **Yoshua Ben** • Berners-Lee's Leo Breiman	- Develop IBM's Deep Blue expert system - Develop speech recognition system(Dragon NaturallySpeaking) - The Internet and AI - Develop Random Forests and Boosting algorithm

2001-2023 Deep learning	• iRobot • Geoffrey Hinton, Yann LeCun • Facebook(Meta) Google	- Design Roomba, a robotic vacuum cleaner - Introduce deep Neural Network - AI becomes widely integrated into mainstream business applications (social media)

1.5 Artificial Intelligence Domains

- **Expert _Systems**: Expert systems are intelligent software programs designed to simulate human expertise in solving specific problems. These systems rely on a knowledge base and rules to find solutions. They also provide end users with explanations and insights into how the system arrived at a particular solution.
  - ✓ **Example** (Patient management system, MYCIN).

- **Robotics:** It is an intelligent machine designed to autonomously perform tasks based on human instructions. Nowadays, robotics is widely appreciated and adopted across the globe to handle industrial and warehouse operations, such as using robotic arms.
  - ✓ Example (autonomous vehicles, collaborative robots (cobots), medical robots, industrial robots, and service robots).

- **Natural Language Processing (NLP):** The NLP is a branch of AI is used to extract useful information and insights from text by transferring

unstructured text into structured text with the aim of identifying meaningful patterns from text.

✓ Example (chatbots, speech recognition, Machine Translation).

- **Generative AI:** Generative AI is an artificial intelligence program can drive text, images, or other media, using generative models. AI chatbots launched in 2023, such as Google Bard and Microsoft Copilot, ChatGPT played a main role in enhancing Productivity and Efficiency.

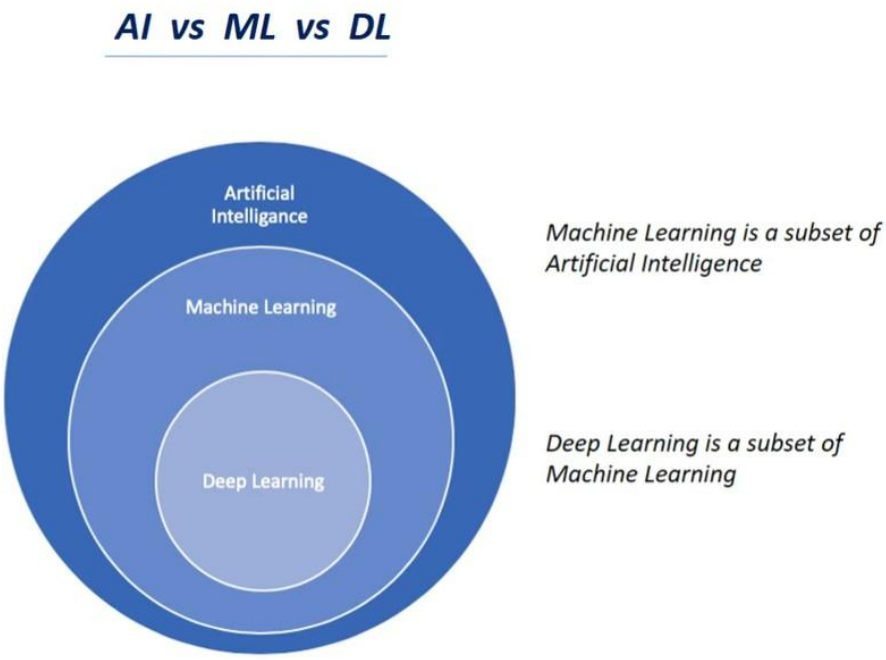
✓ Generative AI is trained based large language models (LLMs)

✓ some models consider the text input which is trained using a vast amount of text data from sources such as Wikipedia, articles, books.

✓ Other multimodal LLMs consider images trained using (Midjourney and the DALL-E image models)

✓ Other multimodal LLMs consider video training using (Runway ML and Synesthesia).

Artificial Intelligence



1.6 Machine Learning

Machine Learning is an application of artificial intelligence that can perform a task without explicit human intervention such as recognition, disease diagnosis and prediction.

- The key feature of machine learning is providing the computer ability to learn from data and make an accurate decision without the need for human assistance
- The machine learning has been applied in various fields such as image processing, natural language processing, time series analysis and recommendation system.
- Traditional programming involves developers manually creating software by defining explicit rules and instructions to solve specific

problems. The efficiency and success of these programs rely heavily on the programmer's expertise.

1.6.1 Data

- Data serves as the fuel for machine learning algorithms, enabling machines to understand relationships between features and make accurate predictions.
- The quality and quantity of data are crucial factors in determining the performance of a machine learning model. A lack of data can result in biased algorithms and inaccurate outcomes.
- Large volumes of data can be collected from various source sources such as sensors, social media, emails, images, and cameras.
- The data feed to machine learning can be categorized into two main formats: numerical or categories.
    - **Numerical data:** It represents measurable data in the form of numbers, such as weight, length, or the number of people.
    - **Categorical data:** It represents qualitative data that can be grouped into categories, such as age groups, gender, or color.

1.7 Deep Learning

- Deep learning is a family of machine learning that is capable of extracting complex features from high dimnational data.
- The ability to learn the various levels of data representations that match hierarchy elements of complex relational architecture is one of the distinctive features of deep learning.
- The Deep learning adopt multi-Layer architecture where each layer consists of several interconnected neural. The hierarchical feature

representation can make the neural network capture the nonlinear relationship between features.

- Deep Learning is widely used in various real applications such as healthcare, text analytics, cybersecurity, visual recognition, self-driving cars, Chatbots and Virtual Assistants.
- There are various types of deep learning architectures these are : Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs),Generative Adversarial Networks (GANs),and Transfer Learning.

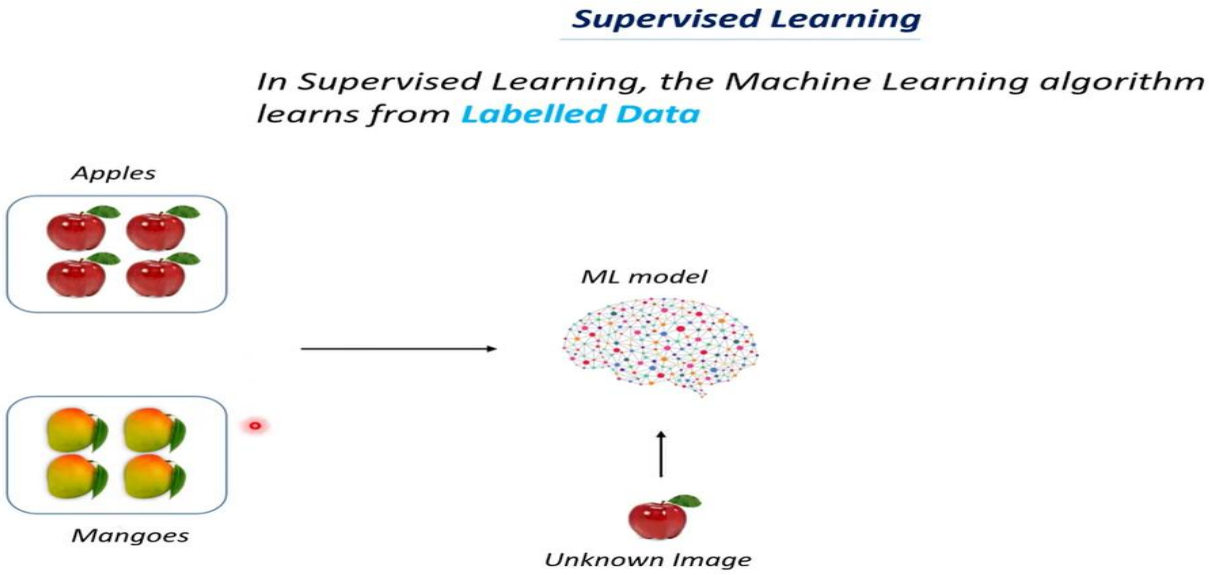
Lecturer 2- Machine Learning

2.1 Machine Learning Types

There are two types of machine learning: supervised machine Learning and unsupervised machine Learning

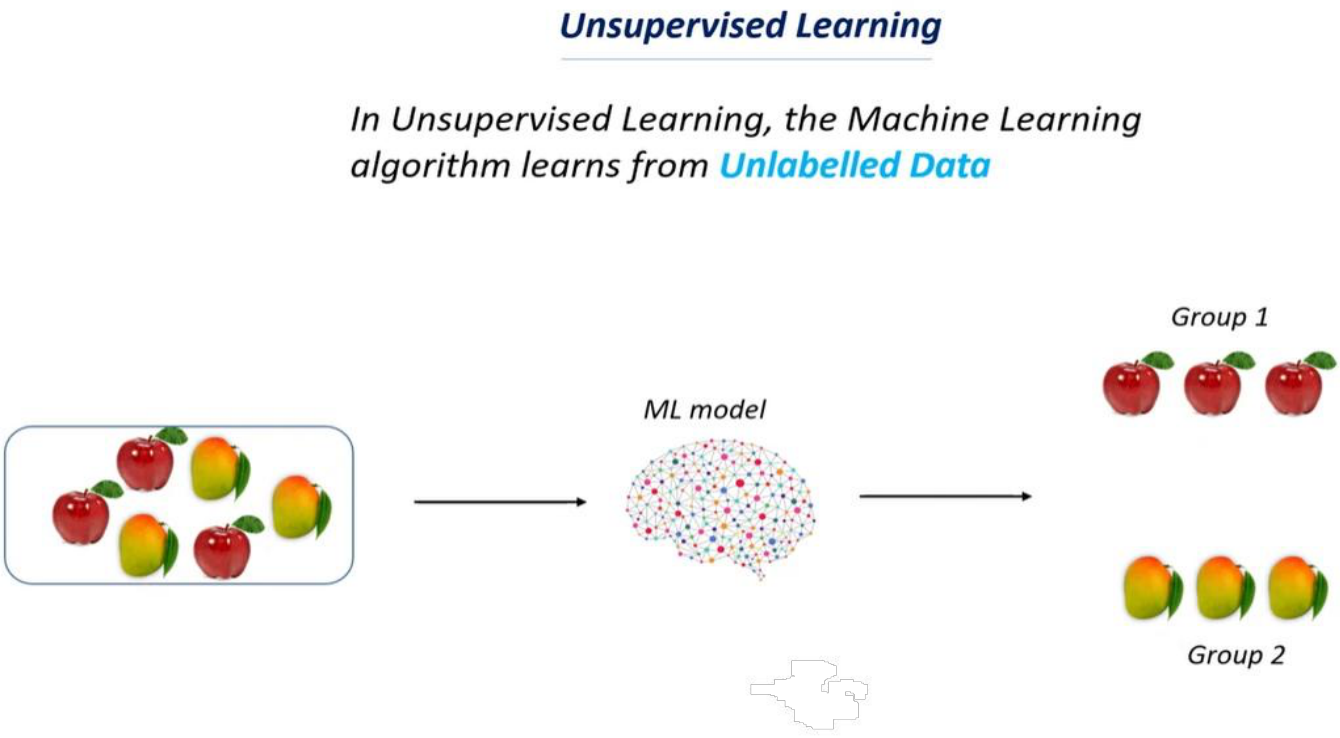
2.1.1 Supervised Machine Learning

- supervised learning, machine learning can learn the task by mapping function from input to output. This approach assumes that training examples contain a pair of input and output
- The learning algorithm is used to map the given examples with actual outputs and generalise new data.
- The main issue with this learning approach is the bias-variance trade-off. It is simultaneously bias and variances error that prevents learning algorithm from making an accurate.



2.1.1 Unsupervised Machine Learning

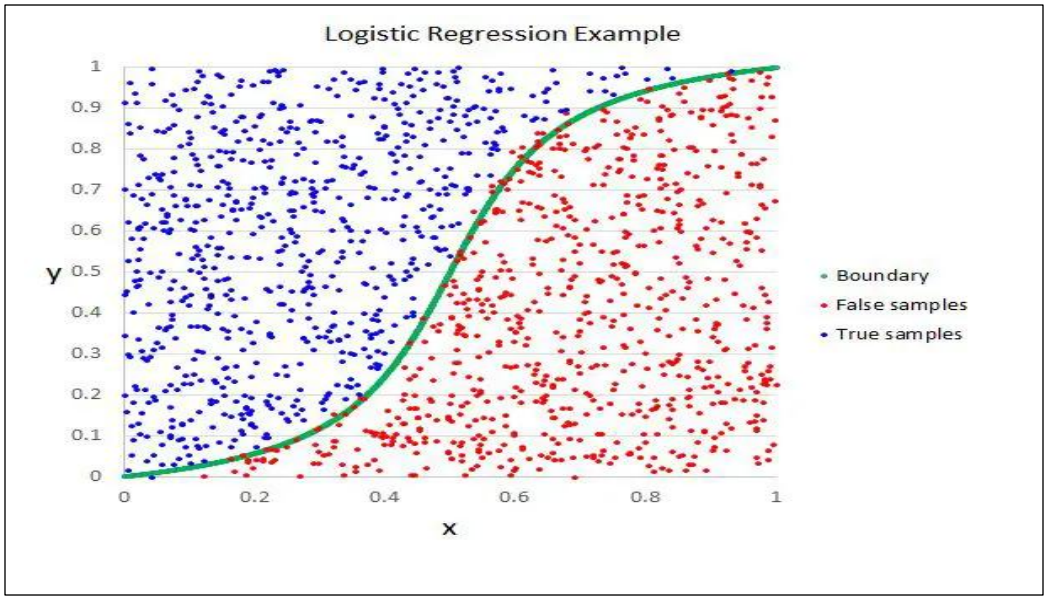
- Unsupervised learning is a type of machine learning that has the capability to infer the pattern within a dataset without providing any corresponding output.
- Unsupervised learning is more complicated than supervised learning as it can learn tasks from unlabelled data without any response variables.
- The purpose of unsupervised machine learning is to discover and draw inference about a similar group of input observations.
- The unsupervised learning has been widely used in a range of domains such as medicine, bioinformatics data, speech recognition, image processing, and finance.



2.2 Types of classification Algorithms

2.2.1 Generalized Linear model

- The generalized linear model is a statistical method that is used for linear mapping between the observed variables and response variables through a specified link function.
- The generalized linear model assumes that the observations follow a particular distribution, namely average, binomial, Poisson and gamma distribution.
- One example of a generalized linear model is logistic regression. The Logistic Regression (LR) can be used in a classification problem when the response variables are discrete values.
- To make a prediction in LR, we assume that the input variables correspond to vector features that are denoted as X_i and response variable represented in the target class.
- Logistic regression is an efficient technique. It can be implemented easily since it doesn't require scale the features and tuning parameters. It shows an advantage over the general linear model.
- Another advantage of logistic regression is that the cost with respect to computational complexity is low. It takes a small amount of time during the learning process.



2.2.2 Decision Tree

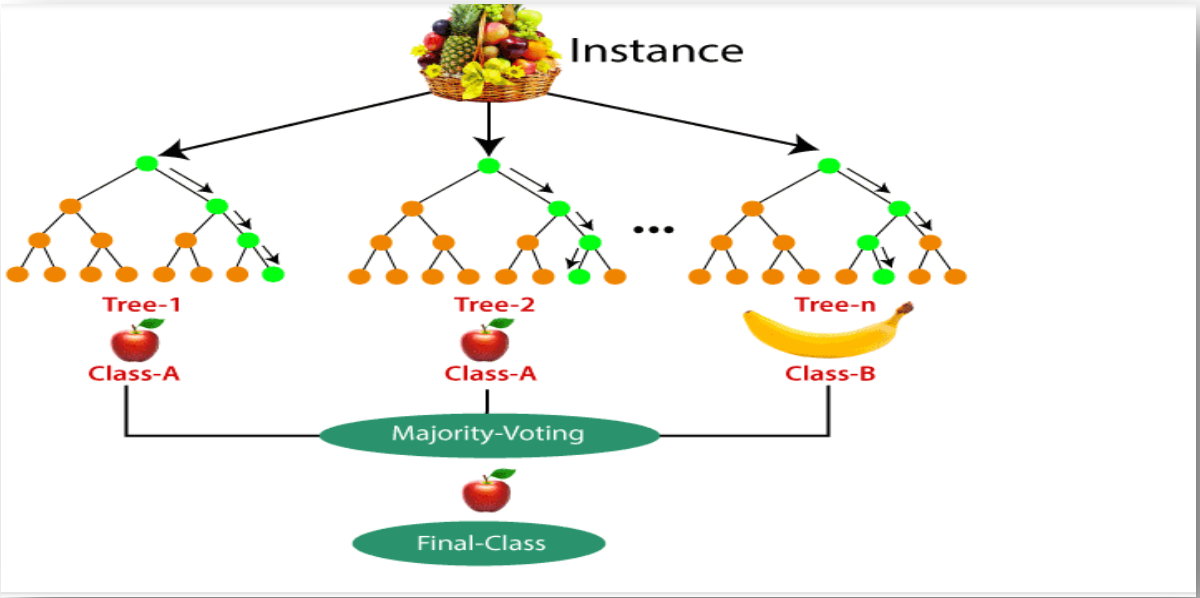
- A decision tree is a hierarchical subtype of directed acyclic graph (DAG), constructed by performing two steps, recursion and partitioning.
- The tree structure consists of three canonical components: a root node, a set of internal nodes, and a set of leaf nodes. Each node acts as a processing element that acts on a subset of the pattern space, performing a logical test on a particular attribute, for which outcomes are propagated by outgoing edge.
- The main advantage of the decision tree is that the output can be easily interpreted, even by non-professionals, as it is represented in graphical form.
- Another benefit is in handling nominal and numeric parameters; it is the nonparametric method which doesn't require normalisation of data. In addition, the decision tree can handle databases that have missing and error values.
- One of the main drawbacks of the decision tree is the overfitting phenomenon. As mentioned, the concept of creating a decision tree model depends on a split dataset, which leads to increasing the number of nodes.

2.2.3 Random Forest

- Random forest is an ensemble method that constructs multiple decision trees during learning time where each tree is generated using random sample vector provided from input features.
- The random forest uses the voting mechanism that selects the most popular classes to classify the target. In theory, the training algorithm of

random forest follows the bootstrap method. Given the training data set consists of N samples and M feature.

- Random forest is considered the most accurate machine learning algorithm due to its capacity to discover the nonlinear association between the features and targets. Also, it can run efficiently in high dimensional data.
- The random forest can handle numerical and categorical values without concern the deletion of observation. When a large amount of data contains the missing value, it can deal with missing data by adopting an imputation algorithm that keeps enhancing accurate prediction results.
- The main drawback of the random forest is its huge computational cost. The computational complexity of training algorithms is high compared with other machine learning models.

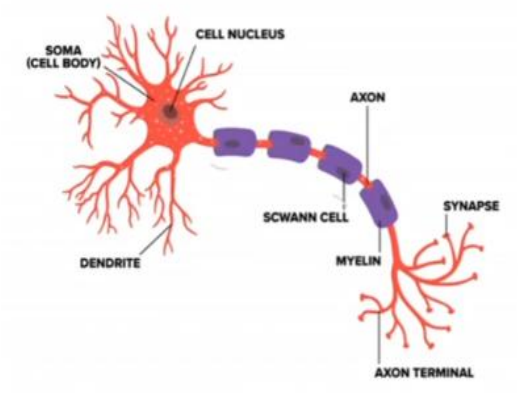


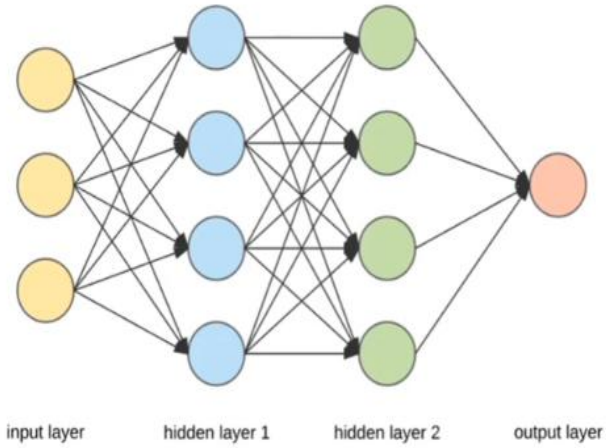
2.2.4 Neural Network

- Neural Networks are a problem-solving methodology grounded in the connectionist paradigm, comprising networks of interconnected simple

units whose adaptive parameters may be tuned to form an emergent solution.

- The simplest type of artificial neural network is Feed-Forward neural network where information transfer is in one direction without cycles; in this network, neurons belong to layer (i), receive input from layer (i-1) and transmit output to layer (i+1) through one hidden layer.
- A Multilayer perception (MLP) is a type of feed-forward neural network; it consists of multi-layers of units. Usually, the MLP consists of three layers: one input, one output layer and at least one hidden layer; each unit fully connects to the other unit in the following layer through a sequence of weighted edges.
- The backpropagation algorithm is typically used to compute and adjust weights in response to some error signal, given some input features. The backpropagation algorithm is employed to compute neural network weight through gradient descent.
- Neural networks can learn and model the complex relationships between features; therefore, they have been used to find accurate solutions of complex problems.
- Another advantage of neural networks is that they can quickly make the correct prediction on unseen data. The new data can be generalised even if it has a high degree of noise.
- The main drawback of neural networks is their black box nature; it could be hard to understand the features that impact the prediction. The interpretability of results could be hard to explain.





2.3. Feature Selection

- The feature selection has been used to reduce noise components and improve the performance of the prediction model.
- In terms of machine learning, feature selection selects a subset of features by eliminating redundant and irrelevant features.
- Features selection approach to classification problems has been proved to enhance the predictive accuracy, decrease the training time and reduce computational complexity.
- There are various features selection methods namely wrapper approach, filter approach and embedded method.

2.3.1 Features Selection Wrapper Approach

- In the Wrapper approach, the classifier model is employed to evaluate the subset of the feature.
- Search algorithms are used to find an optimal number of features heuristically.
- In particular, the dataset is split to train and cross validation, backwards algorithm run with a different number of features on each set, the set with the lowest validation error is selected as the final set.

- Although, the potential of the wrapper approach is in enhancing the predictive model's accuracy, the wrapper approach acts as a black box in the high dimensional dataset.

2.3.2 Features Selection Filter Approach

- With regards to the filter approach, the optimal number of features are selected according to heuristic criteria without considering the classifiers process.
- There are various heuristic criteria to rank features in such a method these including correlation coefficient, Chi-Square, Information Gain, Cross Entropy.
- More specifically, the weight is assigned to features based on these heuristic statistical data and features below the threshold are eliminated.
- As mentioned, the filter approach is independent of any type of classifier consequence, the machine learning algorithms that rely on the filter approach might achieve lower performance than the wrapper approach.
- The filter approach could select the optimal number of features that might exist in the redundant subset. One of the advantages of the ranking method is the low computational cost.

References

[1] Winston, P.H., 1992. Artificial intelligence. Addison-Wesley Longman Publishing Co., Inc.

[2] Larson, E.J., 2021. The myth of artificial intelligence: why computers can't think the way we do. Harvard University Press.

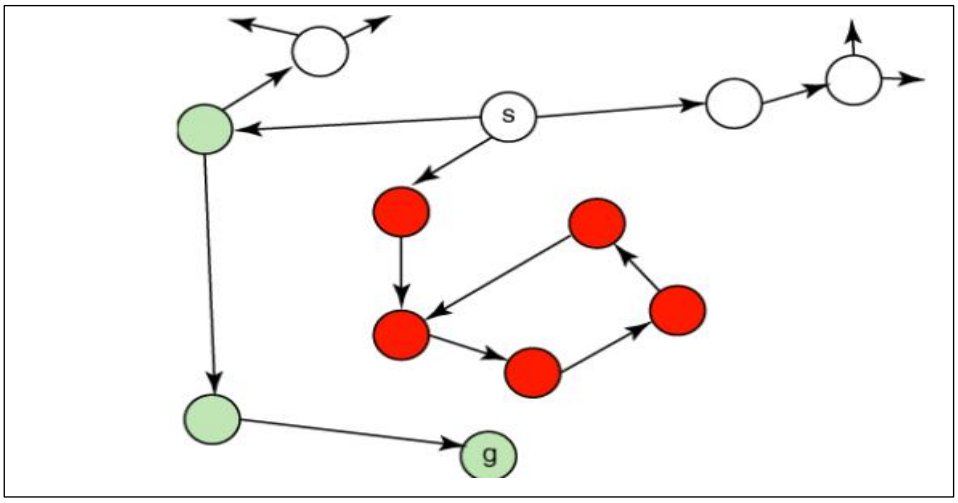
Fundamentals of Artificial Intelligence

Lecturer 4 – Heuristic Search

By Dr.Raghad Alshabandar

4.1What is Heuristic Search

- The informed search is also called a" Heuristic Search" is a type of search method that considers prior information to facilitate the search process in order to achieve an optimal solution efficiently.
- It has been used to solve the complex problem that requires a large search space such as search engine, autonomous car.



Heuristic Search_ Advantage

- It can easily prioritize the best path that leads to an optimal solution.
- It can reduce computational costs.
- It can enhance the performance of the search by narrowing down the search space.

4.2 Comparison Between Blind Search & Heuristic Search

Blind Search	Heuristic search
The search algorithm is unable to use Heuristics	The search algorithm uses heuristics
Less efficient to handle complex problem	More efficiently handle complex problems by prioritizing the promising Solution
The compactional cost is high because of it must explore whole search space	The compactional cost is low because it includes additional information that assist the search algorithm to deliver smart decision

4.3 Hill Climbing Algorithm

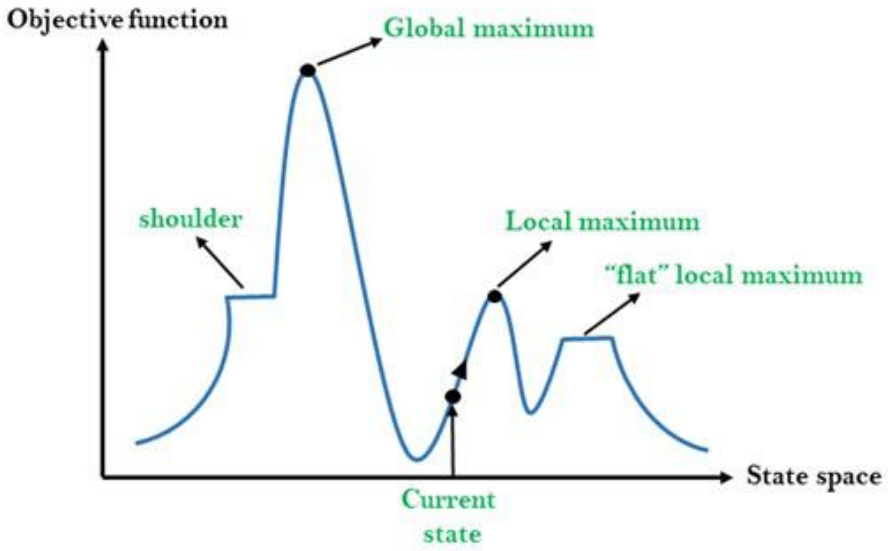
- The Hill Climbing algorithm is a local search algorithm used in mathematical optimization problems to find the optimal solution by continuously seeking to improve the current solution.
- If a change results in a better solution, the algorithm incrementally updates the current solution until no further improvements can be made.

4.3.1 Hill Climbing _Features

- **Test_Varient** : The Hill Climbing is used the objective function to test and evaluate the quality of potential solution
- **Greedy_ Algorithm**: Hill Climbing finds the best solution by considering the immediate neighboring state without evaluating the entire search space.

4.3.2 State Space Diagram

- The state-space landscape is a graphical representation of the hill-climbing algorithm, shows the association between different algorithm states and the objective function.
- **x-axis**: It represents the state space
- **Y-axis**: It represents the objective function
- **Local Maximum**: It represents a peak state which is better than its neighbor states, but there is also another state higher than local maximum.
- **Global Maximum:** It represents the optimal state in space state diagram, where the objective function reaches the best value, the objective value gains the highest value.
- **Current_ state:** It represents the position of state in the diagram while seeking optimal solution.
- **Flat Local Maximum:** represents the flat space of diagram where all neighbors' states have the same objective furcation values
- **Shoulder**: It represents a plateau region on the search space.



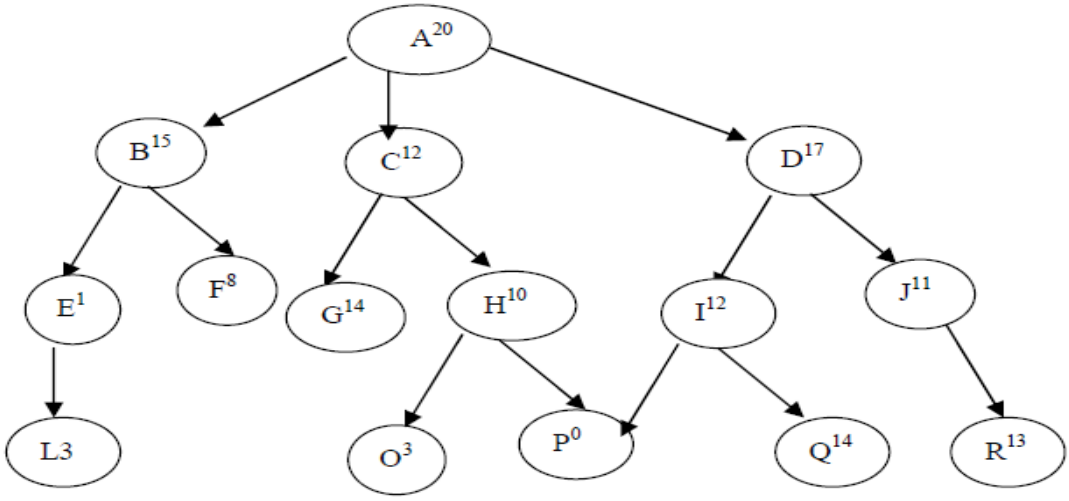
4.3.3 Simple Hill Climbing

- The Simple Hill Climbing algorithm is the easiest approach to implement among climbing algorithms.
- It evaluates only the successor state node and moves to the next state if a better objective value is found. The Simple hill Climbing algorithm **procedure can be found below**.
- Select a starting point (initial state) in the search space.
- Evaluate the one neighbor node of current state using the objective function.
- Move to a new neighboring state if it provides a better solution.
- Repeat the process until no neighboring state offers a satisfactory solution.

  The Simple hill Climbing algorithm **procedure can be found below**.

Example

Using the Simple Hill Climbing search for(P) with Local minimum from (A)



Step	[Open]	[Close]	[Selected_Node]
Step0	[A]	[]	A
Step1	[C,D,B]	[A]	C
Step2	[G,H]	[A,C]	H
Step3	[O,P]	[A,C,H]	P
Step4	[]	[A,C,H,P]	

Final path: A → C → H → P

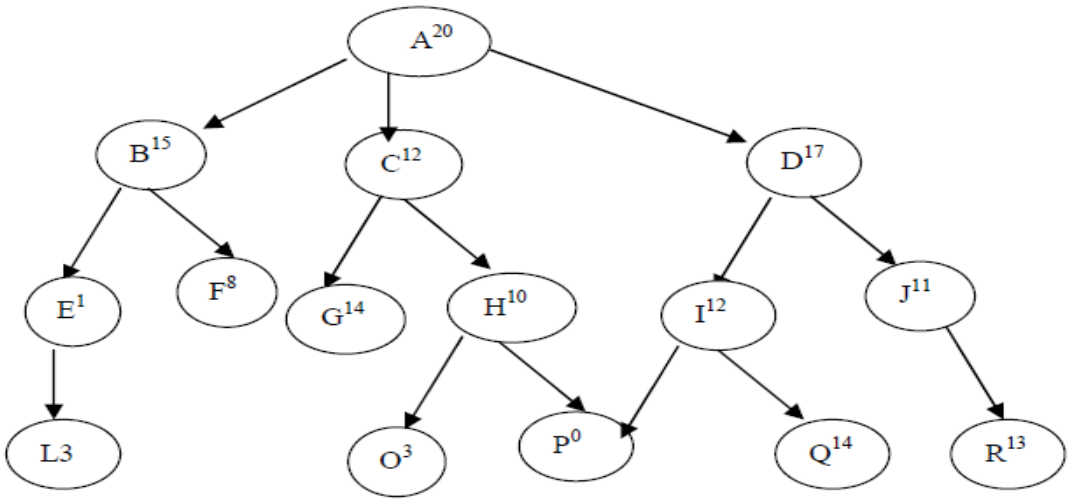
4.3.4 Steepest-Ascent Hill-Climbing

- It is a type of simple hill climbing take examined all nodes to find the optimal solution. It evaluates all nodes of current states and selects the neighbor node with the highest objective value.
- It essentially moves toward nodes that yield to the steepest improvement. The Steepest Ascent Hill-Climbing procedure can be found **below**

- Select a starting point (initial state) in the search space.
- Evaluate all neighboring nodes using the objective function.
- Move to a new neighbor that has the highest objective value.
- Repeat the process until no neighboring state offers a satisfactory solution.

Example

Using the Steepest-Ascent Hill-Climbing search for(P) from (A)



Step	[Open]	[Close]	[Selected_Node]
Step0	[A]	[]	A
Step1	[C,D,B]	[A]	D
Step2	[I,J]	[A,D]	I
Step3	[P,Q]	[A,D,I]	P
Step4	[]	[A,D,IP]	

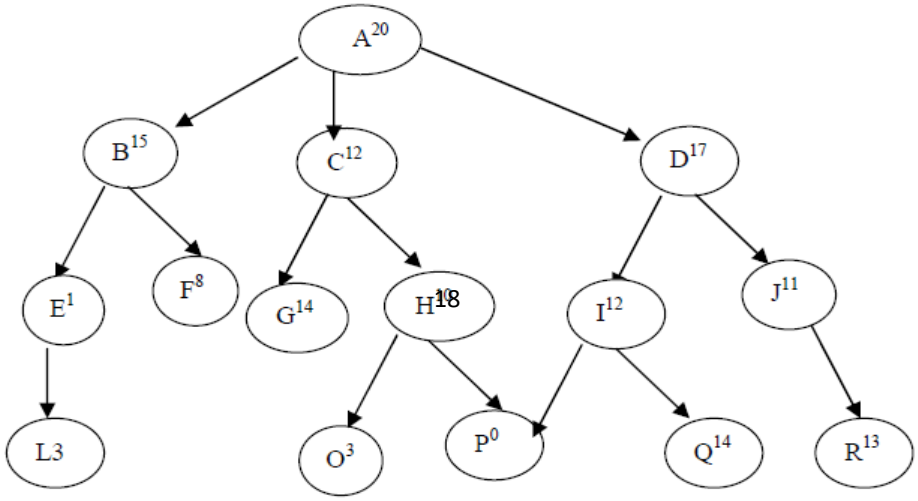
Final path: A → D → I → P

4.3.5 Stochastic Hill Climbing

- The Stochastic Hill Climbing exploring the entire search space by randomly selects one neighbor node and determines whether to accept it as the current state or move on to evaluate another state. The Steepest Stochastic Hill Climbing can be found **below**
- Select a starting point (initial state) in the search space.
- Randomly select one neighbor nodes of current state.
- Comparing neighbor nodes with the current state using objective function.
- Move to a new neighbor if its objective function value of neighbor node better than that of the current state.
- Repeat the process until the best solution is achieved.

Example

Using the **Stochastic Hill Climbing** search for(P) from (A)



Step	[Open]	[Close]	[Selected_ Node]
Step0	[A]	[]	A
Step1	[D,B,C]	[A]	Randomly Select(D)
Step2	[I,G]	[A,D]	Stochastic choice(H)
Step3	[Q,P]	[A,D,H]	Since the H is > P stop

Final path: A → D → H

References

[1] Yokoo, M. and Ishida, T., 1999. *Search algorithms for agents* (pp. 165-199). Multiagent systems, a modern approach to distributed artificial intelligence, MIT Press.

[2]Nilsson, N.J., 2014. Principles of artificial intelligence. Morgan Kaufmann.

[3] Negnevitsky, M., 2005. Artificial intelligence: a guide to intelligent systems. Pearson Education.

Fundamentals of Artificial Intelligence

Lecturer 6 –Travelling salesman problem

By Dr.Raghad Alshabandar

6.1 Shortest Path Problem

- The shortest path problem is a graph theory problem that aims to find the shortest path between two nodes based on a set of metrics assigned to the edges, such as weight, distance, or time. The goal of the shortest path problem is to minimize the total cost of the edges' weights.

- The Shortest path algorithms are successfully applied on web mapping website (Google Maps) to help the driver to find the shortest path between the physical locations.

- The Shortest path algorithms are widely used in social media such as Facebook or Twitter to recommend the friend or connection between two users.

6.1.2 Shortest Path Type

There are different types of Shortest Path algorithm

- **Single-source shortest path (or SSSP) problem**. The problem require finding the shortest path from source node to all other nodes in the graph.
  - ✓ **Example_algorithms**: Depth-First Search (DFS)

    Breadth-First Search (BFS)
- **Single-Pair Shortest Path**: The problem requires to find the shortest path between the induvial pair of nodes in the graph.
  - ✓ **Example_algorithms:** A*Algorithm

    Dijkstra's Algorithm,
- **All-pairs shortest path**: The problem requires finding the shortest path between every pairs of nodes in the graph .
  - ✓ **Example_ algorithms:** Floyd-Warshall Algorithm

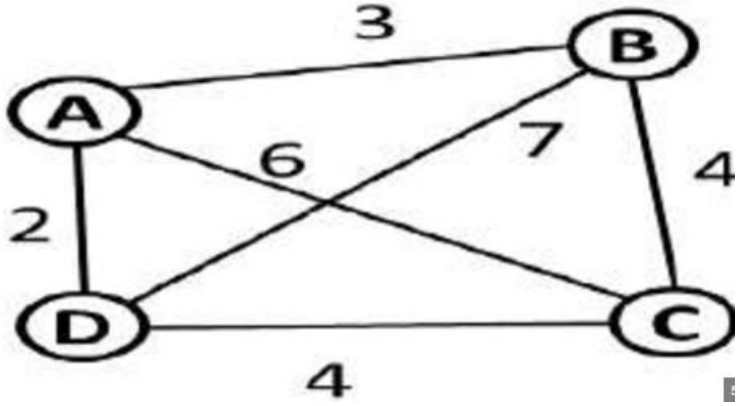
    Johnson's Algorithm

6.2 Travelling Salesman Problem

- The Traveling Salesman Problem (TSP) is a graph problem in the field of computer science and optimization. Its goal is to find the shortest route that visits each city exactly once and returns to the starting point.

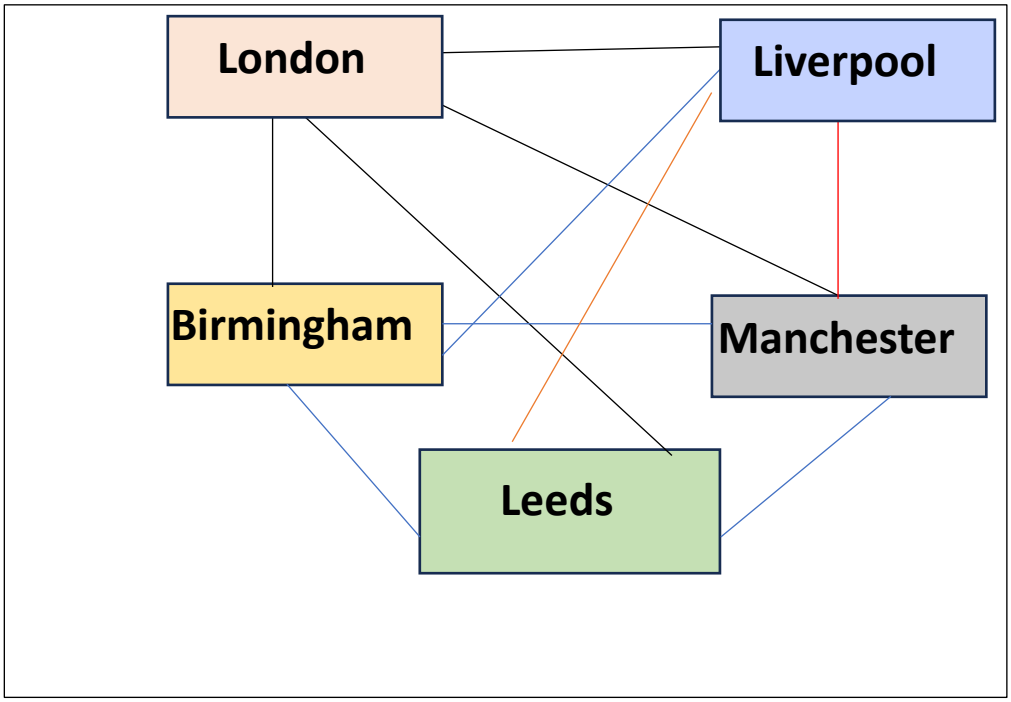
- The TSP can be represented as a weighted graph, where the cities (nodes) are the graph's vertices, and the edges represent the paths between the cities. The travel cost is determined by the weight of the edges. The route starts from a specified node, and it finishes at the same node after visiting other nodes one time.
- There are two types of TSP, namely symmetric and asymmetric. In the symmetric TSP, the distance (weight) between two nodes is the same in both directions, and the graph is represented in an undirected form. In the asymmetric TSP, the distance (weight) between two nodes differs in each direction, and the graph is represented in a directed form.

- Courier Delivery Service (Amazon, UPS) deliver the package to different locations and returned to headquarter office, Bus school and Waste Collection Route Planning are real use case examples of symmetric TSP
- Flight Route (airplane departure and arrival to airport), Traffic congestion are real examples of asymmetric



Example(1):

Please use a Travelling Salesman algorithm to find the shortest route to visit five popular UK cities: London, Manchester, Birmingham, Liverpool, and Leeds. Please note that each city should be exactly once visited returns to the starting city (London). The distance between each pair city can be found in a blow table.



Table(1) Distance between Uk cites

London to Manchester	**335 km**
London to Birmingham	**183 km**
London to Leeds	**350 km**
London to Liverpool	**323 km**
Manchester to Birmingham	**139 km**
Manchester to Leeds	**69 km**
Manchester to Liverpool	**35 km**
Birmingham to Leeds	**160 km**
Birmingham to Liverpool	**127 km**
Leeds to Liverpool	**167 km**

Solution:

Route no	Possible Routes	Total_ Distance
Route(1)	London-Manchester-Birmingham-Liverpool-Leeds	335+139+127+167=768Km
Route(2)	London-Birmingham- Manchester -Liverpool-Leeds	183+139+35+167=524 Km
Route(3)	London-Manchester-Liverpool- Birmingham -Leeds	335+35+127+160=657 Km
Route(4)	London-Liverpool-Leeds-Birmingham- Manchester	323+167+160+139=789 Km
Route(5)	London- Leeds- Manchester-Birmingham-Liverpool	350+69+139+127=685Km
Route (6)	London-Liverpool-Birmingham- Leeds - Manchester	323+127+160+69=679 Km

- The shortest route is: London-Birmingham- Manchester -Liverpool-Leeds= (524 Km)

- Due to the complexity of the travelling salesperson problem, finding the shortest route becomes more difficult as the number of cities increases. Therefore, different approaches have been utilized to tackle this issue.

6.3Travelling Salesman approach

There are various approaches can be used to tackle travelling salesperson Problem. In below more details can be found.

6.3.1 Brute-Force Approach

- The brute-force approach is a straightforward method where all possible routes (permutations) between nodes are explored.

- The total distance for each permutation is computed, and the shortest route selected. (The previous example considers these permutations.)

- This approach becomes inefficient, especially when the number of routes is large as consequences, it cannot be used for real use case seniors with many locations.

6.3.2 Nearest Neighbor Approach

- The Nearest Neighbor algorithm starts at a random node and selects the nearest unvisited city. It visits that city and marks it as the current city.

- Repeat all steps until all cities have been visited. Finally, return to the starting node to complete the entire route.

Example(2):

Please use the Nearest Neighbor Approach to find the shortest route to visit five popular UK cities as menation in example (1).

Solution:

Step_No	Current City	Nearest City	Distance
Step (1)	London	Birmingham	183 Km
Step (2)	Birmingham	Liverpool	127 Km
Step (3)	Liverpool	Manchester	35 Km
Step (4)	Manchester	Leeds	69 Km
Step (5)	Leeds	London	350 km
Total Distance =183+127+35+69+350=764 km			

6.3.3 Greedy Search Algorithm

- The greedy search algorithm is an optimization problem-solving approach that searches for local optima by selecting the shortest edge that minimizes the total distance.

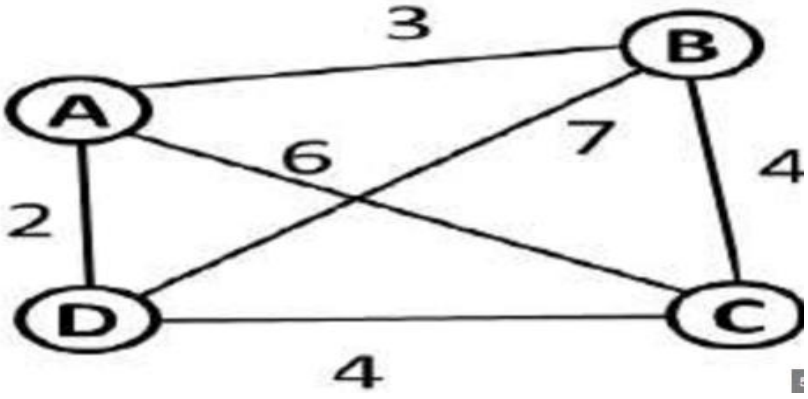
- The algorithm starts with an initial node and sorts all the edges between cities from the shortest to the longest distance.

- It continues searching for the next optimal path of unvisited nodes and it select the best available edge that have lowest cost without forming cycles.
- These steps are repeated until all cities have been visited. Finally, the algorithm returns to the starting node to complete the entire route.

- This algorithm is used by network companies to find the shortest route in a network and determine the amount of data packet delivery over the internet.

Example 3

Please use Greedy Search to find the shortest route to visit four cities (A, B, C,D)



Solution

Step_No	Current Node	Optima edge	Distance
Step (1)	A	A ⟶ D	2
Step (2)	D	D ⟶ C	4
Step (3)	C	C ⟶ B	4
Step (4)	B	B ⟶ A	3
Total Distance =2+4+4+3=13			

References

1-Rich, E., & Knight, K. (1991). Artificial Intelligence.

2-Luger, G. F., & Stubblefield, W. A. (2004). Artificial Intelligence: Structures and Strategies for Complex Problem Solving. Addison Wesley