Quantum Series

ENGINEERING

QUANTUM *Series*

Semester - 7 CS & IT

Artificial Intelligence

- Topic-wise coverage of entire syllabus in Question-Answer form.
- Short Questions (2 Marks)

Session 2019-20 Odd Semester

Includes solution of following AKTU Question Papers

2014-15 • 2015-16 • 2016-17 • 2017-18 • 2018-19

CONTENTS

RCS 702 : Artificial Intelligence

UNIT 1

Introduction

CONTENTS

PART-1

Introduction : Introduction to Artificial Intelligence.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 1.1. What do you mean by artificial intelligence ? Define its goals.

Answer

1. Artificial Intelligence (AI) is an area of computer science that emphasizes the creation of intelligent machines that work and reacts like humans.
2. AI has become an essential part of the technology industry.
3. Research associated with artificial intelligence is highly technical and specialized. The core problems of artificial intelligence include programming computers for certain traits such as :
   - i. Knowledge
   - ii. Reasoning
   - iii. Problem solving
   - iv. Perception
   - v. Learning
   - vi. Planning
   - vii. Ability to manipulate and move objects

Goals of AI :

1. **To create expert systems :** The systems which exhibit intelligent behaviour, learn, demonstrate, explain, and advice its users.
2. **To implement human intelligence in machines :** Creating systems that understand, think, learn, and behave like humans.

Que 1.2. What are the different branches of artificial intelligence ? Discuss some of the branches and progress made in their fields.

AKTU 2014-15, Marks 05

Answer

Different branches of AI and progress made in these fields :

1. **Machine Learning (ML) :**
   i. ML is a method where the target is defined and the steps to reach that target are learned by the machine itself by training.
   ii. For example, to identify a simple object such as an apple or orange. The target is achieved by showing multiple pictures of object and thereby allowing the machine to define the steps to identify it like an apple or an orange.
2. **Natural Language Processing (NLP) :**
   i. NLP is defined as the automatic manipulation of natural language, like speech and text, by software.
   ii. For example, e-mail spam detection which has improved the mail system.
3. **Vision :** Machine vision captures and analyses visual information using a camera, analog-to-digital conversion, and digital signal processing.
4. **Robotics :**
   i. Robotics is a field of engineering focused on the design and manufacturing of robots.
   ii. Robots are used where the tasks are difficult for humans to perform.
   iii. For example, car assembly lines, in hospitals, office cleaner, serving foods and preparing foods in hotels etc.

Que 1.3. Define the role of the machine intelligence in the human life. **AKTU 2017-18, Marks 10**

Answer

1. Machine intelligence is the intelligence provided to the particular machine to achieve the goals of the problems in AI.
2. It is defined as the embedding of intelligence in the machine so that the machine can behave like a human.
3. In human life, machine learning solves many problems of daily purpose of the human.
4. There are many problems which require intelligence such as complex arithmetic which is done by machine very easily.
5. Machine learning plays an important role in following areas :
   i. **Learning :** Learning means to acquire new things from the set of given knowledge or experiences. It refers to the change in subject's behaviour to a given situation brought by repeated experiences in that situation.

ii. **Reasoning :** Reasoning means to infer facts from given facts. Inferences are classified as either deductive or inductive and the reasoning is to draw inferences appropriate to the situation.

iii. **Problem solving :** To solve problem means to move towards the goal. In this, set of rules are defined and a goal is also defined which is to be achieved by using these rules.

iv. **Language understanding :** It means to understand natural language meaning. A language is a system of signs having meaning-by-convention. The meaning-by-convention is distinctive of language and is very different from natural meaning.

Que 1.4. Explain Turing test in detail.

Answer

1. The test conducted for computer intelligence is known as the Turing test.
2. This involves three participants, the computer, a human interrogator and a human foil.
3. The interrogator attempts to determine, by asking questions of the other participants, which is the computer. All communication is via keyboard and screen.
4. The interrogator may ask questions as penetrating and wide-ranging as he or she likes, and the computer is permitted to do everything possible to force a wrong identification.
5. The foil must help the interrogator to make a correct identification.
6. For example, if the human interrogator in Room *C* is not able to identify who is in Room *A* or in Room *B*, then the machine processes intelligence.

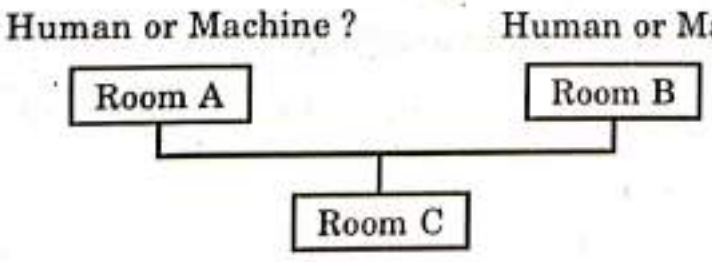


Fig. 1.4.1. Turing test.

Que 1.5. Differentiate between human intelligence and machine intelligence.

Answer

S. No.	Human intelligence	Machine intelligence
1.	Humans perceive by patterns.	The machines perceive by set of rules and data.
2.	Humans store and recall information by patterns.	Machines store and recall information by using searching algorithms.
3.	Humans can figure out the complete object even if some part of it is missing or distorted.	The machines cannot figure out the complete object correctly even if some part of it is missing.

PART-2

Foundation and History of Artificial Intelligence, Applications of Artificial Intelligence.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 1.6. **Write a short note on the foundation of AI.**

AKTU 2014-15, Marks 05

OR

Describe briefly the evolution of artificial intelligence.

AKTU 2018-19, Marks 10

Answer

Foundation of AI :

1. **Beginning of AI (1943) :** The concept of AI began around 1943. AI is not limited to the computer sciences disciplines, but can be seen in various other areas.
2. **AI knowledge-based expert system (1970) :** An AI system often uses a rule-based system to capture knowledge in the form of if-then statements or as decision trees.
3. **Machine learning (1998) :** There are two types of machine learning :
   i. **Formal :** The formal type of machine learning is a computer program that learns from experience in respect to some task and increases performance based on that experience.

ii. **Informal :** The informal involves giving computers the ability to learn without explicitly programming the capability.

4. **Supervised learning (2004) :** The supervised learning is based on giving the correct answers and having the computer mapping inputs to outputs. For example,
   i. **Spam filters :** Software is trained to learn and distinguish between spam and non-spam messages (For example, e-mail filters).
   ii. **Facial recognition :** It is used by cameras to focus and via photo editing software to tag persons (For example, Facebook).
5. **Unsupervised learning (2010) :** Unsupervised learning is the reverse of supervised learning where the correct answers are unknown. For example,
   i. **Clustering algorithm :** Used to find patterns in datasets and then group that data into different coherent clusters.
   ii. **Market segmentation :** Targeting customers based on regions, likes, dislikes, when the consumer makes purchases, etc. This is considered targeted marketing.
6. **Genetic programming (2010) :** Genetic programming is an idea that uses evolutionary process to improve algorithms.
7. **Future of AI (2019 onwards) :** There are many challenges in mimicking human intelligence. Humans acquire common senses that are intuitive but hard to reason rationally. For example, the colour of a blue car is blue.

Que 1.7. Describe the applications of artificial intelligence.

Answer

Applications of artificial intelligence :

1. **Gaming :** AI plays crucial role in strategic games such as chess, poker, tic-tac-toe, etc., where machine can think of large number of possible positions based on heuristic knowledge.
2. **Natural language processing :** It is possible to interact with the computer that understands natural language spoken by human.
3. **Expert systems :** There are some applications which integrate machine, software, and special information to impart reasoning and advising. They provide explanation and advice to the users.
4. **Vision systems :** These systems understand, interpret, and comprehend visual input on the computer.
5. **Speech recognition :** Some intelligent systems are capable of hearing and comprehending the language in terms of sentences and their meanings while a human talks to it. It can handle different accents, slang words, noise in the background, change in human's noise due to cold, etc.

6. **Handwriting recognition :** The handwriting recognition software reads the text written on paper by a pen or on screen by a stylus. It can recognize the shapes of the letters and convert it into editable text.
7. **Intelligent robots :** Robots are able to perform the tasks given by a human. They have sensors to detect physical data from the real world. They have efficient processors, multiple sensors and huge memory, to exhibit intelligence.

PART-3

Intelligent Agent, Structure of Intelligent Agents.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 1.8. **What is intelligent agent ? Describe basic kinds of agent programs.**

AKTU 2015-16, Marks 05

Answer

Intelligent agent : An intelligent agent is an autonomous entity which acts upon an environment using sensors and actuators for achieving goals. An intelligent agent may learn from the environment to achieve their goals.

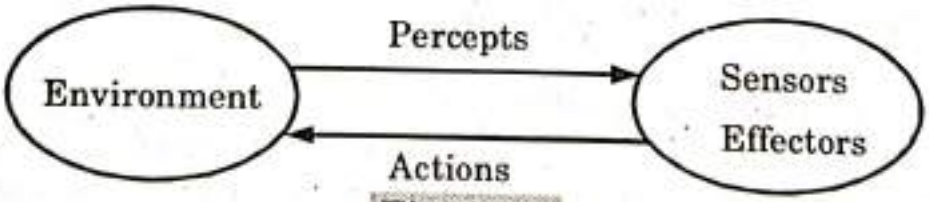


Fig. 1.8.1.

Basic kinds of agent programs are :

1. **Simple reflex agent :**
   i. The simple reflex agents are the simplest agents. These agents take decisions on the basis of the current percepts and ignore the rest of the percept history.
   ii. These agents only succeed in the fully observable environment.
   iii. The simple reflex agent works on condition-action rule, which means it maps the current state to action.

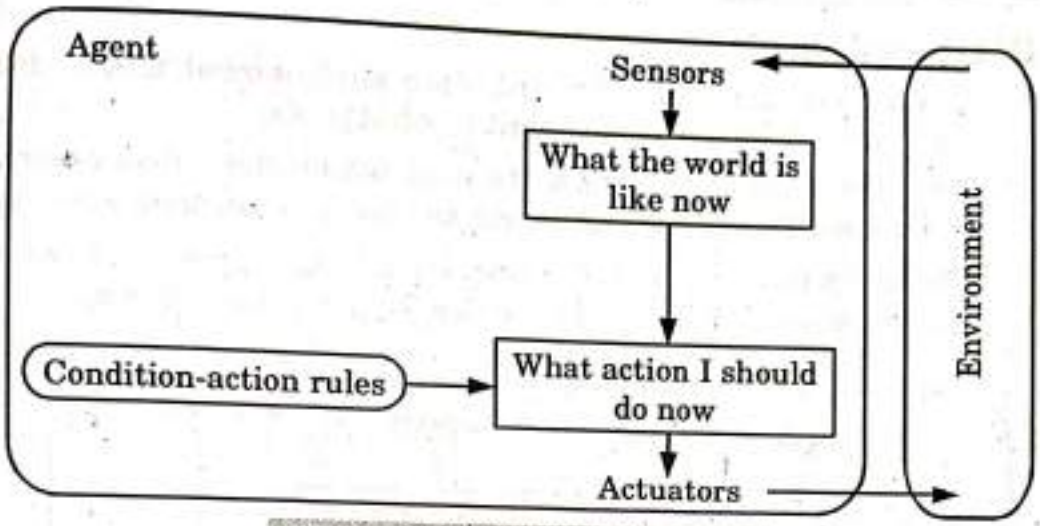


Fig. 1.8.2. Simple reflex agent.

iv. It acts according to a rule whose condition matches the current state, as defined by the percept.

2. Model-based reflex agent :

i. The model-based agent can work in a partially observable environment, and track the situation.

ii. A model-based agent has two important factors :

a. Model : It is knowledge about "how things happen in the world," so it is called model-based agent.

b. Internal state : It is a representation of the current state based on percept history.

iii. These agents have the model, "which is knowledge of the world" and based on the model they perform actions.

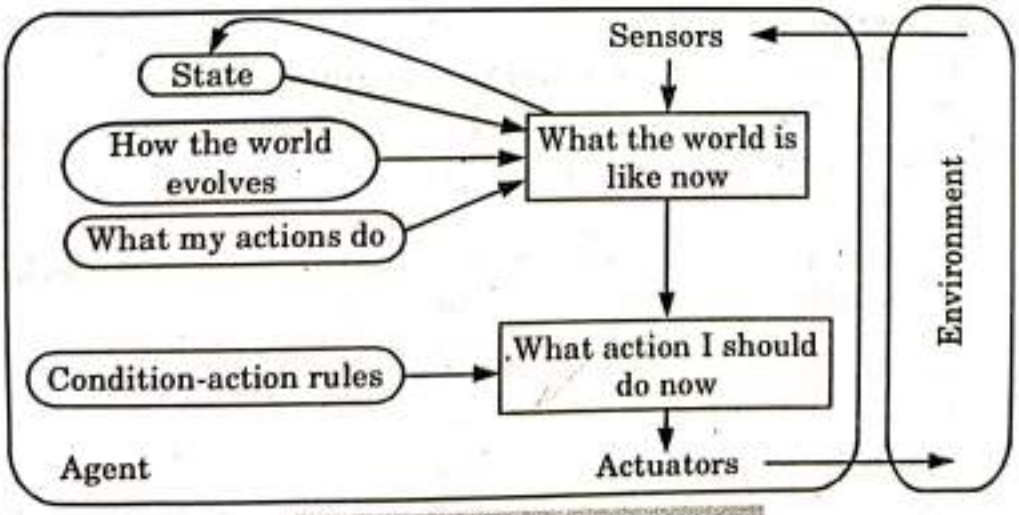


Fig. 1.8.3. Model-based agent.

iv. It keeps track of the current state of the world using an internal model and then chooses an action.

3. Goal-based agents :

i. The knowledge of the current state environment is not always sufficient to decide for an agent to what to do.

ii. The agent needs to know its goal which describes desirable situations. They choose an action, so that they can achieve the goal.

iii. These agents may have to consider a long sequence of possible actions before deciding whether the goal is achieved or not.

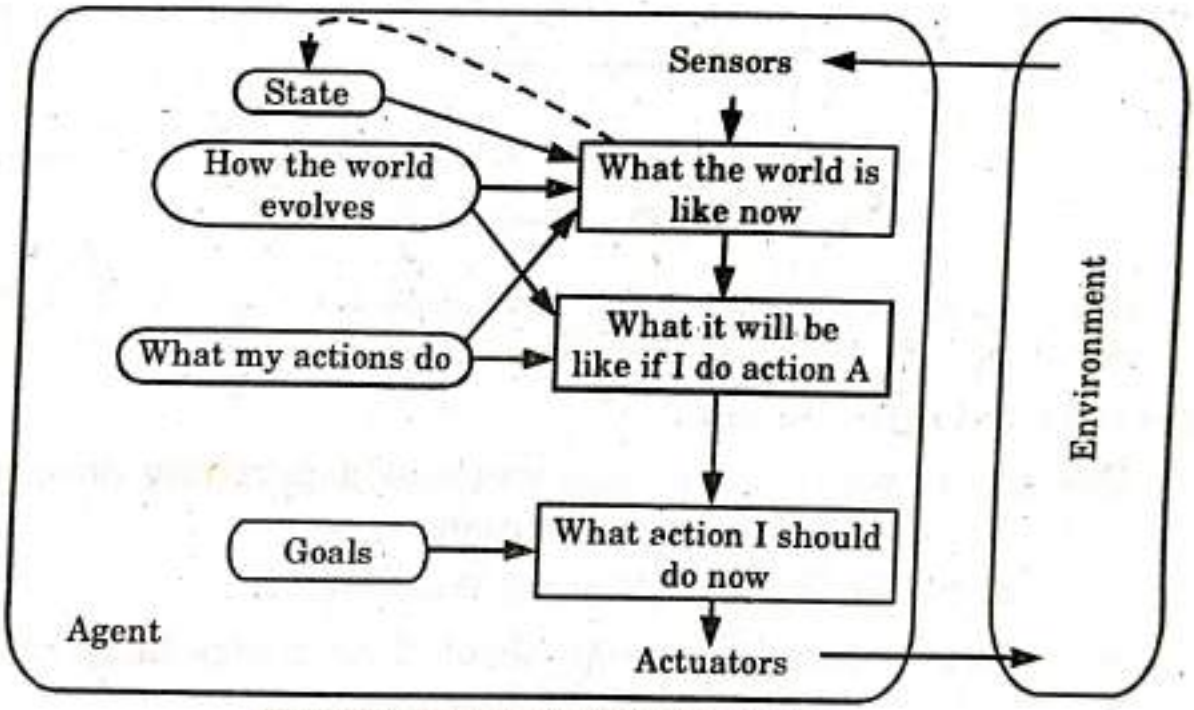


Fig. 1.8.4. Goal-based agent.

iv. It keeps track of the world state as well as a set of goals it is trying to achieve, and chooses an action that will (eventually) lead to the achievement of its goals.

4. Utility-based agents :

i. Utility-based agent is the best way to achieve the goal.

ii. The utility-based agent is useful when there are multiple possible alternatives, and an agent has to choose in order to perform the best action.

iii. The utility function maps each state to a real number to check how efficiently each action achieves the goals.

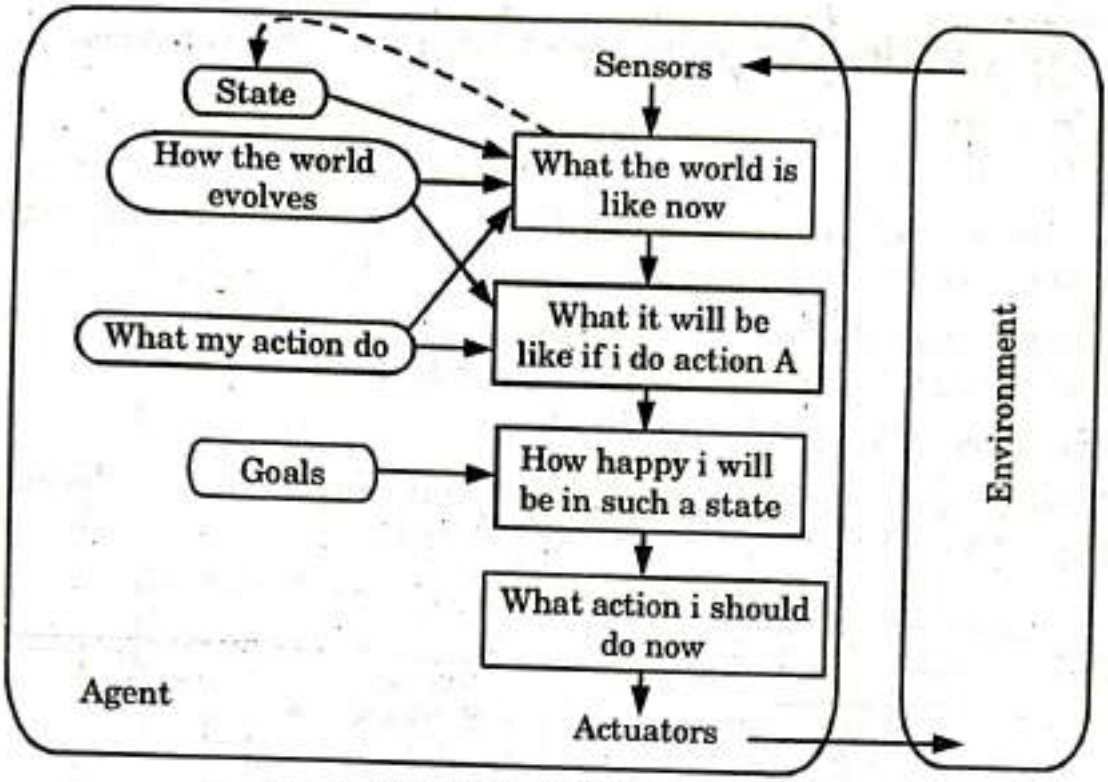


Fig. 1.8.5. Utility-based agent.

iv. It uses a model of the world, along with a utility function that measures its preferences among states of the world. Then it chooses the action that leads to the best expected utility, where expected utility.

Que 1.9. What do you mean by agent and agent program ? How do you ensure that an agent program is an intelligent agent program ?

AKTU 2014-15, Marks 05

Answer

1. An agent can be anything that perceive its environment through sensors and act upon that environment through actuators. An agent runs in the cycle of perceiving, thinking, and acting.
2. To understand the structure of intelligent agents, we should be familiar with architecture and agent program.

   Agent = Architecture + Agent Program
3. Architecture is the machinery that the agent executes on. It is a device with sensors and actuators, for example : a robotic car, a camera, a PC.
4. Agent program is an implementation of an agent function.
5. An agent function is a map from the percept sequence to an action.
6. An agent program is an intelligent agent program if it follows the weak notion (*i.e.*, flexibility, interactivity and autonomy) and strong notion (*i.e.*, information-related states, connotative states and affective states).

Que 1.10. Explain learning agent with its architecture.

Answer

1. A learning agent is a tool in AI that is capable of learning from its experiences.
2. Learning agents are able to perform tasks, analyze performance and look for new ways to improve on those tasks.

Architecture of learning agent :

1. A learning agent can be divided into four components as shown in the Fig. 1.10.1.

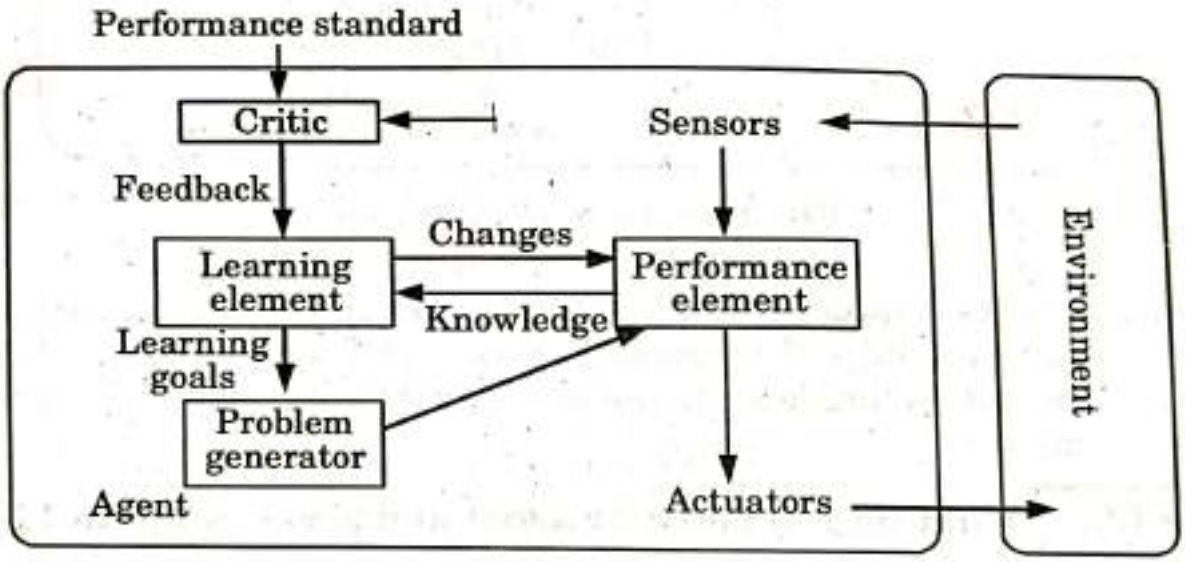


Fig. 1.10.1.

2. Four component of learning agent are :
   - **a. Learning element :**
     - i. The learning element responsible for making improvements.
     - ii. It uses feedback from the critic on how the agent is doing and determines how the performance elements should be modified in the future.
     - ii. The design of learning element depends on the design of performance element.
   - **b. Performance element :** It is responsible for selecting external action.
   - **c. Critic :**
     - i. It tells the learning elements how well the agent is doing with respect to a fixed performance standard.
     - ii. It is necessary because the percepts themselves provide no indication of the agent's success.

d. Problem generator :

i. Problem generator is responsible for suggesting actions that will lead to new and informative experiences.

ii. It also suggests exploratory actions.

Que 1.11. Explain in detail on the characteristics and applications of learning agents. **AKTU 2016-17, Marks 10**

Answer

Characteristics of learning agent :

1. **Situatedness :** When an agent receives some form of sensory input from its environment, it then performs some actions that change its environment in some way.
2. **Autonomy :** This agent characteristic means that an agent is able to act without direct intervention from humans or other agents. This type of agent has almost complete control over its own actions and internal state.
3. **Adaptivity :** This agent characteristic means that it is capable of reacting flexibly to changes within its environment. It is able to accept goal directed initiatives when appropriate and is also capable of learning from its own experiences, environment and interaction with others.
4. **Sociability :** This type of characteristic means that the agent is capable of interacting in a peer-to-peer manner with other agents or humans.

Applications of learning agent :

1. Clustering
2. Classification
3. Prediction
4. Search engines
5. Computer vision
6. Self-driving car
7. Recognition of gestures

Que 1.12. How to specify task environment ? Explain with example.

Answer

PEAS description is used to specify task environment :

PEAS stands for Performance, Environment, Actuators, and Sensors : Based on these properties of an agent, they can be grouped

together or can be differentiated from each other. Each agent has these following properties defines for it.

Performance : The output which we get from the agent. All the necessary results that an agent gives after processing comes under its performance.

Environment : All the surrounding things and conditions of an agent fall in this section. It basically consists of all the things under which the agents work.

Actuators : The devices, hardware or software through which the agent performs any actions or processes any information to produce a result are the actuators of the agent.

Sensors : The devices through which the agent observes and perceives its environment are the sensors of the agent.

Example :

Let us take an example of a self-driven car. As the name suggests, it is a car which drives on its own, by taking all the necessary decisions while driving without any help from the user (customer). In other words, we can say that this car drives on its own and requires no driver. The PEAS description for this agent will be as follows:

Performance : The performance factors for a self-driven car will be the speed, safety while driving (both of the car and the user), Time is taken to drive to a particular location, the comfort of the user, etc.

Environment : The road on which the car is being driven, other cars present on the road, pedestrians, crossings, road signs, traffic signals, etc., all act as its environment.

Actuators : All those devices through which the control of the car is handled are the actuators of the car. For example, the Steering, Accelerator, Breaks, Horn, Music system, etc.

Sensors : All those devices through which the car gets an estimate about its surroundings and it can draw certain perceptions out of it are its sensors. For example, Camera, Speedometer, GPS, Odometer, Sonar, etc.

Que 1.13. State the various properties of environment.

AKTU 2018-19, Marks 10

Answer

Properties of environments :

1. **Discrete / Continuous :** If there are a limited number of distinct, clearly defined, states of the environment, the environment is discrete otherwise it is continuous.
2. **Observable / Partially observable :** If it is possible to determine the complete state of the environment at each time point from the percepts it is observable; otherwise it is only partially observable.

3. **Dynamic / Static :** If the environment is changing for agents action then the environment is dynamic for that agent, otherwise it is static.
4. **Single agent / Multiple agents :** An agent operating by itself in an environment is single agent. Multiple agent is when other agents are present. Other agent is anything that changes from step to step.
5. **Accessible / Inaccessible :** If the agent's sensory apparatus can have access to the complete state of the environment, then the environment is accessible to that agent; otherwise it is inaccessible.
6. **Deterministic / Non-deterministic :** If the next state of the environment is completely determined by the current state and the actions of the agent, then the environment is deterministic; otherwise it is non-deterministic.
7. **Episodic / Non-episodic :** If the agent's experience is divided into atomic episodes and in each episodes the agent receives a percept and then performs a single action, then the environment is episodic; otherwise it is non-episodic.

PART-4

Computer Vision, Natural Language Processing.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 1.14. **Describe the role of computer vision in artificial intelligence.**

AKTU 2014-15, Marks 05

AKTU 2017-18, Marks 10

Answer

1. Computer vision is an interdisciplinary scientific field that deals with how computers can be made to gain high-level understanding from digital images or videos. From the perspective of engineering, it seeks to automate tasks that the human visual system can do.
2. Areas of artificial intelligence deal with autonomous planning for robotical systems to navigate through an environment.
3. A detailed understanding of these environments is required to navigate through them.
4. Information about the environment and the robot could be provided by a computer vision system acting as a vision sensor and providing high level information about the environment and the robot.

Que 1.15. Describe the applications of computer vision.

Answer

Applications of computer vision :

1. Automatic inspection (image-based automated inspection). For example, in manufacturing applications.
2. Assisting humans in identification tasks (to identify object/species using their properties). For example, a species identification system.
3. Controlling processes (in a way of monitoring robots). For example, an industrial robot.
4. Detecting events. For example, for visual surveillance.
5. Modeling objects or environments (using drones can analyses about climatic factors that leads to change in vegetation, etc.). For example, medical image analysis or topographical modeling.
6. Navigation. For example, by an autonomous vehicle or mobile robot.

Que 1.16. Describe the role of artificial intelligence in natural language processing.

AKTU 2014-15, Marks 05

AKTU 2017-18, Marks 10

Answer

1. Natural Language Processing (NLP) refers to AI method of communicating with an intelligent system using a natural language such as English.
2. AI provides computer the ability to accept spoken words as dictation or to follow voice commands by using software.
3. AI programs are able to communicate with humans in a natural way because natural language is one of the most important medium for the communication.
4. To understand the natural language, a program needs a considerable knowledge about the structure of the language including what the words are and how they combine into phrases and sentences.
5. There are three different approaches for the development of natural language understanding programs :
   i. The use of keyword and pattern matching.
   ii. The use of combined syntactic (structural) and semantic directed analysis.
   iii. Comparing and matching the input to real world situation.

Que 1.17. **Discuss various approaches in NLP.**

AKTU 2016-17, Marks 05

Answer

Various approaches in Natural Language Processing (NLP) :

1. **Symbolic approach :**
   i. The symbolic approach to natural language processing is based on human-developed regulations and lexicons.
   ii. The foundation behind this approach is in generally approved regulations of speech within a specific language which is materialised and recorded by experts.
2. **Statistical approach :**
   i. The statistical approach to natural language processing is based on observable and persistent examples of semantic phenomena.
   ii. Models which are based on statistics identify persistent themes through mathematical interpretation of the large text.
   iii. By recognising trends in huge samples of text the computer system can develop its own semantic rules that it will use to interpret future input variables and the development of language output.
3. **Connectionist approach :**
   i. The connectionist approach to natural language processing is a mixture of the symbolic and statistical approaches.
   ii. This approach starts with generally approved rules of language and converts them to specific applications from input procured from statistical inference.

Que 1.18. **Write short note on conceptual dependency.**

AKTU 2016-17, Marks 05

Answer

1. Conceptual Dependency (CD) is a theory of natural language processing which mainly deals with representation of semantics of a language.
2. Knowledge is represented in CD by element which is known as conceptual structure.
3. Conceptual parsing is a strategy for finding both the structure and the meaning of a sentence in one step.

Advantages of CD :

1. These primitives involve fewer inference rules.
2. Many inference rules are already represented in CD structure.

3. The holes in the initial structure help to focus on the points still to be established.

Disadvantages of CD :

1. Knowledge must be decomposed into fairly low level primitives.
2. It is impossible or difficult to find correct set of primitives.
3. A lot of inference may still be required.
4. Representation can be complex even for relatively simple actions.

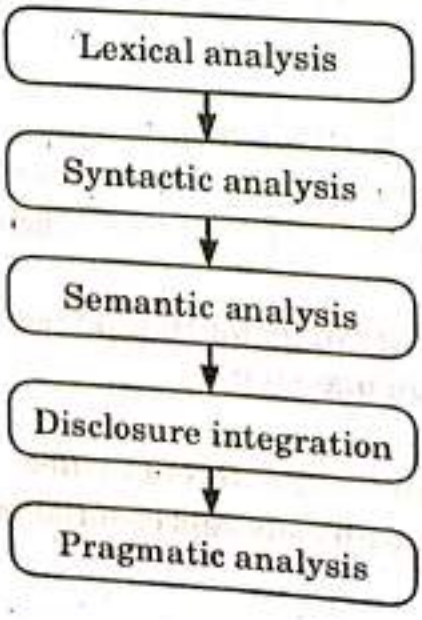
Que 1.19. What is the role of NLP in AI ? Illustrate the various phases in NLP. **AKTU 2018-19, Marks 10**

Answer

Role of NLP : Refer Q. 1.16, Page 1–15A, Unit-1.

There are five phases involved in NLP :

1. **Lexical analysis :** It involves identifying and analyzing the structure of words. Lexical analysis is dividing the whole chunk of text into paragraphs, sentences, and words.
2. **Syntactic analysis (Parsing) :** It involves analysis of words in the sentence for grammar and arranging words in a manner that shows the relationship among the words.
3. **Semantic analysis :** It draws the exact meaning or the dictionary meaning from the text. The text is checked for meaningfulness. It is done by mapping syntactic structures and objects in the task domain.
4. **Discourse integration :** The meaning of any sentence depends upon the meaning of the sentence just before it. In addition, it also brings about the meaning of the sentence that follows it.
5. **Pragmatic analysis :** During pragmatic analysis, what was said is re-interpreted on what it actually meant. It involves deriving those aspects of language which require real world knowledge.



VERY IMPORTANT QUESTIONS

Following questions are very important. These questions may be asked in your SESSIONALS as well as UNIVERSITY EXAMINATION.

Q. 1. Describe the role of computer vision in artificial intelligence.

Ans. Refer Q. 1.4.

Q. 2. Describe the role of artificial intelligence in natural language processing.

Ans. Refer Q. 1.6.

Q. 3. What is intelligent agent ? Describe basic kinds of agent programs.

Ans. Refer Q. 1.8.

Q. 4. Discuss various approaches in NLP.

Ans. Refer Q. 1.17.

Q. 5. What is the role of NLP in AI ? Illustrate the various phases in NLP.

Ans. Refer Q. 1.19.

Q. 6. What are the different branches of artificial intelligence ? Discuss some of the branches and progress made in their fields.

Ans. Refer Q. 1.2.

2 UNIT

Introduction to Search

CONTENTS

PART-1

Introduction to Search : Searching for Solutions, Uninformed Search Strategies; Informed Search Strategies.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 2.1. **Define searching process.**

OR

What are the different parameters used to evaluate a search technique ?

AKTU 2014-15, Marks 05

OR

List the criteria to measure the performance of different search strategies.

AKTU 2018-19, Marks 05

Answer

Searching :

1. Searching is the sequence of steps that transforms the initial state to the goal state.
2. The process of search includes :
   a. Initial state description of the problem.
   b. A set of legal operators that change the state.
   c. The final or goal state.

Following are the parameters used to evaluate a search technique :

1. **Completeness :** By completeness, the algorithm finds an answer in some finite time. So, the algorithm is said to be complete if it is guaranteed to find a solution, if there is one.
2. **Space and time complexity :** With space and time complexity, we address the memory required and the time factor in terms of operations carried out.
3. **Optimality :** Optimality tells us how good the solution is. So, an algorithm or a search process is said to be optimal, if it gives the best solution.

Que 2.2. **Describe the types of search strategies.**

Answer

Types of search strategies :

Uninformed search :

1. Uninformed search means that they have no additional information about states beyond that provided in the problem definition.
2. A uninformed search should proceed in a systematic way by exploring nodes in some predetermined order or simply by selecting nodes at random.
3. Uninformed search is also called brute force search or blind search or exhaustive search.
4. It is of following types :
   i. Breadth First Search
   ii. Depth First Search
   iii. Uniform cost search

Informed search :

1. Informed search algorithm contains an array of knowledge such as how far we are from the goal, path cost, how to reach to goal node, etc. This knowledge helps agents to explore less to the search space and find more efficiently the goal node.
2. Informed search methods often depend on the use of heuristic information.
3. Informed search is also known as heuristic search.
4. Types of informed search are :
   i. Hill climbing
   ii. Best first search
   iii. A* algorithm

Que 2.3. Differentiate between informed search and uninformed search.

Answer

S. No.	Informed search	Uninformed search
1.	It uses knowledge to find the steps to the solution.	No use of knowledge.
2.	It is highly efficient as consumes less time and cost.	Efficiency is mediatory.
3.	Cost is low.	Cost is comparatively high.
4.	It finds solution more quickly.	Its speed is slow.
5.	For example : Best first search, A* algorithm	For example : Breadth First Search, Depth First Search

Que 2.4. **Write a short note on**

i. Breadth First Search (BFS)
ii. Depth First Search (DFS)

Answer

i. Breadth First Search (BFS) :

1. Breadth First Search (BFS) is an algorithm for traversing or searching tree.
2. It starts at the tree root (or some arbitrary node of a graph, sometimes referred to as a 'search key'), and explores all of the neighbour nodes at the present depth prior to moving on to the nodes at the next depth level.

Algorithm :

1. Create a variable called NODE-LIST and set it to the initial state.
2. Until a goal state is found or NODE-LIST is empty :
   a. Remove the first element from NODE-LIST and call it *E*. If NODE-LIST was empty, quit.
   b. For each way that each rule can match the state described in *E* do :
      i. Apply the rule to generate a new state.
      ii. If the new state is a goal state, quit and return this state.
      iii. Otherwise, add the new state to the end of NODE-LIST.

ii. Depth First Search (DFS) :

1. Depth First Search (DFS) is an algorithm for traversing or searching tree.
2. The algorithm starts at the root node (selecting some arbitrary node as the root node in the case of a graph) and explores as far as possible along each branch before backtracking.

Algorithm :

1. If the initial state is a goal state, quit and return success.
2. Otherwise, do the following until success or failure is signaled :
   a. Generate a successor, *E*, of the initial state. If there are no more successors, signal failure.
   b. Call Depth First Search with *E* as the initial state.
   c. If success is returned, signal success. Otherwise continue in this loop.

Que 2.5. **Write the advantages and disadvantages of BFS and DFS.**

Answer

Breadth First Search (BFS) :

Advantages :

1. If there is more than one solution for a given problem, then BFS provide the minimal solution which requires the least number of steps.

Disadvantages :

1. It requires lots of memory since each level of the tree must be saved into memory to expand the next level.
2. BFS needs lots of time if the solution is far away from the root node.

Depth First Search (DFS) :

Advantages :

1. DFS requires very less memory as it only needs to store a stack of the nodes on the path from root node to the current node.
2. It takes less time to reach to the goal node.

Disadvantages :

1. There is the possibility that many states keep re-occurring, and there is no guarantee of finding the solution.
2. DFS algorithm goes for deep down searching and sometime it may go to the infinite loop.

Que 2.6. Prove that breadth first search is a special case of uniform cost search. **AKTU 2014-15, Marks 05**

Answer

It can be proved, in the case where all step costs are equal. It means their value must be equal to 1 or any other constant.

$$f(n) = g(n)$$

$$= 1 \times (\text{Depth of } n)$$

Here, $g(n)$ is a multiple of depth n.

Thus, breadth-first search is a special case of uniform cost search.

Que 2.7. Differentiate between Breadth First Search (BFS) and Depth First Search (DFS).

Answer

S.No.	Breadth First Search (BFS)	Depth First Search (DFS)
1.	BFS uses queue data structure for finding the shortest path.	DFS uses stack data structure for finding the shortest path.
2.	BFS is more suitable for searching vertices which are closer to the given source.	DFS is more suitable when there are solutions away from source.
3.	BFS consider all neighbours first and therefore not suitable for decision making trees used in games or puzzles.	DFS is more suitable for game or puzzle problems.
4.	The time complexity of BFS is $O(V + E)$, where V stands for vertices and E stands for edges.	The time complexity of DFS is also $O(V + E)$, where V stands for vertices and E stands for edges.

Que 2.8. What is production system ? Explain the various types of production system. **AKTU 2014-15, Marks 05**

Answer

Production systems :

1. Search forms the core of most intelligent processes, hence AI programs are structured in such a way that describing and performing the search process becomes easy. Production systems provide such structures.
2. A production system consists of :
   i. A set of rules, each consisting of a left side that determines the applicability of the rule and a right side that describe the operation to be performed if the rule is applied.
   ii. One or more knowledge/databases that contain whatever information is appropriate for the particular task.
   iii. A control strategy that specifies the order in which the rules will be compared to the database and a way of revolving the conflicts that arise when several rules match at once.
   iv. A rule applier.

Types of production systems :

1. **Monotonic production system :** It is a production system in which the application of a rule never prevents the later application of another rule that could also have been applied at the time the first rule was selected.
2. **Non-monotonic production system :** It is one in which the application of a rule prevents the later application of another rule that could also have been applied at the time the first rule was selected.
3. **Partially commutative production system :** It is a production system with the property that if the application of a particular sequence of rules transforms state x into state y, then any permutation of those rules that is allowable also transforms state x into state y.
4. **Commutative production system :** It is a production system that is both monotonic and partially commutative.

Que 2.9. What are various production system characteristics ? Define the various issues involved in the design of search programs.

AKTU 2016-17, Marks 10

Answer

Various production system characteristics :

1. **Simplicity :**
   i. The structure of each sentence in a production system is unique and uniform as they use "IF-THEN" structure.
   ii. This structure provides simplicity in knowledge representation.
   iii. This feature of production system improves the readability of production rules.
2. **Modularity :**
   i. This means production rule code the knowledge available in discrete pieces.
   ii. Information can be treated as a collection of independent facts which may be added or deleted from the system with essentially no deleterious side effects.
3. **Modifiability :**
   i. This means the facility of modifying rules.
   ii. It allows the development of production rules in a skeletal form first and then it is accurate to suit a specific application.
4. **Knowledge intensive :**
   i. The knowledge base of production system stores pure knowledge.

ii. This part does not contain any type of control or programming information.

iii. Each production rule is normally written as an English sentence; the problem of semantics is solved by the structure of the representation.

Issues involved in the design of search programs :

1. **The direction in which to conduct the search (forward or backward reasoning) :** The tree can be searched forward from the initial node to the goal state or backwards from the goal state to the initial state.
2. **How to select applicable rules (matching) :** To select applicable rules, it is critical to have an efficient procedure for matching rules against states.
3. **How to represent each node of the search process :** This is the knowledge representation problem or the frame problem.

   For problems like chess, a node can be fully represented by a simple array. In more complex problem solving, however, it is impossible to represent all of the facts.

PART-2

Local Search Algorithms and Optimistic Problems.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 2.10. **What is heuristic function ? Differentiate between blind search and heuristic search strategies.** **AKTU 2018-19, Marks 05**

OR

What is heuristic search ? Give the desirable properties of heuristic search algorithm.

Answer

Heuristic function :

1. A heuristic function is a function that ranks alternatives in search algorithms at each branching step based on available information to decide which branch to follow.

2. It takes the current state of the agent as its input and produces the estimation of how close agent is from the goal.
3. The heuristic method might not always give the best solution, but it guarantees to find a good solution in reasonable time.
4. Heuristic function estimates how close a state is to the goal.
5. It calculates the cost of an optimal path between the pair of states. The value of the heuristic function is always positive.

Properties of heuristic search are :

1. **Admissibility condition :** Algorithm A is admissible if it is generated to return as optimal solution.
2. **Completeness condition :** Algorithm A is complete if it always terminates with a solution.
3. **Dominance properties :** Let A_1 and A_2 be admissible algorithms with heuristic estimation function h_1 and h_2 respectively. A_1 is said to be more informed than A_2 whenever $h_1'(n) > h_2'(n)$ for all n, A_1 is said to dominate A_2.
4. **Optimality property :** Algorithm A is optimal over a class of algorithms if A dominates all members of the class.

Difference :

S. No.	Blind search strategies	Heuristic search strategies
1.	Blind search technique has access only to the problem definition.	Heuristic search technique has access to heuristic function as well as problem definition.
2.	Blind search is less efficient.	Heuristic search is more efficient.
3.	Every action is equally good in blind search.	Every action is not equally good in heuristic search.
4.	Many problems are not solved by blind search.	Most of the problems are solved by heuristic search.
5.	Blind search is known as uninformed search.	Heuristic search is known as informed search.
6.	Blind search use more computation.	Heuristic search use less computation.
7.	Blind search techniques include Breadth-first, Depth-first search, etc.	Heuristic search techniques include Best-first, A*, etc.

Que 2.11. Why heuristic search is better than the blind search ?

Answer

1. The algorithms and strategies for exhaustive search are sometimes called the blind search methods because they lack the power of knowledge-guided search.
2. Blind search methods do not need knowledge. They take $O(b^d)$ time, and O(d) space.
3. In order to solve problems efficiently it is often necessary to construct a control structure that is no longer guaranteed to find the best answer, but will almost always find a very good answer. This is heuristic search methods.
4. Heuristic search guarantees to find a good solution in reasonable time.
5. By sacrificing completeness, it increases efficiency.
6. It is useful in solving such problems which could not be solved by any other way.
7. Hence, heuristic search is better than the blind search because it make use of domain knowledge.

Que 2.12. Explain about the hill climbing algorithm with its drawback and how it can be overcome ?

AKTU 2015-16, Marks 10

OR

Discuss the problems of hill climbing algorithm.

AKTU 2018-19, Marks 10

Answer

1. The hill climbing search algorithm is simply a loop that continually moves in the direction of increasing values, that is, uphill (the goal).
2. It terminates when it reaches a "peak" where no neighbour has a higher value.
3. The algorithm does not maintain a search tree, so the current node data structure only records the state and its objective function value.

Hill climbing algorithm :

1. Evaluate the initial state. If it is also a goal state, then return it and quit. Otherwise, continue with the initial state as the current state.
2. Loop until a solution is found or until there are no new operators left to be applied in the current state.
   a. Select an operator that has not yet been applied to the current state and apply it to produce a new state.
   b. Evaluate the new state.
      i. If it is goal state, then return it and quit.
      ii. If it is not a goal state but it is better than the current state, then make it the current state.
      iii. If it is not better than the current state, then continue in the loop.

Drawbacks of hill climbing algorithm :

1. **Local maxima :**
   i. A local maximum is a peak that is higher than each of its neighboring states, but lower that the global maximum.
   ii. Hill climbing algorithms that reach the vicinity of a local maximum will be drawn upwards towards the peak, but will then be stuck with nowhere else to go.
2. **Plateau :**
   i. A plateau is a flat area of the search space in which a whole set of neighboring states has the same value.
   ii. A hill climbing search might be unable to find its way off the plateau.
3. **Ridges :** Ridges result in a sequence of local maxima that is very difficult for greedy algorithms to navigate.

To overcome the drawbacks of hill climbing algorithm :

1. **Local maxima :**
   i. Backtrack to some earlier node and try going in a different direction.
   ii. To implement this strategy, maintain a list of paths almost taken and go back to one of them if the path that was taken leads to a dead end.
2. **Plateau :**
   i. Make a big jump in some direction to try to get to a new section of the search space.
   ii. If the only rules available describe single small steps, apply them several times in the same direction.
3. **Ridges :**
   i. Apply two or more rules before doing the test.
   ii. This corresponds to moving in several directions at once.

Que 2.13. Write a short note on heuristic search.

OR

What are the heuristic search techniques in AI ? Explain any one in detail. **AKTU 2016-17, Marks 05**

Answer

Heuristic search : Refer Q. 2.10, Page 2–8A, Unit-2.

Heuristic search techniques :

1. Generate-and-test
2. Hill climbing algorithm
3. Best-first search
4. Problem reduction (AO* search)
5. A* algorithm
6. Breadth first search
7. Depth first search

Explanation of hill climbing algorithm : Refer Q. 2.12, Page 2–10A, Unit-2.

Que 2.14. **Discuss the problem of water jug with heuristic search techniques.** **AKTU 2015-16, Marks 10**

Answer

Water jug problem :

1. You are given two jugs, a 4 litres one and a 3 litres one. Neither has any measuring marker on it. There is a pump that can be use to fill the jugs with water. How can you get exactly 2 litres of water into the 4 litres jug ?
2. The state space for this problem can be represented by ordered pairs of integers (x, y) such that $x = 0$, represents the quantity of water in the 3 litres jug.
3. The start state is (0, 0)
4. The goal state is $(2, n)$ for any value of n.

Production rules :

1.	$(x, y) \rightarrow (4, y)$ If $x < 4$	Fill the 4 litres jug.
2.	$(x, y) \rightarrow (x, 3)$ If $y < 3$	Fill the 3 litres jug.
3.	$(x, y) \rightarrow (x - d, y)$ If $x > 0$	Pour some water out of the 4 litres jug.
4.	$(x, y) \rightarrow (x, y - d)$ If $y > 0$	Pour some water out of the 3 litres jug.
5.	$(x, y) \rightarrow (0, y)$ If $x > 0$	Empty the 4 litres jug on the ground.
6.	$(x, y) \rightarrow (x, 0)$ If $y > 0$	Empty the 3 litres jug on the ground.
7.	$(x, y) \rightarrow (4, y - (4 - x))$ If $x - y \geq 0$ and $y > 0$	Pour water from the 3 litres jug into the 4 litres jug until the 4 litres jug is full.
8.	$(x, y) \rightarrow (x - (3 - y), 3)$ If $x + y \geq 3$ and $x > 0$	Pour water from the 4 litres jug into the 3 litres jug until the 3 litres jug is full.
9.	$(x, y) \rightarrow (x - y, 0)$ If $x + y \leq 4$ and $y > 0$	Pour all the water into the 4 litres jug.
10.	$(x, y) \rightarrow (0, x + y)$ If $x + y \leq 3$ and $x > 0$	Pour all the water from the 4 litres jug into the 3 litres jug.

11.	$(0, 2) \rightarrow (2, 0)$	Pour the 2 litres from the 3 litres jug into the 4 litres jug.
12.	$(2, y) \rightarrow (0, y)$	Empty the 2 litres in the 4 litres jug on the ground.

Que 2.15. What is problem space ? How problem can be defined as state space search ? **AKTU 2016-17, Marks 05**

Answer

1. Problem space refers to the entire range of components that exist in the process of finding a solution to a problem. This range starts with "defining the problem," then proceeds to the intermediate stage of "identifying and testing possible solutions" and ends with the final stage of "choosing and implementing a solution".
2. Problem is the question which is to be solved. For solving the problem it needs to be precisely define, which means defining the start state, goal state, other valid states and transitions.
3. A state space representation allows for the formal definition of a problem which makes the movement from initial state to the goal state quite easily.

For example :

The eight tile puzzle consist of a 3 by 3 (3*3) square frame board which holds 8 movable tiles numbered 1 to 8. One square is empty, allowing the adjacent tiles to be shifted. The objective of the puzzle is to find a sequence of tile movements that leads from a starting configuration to a goal configuration.

3	8	1
6	2	5
	4	7

to be transformed →

1	2	3
8		4
7	6	5

The states of 8 tile puzzle are the different permutations of the tiles within frame.

Standard formulation of the problem :

States : It specifies the location of each of the 8 tiles and the blank in one of the nice squares.

Initial state : Any state can be designated as the initial state.

Goal : Many goal configurations are possible.

Legal moves (or state) : They generate legal states that result from trying the four actions :

1. Blank moves left

2. Blank moves right
3. Blank moves up
4. Blank moves down

Path cost : Each step costs 1, so the path cost is the number of steps in the path.

Que 2.16. Prove that breadth first search and depth first search are the special cases of best first search.

AKTU 2017-18, Marks 10

Answer

1. Best first search is a combination of depth first and breadth first searches.
2. Depth first is good because a solution can be found without computing all nodes and breadth first is good because it does not get trapped in dead ends.
3. The best first search allows us to switch between paths thus gaining the benefit of both approaches. At each step the most promising node is chosen.
4. If one of the nodes chosen generates nodes that are less promising it is possible to choose another at the same level and in effect the search changes from depth to breadth.
5. If on analysis these are no better than the search method reverts to the descendants of the first choice and proceeds the backtracking as it were.

PART-3

Adversarial Search, Search for Games, Alpha-Beta Pruning.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 2.17. Write a short note on adversarial search.

Answer

1. Adversarial search is a search, where we examine the problem which arises when we try to plan ahead of the world and other agents are planning against us.
2. The environment with more than one agent is termed as multi-agent environment, in which each agent is an opponent of other agent and

playing against each other. Each agent needs to consider the action of other agent and effect of that action on their performance.

3. So, searches in which two or more players with conflicting goals are trying to explore the same search space for the solution are called adversarial searches, often known as Games.

Que 2.18. How branch and bound techniques could be used to find the shortest path solution to the travelling salesman problem. Discuss.

AKTU 2017-18, Marks 05

Answer

The branch and bound strategy is a system for solving a sequence of subproblems each of which may have multiple possible solutions and where the solution chosen for one sub-problem may affect the possible solutions of later sub-problems.

Principle :

1. Suppose it is required to minimize an objective function.
2. Suppose that we have a method for getting a lower bound on the cost of any solution among those in the set of solutions represented by some subset.
3. If the best solution found so far costs less than the lower bound for this subset :

   Let S be some subset for solution. Let
   $L(S)$ = a lower bound on the cost of any solution belonging to S
   C = cost of the best solution found so far
   If $C \leq L(S)$, there is no need to explore S because it does not contain any better solution.
   If $C > L(S)$, then we need to explore S because it may contain a better solution.

A lower bound for a TSP :

Cost of any tour = $\frac{1}{2}\sum_{v \in V}$ (Sum of the costs of the two tour edges adjacent to v)

Now,

The sum of the two tour edges adjacent to a given vertex $v \geq$ sum of the two edges of least cost adjacent to v.

Therefore,

Cost of any tour $\geq \frac{1}{2}\sum_{v \in V}$ (Sum of the costs of the two least cost edges adjacent to v).

Example of complete graph with five vertices :

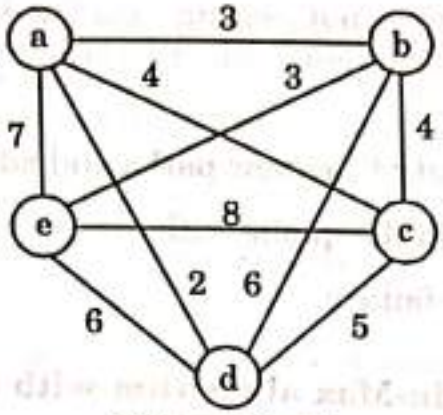


Fig. 2.18.1.

Node	Least Cost Edges	Total Cost
a	(a, d), (a, b)	5
b	(a, b), (b, e)	6
c	(c, b), (c, a)	8
d	(d, a), (d, c)	7
e	(e, b), (e, d)	9

Thus, a lower bound on the cost of any tour

$$= \frac{1}{2}(5 + 6 + 8 + 7 + 9) = 17.5$$

Que 2.19. Discuss branch and bound search.

Answer

1. The branch and bound search strategy applies to problems having a graph search space where more than one alternative path may exist between two nodes.
2. This strategy saves all path lengths (or costs) from a node to all generated nodes and chooses the shortest path for further expansion.
3. It then compares the new path lengths with all old ones and again chooses the shortest path for expansion.
4. In this way, any path to a goal node is certain to be a minimal length path.

Branch and bound search algorithm :

1. Place the start node of zero path length on the queue.
2. Until the queue is empty or a goal node has been found :
   a. Determine if the first path in the queue contains a goal node,

b. If the first path contains a goal node, then exit with success,

c. If the first path does not contain a goal node, remove the path from the queue and form new paths by extending the removed path by one step,

d. Compute the cost of the new paths and add them to the queue,

e. Sort the paths on the queue with lowest-cost paths in front.

3. Otherwise, exit with failure.

Que 2.20. Explain Min-Max algorithm with example.

AKTU 2016-17, Marks 05

Answer

Min-Max algorithm :

Step 1: Set FINAL_VALUE to be minimum as possible.

Step 2 : If limit of search has been reached, then FINAL_VALUE = GOOD_VALUE of the current position.

Step 3 : Else do.

Step 3.1 : Generate the successors of the position.

Step 3.2 : Recursively call MIN-MAX again with the present position with depth incremented by unity.

Step 4 : Evaluate the GOOD_VALUE.

Step 5 : If GOOD_VALUE > FINAL_VALUE then FINAL_VALUE = GOOD_VALUE.

For example :

1. Consider a game which has four final states and paths to reach final state are from root to four leaves of a perfect binary tree as shown Fig. 2.20.1.

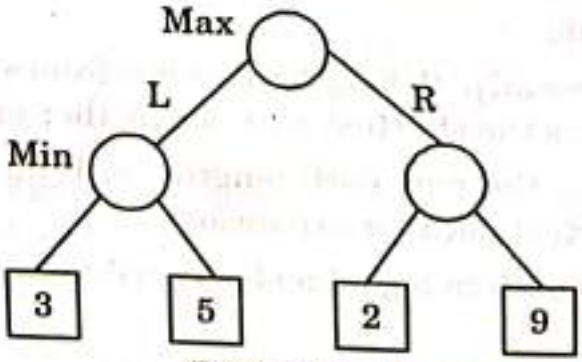


Fig. 2.20.1.

2. Assume that maximizing player get the first chance to move, *i.e.*, maximizer at the root and opponent at next level.

3. As this is a backtracking based algorithm, it tries all possible moves, then backtracks and makes a decision :

i **Maximizer goes LEFT :** It is now the minimizers turn. The minimizer now has a choice between 3 and 5. Being the minimizer it will definitely choose the least among both, that is 3.

ii **Maximizer goes RIGHT :** It is now the minimizers turn. The minimizer now has a choice between 2 and 9. He will choose 2 as it is the least among the two values.

4. Being the maximizer he would choose the larger value that is 3. Hence the optimal move for the maximizer is to go LEFT and the optimal value is 3.
5. Fig. 2.18.2 shows two possible scores when maximizer makes left and right moves.

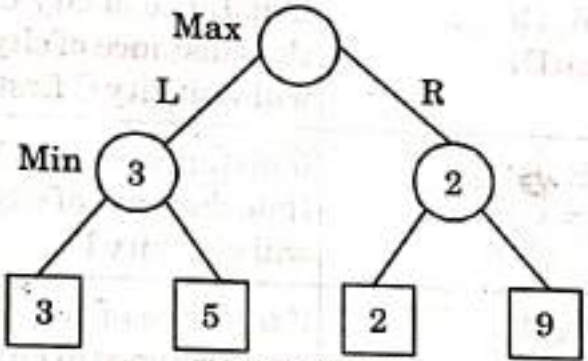


Fig. 2.20.2.

Que 2.21. Give the production rules for travelling salesman problem. Consider that four cities A, B, C and D are given.

AKTU 2014-15, Marks 05

Answer

1. To solve this problem, we can apply the nearest neighbour heuristic, which works by selecting the locally superior alternative at each
2. We can produce the following procedure :
   a. Arbitrarily select a starting city.
   b. To select the next city, look at all cities not yet visited and select the one closest to the current city. Go to it next.
   c. Repeat step *b* until all cities have been visited.

Production rules :

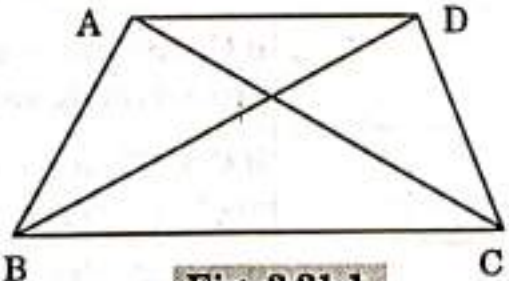


Fig. 2.21.1.

Consider four cities A, B, C, D

Starting point = A

1.	(A, B, C, D) → (A, B) if dist(B) < dist(C) and dist(D)	if distance of city B from city A is less than distance of city C and city D. Then salesman will visit city B first.
2.	(A, B, C, D) → (A, B) if dist(C) < dist(B) and dist(D)	if distance of city C from city A is less than distance of city B and city D. Then salesman will visit city C first.
3.	(A, B, C, D) → (A, B) if dist(D) < dist(B) and dist(C)	if distance of city D from city A is less than distance of city C and city D. Then salesman will visit city D first.
4.	(B, C, D) → (B, C) if dist(C) < dist(D)	if distance of city C from city B is less than distance of city D. Then salesman will visit city C first.
5.	(B, C, D) → (B, D) if dist(D) < dist(C)	if distance of city D from city B is less than distance of city C. Then salesman will visit city D.
6.	(C, B, D) → (C, B) if dist(B) < dist(D)	if distance of city B from city C is less than distance of city D. Then salesman will visit city B.
7.	(C, B, D) → (C, B) if dist(D) < dist(B)	if distance of city D from city C is less than distance of city B. Then salesman will visit city D.
8.	(D, C, B) → (D, C) if dist(C) < dist(B)	if distance of city C from city D is less than distance of city B. Then salesman will visit city C.
9.	(D, C, B) → (D, B) if dist(B) < dist(C)	if distance of city B from city D is less than distance of city C. Then salesman will visit city B.
10.	(B, D) → (B, D) if visited(A, C)	if B is 3rd city and city A and city C are visited then salesman will visit city D.
11.	(B, C) → (B, C) if visited(A, D)	if B is 3rd city and city A and city D are visited then salesman will visit city C.
12.	(C, D) → (C, D) if visited(A, B)	if C is 3rd city and city A and city B are visited then salesman will visit city B.
13.	(C, B) → (C, B) if visited(A, D)	if C is 3rd city and city A and city B are visited then salesman will visit city C.
14.	(D, C) → (D, C) if visited(A, B)	if D is 3rd city and city A and city B are visited then salesman will visit city C.

15.	(D, B) → (D, B) if visited(A, C)	if D is 3rd city and city A and city C are visited then salesman will visit city B.
16.	(B, A) → (A)	if 4th city is B then salesman will return to first city A.
17.	(C, A) → (A)	if 4th city is B then salesman will return to first city A.
18.	(D, A) → (A)	if 4th city is C then salesman will return to first city A.
19.	(B, A) → (A)	if 4th city is D then salesman will return to first city A.

Que 2.22. Describe AO* search technique.

AKTU 2015-16, Marks 05

Answer

AO* algorithm :

1. Initialise the graph to start node.
2. Traverse the graph following the current path accumulating nodes that have not yet been expanded or solved.
3. Pick any of these nodes and expand it and if it has no successors call this value FUTILITY otherwise calculate only f' for each of the successors.
4. If f' is 0 then mark the node as SOLVED.
5. Change the value of f' for the newly created node to reflect its successors by back propagation.
6. Wherever possible use the most promising routes and if a node is marked as SOLVED then mark the parent node as SOLVED.
7. If starting node is SOLVED or value greater than FUTILITY, stop, else repeat from 2.

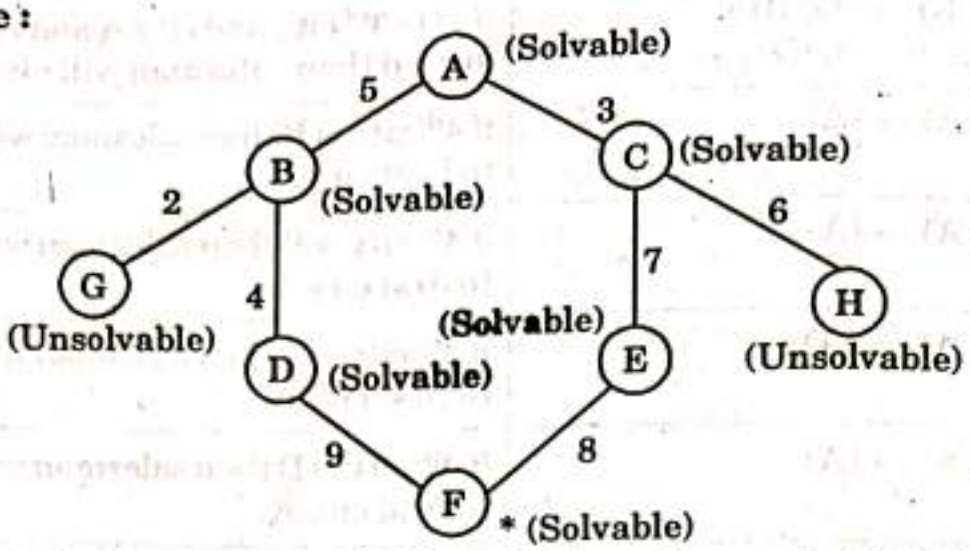
Que 2.23. Explain AO* algorithm with a suitable example. Also the limitations in the algorithm.

AKTU 2016-17, Marks 05

Answer

AO* algorithm : Refer Q. 2.22, Page 2–20A, Unit-2.

Example :

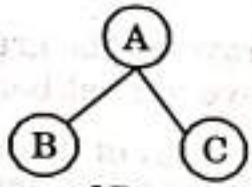


Step 1 : In the given graph, the solvable nodes are *A*, *B*, *C*, *D*, *E*, *F* and the unsolvable nodes are *G*, *H*. Take *A* as the starting node. So, place *A* into OPEN.

i.e., OPEN = | A | CLOSE = (NULL) | ϕ |

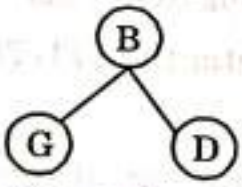
Step 2 : The children of *A* are *B* and *C* which are solvable. So, place them into OPEN and place *A* into the CLOSE.

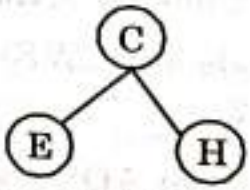
i.e., OPEN = | B | C | CLOSE = | A |



Step 3 : Now process the nodes *B* and *C*. The children of *B* and *C* are to be placed into OPEN. Also remove *B* and *C* from OPEN and place them into CLOSE.

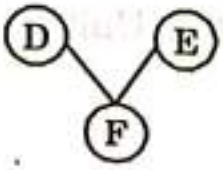
So, OPEN = | G | D | E | CLOSE = | A | B | C |





Step 4 : As the nodes *G* and *H* are unsolvable, so place them into CLOSE directly and process the nodes *D* and *E*.

i.e., OPEN = | A | CLOSE = | A | B | C | $G^{(O)}$ | D | E | $H^{(O)}$ |



'*O*' indicates that the nodes *G* and *H* are unsolvable.

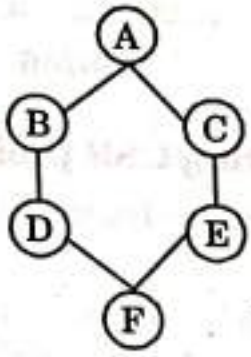
Step 5 : Now we have been reached at our goal state. So, place *F* into CLOSE.

i.e., CLOSE =

A	B	C	$G^{(O)}$	D	E	$H^{(O)}$	F

Step 6 : Success and exit.

AO* Graph :



Limitations of AO* algorithm :

1. The algorithm inserts all the node's ancestors into the set, which may result in the propagation of the cost change back up through a large number of paths which is not good.
2. Sometimes, it cannot find optimal path for unsolvable nodes.

Que 2.24. Explain the constraint satisfaction procedure to solve the cryptarithmetic problem.

CROSS + ROADS = DANGER. **AKTU 2016-17, Marks 10**

Answer

Constraint satisfaction is a search procedure that operates in a space of constraint sets. The initial state contains the constraints that are originally given in the problem description.

The general form of the constraint satisfaction procedure is :

1. Select an unexpected node of the search graph.
2. To generate all possible new constraints, apply the constraint inference rule to the selected node.
3. If the set of constraints contain a contradiction, then report that this path is a dead end.
4. To report success, if the set of constraints, a contradiction describes a complete solution.
5. If neither a contradiction nor a complete solution has been found then apply the problem space rules to generate new partial solutions that are consistent with the current set of constraints. Insert these partial solutions into the search graph.

Numerical :

Constraints :

1. D = 1

2. $2S = R$
3. $C + R > 9$ or 8
4. $S + 1 = E => S = E - 1$

Hence, after guessing

```
  CROSS        96233
+ ROADS      + 62513
-----------  --------
 DANGER       158746
```

Que 2.25. **Solve the following CSP problem of crypt arithmetic.**

Problem :

```
  SEND
+ MORE
------
MONEY
```

AKTU 2017-18, Marks 05

AKTU 2015-16, Marks 10

Answer

1. Initial guess, $M = 1$ because the sum of two single digits can generate atmost a carry of '1'.
2. If $M = 1$, then S should be either 8 or 9 because $S + M$ gives a two-digit number. This also considers the carry digit.
3. When $M = 1$ and $S = 8/9$, the two digit result of $M + S$ can either be 10 or 11. That is, O will be either 0 or 1. But, 1 is already assigned to M so it can not be assigned to any other digit. Thus, $O = 0$, $(M + S) = 10$. S can be 8/9 depending on the carry.
4. Now, $E + 0 = N$, which is only possible if there's a carry of 1 because otherwise, $E + 0 = E$. Thus, $N = E + 1$ and $C2 = 1$.
5. So far, we have $M = 1$, $S = 8/9$, $O = 0$, $C2 = 1$. We are struck here because we do not know the value of E. Thus, we will try different possible values of E.
6. For $E = 5$, $N = 6$, $(N + R) = 15$, $R = 8/9$, $(D + 2) = Y$. Again, we do not know the value of $C1$. So, we will assume it.
7. For $C1 = 1$, $R = 8$, $S = 9$, $(D + 5) = 10 + Y$. Maximum value of D can be 7. If $D = 7$, $Y = 2$.
8. So, $M = 1$, $S = 9$, $O = 0$, $E = 5$, $N = 6$, $R = 8$, $D = 7$, $E = 5$, $Y = 2$.

Thus, we have our probable solution as follows :

```
  SEND        9567
+ MORE      + 1085
-----------  -------
MONEY        10652
```

Que 2.26. **Write a short note on backtracking.**

AKTU 2015-16, Marks 05

Answer

1. The backtracking is an algorithmic method to solve a problem with an additional way.
2. It uses a recursive approach to explain the problems. The backtracking is needed to find all possible combination to solve an optimization problem.
3. It is a systematic way of trying out different sequences of decisions until we find one that works.
4. It is considered an important technique to solve constraint satisfaction issues and puzzles. It is also considered a great technique for parsing and also forms the basis of many logic programming languages.

Que 2.27. **Why is game playing good candidate of AI ? Explain.**

AKTU 2014-15, Marks 05

Answer

1. A good candidate for AI has following characteristics :
   i. It contains a large amount of domain-specific knowledge.
   ii. It contains computationally complex problems.
   iii. It can be developed as a repository for the knowledge of several experts.
2. Let us consider an example of game playing, an intelligent system that plays chess.
   i. The rules of chess are easy to learn, but to play this game at an expert level is not easy because it has 10^{120} possible games. This 10^{120} possible games of chess satisfy the first characteristics of good candidate for AI *i.e.*, large amount of domain-specific knowledge.
   ii. These 10^{120} possible games of chess have equally large and complex moves by various chess pieces (*i.e.*, pawns, rooks, king etc). These are computationally complex problems which cannot be solved by straightforward algorithms. This satisfies the second characteristics of good candidate for AI.
   iii. The chess program is build based on the inputs from several expert chess players. It has enormous amount of knowledge about chess (domain-specific knowledge) that it uses as part of its decision making process. This satisfies the third characteristics of good candidate for AI.

Hence we can say that game playing is good candidate for AI.

Que 2.28. Give an example of game tree. What is the purpose of minimax procedure in a game tree ? **AKTU 2014-15, Marks 05**

Answer

1. A game tree is a directed graph whose nodes are positions in a game and whose edges are moves.
2. The complete game tree for a game is the game tree starting at the initial position and containing all possible moves from each position.
3. Fig. 2.28.1 shows the first two levels in the game tree for tic-tac-toe.
4. The rotations and reflections of positions are equivalent, so the first player has three choices of move : in the center, at the edge, or in the corner.
5. The second player has two choices for the reply if the first player played in the center, otherwise five choices. And so on.

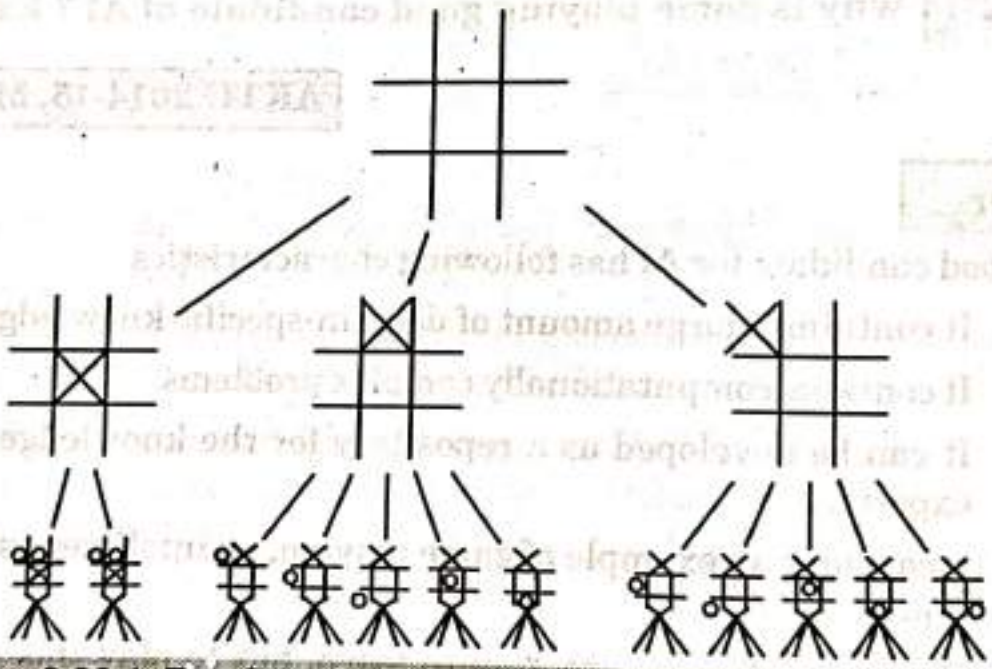


Fig. 2.28.1. The first two levels of the game tree for tic-tac-toe.

6. The number of leaf nodes in the complete game tree is the number of possible different ways the game can be played.
7. For example, the game tree for tic-tac-toe has 255,168 leaf nodes.

Purpose of minimax procedure :

1. For many complex games such as chess, search to termination is impossible, *i.e.*, a win or draw cannot be obtained.
2. Our goal in searching such a game tree might be to find a good first move.
3. This good first move can be extracted by minimax procedure. This is the purpose of minimax procedure in a game tree.

Que 2.29. **Explain Min-Max procedure. Describe alpha-beta pruning and give the other modifications to the Min-Max procedure to improve its performance.** **AKTU 2015-16, Marks 15**

Answer

Min-Max procedure :

1. The min-max algorithm removes unwanted nodes and selects the optimal node for progressing further.
2. The min-max algorithm computes the min-max decision from the current state.
3. It uses a simple recursive computation of the min-max values of each successor state.

Min-Max algorithm : Refer Q. 2.20, Page 2–17A, Unit-2.

Alpha-beta pruning :

1. Alpha-beta pruning is a modified version of the min-max algorithm. It is an optimization technique for the min-max algorithm.
2. In min-max search algorithm, the number of nodes (game states) that it has to examine is exponential in depth of the tree. Now we cannot eliminate the exponent completely, but we can cut it to half.
3. There is a technique by which without checking each node of the game tree we can compute the correct min-max decision, and this technique is called pruning. This involves two threshold parameter alpha and beta for future expansion, so it is called alpha-beta pruning.
4. Alpha-beta pruning can be applied at any depth of a tree, and sometime it not only prunes the tree leaves but also entire sub-tree.
5. The two parameters can be defined as :
   a. **Alpha :** The best (highest-value) choice we have found so far at any point along the path of Maximizer. The initial value of alpha is $-\infty$.
   b. **Beta :** The best (lowest-value) choice we have found so far at any point along the path of Minimizer. The initial value of beta is $+\infty$.

Modifications to min-max : There are some heuristic search methods other than alpha-beta pruning method which are used to improve the performance of min-max procedure. They are :

1. Greedy hill climbing method
2. Artificial immune algorithm

VERY IMPORTANT QUESTIONS

Following questions are very important. These questions may be asked in your SESSIONALS as well as UNIVERSITY EXAMINATION.

Q. 1. What are the different parameters used to evaluate a search technique ?

Ans: Refer Q. 2.1.

Q. 2. Prove that breadth first search is a special case of uniform cost search.

Ans: Refer Q. 2.6.

Q. 3. Why heuristic search is better than the blind search ?

Ans: Refer Q. 2.11.

Q. 4. Explain AO* algorithm with a suitable example. State the limitations in the algorithm.

Ans: Refer Q. 2.23.

Q. 5. Discuss the problem of water jug with heuristic search techniques.

Ans: Refer Q. 2.14.

Q. 6. Explain about the hill climbing algorithm with its drawback and how it can be overcome ?

Ans: Refer Q. 2.12.

Q. 7. Solve the cryptarithmetic problem :

```
  SEND
+ MORE
--------
 MONEY
--------
```

Ans: Refer Q. 2.25.

Q. 8. Explain Min-Max algorithm with example.

Ans: Refer Q. 2.20.

Q. 9. How branch and bound techniques could be used to find the shortest path solution to the travelling salesman problem. Discuss.

Ans: Refer Q. 2.18.

Q. 10. What is production system ? Explain the various types of production system.

Ans: Refer Q. 2.8.

3 UNIT

Knowledge Representation and Reasoning

CONTENTS

PART-1

Knowledge Representation and Reasoning.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 3.1. **Briefly describe the meaning of knowledge representation and knowledge acquisition. What procedure is followed for knowledge acquisition ? Explain.**

AKTU 2014-15, Marks 10

Answer

Knowledge representation :

1. Knowledge representation is the study of ways of picturization of knowledge and how effectively it resembles the representation of knowledge in human brain.
2. Knowledge representation has two entities :
   a. **Facts :** Facts are the truth in some relevant world.
   b. **Representation :** Representation is the presentation of facts in some chosen formalism.
3. For example :

   Fact : Charlie is a dog.

   Representation of fact using mathematical logic : Dog (Charlie)
4. Knowledge representation should possess following characteristics :
   a. Representation scheme should have a set of well defined syntax and semantic.
   b. It should have good expressive capacity.
   c. It must be effective.

Knowledge acquisition :

1. Knowledge acquisition is the process of acquiring knowledge from a human expert for an expert system, which must be carefully organized into IF-THEN rules or some other form of knowledge representation.
2. Knowledge acquisition is the process of absorbing and storing new information in memory, the success of which is determined by how well the information can later be retrieved from memory.
3. The process of storing and retrieving information depends heavily on the representation and organization of the information.

iv. Semantic networks can categorize the object in different forms and can also link those objects.

3. **Frame representation :**

i. A frame is a record like structure which consists of a collection of attributes and its values to describe an entity in the world.

ii. Frames are the AI data structure which divides knowledge into substructures by representing stereotypes situations.

iii. It consists of a collection of slots and slot values. These slots may be of any type and sizes. Slots have names and values which are called facets.

4. **Production rules :**

i. Production rules system consist of (condition, action) pairs which mean, "If condition then action".

ii. In production rules agent checks for the condition and if the condition exists then production rule and corresponding action is carried out.

iii. The condition part of the rule determines which rule may be applied to a problem. And the action part carries out the associated problem solving steps. This complete process is called a recognize-act cycle.

Que 3.3. What are the desirable properties of good knowledge representation schemes ? **AKTU 2015-16, Marks 10**

Answer

A good system for the representation of knowledge in a particular domain should possess the following four properties :

1. **Representational adequacy :** The ability to represent all kinds of knowledge that are needed in that domain.

2. **Inferential adequacy :** The ability to manipulate the representational structures in such a way as to derive new structures corresponding to new knowledge inferred from old.

3. **Inferential efficiency :** The ability to incorporate into the knowledge structure, additional information that can be used to focus the attention of the inference mechanisms in the most promising directions.

4. **Acquisitional efficiency :** The ability to acquire new information easily. The simplest case involves direct insertion, by a person, of new knowledge into the database. Ideally, the program itself would be able to control knowledge acquisition.

PART-2

Propositional Logic, Theory of First Order Logic, Inference in First Order Logic.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 3.4. What is propositional logic ? Describe the types of propositions.

Answer

Propositional logic :

1. Propositional logic (PL) is the simplest form of logic where all the statements are made by propositions. A proposition is a declarative statement which is either true or false. It is a technique of knowledge representation in logical and mathematical form.
2. Propositional logic is also called Boolean logic as it works on 0 and 1.
3. In propositional logic, we use symbolic variables to represent the logic.
4. Propositions can be either true or false, but it cannot be both.

Two types of propositions :

1. **Atomic proposition :** Atomic propositions are the simple propositions. It consists of a single proposition symbol. These are the sentences which must be either true or false. For example :
   i. 2 + 2 is 4, it is an atomic proposition as it is a true fact.
   ii. "The Sun is cold" is also a proposition as it is a false fact.
2. **Compound proposition :** Compound propositions are constructed by combining simpler or atomic propositions, using parenthesis and logical connectives. For example :
   i. "It is raining today, and street is wet."
   ii. "Ankit is a doctor, and his clinic is in Mumbai."

Que 3.5. Explain logical connectives.

Answer

Logical connectives are used to connect two simpler propositions or representing a sentence logically. We can create compound propositions

with the help of logical connectives. There are mainly five connectives which are given as follows :

1. **Negation :** A sentence such as $\neg P$ is called negation of P. A literal can be either positive literal or negative literal.
2. **Conjunction :** A sentence which has $\wedge$ connective such as, $P \wedge Q$ is called a conjunction.
3. **Disjunction :** A sentence which has $\vee$ connective, such as $P \vee Q$ is called disjunction, where P and Q are the propositions.
4. **Implication :** A sentence such as $P \rightarrow Q$, is called an implication. Implications are also known as if-then rules. It can be represented as

   If it is raining, then the street is wet.

   Let P = It is raining and Q = Street is wet, so it is represented as $P \rightarrow Q$
5. **Biconditional :** A sentence such as $P \Leftrightarrow Q$ is a Biconditional sentence, example, If I am breathing, then I am alive

   P = I am breathing, Q = I am alive, it can be represented as $P \Leftrightarrow Q$.

Que 3.6. What is propositional logic ? Define the various inference rules with the help of example.

AKTU 2018-19, Marks 10

Answer

Propositional logic : Refer Q. 3.4, Page 3–5A, Unit-3.

Various inference rules :

1. **Modus Ponens :** The Modus Ponens rule states that if P and $P \rightarrow Q$ is true, then we can infer that Q will be true.

Example :

Statement-1 : "If I am sleepy then I go to bed". $(P \rightarrow Q)$

Statement-2 : "I am sleepy". (P)

Conclusion : "I go to bed". (Q)

Hence, we can say that, if $P \rightarrow Q$ is true and P is true then Q will be true.

Proof by truth table :

P	Q	$P \rightarrow Q$
0	0	0
0	1	1
1	0	0
1	1	1 ←

2. **Modus Tollens :** The Modus Tollens rule state that if $P \rightarrow Q$ is true and Q is true, then $\neg P$ will also true.

Example :

Statement-1 : "If I am sleepy then I go to bed". $(P \rightarrow Q)$

Statement-2 : "I do not go to the bed". $(\sim Q)$

Statement-3 : Which infers that "I am not sleepy". $(\sim P)$

Proof by truth table :

P	Q	$\sim P$	$\sim Q$	$P \rightarrow Q$
0	0	1	1	1 ←
0	1	1	0	1
1	0	0	1	0
1	1	0	1	1

3. **Hypothetical Syllogism :** The Hypothetical Syllogism rule state that if $P \rightarrow R$ is true whenever $P \rightarrow Q$ is true and $Q \rightarrow R$ is true.

Example :

Statement-1 : If you have my home key then you can unlock my home. $(P \rightarrow Q)$

Statement-2 : If you can unlock my home then you can take my money. $(Q \rightarrow R)$

Conclusion : If you have my home key then you can take my money. $(P \rightarrow R)$

Proof by truth table :

P	Q	R	$P \rightarrow Q$	$Q \rightarrow R$	$P \rightarrow R$
0	0	0	1	1	1 ←
0	0	1	1	1	1
0	1	0	1	0	1
0	1	1	1	1	1 ←
1	0	0	0	1	1
1	0	1	0	1	1
1	1	0	1	0	0
1	1	1	1	1	1 ←

4. **Disjunctive Syllogism :** The Disjunctive Syllogism rule state that if $P \vee Q$ is true, and $\neg P$ is true, then Q will be true.

Example :

Statement-1 : Today is Sunday or Monday. $(P \wedge Q)$

Statement-2 : Today is not Sunday. $(\neg P)$

Conclusion : Today is Monday. (Q)

Proof by truth-table :

P	Q	$\neg P$	$P \vee Q$	
0	0	1	0	
0	1	1	1	←
1	0	0	1	
1	1	0	1	

5. **Addition :** The Addition rule states that If P is true, then $P \vee Q$ will be true.

Example :

Statement-1 : I have a vanilla ice-cream. (P)

Statement-2 : I have Chocolate ice-cream. (Q)

Conclusion : I have vanilla or chocolate ice-cream. ($P \vee Q$)

Proof by truth-table :

P	Q	$P \vee Q$	
0	0	0	
1	1	1	←
0	1	1	
1	1	1	←

6. **Simplification :** The simplification rule state that if $P \wedge Q$ is true, then Q or P will also be true.

Proof by truth-table :

P	Q	$P \wedge Q$	
0	0	0	
1	0	0	←
0	1	0	
1	1	1	←

7. **Resolution :** The resolution rule state that if $P \vee Q$ and $\neg P \wedge R$ is true, then $Q \vee R$ will also be true.

Proof by truth-table :

P	$\neg P$	Q	R	$P \vee Q$	$\neg P \wedge R$	$Q \vee R$	
0	1	0	0	0	0	0	
0	1	0	1	0	0	1	
0	1	1	0	1	1	1	←
0	1	1	1	1	1	1	←
1	0	0	0	1	0	0	
1	0	0	1	1	0	1	
1	0	1	0	1	0	1	
1	0	1	1	1	0	1	←

Que 3.7. **Explain the conversion procedure of given formula into normal form.** **AKTU 2017-18, Marks 10**

Answer

The steps used to convert a given formula into its normal form are as follows :

Step 1 : Eliminate implications and biconditionals. For this, use the laws :

$$(A \rightarrow B) = \sim A \vee B$$

$$(A \leftrightarrow B) = (A \rightarrow B) \wedge (B \rightarrow A)$$

$$= (\sim A \vee B) \wedge (\sim B \vee A)$$

Step 2 : Reduce the NOT symbol by the formula $(\sim(\sim A)) = A$ and apply De Morgan's theorem to bring negations before the atoms.

$$\sim(A \vee B) = \sim A \wedge \sim B$$

$$\sim(A \wedge B) = \sim A \vee \sim B$$

Step 3 : Use distributive laws to obtain the normal form.

$$A \wedge (B \vee C) = (A \wedge B) \vee (A \wedge C)$$

$$A \vee (B \wedge C) = (A \vee B) \wedge (A \vee C)$$

Que 3.8. **Prove that following sentence is valid :**

"If prices fall then sell increases. If sell increases then John makes the whole money. But John doesn't make the whole money. Therefore, prices do not fall." **AKTU 2017-18, Marks 10**

Answer

1. If prices fall then sell increases.
   The logical form of the sentence is : If P then Q.
2. If sell increases then John makes the whole money.
   The logical form of the sentence is : If Q then R.
3. We have to check the validity of the sentence.
   John doesn't make the whole money. Therefore, Prices do not fall.

$$\neg R \rightarrow \neg P$$

Hence, from (1) and (2) and by using Modus Ponens rule we get :

$P \rightarrow Q$ (Price falls $\rightarrow$ Sell increases)

$Q \rightarrow R$ (Sell increases $\rightarrow$ John makes whole money)

$P \rightarrow R$ (Price falls $\rightarrow$ John makes whole money)

From transposition rule,

$$P \rightarrow R \text{ infer } \neg R \rightarrow \neg P$$

We conclude that John does not make the whole money, therefore prices do not fall.

Hence, the given argument is valid.

Que 3.9. Describe first order logic in artificial intelligence.

Answer

1. First order logic is also known as Predicate logic or First order predicate logic. First order logic is a powerful language that develops information about the objects in a more easy way and can also express the relationship between those objects.
2. First order logic is an extension to propositional logic.
3. First order logic is sufficiently expressive to represent the natural language statements in a concise way.
4. First-order logic (like natural language) does not only assume that the world contains facts like propositional logic but also assumes objects, relations, function.
5. As a natural language, first-order logic also has two main parts syntax and semantics.

Que 3.10. Explain inference rules in first order logic.

Answer

Inference rules for first order logic :

1. Universal generalization :

i. Universal generalization is a valid inference rule which states that if premise $P(c)$ is true for any arbitrary element c in the universe of discourse, then we can have a conclusion as $\forall \times P(x)$.

ii. It can be represented as :

$$= \frac{P(c)}{\forall \times P(x)}$$

iii. This rule can be used if we want to show that every element has a similar property.

2. Universal instantiation :

i. Universal instantiation is also called as universal elimination or UI is a valid inference rule. It can be applied multiple times to add new sentences.

ii. As per UI, we can infer any sentence obtained by substituting a ground term for the variable.

iii. The UI rule state that we can infer any sentence $P(c)$ by substituting a ground term c (a constant within domain x) from $\forall \times P(x)$ for any object in the universe of discourse.

iv. It can be represented as :

$$= \frac{\forall \times P(x)}{P(c)}$$

3. Existential instantiation :

i. Existential instantiation is also called as Existential elimination, which is a valid inference rule in first order logic.

ii. It can be applied only cnce to replace the existential sentence.

iii. This rule states that one can infer $P(c)$ from the formula given in the form of $\exists \times P(x)$ for a new constant symbol c.

iv. The restriction with this rule is that c used in the rule must be a new term for which $P(c)$ is true.

v. It can be represented as :

$$= \frac{\exists \times P(x)}{P(c)}$$

4. Existential introduction :

i. An existential introduction is also known as an existential generalization, which is a valid inference rule in first order logic.

ii. This rule states that if there is some element c in the universe of discourse which has a property P, then we can infer that there exists something in the universe which has the property P.

iii. It can be represented as :

$$= \frac{P(c)}{\exists \times P(x)}$$

Que 3.11. Translate the following sentences into formulas in predicate logic and clausal form :

i. John likes all kind of food.
ii. Apples are food.
iii. Chicken is food.
iv. Anything any one eats and is not killed by is food.
v. Bill eats peanuts and is still alive.
vi. Sue eats everything Bill eats.

AKTU 2014-15, Marks 10

Answer

Predicate logic :

i. $\forall x$ food (x) $\rightarrow$ likes (John, x)

ii. food (apples)

iii. food (chicken)

iv. $\forall x \ \forall y$ Eats (x, y) $\wedge \neg$ killed (y) $\rightarrow$ food (x)

v. Eats (Peanuts, Bill) $\wedge$ alive (Bill)

vi. $\forall x$, Eats (x, Bill) $\rightarrow$ Eats (x, sue)

Clause form :

i. $\neg$ food (x) $\vee$ likes (John, x)

ii. food (apple)

iii. food (chicken)

iv. $\neg$ Eats (x, y) $\vee$ killed (y) $\vee$ food (x)

v.

a. Eats (Peanut, Bill)

b. Alive (Bill)

vi. $\neg$ Eats (x, Bill) $\vee$ Eats (x, sue)

Que 3.12. **Translate following sentences in formulas in predicate logic and casual form :**

a. Mutton is food.

b. Anything one eats and it does not kill is a food.

c. Rajiv eats everything that Sue eats.

d. Kin eats peanuts and is still alive.

e. John will marry Mary if Mary loves John.

AKTU 2015-16, Marks 10

Answer

Predicate logic :

a. Food (Mutton)

b. $\forall x \ \forall y$ Eats $(x, y) \wedge \neg$ killed (y) $\rightarrow$ food (x)

c. Eats (sue, x) $\rightarrow$ eats (Rajiv, x)

d. Eats (peanuts, kin) $\rightarrow$ alive (kin)

e. $\forall x \ \forall y$ Marry $(x, y) \rightarrow$ loves (y, x)

Casual form :

a. Food (Mutton)

b. $\neg$ (Eats $(x, y) \wedge$ killed $(y) \vee$ food (x)

c. $\neg$ Eats $(x,$ Sue$) \vee$ Eats $(x,$ Rajiv)

d. i. Eats (Peanut, kin)

ii. Alive (kin)

e. Marry (John, Mary) $\rightarrow$ loves (Mary, John)

Que 3.13. **What do you mean by representing instance and ISA relationship ?**

AKTU 2016-17, Marks 10

Answer

1. Instance and ISA plays an important role in a useful form of reasoning, property inheritance.
2. The first part of the Fig. 3.13.1 contains class membership with unary predicates (such as Roman), each of which corresponds to a class. Asserting that $P(x)$ is true is equivalent to asserting that x is an instance (or element) of P.
3. The second part of the Fig. 3.13.1 contains representations that use the instance predicate explicitly.
4. The predicate instance is a binary one, whose first argument is an object and whose second argument is a class to which the object belongs. But these representations do not use an explicit ISA predicate.
5. The third part of Fig. 3.13.1 contains representations that use both the instance and ISA predicates explicitly.

1. man(Marcus)
2. Pompeian(Marcus)
3. $\forall x$: Pompeian$(x) \rightarrow$ Roman(x)
4. ruler(Caesai)
5. $\forall x$: Roman$(x) \rightarrow$ loyalto$(x,$ Caesar$) \vee$ hate$(x,$ Caesar)

1. instance(Marcus, man)
2. instance(Marcus, Pompeian)
3. $\forall x$: instance$[x,$ Pompeian$) \rightarrow$ instance$(x,$ Roman)
4. instance(Caesar, ruler)
5. $\forall x$: instance$(x,$ Roman$) \rightarrow$ loyalto$(x,$ Caesar$) \vee$ hate$(x,$ Caesar)

1. instance(Marcus, man)
2. instance(Marcus, Pompeian)
3. isa(Pompeian, Roman)
4. instance(Caesar, ruler)
5. $\forall x$: instance(x, Roman) $\rightarrow$ loyalto(x, Caesar) $\vee$ hate(x, Caesar)

Fig. 3.13.1. Three ways of representing class membership.

Que 3.14. **Consider the argument, "All dogs bark. Some animals are dogs. Therefore, some animals bark". Determine whether the conclusion is a valid consequence of the premises.**

Answer

$\forall x$ Dogs (x) $\rightarrow$ Bark ...(3.14.1)

$\exists$ Animals (x) $\rightarrow$ Dogs (x) ...(3.14.2)

From (3.14.1) and (3.14.2) we conclude that,

$\exists$ Animals (x) $\rightarrow$ Bark

Hence, this conclusion is a valid sequence of premises.

Que 3.15. **Determine whether the following argument is valid. "If I work whole night on this problem, then I can solve it. If I solve the problem, then I will understand the topic. Therefore, I will work whole night on this problem, then I will understand the topic."**

Answer

We use the following symbols for each sentence :

A : If I work whole night on this problem

B : I can solve it

C : I will understand the topic

The translation for each line of the argument is as follows :

$(A \supset B) \cup (B \supset C)$: If I work whole night on this problem, then I can solve it. If I solve the problem, then I will understand the topic.

$(A \supset C)$: I will work whole night on this problem, then I will understand the topic.

The sentence corresponding to the argument is :

$$((A \supset B) \cup (B \supset C)) \supset (A \supset C)$$

To see if the sentence is valid, we try to find an interpretation that falsify $A \supset C$ while satisfying $((A \supset B) \cup (B \supset C))$. Falsifying $A \supset C$ requires A- true and C-true. To satisfy $((A \supset B) \cup (B \supset C))$ when A-true and B-true requires C-true and C is true. Thus, the argument is valid.

Que 3.16. **Convert the formula (A $\rightarrow$ ((B $\wedge$ C) $\rightarrow$ D)) into DNF.**

Answer

Eliminate implications :

Equation $(A \rightarrow ((B \wedge C)\ D))$ becomes :

$$= \sim A \vee ((B \wedge C) \rightarrow D)$$
$$= \sim A \vee (\sim(B \wedge C) \vee D)$$
$$= \sim A \vee (\sim B \vee \sim C \vee D)$$
$$= \sim A \vee \sim B \vee \sim C \vee D$$

Hence, $\sim A \vee \sim B \vee \sim C \vee D$ is the DNS of given equation because the literals are separated by OR condition.

Que 3.17. **Convert $((A \rightarrow B) \rightarrow C)$ into CNF.**

Answer

Eliminate implications :

Equation $((A \rightarrow B) \rightarrow C)$ becomes :

$$= \sim (A \rightarrow B) \vee C$$
$$= \sim (\sim A \vee B) \vee C$$
$$= (A \wedge \sim B) \vee C$$
$$= (A \vee C) \wedge (\sim B \vee C)$$

Hence, $(A \vee C) \wedge (\sim B \vee C)$ is the CNF of $((A \rightarrow B) \rightarrow C)$.

PART-3

Forward and Backward Chaining.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 3.18. **Elaborate forward and backward chaining.**

AKTU 2016-17, Marks 05

Answer

Forward chaining :

1. Forward chaining is a method of reasoning when using inference rules in artificial intelligence.

2. Forward chaining starts with the available data and uses inference rules to extract more data (from an end user) until an optimal goal is reached.
3. An inference engine using forward chaining searches the inference rules until it finds one where the If clause is known to be true.
4. When found it can conclude, or infer, the Then clause, resulting in the addition of new information to its dataset.
5. Inference engines will often cycle through this process until an optimal goal is reached.
6. For example, suppose that the goal is to conclude the colour of my pet Bruno given that he croaks and eats flies, and that the rule base contains the following two rules :
   a. If X croaks and eats flies - Then X is a frog.
   b. If X is a frog - Then X is red.

Backward chaining :

1. Backward chaining starts with a list of goals (or a hypothesis) and works backwards to see if there is data available that will support any of these goals.
2. An inference engine use backward chaining would search the inference rules until it finds one which has a Then clause that matches a desired goal.
3. If the If clause of that inference rule is not known to be true, then it is added to the list of goals.
4. For example, suppose that the goal is to conclude the colour of my pet Bruno given that he croaks and eats flies, and that the rulebase contains the following two rules :
   a. If X croaks and eats flies - Then X is a frog.
   b. If X is a frog - Then X is red.

Que 3.19. Write down the properties of forward chaining and backward chaining.

Answer

Properties of forward chaining :

1. It is a down-up approach, as it moves from bottom to top.
2. It is a process of making a conclusion based on known facts or data, by starting from the initial state and reaches the goal state.
3. Forward chaining approach is also called as data-driven as we reach to the goal using available data.
4. Forward chaining approach is commonly used in the expert system, such as business, and production rule systems.

Properties of backward chaining :

1. It is known as a top-down approach.
2. Backward chaining is based on Modus Ponens inference rule.
3. In backward chaining, the goal is broken into sub-goal or sub-goals to prove the facts true.
4. It is called a goal-driven approach, as a list of goals decides which rules are selected and used.
5. Backward chaining algorithm is used in game theory, automated theorem proving tools, inference engines, proof assistants, and various AI applications.
6. The backward chaining method mostly used a depth-first search strategy for proof.

Que 3.20. Differentiate between forward and backward chaining of inference with the help of an example.

AKTU 2018-19, Marks 10

Answer

S. No.	Forward chaining	Backward chaining
1.	Forward chaining is a data driven method.	Backward chaining is a goal driven method.
2.	It uses planning, monitoring and controlling method.	It uses diagnosis method.
3.	It uses bottom-up processing.	It uses top-down processing.
4.	Forward chaining finds possible conclusions supported by given facts.	Backward chaining finds facts that support a given hypothesis.
5.	Forward chaining is similar to breadth-first search.	Backward chaining is similar to depth-first search.
6.	For example : CLIPS.	For example : PROLOG.

PART-4

Resolution.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 3.21. **Write a short note on resolution.**

OR

What is resolution ? Discuss the role of resolution in predicate logic. **AKTU 2016-17, Marks 10**

Answer

1. Resolution is a proof, procedure that carries out a single operation, the variety or processes involved in reasoning with statements in predicate logic.
2. Resolution operates on statements that have not been converted into a very convenient form.
3. Resolution procedure is proved by reputation, in other words, to prove a statement resolution attempts to show that negation of statement produces a contradiction with the known fact, *i.e.*, it is unsatisfiable.

Role of resolution in predicate logic :

1. Resolution is an inference step required to build a complete inference system for predicate logic in clause form.
2. Applying the rule of resolution to the clauses makes the resolution inference highly suitable for computer implementation.

Que 3.22. **Discuss resolution in propositional and predicate logic.**

Answer

Resolution of propositional logic : In resolution of propositional logic, the procedure for producing a proof by resolution of propositional S with respect to a set of axioms F is as follows :

1. Convert all the propositions of F to clause form.
2. Negate S and convert the result to clause form. Add it to the set of clauses obtained in step 1.
3. Repeat until either a contradiction is found or no progress can be made :
   i. Select two clauses, call these parent clauses.
   ii. Resolve them together. The resulting clause called the resolvent will be the disjunction of all of the literals of both of the parent clauses with the following exception : If there are any pairs of literals L and $\sim L$, such that one of the parent clause contains L and other $\sim L$, then select one such pair and eliminate both L and $\sim L$ from the resolvent.
   iii. If the resolvent is the empty clause, then a contradiction has been found. If it is not, then add it to the set of clauses available to the procedure.

Resolution in predicate logic :

1. Convert all the statements of F to clause form.
2. Negate S and convert the result to clause form. Add it to the set of clauses obtained in step 1.
3. Repeat until either a contradiction is found, no progress can be made or a predetermined amount of effort has been expended :
   i. Select two clauses, call these parent clauses.
   ii. Resolve them together. The resolvent will be disjunction of all literals of both of parent clause with appropriate substitutions performed and with the following exception : If there is one pair of literals T_1 and $\sim T_2$ such that one of the parent clauses contains T_1 and the other contains T_2 and if T_1 and T_2 are unifiable, then neither T_1 nor T_2 should appear in the resolvent.
   iii. If the resolvent is the empty clause, then a contradiction has been found. If it is not, then add it to the set of clauses available to the procedure.

PART-5

Probablistic Reasoning, Utility Theory.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 3.23. What is reasoning ? Explain its types.

Answer

1. Reasoning is the mental process of looking for reasons for beliefs, conclusions, actions or feelings.
2. Reasoning means infer facts from existing data.
3. Types of reasoning are :
   a. Deductive reasoning :
      i. Deductive arguments are intended to have reasoning that is valid.
      ii. Reasoning is an argument that is valid if the argument's conclusion must be true when the premises (the reasons given to support that conclusion) are true.
   b. Inductive reasoning :
      i. Inductive reasoning contrasts strongly with deductive reasoning.

ii. Even in the best, or strongest cases of inductive reasoning, the truth of the premises does not guarantee the truth of the conclusion.

iii. Instead, the conclusion of an inductive argument follows with some degree of probability.

iv. The conclusion of an inductive argument contains more information than is already contained in the premises. Thus, this method of reasoning is ampliative.

c. **Abductive reasoning :** Abductive reasoning, or argument to the best explanation often involves both inductive and deductive arguments.

Que 3.24. **Write a short note on probabilistic reasoning.**

OR

What is probabilistic reasoning ? Also, describe the role HMM in probabilistic reasoning. **AKTU 2014-15, Marks 10**

Answer

Probabilistic reasoning is a method of representation of knowledge where the concept of probability is applied to indicate the uncertainty in knowledge.

Probabilistic reasoning is used in AI :

1. When we are unsure of the predicates.
2. When the possibilities of predicates become too large to list down.
3. When it is known that an error occurs during an experiment.

Role of HMM in probabilistic reasoning :

1. Probabilistic reasoning with Bayesian networks is use to calculate the conditional probability of a variable given the values of a set of other variable (the evidence).
2. If there is no evidence, it is the marginal probabilities that need to be computed.
3. These networks are very useful for modeling joint distributions.
4. But they have their limitations *i.e.*, they cannot account for temporal/ sequence models.
5. HMM models a set observation with a set of hidden states *i.e.*, hidden states generate observations.
6. Thus, HMM helps in determining the probabilistic nature of observation by using the unobservable *i.e.*, hidden states.
7. For example, robot movement :

   Observations : range sensor, visual sensor

   Hidden states : location (on a map)

Que 3.25. Describe utility theory.

Answer

1. Utility theory is the discipline that lays the foundation to create and evaluate utility functions.
2. Utility theory uses the notion of Expected Utility (EU) as a value that represents the average utility of all possible outcomes of a state, weighted by the probability that the outcome occurs.
3. The key concept of utility theory is known as the Principle of Maximum Expected Utility (MEU) which states that any rational agent should choose to maximize the agent's EU.
4. Utility theory is often combined with probabilistic theory to create what we know as decision-theoretic agents.
5. Utility functions are a product of utility theory which is one of the disciplines that helps to address the challenges of building knowledge under uncertainty.

Que 3.26. Explain unification algorithm used for reasoning under predicate logic with an example. **AKTU 2016-17, Marks 10**

Answer

1. In order to do resolution for expressions in the predicate logic, we use unification algorithm to locate pairs of literals that cancel out.
2. We need to use the unifier produced by the unification algorithm to generate the resolvent clause.

For example, we want to resolve following two clauses :

a. man (Marcus)

b. $\neg$ Man (x) $\vee$ mortal (x)

3. The literal man (Marcus) can be unified with the literal man (x) with the substitution Marcus / x, telling us that for x = Marcus, $\neg$ man (Marcus) is false.
4. Now we cannot simply cancel out the two man literals as in case of propositional logic and generate logic we now conclude only that mortal (Marcus) must be true which we get by applying the result of the unification process to the resolvent.
5. The resolution process can then proceed to discover whether mortal (Marcus) leads to a contradiction with other available clauses.

Unification algorithm : Verify (U, V)

1. If U and V are both variables or constants then

a. If U and V are identical, then return null.

b. If U is a variable then if U occurs in V, then return {FAIL} else return $\{U/V\}$

c. If V is a variable then if V occurs in U, return {FAIL} else return $\{U/V\}$.

d. Return {FAIL}.

2. If the initial predicate symbols in U and V are not identical, return {FAIL}.

3. If U and V have a different number of arguments, then return {FAIL}.

4. Set SUBSET to NULL.

5. For $i \leftarrow l$ to the number of arguments of U.

   a. Call unify with the i^{th} argument of U and the i^{th} argument of V putting the result in S.

   b. If S contain {FAIL}, return {FAIL}.

   c. If S is not equal to NULL

      i. Apply step 5 to be the remainder of both U and V.

      ii. Set subset equal to APPENDS (S, SUBSET).

6. Return SUBSET.

PART-6

Hidden Markov Models (HMM).

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 3.27. Define Hidden Markov Model (HMM).

OR

Define Hidden Markov Model (HMM). Illustrate how HMMs are used for speech recognition. **AKTU 2017-18, Marks 10**

OR

Explain the use of Hidden Markov models in speech recognition. **AKTU 2016-17, Marks 05**

Answer

Hidden Markov Model (HMM) :

1. Hidden Markov Model (HMM) is a statistical Markov model in which the system being modeled is assumed to be a Markov process with unobserved (*i.e.*, hidden) states.

Use of Hidden Markov Model in speech recognition :

1. The Hidden Markov model is a statistical model that uses a finite number of states and the associated state transitions to jointly model the temporal and spectral variations of signals.
2. It has been used extensively to model fundamental speech units in speech recognition because the HMM can adequately characterise both the temporal and spectral varying nature of the speech signal.
3. The use of Hidden Markov models to model speech units, the other key contribution of speech research is the use of data structures for optimally decoding speech into text.

Que 3.28. Distinguish between Markov Model and Hidden Markov Model (HMM).

AKTU 2015-16, Marks 10

Answer

S. No.	Markov models	Hidden Markov models (HMM)
1.	Markov models are a mathematical tool. It's the theory about particular family of functions and its various properties.	HMMs are a special way of using Markov models. It's the application.
2.	Markov models defines how to compute the function, it's not machine learning.	HMM defines about how to learn the weights in the Markov matrix.
3.	Markov model is a state machine with the state changes being probabilities.	In a Hidden Markov model, you don't know the probabilities, but you know the outcomes.
4.	States are observed directly in Markov model.	There are hidden states in HMM.
5.	Example : Language modeling.	Example : Speech Recognition.

Que 3.29. **Draw diagram of HMM and show what is the hidden part of it that we refer to ?** **AKTU 2015-16, Marks 10**

Answer

1. Hidden Markov Model (HMM) is a statistical Markov model in which the system being modeled is assumed to be a Markov process with unobserved (*i.e.*, hidden) states.
2. The term hidden refers to the first order Markov process behind the observation. Observation refers to the data we know and can observe.
3. Markov process is shown by the interaction between "Rainy" and "Sunny" in the Fig. 3.29.1 and each of these are hidden states.
4. Observations are known data and refer to "Walk", "Shop", and "Clean" in the Fig. 3.29.1.

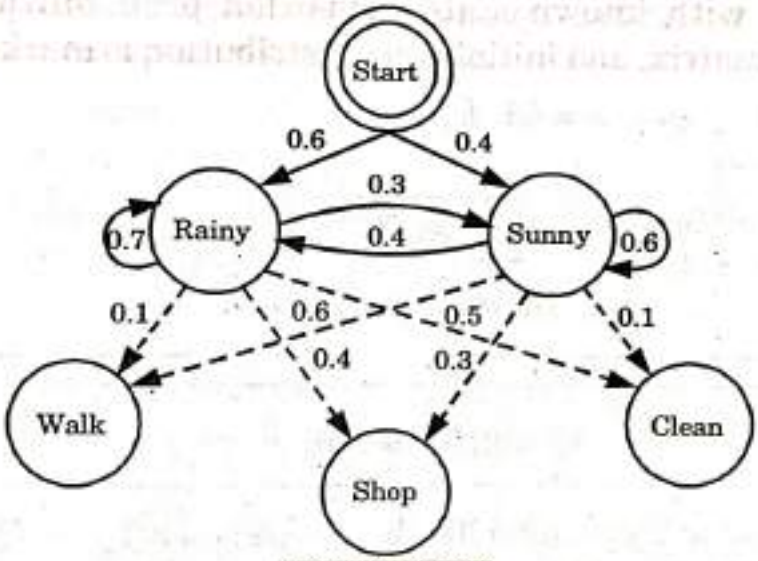


Fig. 3.29.1.

5. In machine learning sense, observation is our training data, and the number of hidden states is our hyper parameter for our model.

T = length of the observation sequence

N = number of states in the model

M = number of observation symbols

$Q = \{q_0, q_1, \ldots, q_{N-1}\}$ = distinct states of the Markov process

$V = \{0, 1, \ldots, M-1\}$ = set of possible observations

A = state transition probabilities

B = observation probability matrix

π = initial state distribution

$O = (O_0, O_1, \ldots, O_{T-1})$ = observation sequence

T = don't have any observation yet, $N = 2$, $M = 3$, Q = ("Rainy", "Sunny"), V = ("Walk", "Shop", "Clean")

6. State transition probabilities are the arrows pointing to each hidden state.
7. Observation probability matrix are the dotted arrows pointing to each observations from each hidden state.
8. The matrix are row stochastic meaning the rows add up to 1.

$$a_{ij} = P(\text{state } q_j \text{ at } t+1 \mid \text{state } q_i \text{ at } t)$$

$$b_j(k) = P(\text{observation } k \text{ at } t \mid \text{state } q_j \text{ at } t)$$

9. The matrix explains what the probability is from going to one state to another, or going from one state to an observation.
10. Initial state distribution gets the model going by starting at a hidden state.
11. Full model with known state transition probabilities, observation probability matrix, and initial state distribution is marked as,

$$\lambda = (A, B, \pi)$$

PART-7

Bayesian Network.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 3.30. Write a short note on Bayesian network.

OR

Explain Bayesian network by taking an example. How is the Bayesian network powerful representation for uncertainty knowledge ?

AKTU 2015-16, Marks 10

Answer

1. A Bayesian network is a directed acyclic graph in which each node is annotated with quantitative probability information.
2. The full specification is as follows :
   i. A set of random variables makes up the nodes of the network variables may be discrete or continuous.
   ii. A set of directed links or arrows connects pairs of nodes. If there is an arrow from x to node y, x is said to be a parent of y.

iii. Each node x_i has a conditional probability distribution $P(x_i \mid \text{parent}(x_i))$ that quantifies the effect of parents on the node.

iv. The graph has no directed cycles (and hence is a directed acyclic graph or DAG).

3. A Bayesian network provides a complete description of the domain. Every entry in the full joint probability distribution can be calculated from the information in the network.

4. Bayesian networks provide a concise way to represent conditional independence relationships in the domain.

5. A Bayesian network is often exponentially smaller than the full joint distribution.

For example :

1. Suppose we want to determine the possibility of grass getting wet or dry due to the occurrence of different seasons.

2. The weather has three states : Sunny, Cloudy, and Rainy. There are two possibilities for the grass : Wet or Dry.

3. The sprinkler can be on or off. If it is rainy, the grass gets wet but if it is sunny, we can make grass wet by pouring water from a sprinkler.

4. Suppose that the grass is wet. This could be contributed by one of the two reasons - Firstly, it is raining. Secondly, the sprinklers are turned on.

5. Using the Baye's rule, we can deduce the most contributing factor towards the wet grass.

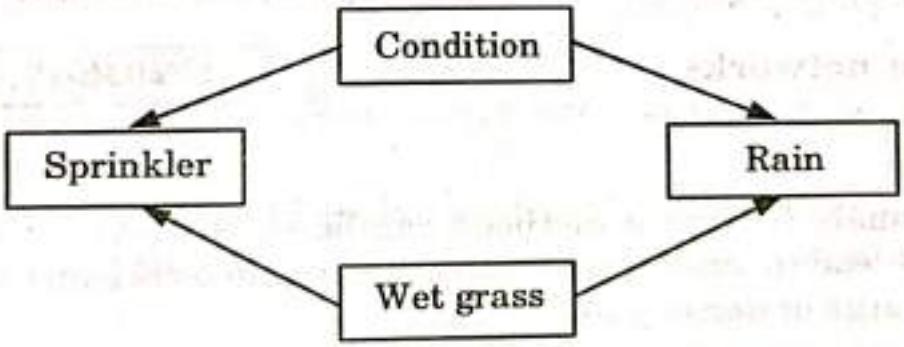


Fig. 3.30.1.

Bayesian network possesses the following merits in uncertainty knowledge representation :

1. Bayesian network can conveniently handle incomplete data.

2. Bayesian network can learn the casual relation of variables. In data analysis, casual relation is helpful for field knowledge understanding, it can also easily lead to precise prediction even under much interference.

3. The combination of bayesian network and bayesian statistics can take full advantage of field knowledge and information from data.

4. The combination of bayesian network and other models can effectively avoid over-fitting problem.

Que 3.31. **What is Bayesian theory ? Explain the role of prior probability and posterior probability in bayesian classification.**

AKTU 2018-19, Marks 10

Answer

Bayesian theory : Refer Q. 3.30, Page 3–25A, Unit-3.

Role of prior probability :

1. The prior probability is used to compute the probability of the event before the collection of new data.
2. It is used to capture our assumptions / domain knowledge and is independent of the data.
3. It is the unconditional probability that is assigned before any relevant evidence is taken into account.

Role of posterior probability :

1. Posterior probability is used to compute the probability of an event after collection of data.
2. It is used to capture both the assumptions / domain knowledge and the pattern in observed data.
3. It is the conditional probability that is assigned after the relevant evidence or background is taken into account.

Que 3.32. **Explain the method of handling approximate inference in Bayesian networks.**

AKTU 2016-17, Marks 10

Answer

1. Approximate inference methods can be used when exact inference methods lead to unacceptable computation times because the network is very large or densely connected.
2. Methods handling approximate inference :
   i. **Simulation methods :** This method use the network to generate samples from the conditional probability distribution $\Pr(V'' \mid e)$ and estimate conditional probabilities of interest when the number of samples is sufficiently large.
   ii. **Variational methods :** This method express the inference task as a numerical optimization problem and then find upper and lower bounds of the probabilities of interest by solving a simplified version of this optimization problem.

Que 3.33. What are planning graphs ? Explain the methods of planning and acting in the real world. **AKTU 2016-17, Marks 10**

Answer

1. Planning graph is a data structure that encodes the information contained in the states regarding their reachability.
2. The basic idea is to construct smaller size graphs that would contain partial information about the reachability.
3. A planning graph is a layered graph comprising states and actions that form a layer and appear alternately. So, it can be represented as follows :

(State 1, Action 1, State 2, Action 2, ..., State n, Action n)

For each operator o that belongs to action, there is an edge from literal in state x. This literal is the pre-condition for operator o.

Methods of planning and acting in the real world :

1. Hierarchical task network (HTN) planning allows the agent to take advice from the domain designer in the form of decomposition rules. This makes it feasible to create the very large plans required by many real-world applications.
2. Conditional plans allow the agent to sense the world during execution to decide what branch of the plan to follow.
3. Incorrect information results in unsatisfied preconditions for actions and plans. Execution monitoring detects violations of the preconditions for successful completion of the plan.
4. A continuous planning agent creates new goals as it goes and reacts in real time.
5. Multiagent planning is necessary when there are other agents in the environment with which to cooperate, compete, or coordinate.

VERY IMPORTANT QUESTIONS

Following questions are very important. These questions may be asked in your SESSIONALS as well as UNIVERSITY EXAMINATION.

Q. 1. Translate the following sentences into formulas in predicate logic and clausal form :

i. John likes all kind of food.

ii. Apples are food.

iii. Chicken is food.

iv. **Anything any one eats and is not killed by is food.**
v. **Bill eats peanuts and is still alive.**
vi. **Sue eats everything Bill eats.**

Ans: Refer Q. 3.11.

Q. 2. What is probabilistic reasoning ? Also, describe the role HMM in probabilistic reasoning.

Ans: Refer Q. 3.24.

Q. 3. Distinguish between Markov Model and Hidden Markov Model (HMM).

Ans: Refer Q. 3.28.

Q. 4. Define Hidden Markov Model (HMM). Illustrate how HMMs are used for speech recognition.

Ans: Refer Q. 3.27.

Q. 5. What are the desirable properties of good knowledge representation schemes ?

Ans: Refer Q. 3.3.

Q. 6. Explain Bayesian network by taking an example. How is the Bayesian network powerful representation for uncertainty knowledge ?

Ans: Refer Q. 3.30.

Q. 7. Prove that following sentence is valid :
"If prices fall then sell increases. If sell increases then John makes the whole money. But John doesn't make the whole money. Therefore, prices do not fall."

Ans: Refer Q. 3.8.

4

UNIT

Machine Learning

CONTENTS

PART-1

Machine Learning : Supervised and Unsupervised Learning.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 4.1. Describe machine learning.

Answer

1. Machine learning is primarily concerned with the design and development of algorithms that allow the system to learn from historical data.
2. Machine learning is based on the idea that machines can learn from past data, identify patterns, and make decisions using algorithms.
3. Machine learning algorithms are designed in such a way that they can learn and improve their performance automatically.
4. Machine learning helps in discovering patterns in data.
5. Machine learning can be subdivided into three types :
   i. Supervised learning
   ii. Reinforcement learning
   iii. Unsupervised learning

Que 4.2. Differentiate between artificial intelligence and machine learning.

Answer

S. No.	Artificial intelligence	Machine learning
1.	Artificial Intelligence (AI) is an area of computer science that emphasizes the creation of intelligent machines that work and reacts like humans.	Machine learning is based on the idea that machines can learn from past data, identify patterns, and make decisions using algorithms.
2.	The aim is to increase chance of success and not accuracy.	The aim is to increase accuracy, but it does not care about success.

3.	It works as a computer program that does smart work.	It is a simple concept machine that takes data and learns from data.
4.	The goal is to simulate natural intelligence to solve complex problem.	The goal is to learn from data on certain task to maximize the performance of machine on this task.
5.	Artificial intelligence is decision making.	Machine learning allows system to learn new things from data.

Que 4.3. Describe supervised learning and unsupervised learning.

Answer

Supervised learning :

1. Supervised learning is also known as associative learning, in which the network is trained by providing it with input and matching output patterns.
2. Supervised training requires the pairing of each input vector with a target vector representing the desired output.
3. The input vector together with the corresponding target vector is called training pair.
4. To solve a problem of supervised learning following steps are considered :
   a. Determine the type of training examples.
   b. Gathering of a training set.
   c. Determine the input feature representation of the learned function.
   d. Determine the structure of the learned function and corresponding learning algorithm.
   e. Complete the design.
5. Supervised learning can be classified into two categories :
   i. Classification ii. Regression

Unsupervised learning :

1. Unsupervised learning, an output unit is trained to respond to clusters of pattern within the input.
3. In this method of training, the input vectors of similar type are grouped without the use of training data to specify how a typical member of each group looks or to which group a member belongs.
3. Unsupervised training does not require a teacher, it requires certain guidelines to form groups.
4. Unsupervised learning can be classified into two categories :

i. Clustering ii. Association

Que 4.4. Differentiate between supervised learning and unsupervised learning.

Answer

Difference between supervised and unsupervised learning:

S. No.	Supervised learning	Unsupervised learning
1.	It uses known and labeled data as input.	It uses unknown data as input.
2.	Computational complexity is very complex.	Computational complexity is less.
3.	It uses offline analysis.	It uses real time analysis of data.
4.	Number of classes are known.	Number of classes are not known.
5.	Accurate and reliable results.	Moderate accurate and reliable results.

Que 4.5. Consider the problem of learning to play tennis. Are there aspects of this learning that are supervised learning? Is this supervised learning or reinforcement learning?

AKTU 2016-17, Marks 05

Answer

1. **The requisite skills can be divided into movement, playing strokes, and strategy.**
2. **The environment consists of the court, ball, opponent, and one's own body.**
3. **The relevant sensors are the visual system and proprioception.**
4. **The effectors are the muscles involved in moving to the ball and hitting the stroke.**
5. **The learning process involves both supervised learning and reinforcement learning.**
6. Supervised learning occurs in acquiring the predictive transition models, for example, where the opponent will hit the ball, where the ball will land, and what trajectory the ball will have after one's own stroke.

7. Reinforcement learning occurs when points are won and lost, this is particularly important for strategic aspects of play such as shot placement and positioning.
8. In the early stages, reinforcement also occurs when a shot succeeds in clearing the net and landing in the opponent's court.

Que 4.6. **What is regression ? Compare between linear regression and non-linear regression.** **AKTU 2018-19, Marks 10**

Answer

Regression :

1. Regression is defined as the parametric technique that allows us to make decisions / predictions based upon data by learning the relationship between input and output variables.
2. In regression, the relationship between input and output variables helps us in understanding how the value of the output variable changes with the change of input variable.
3. Regression is one of the most important statistical and machine learning tools.
4. Regression is frequently used for prediction of prices, economics, variations, and so on.

Comparison :

S. No.	Linear regression	Non-linear regression
1.	Linear regression is a method of finding the relationship between one dependent variable and one or more independent variable.	Non-linear regression is a method of finding the relationship between the dependent variable and set of independent variable.
2.	Overfitting occurs in linear regression.	Overfitting does not occur in non-linear regression.

PART-2

Decision Tree.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 4.7. **Illustrate decision trees technique using a suitable example.** **AKTU 2017-18, Marks 10**

OR

Describe decision tree learning technique.

Answer

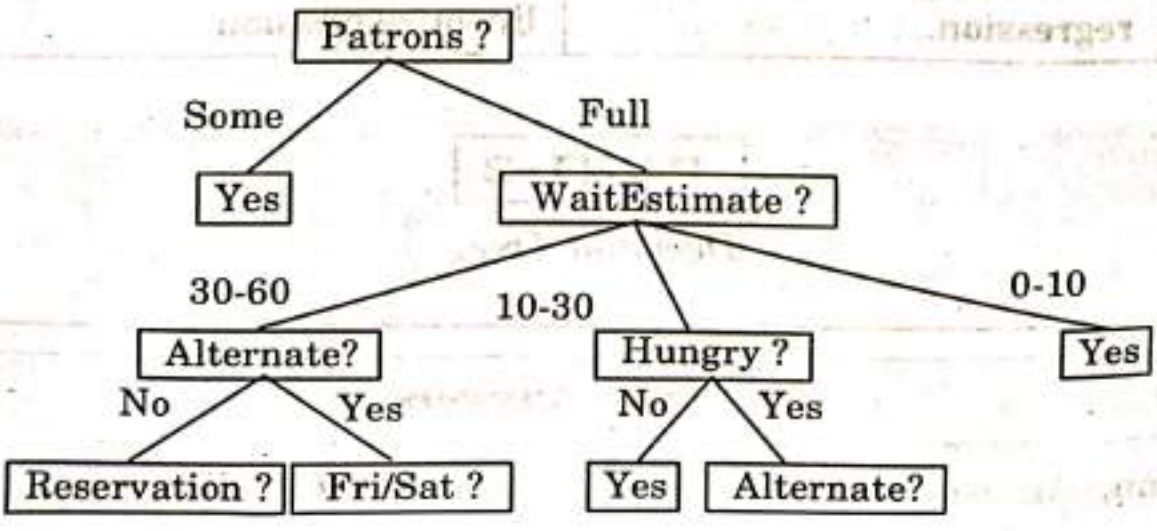
1. A decision tree takes input as an object or situation described by a set of attributes and returns a "decision", the predicted output value for the input. The input attributes and output value can be discrete or continuous.
2. A decision tree reaches its decision by performing a sequence of tests.
3. Each internal node in the tree corresponds to a test of the value of one of the properties and the branches from the node are labeled with the possible values of the test.
4. Each leaf node in the tree specifies the value to be returned if that leaf is reached.

For example : The problem of whether to wait for a table at a restaurant.

i. The aim here is to learn a definition for the goal predicate WillWait.

ii. Let's suppose we decide on the following list of attributes :

1. **Alternate :** whether there is a suitable alternative restaurant nearby.
2. **Fri/Sat :** true on Fridays and Saturdays.
3. **Hungry :** whether we are hungry.
4. **Patrons :** how many people are in the restaurant (values are Some and Full).
5. **Reservation :** whether we made a reservation.
6. **Type :** the kind of restaurant (French, Italian, Thai).
7. **WaitEstimate :** the wait estimated by the host (0-10, 10-30, 30-60 minutes).



Que 4.8. **Explain attribute selection.**

OR

What do you understand by information gain ? How it is calculated ?

AKTU 2018-19, Marks 10

Answer

In decision tree, the major challenge is the identification of the attribute for the root node in each level. This process is known as attribute selection. We have two attribute selection measures :

1. **Information gain :**
   i. Information gain is defined as the reduction in entropy. The entropy of a set of training data, S, is defined as

$$H(S) = -p_1 \log_2 p_1 - p_0 \log_2 p_0$$

   where, p_1 is defined as the proportion of the training data that includes positive examples, and p_0 is defined as the proportion that includes negative examples.
   ii. The entropy of S is zero when all the examples are positive, or when all the examples are negative.
   iii. The entropy reaches its maximum value of 1 when exactly half of the examples are positive and half are negative.
   iv. The information gain of a particular feature tells us how closely that feature represents the entire target function, and so at each stage, the feature that gives the highest information gain is chosen to turn into a question.
2. **Gini index :**
   i. Gini index is a metric to measure how often a randomly chosen element would be incorrectly identified.
   ii. It means an attribute with lower Gini index should be preferred.
   iii. The formula for the calculation of the of the gini index is :

$$\text{Gini index} = 1 - \sum_j p_j^2$$

PART-3

Statistical Learning Models, Learning with Complete Data.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 4.9. Explain the following terms :

i. Maximum a posteriori (MAP)

ii. Maximum likelihood hypothesis **AKTU 2014-15, Marks 10**

Answer

i. Maximum a posteriori (MAP) :

1. To make predictions based on a single most probable hypothesis, that is, an h_i that maximizes $P(h_i | d)$ is called a maximum a posteriori.
2. Predictions made according to an MAP hypothesis h_{MAP} are approximately bayesian to the extent that $P(x | d) \approx P(x | h_{MAP})$.
3. In MAP learning, the hypothesis prior $P(h_i)$ plays an important role. Bayesian and MAP learning methods use the prior to penalize complexity.
4. Typically, more complex hypothesis has a lower prior probability because there are usually many more complex hypothesis than simple hypothesis. Though, they have a greater capacity to fit the data.
5. Hence, the hypothesis prior embodies a trade-off between the complexity of a hypothesis and its degree to fit to the data.
6. In the case of trade-off, $P(d | h_i)$ is 1, if h_i is consistent and 0 otherwise.
7. Let D represent all the data, with observed value d; then the probability of each hypothesis is obtained by Bayes rule :

$$P(h_i | d) = \alpha P(d | h_i) P(h_i)$$

here h_{MAP} will be the simplest logical theory that is consistent with the data.

8. Hence, MAP learning is choosing the hypothesis that provides maximum compression of the data.

ii. Maximum likelihood hypothesis :

1. Maximum likelihood method views the parameters as quantities whose values are fixed but unknown.
2. The best estimate of their value is defined to be the one that maximizes the probability of obtaining the samples actually observed.
3. Let us consider an M-class problem with feature vectors distributed according to $p(x|\omega_i)$, $i = 1, 2, ..., M$.
4. Assume that these likelihood functions are given in a parametric form and the corresponding parameters form the vectors θ_i which are unknown *i.e.*, $p(x|\omega_i)$ has a known parametric form and is uniquely determined by the value of parameter θ_i. To show the dependence on θ_i, it is written as $p(x|\omega_i ; \theta_i)$.

5. If the data from one class do not affect the parameter estimation of the others, then the problem can be formulated independent of classes.
6. Thus the problem can be solved for each class independently. Let x_1, x_2, x_3, ..., x_N be random samples drawn from probability density functions $p(x\ ;\ \theta)$. We form the joint probability density function $p(X\ ;\ \theta)$, where $X = \{x_1, x_2, ..., x_N\}$ is the set of the samples.
7. Assuming statistical independence between the different samples, we have

$$p(X;\theta) \equiv p(x_1, x_2, ..., x_N;\theta) = \prod_{k=1}^{N} p(x_k;\theta)$$

8. It is a function of θ and is also known as the likelihood function of θ with respect to X.
9. The maximum likelihood (ML) method estimates θ so that the likelihood function takes its maximum value, *i.e.*,

$$\hat{\theta}_{ML} = arg \max_{\theta} \prod_{k=1}^{N} p(x_k;\ \theta)\ .$$

10. The gradient of the likelihood function with respect to θ should be zero to satisfy the condition that $\hat{\theta}_{ML}$ is maximum, *i.e.*,

$$\frac{\partial \prod_{k=1}^{n} p(x_k\ ;\ \theta)}{\partial \theta}\ .$$

Que 4.10. Describe learning with complete data – discrete models, Naive Bayes model, continuous model.

OR

Write short notes on :

i. Discrete model maximum – likelihood parameter learning.

ii. Continuous model.

AKTU 2017-18, Marks 10

Answer

Discrete model :

1. Suppose we buy a bag of lime and cherry candy from a manufacturer whose lime-cherry proportions are completely unknown, that is, the fraction could be anywhere between 0 and 1.
2. In that case, we have a continuum of hypothesis.
3. The parameter in this case, which we call θ, is the proportion of cherry candies, and the hypothesis is h_θ (the proportion of limes is just $1 - \theta$).

4. If we assume that all proportions are equally likely a priori, then a maximum-likelihood approach is reasonable.
5. If we model the situation with a bayesian network, we need just one random variable, flavor (the flavor of a randomly chosen candy from the bag).
6. It has values cherry and lime, where the probability of cherry is θ. Suppose we unwrap N candies, of which c are cherries and $l = N - c$ are limes.
7. According to equation,

$$P(d \mid h_i) = \prod_j P(d_j \mid h_i)$$

8. The likelihood of the particular data set is,

$$P(d \mid h_\theta) = \prod_{j=1}^{N} P(d_j \mid h_\theta) = \theta^c \cdot (1-\theta)^l$$

9. The maximum-likelihood hypothesis is given by the value of θ that maximizes this expression. The same value is obtained by maximizing the log likelihood,

$$L(d \mid h_\theta) = \log P(d \mid h_\theta)$$

$$= \sum_{j=1}^{N} \log P(d_j \mid h_\theta) = c \log \theta + l \log(1-\theta)$$

10. By taking logarithms, we reduce the product to a sum over the data, which is usually easier to maximize. To find the maximum-likelihood value of θ, we differentiate L with respect to θ and set the resulting expression to zero :

$$\frac{dL(d \mid h_\theta)}{d\theta} = \frac{c}{\theta} - \frac{l}{1-\theta} = 0$$

$$\Rightarrow \quad \theta = \frac{c}{c+l} = \frac{c}{N}$$

Naive Bayes model :

1. Naive Bayes model is the most common bayesian network model used in machine learning.
2. Here the class variable C is the root which is to be predicted and the attribute variables X_i are the leaves.
3. The model is naive because it assumes that the attributes are conditionally independent of each other, given the class.
4. Assuming boolean variables, the parameters are :

$$\theta = P(C = \text{true}), \theta_{i1} = P(X_i = \text{true} \mid C = \text{true}),$$
$$\theta_{i2} = P(X_i = \text{true} \mid C = \text{false})$$

5. Naive Bayes models can be viewed as bayesian networks in which each X_i has C as the sole parent and C has no parents.
6. A Naive Bayes model with gaussian $P(X_i \mid C)$ is equivalent to a mixture of gaussians with diagonal covariance matrices.

7. The use of mixtures of gaussians in Naive Bayes models allow for very efficient inference of marginal and conditional distributions.
8. Naive Bayes learning has no difficulty with noisy data and can give more appropriate probabilistic predictions.

Continuous model :

1. Let us consider, learning the parameters of a gaussian density function on a single variable. That is, the data are generated as follows :

$$P(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

2. The parameters of this model are the mean μ and the standard deviation σ. Let the observed values be $x_1, \ldots\ldots, x_N$. Then the log likelihood is,

$$L = \sum_{j=1}^{N} \log \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{(x_j-\mu)^2}{2\sigma^2}}$$

$$= N(-\log\sqrt{2\pi} - \log\sigma) - \sum_{j=1}^{N} \frac{(x_j-\mu)^2}{2\sigma^2}$$

3. Setting the derivatives to zero, we obtain,

$$\frac{\partial L}{\partial \mu} = -\frac{1}{\sigma^2}\sum_{j=1}^{N}(x_j-\mu) = 0 \Rightarrow \mu = \frac{\sum_j x_j}{N}$$

$$\frac{\partial L}{\partial \sigma} = -\frac{N}{\sigma} + \frac{1}{\sigma^3}\sum_{j=1}^{N}(x_j-\mu)^2 = 0 \Rightarrow \sigma = \sqrt{\frac{\sum_j (x_j-\mu)^2}{N}}$$

4. The maximum-likelihood value of the mean is the sample average and the maximum-likelihood value of the standard deviation is the square root of the sample variance.
5. Now consider a linear gaussian model with one continuous parent X and a continuous child Y.
6. Y has a gaussian distribution whose mean depends linearly on the value of X and whose standard deviation is fixed.
7. To learn the conditional distribution $P(Y|X)$, we can maximize the conditional likelihood,

$$P(y|x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{(y-(\theta_1 x+\theta_2))^2}{2\sigma^2}} \quad \ldots(4.10.1)$$

8. Here, the parameters are θ_1, θ_2, and σ. The data is a collection of (x_j, y_j) pairs.
9. If we consider just the parameters θ_1 and θ_2 that define the linear relationship between x and y, it becomes clear that maximizing the log

likelihood with respect to these parameters is the same as minimizing the numerator in the exponent of equation (4.10.1).

$$E = \sum_{j=1}^{N} (y_j - (\theta_1 x_j + \theta_2))^2$$

10. The quantity $(y_j - (\theta_1 x_j + \theta_2))$ is the error for (x_j, y_j) that is, the difference between the actual value y_j and the predicted value $(\theta_1 x_j + \theta_2)$, so E, is well-known sum of squared errors.
11. This is the quantity that is minimized by the standard linear regression procedure.
12. That's why minimizing the sum of the squared errors gives the maximum-likelihood straight-line model, provided that the data are generated with gaussian noise of fixed variance.

PART-4

Learning with Hidden Data – EM Algorithm.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 4.11. Describe learning with hidden data – EM algorithm.

OR

Write a short note on EM algorithm. **AKTU 2015-16, Marks 05**

OR

Explain learning with complete data – Navie Bayes models and learning with hidden data – EM algorithm.

AKTU 2014-15, Marks 10

Answer

Learning with complete data – Navie Bayes models : Refer Q. 4.10, Page 4–9A, Unit-4.

Learning with hidden data – EM algorithm :

1. The EM algorithm is an efficient iterative procedure to compute the maximum likelihood (ML) estimate in the presence of missing or hidden data.
2. In ML estimation, we wish to estimate the model parameter for which the observed data are the most likely.

3. Each iteration of the EM algorithm consists of two processes : the E-step, and the M-step.
4. In the expectation or E-step the missing data are estimated using observed data as current estimate of the model parameters.
5. This is achieved using the conditional expectation, explaining the choice of terminology.
6. In the M-step, the likelihood function is maximized under the assumption that the missing data are known.
7. The estimation of the missing data from the E-step is used in lieu of the actual missing data.
8. Convergence is assured since the algorithm is guaranteed to increase the likelihood at each iteration.

Que 4.12. Describe statistical learning model in detail.

AKTU 2017-18, Marks 10

Answer

Statistical learning model are :

1. **Maximum likelihood hypothesis :** Refer Q. 4.9, Page 4–8A, Unit-4.
2. **Maximum a posteriori (MAP) :** Refer Q. 4.9, Page 4–8A, Unit-4.
3. **Learning with complete data – Naive bayes model :** Refer Q. 4.10, Page 4–9A, Unit-4.
4. **Learning with hidden data – EM algorithm :** Refer Q. 4.11, Page 4–12A, Unit-4.
5. **Discrete model :** Refer Q. 4.10, Page 4–9A, Unit-4.
6. **Continuous model :** Refer Q. 4.10, Page 4–9A, Unit-4.

Que 4.13. Define overfitting. Discuss the techniques which are used to reduce the overfitting.

Answer

Overfitting occurs even when the target function is not at all random. It afflicts every kind of learning algorithm not just decision trees.

Techniques to reduce overfitting are :

a. **Decision tree pruning :** Pruning works by preventing recursive splitting on attributes that are not clearly relevant, when the data at that node in the tree are not uniformly classified.

b. **Cross validation :** It can be applied to any learning algorithm, not just decision tree learning. The basic idea is to estimate how well each hypothesis will predict unseen data. This is done by setting some fraction

of the known data and using it to test the prediction performance of a hypothesis induced from the remaining data.

PART-5

Reinforcement Learning.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 4.14. What is reinforcement learning ? Differentiate between active and passive reinforcement learning.

AKTU 2018-19, Marks 10

OR

Describe reinforcement learning (RL). Write its applications.

Answer

Reinforcement learning :

1. Reinforcement learning is the study of how animals and artificial systems can learn to optimize their behaviour in the face of rewards and punishments.
2. The task of reinforcement learning is to use observed rewards to learn an optimal policy for the environment. An optimal policy is a policy that maximizes the expected total reward.
3. Without some feedback about what is good and what is bad, the agent will have no grounds for deciding which move to make.
4. The agents needs to know that something good has happened when it wins and that something bad has happened when it loses.
5. This kind of feedback is called a reward or reinforcement.
6. Reinforcement learning is very valuable in the field of robotics, where the tasks to be performed are frequently complex enough to defy encoding as programs and no training data is available.

Application of reinforcement learning :

1. RL can be used in robotics for industrial automation.
2. RL can be used in machine learning and data processing
3. RL can be used to create training systems that provide custom instruction and materials according to the requirement of students.

Difference :

S. No.	Active reinforcement learning (RL)	Passive reinforcement learning (RL)
1.	Active reinforcement learning does not operate with fixed policies. Instead, the AI agent decides which actions to take on any given stage.	In passive reinforcement learning, the policies for AI agent are fixed.
2.	The goal of active RL is to act and learn an optimal policy.	The goal of passive RL is to execute a fixed policy (sequence of actions).

Que 4.15. Differentiate between reinforcement learning and supervised learning.

Answer

S. No.	Reinforcement learning	Supervised learning
1.	Reinforcement learning is all about making decisions sequentially. We can say that the output depends on the state of the current input and the next input depends on the output of the previous input.	In supervised learning, the decision is made on the initial input or the input given at the start.
2.	In reinforcement learning, decision is dependent. So we give labels to sequences of dependent decisions.	In supervised learning, the decisions are independent of each other so labels are given to each decision.
3.	Example : Chess game.	Example : Object recognition

VERY IMPORTANT QUESTIONS

Following questions are very important. These questions may be asked in your SESSIONALS as well as UNIVERSITY EXAMINATION.

Q. 1. Write short notes on :
i. Discrete model maximum – likelihood parameter learning.
ii. Continuous model.

Ans. Refer Q. 4.10.

Q. 2. Explain learning with complete data – Navie Bayes models and learning with hidden data – EM algorithm.

Ans. Refer Q. 4.11.

Q. 3. Describe statistical learning model in detail.

Ans. Refer Q. 4.12.

Q. 4. Explain the following terms :
i. Maximum a posteriori (MAP)
ii. Maximum likelihood hypothesis

Ans. Refer Q. 4.9.

Q. 5. Describe decision tree learning technique.

Ans. Refer Q. 4.7.

Q. 6. What is reinforcement learning ? Differentiate between active and passive reinforcement learning.

Ans. Refer Q. 4.14.

☺☺☺

5

UNIT

Pattern Recognition

CONTENTS

PART-1

Pattern Recognition : Introduction, Design Principles of Pattern Recognition System.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 5.1. Write a short note on pattern recognition.

Answer

1. Pattern recognition is the process of recognizing patterns by using machine learning algorithm.
2. Pattern recognition can be defined as the classification of data based on knowledge already gained or on statistical information extracted from patterns and/or their representation.
3. In a pattern recognition application, the raw data is processed and converted into a form that is amenable for a machine to use.
4. **Examples :** Speech recognition, speaker identification, Multimedia Document Recognition (MDR), automatic medical diagnosis.
5. **Pattern recognition involves :**
   - i. **Classification :** In classification, an appropriate class label is assigned to a pattern based on an abstraction that is generated using a set of training patterns or domain knowledge. Classification is used in supervised learning.
   - ii. **Clustering :** It generates a partition of the data which helps decision making, the specific decision making activity of interest to us. Clustering is used in an unsupervised learning.

Que 5.2. Write down the applications of pattern recognition.

Answer

Applications of pattern recognition :

1. **Image processing, segmentation and analysis :** Pattern recognition is used to give human recognition intelligence to machine which is required in image processing.

2. **Computer vision :** Pattern recognition is used to extract meaningful features from given image/video samples and is used in computer vision for various applications like biological and biomedical imaging.
3. **Seismic analysis :**
   i. Pattern recognition approach is used for the discovery, imaging and interpretation of temporal patterns in seismic array recordings.
   ii. Statistical pattern recognition is implemented and used in different types of seismic analysis models.
4. **Radar signal classification/analysis :** Pattern recognition and signal processing methods are used in various applications of radar signal classifications like AP mine detection and identification.
5. **Speech recognition :** Speech recognition is used in various algorithms of speech recognition which tries to avoid the problems of using a phoneme level of description and treats larger units such as words as pattern.
6. **Finger print identification :**
   i. The fingerprint recognition technique is a dominant technology in the biometric market.
   ii. A number of recognition methods have been used to perform fingerprint matching out of which pattern recognition approaches is widely used.

Que 5.3. Discuss the classification approach of pattern recognition.

AKTU 2017-18, Marks 10

Answer

Two basic approaches of pattern recognition :

1. **Statistical approach (decision theoretic approach) :**
   i. The statistical approach is based on the use of decision functions to classify objects.
   ii. A decision function maps pattern vectors X into decision regions of D.
   iii. The problem can be stated as follows :
      a. Given objects $o = \{o_1, o_2, \ldots\ldots, o_n\}$, let each o_i have k observable attributes and relations expressible as a vector,
      $$V = (v_1, v_2, \ldots v_k).$$
      b. Determine :
         i. a subset of $m \leq k$ of the v_i, say $X = (x_1, x_2, \ldots\ldots, x_m)$ whose values uniquely characterize the o_i, and

ii. $c \geq 2$ groupings or classifications of the o_i which exhibit high intraclass and low interclass similarities such that a decision function $d(X)$ can be found which partitions D into c disjoint regions.

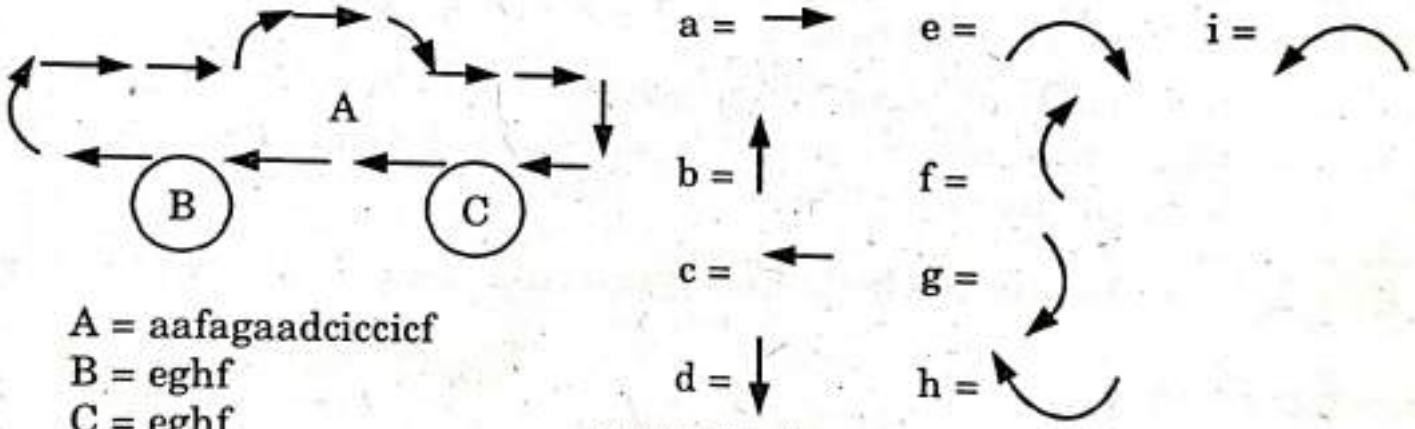
c. The regions are used to classify each o_i as belonging to at most one of the c classes.

2. Syntactic (structured) approach :

i. The syntactic approach is based on the uniqueness of syntactic structure among the object classes.

ii. Instead of defining the grammar in terms of an alphabet of characters or terminal words, the vocabulary is based on shape primitives.

iii. For example, the objects depicted in Fig. 5.3.1, could be defined using the grammar $G(v_n, v_t, p, s)$, where the terminals v_t consists of the following shape primitives.



A = aafagaadciccicf
B = eghf
C = eghf

Fig. 5.3.1.

iv. Using syntactic analysis, that is, parsing and analyzing the string structures, classification is accomplished by assigning an object to class C_i, when the string describing it has been generated by the grammar G_i.

v. This requires that the string be recognized as a member of the language $L(G_i)$.

Que 5.4. Describe the steps in designing pattern recognition system.

OR

What are the components of pattern recognition ?

Answer

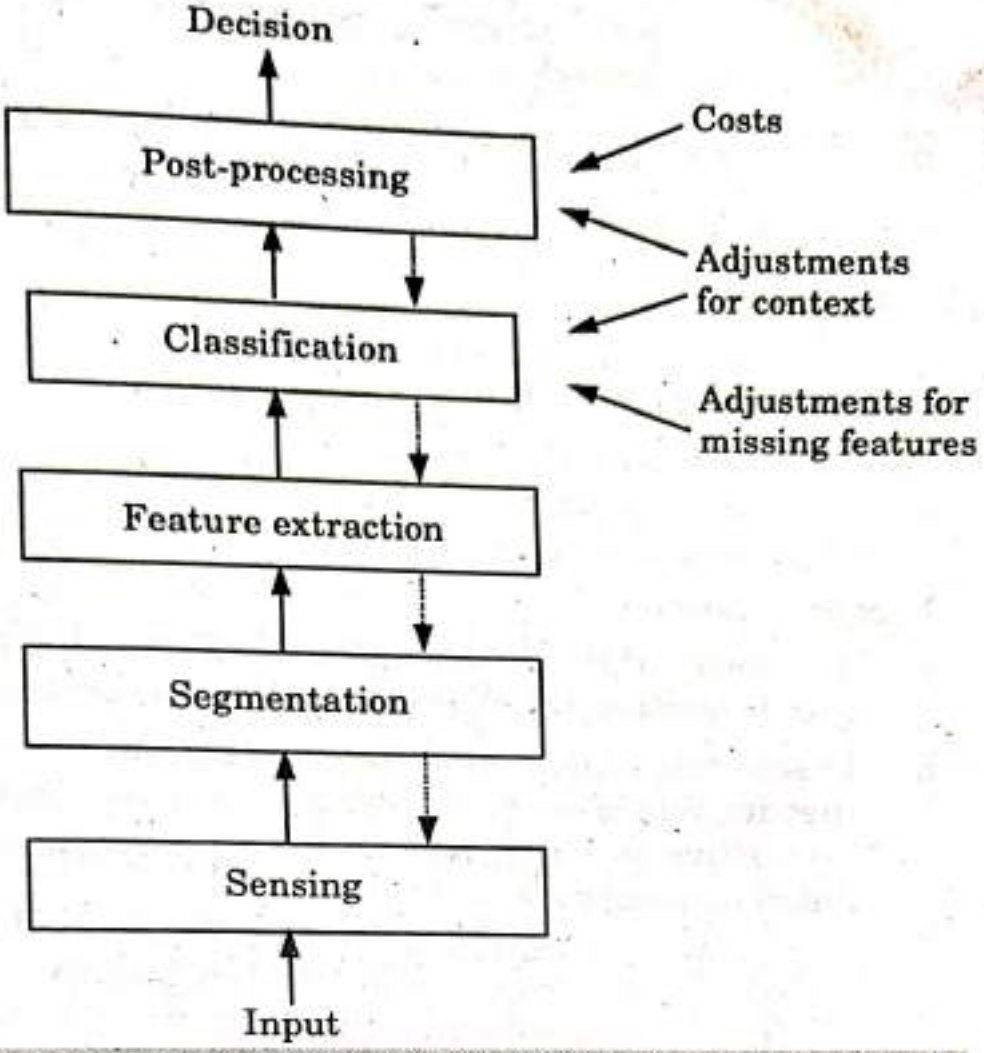


Fig. 5.4.1. Components used to design a pattern recognition system.

1. **Sensing :**
   a. A sensor converts images or sounds or other physical inputs into signal data.
   b. The input to a pattern recognition system is often some kind of a transducer, such as a camera, or a microphone array.
2. **Segmentation and grouping :**
   a. Segmentation refers to the process of partitioning a digital image or speech in multiple segments.
   b. The segmentor isolates sensed objects from the background or from other objects.
3. **Feature extraction :**
   a. A feature extractor measures object properties that are useful for classification.
4. **Classification :**
   a. The task of the classifier component of a full system is to use the feature vector provided by the feature extractor to assign the object to a category.

5. Post processor :

a. The post processor uses the output of the classifier to decide on the recommended action.

Que 5.5. Write down the activities used to design a pattern recognition system.

Answer

The design of pattern recognition system usually entails the repetition of a number of different activities :

1. **Data collection :** Data must be collected, both to train and to test the system. Data collection can account for large part of the cost for developing a pattern recognition system.
2. **Feature choice :**
   a. The choice of the distinguishing features is a critical design step and depends on the characteristics of the problem domain.
   b. In selecting or designing features, we would like to find features that are simple to extract, invariant to irrelevant transformations, insensitive to noise, and useful for discriminating patterns in different categories.

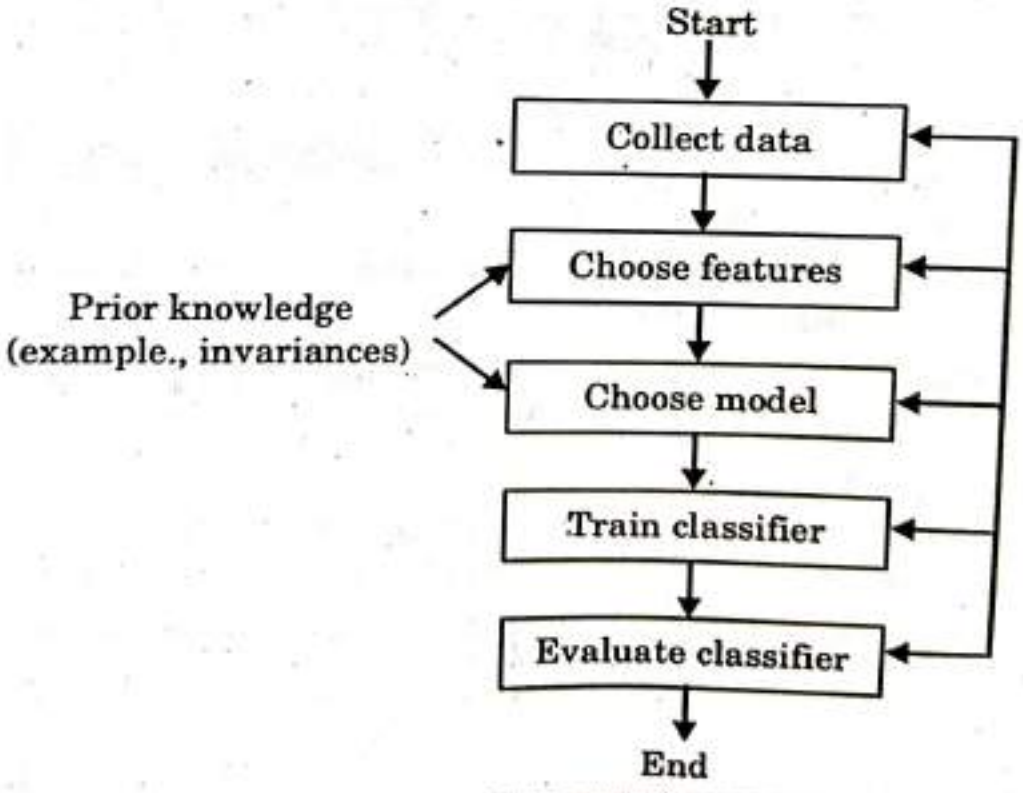


Fig. 5.5.1. Design cycle.

3. **Model choice :** The characteristic of the data affect the choice of models for the different categories.
4. **Training :** The process of using data to determine the classifier is referred to as training the classifier.

5. **Evaluation :** Evaluation is important for both, to measure the performance of the system and to identify the need for improvement in its components.

Que 5.6. **Describe the approaches of pattern recognition.**

Answer

Approaches of pattern recognition :

i. Statistical pattern recognition approach :

1. Statistical pattern recognition approach is an approach in which results can be drawn out from established concepts in statistical decision theory in order to discriminate among data based upon quantitative features of the data from different groups. For example : Mean, Standard Deviation.
2. The comparison of quantitative features is done among multiple groups.
3. The various statistical approaches used are :
   a. Bayesian decision theory
   b. Normal density
   c. Discriminant function

ii. Structural pattern recognition approach :

1. A structural approach is an approach in which results can be drawn out from established concepts in structural decision theory in order to check interrelations and interconnections between objects inside single data sample.
2. Sub-patterns and relations are the structural features while applying a structural approach.
3. **For example :** Graphs.

PART-2

Statistical Pattern Recognition, Parameter Estimation Methods : Principle Component Analysis (PCA) and Linear Discriminant Analysis (LDA).

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 5.7. **Explain statistical pattern recognition.**

Answer

1. Statistical pattern recognition is a problem used to cover all stages from problem formulation and data collection through discrimination and classification, assessment of results and interpretation.
2. In statistical pattern recognition, recognition is done by classifying the input into predefined categories.
3. In this approach, each pattern is represented in terms of features and is viewed as a point in a d-dimensional space.
4. The goal is to choose those features that allow pattern vectors belonging to different categories and disjoint regions in a d-dimensional feature space.
5. The effectiveness of the feature space is determined by how well patterns from different classes can be separated.

Que 5.8. Write a short note on Principle Component Analysis (PCA).

Answer

1. PCA is one of the most widely used statistical methods. Principal Component Analysis (PCA) technique is adopted for dimensionality reduction.
2. Using the PCA technique, a higher dimensional data space can be transformed onto a lower dimensional space. This transformation is also known as Hotelling transform.
3. PCA linearly transforms a high-dimensional input vector into a low-dimensional one whose components are uncorrelated through the calculation of eigen vectors of the covariance matrix of the original inputs.
4. The primary advantages of the PCA are the reduction of the dimensionality of the data set and the identification of new meaningful underlying variables.

Que 5.9. Write steps involved in making principle components to do a classification of given data. **AKTU 2015-16, Marks 05**

Answer

PCA : Refer Q. 5.8, Page 5–8A, Unit-5.

PCA algorithm :

Step 1 : Get data.

Step 2 : Subtract the mean.

Step 3 : Calculate the covariance matrix.

Step 4 : Calculate the eigen vectors and eigen values of the covariance matrix.

Step 5 : Choosing components and forming a feature vector.

Step 6 : Deriving the new data set, this is the final step in PCA.

Que 5.10. **Determine 2 Principle components of the following set of observations of 2-dimensional data having 5 examples.**

AKTU 2015-16, Marks 05

S. No.	X	Y
1	–1.3	–1.8
2	– 0.6	– 0.9
3	0	0
4	0.6	0.9
5	1.3	1.8

Answer

X	Y
– 1.3	– 1.8
– 0.6	– 0.9
0	0
0.6	0.9
1.3	1.8

Covariance matrix :

$$\text{Let } A = \begin{bmatrix} -1.3 & -1.8 \\ -0.6 & -0.9 \\ 0 & 0 \\ 0.6 & 0.9 \\ 1.3 & 1.8 \end{bmatrix}$$

Mean of matrix $A = \bar{X} = [0, 0]$

Matrix of deviation from the mean :

$$\begin{bmatrix} -1.3 & -1.8 \\ -0.6 & -0.9 \\ 0 & 0 \\ 0.6 & 0.9 \\ 1.3 & 1.8 \end{bmatrix} - \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} -1.3 & -1.8 \\ -0.6 & -0.9 \\ 0 & 0 \\ 0.6 & 0.9 \\ 1.3 & 1.8 \end{bmatrix}$$

$$\text{Standard deviation} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})(X_i - \bar{X})^T$$

$$\frac{1}{(5-1)}\begin{bmatrix} -1.3 & -0.6 & 0 & 0.6 & 1.3 \\ -1.8 & -0.9 & 0 & 0.9 & 1.8 \end{bmatrix}\begin{bmatrix} -1.3 & -1.8 \\ -0.6 & -0.9 \\ 0 & 0 \\ 0.6 & 0.9 \\ 1.3 & 1.8 \end{bmatrix}$$

$$= \begin{bmatrix} 1.025 & 1.44 \\ 1.44 & 2.025 \end{bmatrix}$$

$$\text{Now, eigen values} = \begin{bmatrix} 1.025-\lambda & 1.44 \\ 1.44 & 2.025-\lambda \end{bmatrix} = 0$$

$$= [(1.025-\lambda)(2.025-\lambda) - 1.44 \times 1.44] = 0$$

$$= 2.0756 - 1.025\lambda - 2.025\lambda + \lambda^2 - 2.0735 = 0$$

$$= \lambda^2 - 3.05\lambda - 0.002 = 0$$

$$= 100\lambda^2 - 305\lambda - 0.2 = 0$$

$$\lambda_1 = 3.05065597,\ \lambda_2 = -0.0006555967$$

$$\text{Eigen values} = \begin{bmatrix} -0.0006555 \\ 3.0506559 \end{bmatrix}$$

λ_1 is very small. Hence neglect.

$$\text{Eigen vector} = \begin{bmatrix} 1.025 - 3.0506559 & 1.44 \\ 1.44 & 2.025 - 3.0506559 \end{bmatrix}$$

$$= \begin{bmatrix} -2.0256 & 1.44 \\ 1.44 & -1.02565 \end{bmatrix}$$

$$\text{Final data} = \begin{bmatrix} -2.0256 & 1.44 \\ 1.44 & -1.02565 \end{bmatrix}\begin{bmatrix} -1.3 & -0.6 & 0 & 0.6 & 1.3 \\ -1.8 & -0.9 & 0 & 0.9 & 1.8 \end{bmatrix}$$

$$= \begin{bmatrix} 0.04128 & -0.08064 & 0 & 0.08064 & 0.04128 \\ -0.02583 & 0.059085 & 0 & -0.059085 & 0.02583 \end{bmatrix}$$

Que 5.11. **Explain how PCA is used in pattern recognition. Describe parameter estimation methods in pattern recognition.**

AKTU 2014-15, 2017-18; Marks 10

OR

Write a note on Linear Discriminant Analysis (LDA).

AKTU 2017-18, Marks 10

Answer

PCA use in pattern recognition :

1. Principle Component Analysis (PCA) is an application in pattern recognition. PCA is able to recognise underlying patterns in high dimensional data and the outputs of PCA can be use to highlight both the similarities and differences within a dataset.
2. It can be hard to describe patterns in data with high dimensions, the results of PCA lead to simple interpretations.

Parametric estimation methods in pattern recognition are :

i. **PCA :** Refer Q. 5.8, Page 5–8A, Unit-5.

ii. **LDA :**

1. Linear Discriminant Analysis (LDA) is commonly used technique for data classification and dimensionality reduction.
2. LDA easily handles the case where the values within class frequencies are unequal and their performances have been examined on randomly generated test data.
3. LDA works when the measurement made on independent variables for each observation are continuous quantities.
4. The use of LDA for data classification is applied to a classification problem in speech recognition.

Que 5.12. **Describe in brief the various feature extraction and selection methods in pattern recognition.**

AKTU 2014-15, Marks 10

OR

Analyze the various feature extraction and selection methods in pattern recognition.

AKTU 2018-19, Marks 10

Answer

Various feature extraction and selection methods are :

1. **PCA :** Refer Q. 5.8, Page 5–8A, Unit-5.

2. Fisher's LDA :

i. Fisher's linear discriminant is a classification method that projects high-dimensional data onto a line and performs classification in this one-dimensional space.

ii. The projection maximizes the distance between the means of the two classes while minimizing the variance within each class.

iii. This defines the Fisher criterion, which is maximized over all linear projections, w:

$$J(w) = \frac{|m_1 - m_2|^2}{s_1^2 + s_2^2}$$

where, m represents a mean,

s^2 represents a variance and the subscripts denote the two classes.

3. Non-linear PCA :

i. Non-linear Principal Component Analysis (NLPCA) is a non-linear generalization of standard principal component analysis (PCA).

ii. It generalizes the principal components from straight lines to curves (nonlinear).

iii. Thus, the subspace in the original data space which is described by all non-linear components is also curved.

Que 5.13. What do you mean by dimension reduction ? Discuss Principal Component Analysis (PCA) for dimension reduction.

AKTU 2018-19, Marks 10

Answer

Dimension reduction :

1. Dimensionality reduction is the process of reducing the number of random variables under consideration, by obtaining a set of principal variables.
2. It can be divided into feature selection and feature extraction.

Components of dimensionality reduction :

1. **Feature selection :** In this, we try to find a subset of the original set of variables, or features, to get a smaller subset which can be used to model the problem.
2. **Feature extraction :** This reduces the data in a high dimensional space to a lower dimension space, *i.e.*, a space with lesser number of dimensions.

PCA : Refer Q. 5.8, Page 5–8A, Unit-5.

Que 5.14. **Write down the advantages and disadvantages of LDA and PCA ?**

Answer

Advantages of Linear Discriminant Analysis :

1. It is suitable for larger dataset.
2. The calculation of scatter matrix in LDA is much easy as compared to co-variance matrix.

Disadvantages of Linear Discriminant Analysis :

1. More redundancy in data.
2. Memory requirement is high.
3. More noisy.

Advantages of Principal Component Analysis :

1. Less redundancy in data.
2. Less noise reduction.
3. Efficient for smaller.

Disadvantages of Principal Component Analysis :

1. The calculation of exact co-variance matrix is very difficult.
2. It is not suitable for larger data sets.

PART-3

Classification techniques-Nearest Neighbour (NN) Rule, Bayes Classifier.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 5.15. **Write a short note on**

i. Nearest Neighbour (NN) rule

ii. Bayes classifier

Answer

i. Nearest Neighbour rule :

1. In pattern recognition, the k-Nearest Neighbours algorithm (k-NN) is a method for classifying objects based on closest training examples in the feature space.

2. k-NN is a type of instance-based learning, or lazy learning where the function is only approximated locally and all computation is deferred until classification.
3. The k-nearest neighbour algorithm is the simplest of all machine learning algorithms in which an object is classified by a majority vote of its neighbours, with the object being assigned to the class most common amongst its k-nearest neighbours (k is a positive integer, typically small). If k = 1, then the object is simply assigned to the class of its nearest neighbour.
4. Nearest neighbour rules in effect compute the decision boundary in an implicit manner.

Nearest Neighbour algorithm :

1. Begin
2. Initialize c; $c' = n$; $D_i = \{x_i\}$; $i = 1, \ldots\ldots, n$
3. Do
4. $c' = c' - 1$
5. Find nearest clusters D_i and D_j
6. Merge D_i and D_j
7. Until $c = c'$
8. Return c clusters
9. End

ii. Bayes classifier :

1. A Bayes classifier is a simple probabilistic classifier based on applying Bayes theorem (from Bayesian statistics) with strong (Naive) independence assumptions.
2. In simple terms, a Naive Bayes classifier assumes that the presence (or absence) of a particular feature of a class is unrelated to the presence (or absence) of any other feature.
3. Depending on the precise nature of the probability model, Naive Bayes classifiers can be trained very efficiently in a supervised learning setting.
4. An advantage of the Naive Bayes classifier is that it requires a small amount of training data to estimate the parameters (means and variances of the variables) necessary for classification.

PART-4

Support Vector Machine (SVM), K-Means Clustering.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 5.16. **Write a short note on support vector machine.**

AKTU 2015-16, Marks 05

OR

Describe SVM in detail

Answer

1. A Support Vector Machine (SVM) is machine learning algorithm that analyzes data for classification and regression analysis.
2. SVM is a supervised learning method that looks at data and sorts it into one of two categories.
3. An SVM outputs a map of the sorted data with the margins between the two as far apart as possible.
4. Applications of SVM :
   i. Text and hypertext classification
   ii. Image classification
   iii. Recognizing handwritten characters
   iv. Biological sciences, including protein classification

Que 5.17. **What is clustering ? Describe *k*-mean clustering technique.**

AKTU 2014-15, Marks 10

OR

Apply K-means algorithm for clustering data with the help of example.

AKTU 2018-19, Marks 10

Answer

Clustering :

1. Clustering is the process or grouping of classifying objects on the basis of a close association or shared characteristics.
2. The objects can be physical or abstract entities, and the characteristics can be attribute values, relations among the objects, and combinations of both.
3. Clustering is essentially a discovery learning process in which similarity patterns are found among a group of objects.

K-mean clustering :

1. K-mean is the simplest unsupervised learning algorithms that solve the well known clustering problem.
2. K-mean clustering is a method of vector quantization, originally from signal processing, that is popular for cluster analysis in data mining.
3. The procedure follows a simple way to classify a given data set through a certain number of clusters (assume k clusters) fixed initially.
4. The algorithm aims at minimizing an objective function known as squared error function given by :

$$J(V) = \sum_{i=1}^{c} \sum_{j=1}^{c_i} (\|x_i - v_j\|)^2$$

where, $\|x_i - v_j\|$ = Euclidean distance between x_i and v_j

c_i = number of data points in i^{th} cluster

c = number of cluster centers

Algorithmic steps for K-mean clustering :

Let $X = \{x_1, x_2, x_3, ..., x_n\}$ be the set of data points and

$V = \{v_1, v_2, ..., v_c\}$ be the set of centers.

1. Randomly select 'c' cluster centers.
2. Calculate the distance between each data point and cluster centers.
3. Assign the data point to the cluster center whose distance from the cluster center is minimum of all the cluster centers.
4. Recalculate the new cluster center using :

$$v_i = (1/c_i) \sum_{j=1}^{c_i} x_i$$

where, c_i represents the number of data points in i^{th} cluster.

5. Recalculate the distance between each data point and new obtained cluster centers.
6. If no data point was reassigned then stop, otherwise repeat from step 3.

Que 5.18. Explain speech recognition in detail. Write its application.

AKTU 2014-15, Marks 10

Answer

Speech recognition :

1. Speech recognition is the process of converting an acoustic signal, captured by a microphone or a telephone, to a set of words.

2. The recognized words can be the final results, as for applications such as commands and control, data entry and document preparation.
3. Speech recognition is a difficult problem, because of the many sources of variability associated with the signal.
4. Fig. 5.18.1 shows the major components of a typical speech recognition system.

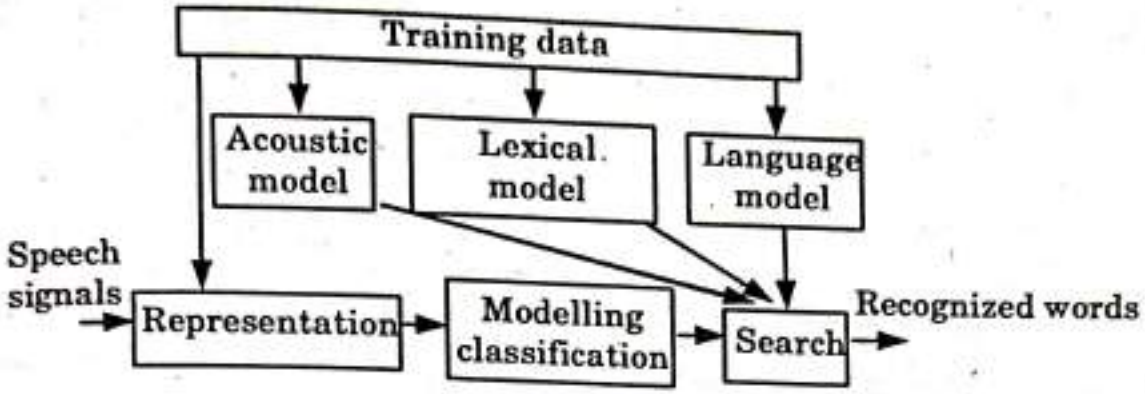


Fig. 5.18.1.

5. The speech signal is first transformed into a set of useful measurements or features at a fixed rate, typically once every 10-20 msec.
6. These measurements are then used to search for the most likely word candidate, making use of constraints imposed by the acoustic, lexical, and language models.

Applications of speech recognition :

1. **In the workplace :** Speech recognition technology in the workplace has evolved into incorporating simple tasks to increase efficiency, as well as beyond tasks that have traditionally needed humans to be performed. For example,
   i. Search for reports or documents on your computer
   ii. Start video conferences
   iii. Schedule meetings
   iv. Record minutes
2. **In banking :** The aim of the banking and financial industry for speech recognition's to reduce friction for the customer. Voice-activated banking could largely reduce the need for human customer service. For example,
   i. Request information regarding your balance, transactions, and spending habits.
   ii. Make payments.
3. **In healthcare :** In healthcare, immediate access to information can have a significantly positive impact on patient safety and medical efficiency. For example,
   i. Quickly finding information from medical records.

ii. Nurses can be reminded of processes or given specific instructions.

iii. Nurses can ask for administrative information, such as the number of patients on a floor and the number of available units.

Que 5.19. Write a short note on speech processing.

Answer

1. Speech processing is the process by which speech signals are interpreted, understood, and acted upon.
2. It specifically refers to the processing of human speech by computerized systems, as in voice recognition software or voice-to-text programs.
3. Speech processing is important to both theoretical and practical uses, ranging from voice activation and control in phones lead to development of functional artificial intelligence in computer science.
4. Interpretation and production of coherent speech are both important in the processing of speech; some concerns do favour one over the other, however, as the application needs of speech processing are very diverse.
5. Speech recognition is one of the most important aspects of speech processing because the overall aim of processing speech is to comprehend and to act on spoken language.
6. One commonly used application of speech recognition is simple speech-to-text conversion, which is used in many word processing programs.
7. Another component of speech processing is voice recognition, which is essentially a combination of speech and speaker recognition.
8. Voice recognition occurs when speech recognition programs process the speech of a known speaker; such programs can generally interpret the speech of a known speaker with much greater accuracy than that of a random speaker.
9. Voice analysis differs from other topics in speech processing because it is not really concerned with the linguistic content of speech.
10. It is primarily concerned with speech patterns and sounds. Voice analysis could be used to diagnose problems with the vocal cords or other organs related to speech by noting sounds that are indicative of disease or damage.
11. Sound and stress patterns could also be used to determine if an individual is telling the truth, though this use of voice analysis is highly controversial.

VERY IMPORTANT QUESTIONS

Following questions are very important. These questions may be asked in your SESSIONALS as well as UNIVERSITY EXAMINATION.

Q. 1. Explain how PCA is used in pattern recognition. Describe parameter estimation methods in pattern recognition.

Ans. Refer Q. 5.11.

Q. 2. Describe in brief the various feature extraction and selection methods in pattern recognition.

Ans. Refer Q. 5.12.

Q. 3. What is clustering ? Describe k-mean clustering technique

Ans. Refer Q. 5.17.

Q. 4. Write a short note support vector machine.

Ans. Refer Q. 5.16.

Q. 5. Write a note on linear discriminant analysis.

Ans. Refer Q. 5.11.

Introduction (2 Marks Questions)

1.1. What do you mean by intelligent agent ?

AKTU 2015-16, Marks 02

Ans An intelligent agent is an autonomous entity which acts upon an environment using sensors and actuators for achieving goals. An intelligent agent may learn from the environment to achieve their goals.

1.2. Describe the role of computer vision.

AKTU 2015-16, Marks 02

Ans

1. Computer vision in the perspective of engineering, seeks to automate tasks that the human visual system can do.
2. Areas of artificial intelligence deal with autonomous planning for robotical systems to navigate through an environment.
3. A detailed understanding of these environments is required to navigate through them.
4. Information about the environment and the robot could be provided by a computer vision system acting as a vision sensor.

1.3. Describe how can we use artificial intelligence in natural language processing ?

AKTU 2015-16, Marks 02

Ans

1. AI provides computer the ability to accept spoken words as dictation or to follow voice commands by using software.
2. AI programs are able to communicate with humans in a natural way because natural language is one of the most important medium for the communication.
3. To understand the natural language, a program needs a considerable knowledge about the structure of the language including what the words are and how they combine into phrases and sentences.

1.4. Describe the role of rational agent.

AKTU 2015-16, Marks 02

Ans: Role of rational agent :

1. Rational agents are autonomous programs that are capable of goal directed behavior.
2. It carries out an action with the best outcome after considering past and current percepts.
3. An AI system is composed of a rational agent and its environment. The rational agents act in their environment.

1.5. List various criterion for success in AI.

AKTU 2016-17, Marks 02

Ans: To determine if research work in artificial intelligence is successful, we should ask if we have constructed machine that is intelligent. For this we use a method known as Turing test, to conduct this test we require two person and the machine which needs to be evaluated.

One person acts as interrogator, who is in separate room from the machine and the other person.

The interrogator can ask question of either the person or the machine by typing questions and receiving typed answers.

However, the interrogator knows them only as *X* and *Y* and aims to determine which is person and which is machine. The goal of machine is to fool the interrogator into believing that it is the person. If the machine can do this then we will conclude that the machine can think.

1.6. What are goals of AI ?

AKTU 2017-18, Marks 02

Ans: Goals of AI are :

1. **To create expert systems :** The systems which exhibit intelligent behavior, learn, demonstrate, explain, and advice its users.
2. **To implement human intelligence in machines :** Creating systems that understand, think, learn, and behave like humans.

1.7. What is Turing test ?

AKTU 2017-18, Marks 02

Ans: Turing test is a test for intelligence in a computer, requiring that a human being should be unable to distinguish the machine from another human being by using the replies to questions put to both. Turning test is the ultimate test a machine must pass in order to be called as intelligent.

1.8. Define learning agent with the help of architecture.

AKTU 2018-19, Marks 02

Ans: A learning agent is a tool in AI that is capable of learning from its experiences. Learning agents are able to perform tasks, analyze performance and look for new ways to improve on those tasks.

Architecture of learning agent :

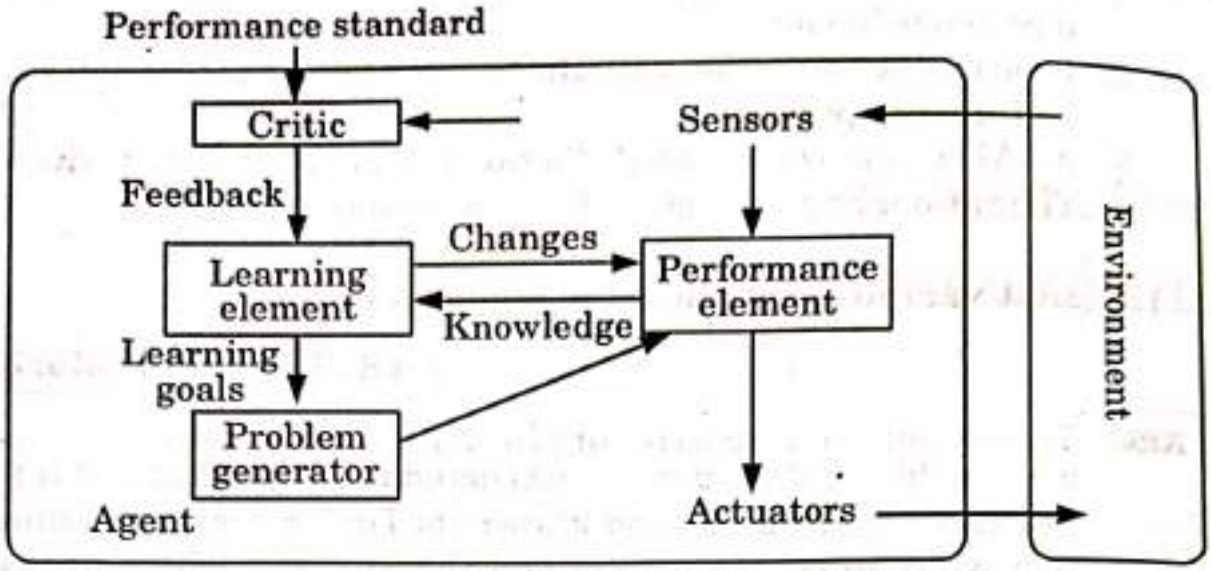


Fig. 1.

1.9. What is computer vision ? **AKTU 2018-19, Marks 02**

Ans. Computer vision is an interdisciplinary scientific field that deals with how computers can be made to gain high-level understanding from digital images or videos. From the perspective of engineering, it seeks to automate tasks that the human visual system can do.

1.10. Differentiate between strong AI and weak AI.

Ans.

S. No.	Strong AI	Weak AI
1.	Strong AI supposes that it is possible for machines to become human or self-aware, but may or may not exhibit human-like thought processes.	Weak AI refers to the use of software to study or accomplish specific problem solving or reasoning tasks that do not encompass all human cognitive abilities.
2.	Strong AI claims that machine can act intelligently, has mind and understanding.	Weak AI claims that machines can act intelligently/or merely specific problem solver.

1.11. What are major components of AI ?

Ans. Major components of AI are :

1. AI hardware
2. AI software
3. Knowledge representation
4. Heuristic search

1.12. What is conceptual dependency (CD) ?

Ans. Conceptual dependency (CD) is a theory of natural language processing which deals with representation of semantics of a language. Knowledge is represented in CD by elements called as conceptual structures.

1.13. Design the PEAS measure for "Satellite Agent".

AKTU 2018-19, Marks 02

Ans. **The PEAS measures for satellite agent are :**

Performance measure : Correct image categorization.

Environment : Downlink from orbiting satellite.

Actuators : Display categorization of scene.

Sensors : Colour pixel arrays.

2 UNIT

Introduction to Search (2 Marks Questions)

2.1. State the significance of using heuristic functions.

AKTU 2016-17, Marks 02

Ans. **Significance of heuristic function :**

1. The heuristic function is a way to inform the search about the direction to a goal. It provides an informed way to guess which neighbour of a node will lead to a goal.
2. The heuristic function helps to guide the search process in the most profitable directions, by suggesting which path to follow first when more than one path is available.

2.2. Distinguish between state space search and plan space search.

AKTU 2016-17, Marks 02

Ans.

S. No.	State space search	Plan space search
1.	In state space search, there is commitment in plan step ordering.	In plan space search, there is no-commitment in plan step ordering.
2.	State space search can suffer with goal orderings.	Plan space search do not suffer from goal orderings.
3.	State space search handles goal interactions poorly.	Plan space search handles goal interactions efficiently.

2.3. What do you mean by local maxima with respect to search technique ?

AKTU 2016-17, Marks 02

Ans. Local maxima is a state better than the local region or neighbouring states, but not a global maximum. This occurs since a better solution exists which is not in the vicinity of the present state.

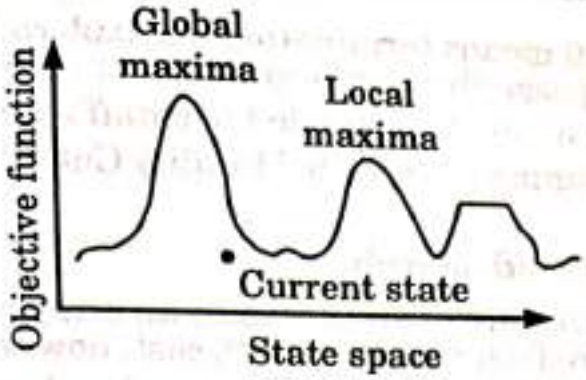


Fig. 1.

2.4. Define uninformed search. **AKTU 2017-18, Marks 02**

Ans. Uninformed search is a used by search procedures which explore all the alternatives during the search process. They do not have domain specific knowledge. All they need are the initial state, the final state, and set of legal operators.

2.5. Write a short note on horizon effect.

AKTU 2017-18, Marks 02

Ans. Horizon effect is a problem that involves an extremely long sequence of moves that clearly lead to a strong advantage for one player here the sequence of moves, takes more moves than is allowed by the bounded search.

2.6. Write down the time and space complexity of DFS search strategies. **AKTU 2018-19, Marks 02**

Ans. The time complexity of the depth first search (DFS) is $O(b^d)$. The space complexity of depth first search (DFS) is $O(d)$.

2.7. Give the desirable properties of search algorithms.

Ans. **Properties of heuristic search algorithms are :**

1. Completeness
2. Space complexity
3. Time complexity
4. Optimality

2.8. What are problems in hill climbing ?

Ans. **Problems in hill climbing :**

i. Local maxima
ii. Plateau
iii. Ridge

2.9. What do you mean by futility cutoff ?

Ans: Futility cutoff means terminating the exploration of a subtree that offers little possibility for improvement over other known path. Alpha-beta can also be extended to cutoff search on paths that are only slightly improved is called Futility Cutoff.

2.10. Define informed search.

Ans: Informed search algorithm contains an array of knowledge such as how far we are from the goal, path cost, how to reach to goal node, etc. This knowledge helps agents to explore less to the search space and find more efficiently the goal node

2.11. Write down the types of production system.

Ans: **Types of production system :**

1. Monotonic production system
2. Non-monotonic production system
3. Partially commutative production system
4. Commutative production system

2.12. What are characteristics of production system ?

Ans: **Characteristics of production system :**

1. Simplicity
2. Modularity
3. Modifiability
4. Knowledge intensive

2.13. Define adversarial search.

Ans: Adversarial search is the search in which two or more players with conflicting goals are trying to explore the same search space for the solution.

2.14. What is constraint satisfaction procedure ?

Ans: Constraint satisfaction is a search procedure that operates in a space of constraint sets. The initial state contains the constraints that are originally given in the problem description.

2.15. Define alpha-beta pruning.

Ans: Alpha-beta pruning is a modified version of the min-max algorithm. It is an optimization technique for the min-max algorithm.

3 UNIT

Knowledge Representation and Reasoning (2 Marks Questions)

3.1. Define Modus Ponen's rule in propositional logic.

AKTU 2015-16, Marks 02

Ans: The logical rule of inference, "From P and $(P \rightarrow Q)$ infer Q". Here P and Q can be any well-formed propositional. This is written as,

$$\frac{P \quad P \rightarrow Q}{Q}$$

This rule says that if $P \rightarrow Q$ is true, and P is true, then Q is necessarily true.

3.2. Define informational equivalence and computational equivalence.

AKTU 2015-16, Marks 02

Ans: **Informational equivalence :** Two representation are informationally equivalent if the transformation from one to the other demand no loss of information, *i.e.*, if each can be constructed from the other.

Computational equivalence : Two representations are computationally equivalent if the same information can be extracted from each (the same inferences drawn) with about the same amount of computation.

3.3. List two applications of Hidden Markov model.

AKTU 2016-17, Marks 02

Ans: **Two applications of Hidden Markov Model (HMM) are :**

1. Bioinformatics
2. Speech recognition

3.4. What is semantic analysis ? Explain.

AKTU 2016-17, Marks 02

Ans: Semantic analysis is the study of semantics, or the structure and meaning of speech. It is the job of a semantic analyst to discover grammatical patterns, the meanings of colloquial speech, and to uncover specific meanings to words in foreign languages.

Semantic analysis draws the exact meaning or the dictionary meaning from the text. The text is checked for meaningfulness. It is done by mapping syntactic structures and objects in the task domain.

3.5. List various issues in knowledge representation.

AKTU 2016-17, Marks 02

Ans: Issues in knowledge representation :

1. Important attributes
2. Relationship among attributes
3. Choosing granularity
4. Set of objects
5. Finding right structure

3.6. List down two applications of temporal probabilistic models.

AKTU 2016-17, Marks 02

Ans: Two applications of temporal probabilistic models are :

1. Speech recognition
2. Tracking

3.7. What are the limitations in using propositional logic to represent the knowledge base?

AKTU 2016-17, Marks 02

Ans: Limitations of propositional logic to represent knowledge base :

1. Lots of propositional variables
2. Lots of rules
3. Inference space hungry

3.8. List various schemes of knowledge representation.

AKTU 2017-18, Marks 02

Ans: Various schemes of knowledge representation :

1. Simple relational knowledge
2. Inheritable knowledge
3. Inferential knowledge
4. Procedural knowledge

3.9. Define inference.

AKTU 2017-18, Marks 02

Ans: In artificial intelligence, we need intelligent computers which can create new logic from old logic or by evidence, so generating the conclusions from evidence and facts is termed as inference. The standard patterns of inference that can be applied to derive chains of conclusions that lead to the desired goal are called inference rules. The best known rule is called Modus Ponens.

3.10. State soundness property of inference.

AKTU 2018-19, Marks 02

Ans: An inference procedure $\vdash$ is sound if whenever $p \vdash q$ then it is also the case that $p \models q$.

The idea of soundness is applied in logic. Whenever we create a knowledge based program we use the syntax of the knowledge representation language, we assign semantics in some way and the reasoning mechanism defines the inference procedures.

3.11. Discuss the normal forms in propositional logic.

Ans: The normal forms in propositional logic are as follows :

i. **Conjunctive Normal Form (CNF) :** A formula is said to be in CNF if it has the form

$A = A_1 \cap A_2 \cap A_3 \ldots\ldots\ldots\ldots\ldots \cap A_n; n >= 1$

where each $A_1, A_2, A_3, \ldots, A_n$ is a disjunction of an atom or negation of an atom.

ii. **Disjunctive Normal Form (DNF) :** A formula is said to be in DNF if it has the form

$A = A_1 \cap A_2 \cap A_3 \ldots\ldots \cap A_n;\ n >= 1$

where each A_1, A_2, A_3 ,....., A_n is a conjunction of an atom or negation of an atom.

3.12. What is Universal quantifier ?

Ans: Universal quantifier is a symbol of logical representation, which specifies that the statement within its range is true for everything or every instance of a particular thing. The Universal quantifier is represented by a symbol $\forall$, which resembles an inverted A.

3.13. What is Existential quantifier ?

Ans: Existential quantifiers are the type of quantifiers, which express that the statement within its scope is true for atleast one instance of something. It is denoted by the logical operator $\exists$, which resembles as inverted E. When it is used with a predicate variable then it is called as an Existential quantifier.

3.14. What are the types of variable in FOC ?

Ans: **There are two types of variables in First order logic :**

1. **Free variable :** A variable is said to be a free variable in a formula if it occurs outside the scope of the quantifier. For example: $\forall x \exists (y)[P(x, y, z)]$, where z is a free variable.
2. **Bound variable :** A variable is said to be a bound variable in a formula if it occurs within the scope of the quantifier. For example: $\forall x\ [A(x)\ B(y)]$, here x and y are the bound variables.

4 UNIT

Machine Learning (2 Marks Questions)

4.1. Define inductive learning. How the performance of inductive learning algorithms can be measured ?

AKTU 2015-16, Marks 02

Ans:

1. Inductive learning is the process of acquiring generalized knowledge from examples or instances of same class. This form of learning is accomplished through inductive inference.
2. Performance of inductive learning algorithm is measured by their learning curve, which shows the prediction accuracy as a function of the number of observed examples.

4.2. State the factors that play a role in the design of a learning system.

AKTU 2015-16, Marks 02

Ans: Factors that play a role in design of a learning system :

1. Data
2. Model selection
3. Learning
4. Application (Evaluation)

4.3. Define reinforcement learning.

AKTU 2016-17, Marks 02

Ans: Reinforcement learning is the study of how animals and artificial systems can learn to optimize their behavior in the face of rewards and punishments.

4.4. List out performance measure for learning.

AKTU 2017-18, Marks 02

Ans: Performance measures for learning are :

i. Generality
ii. Efficiency
iii. Robustness
iv. Efficacy
v. Ease of implementation

4.5. What are the types of nodes in decision tree ?

AKTU 2017-18, Marks 02

Ans: Types of nodes in decision tree :

1. **A root node (decision node) :** It represents a choice that will result in the subdivision of all records into two or more mutually exclusive subsets.
2. **Internal node (chance node) :** It represents one of the possible choices available at that point in the tree structure; the top edge of the node is connected to its parent node and the bottom edge is connected to its child nodes or leaf nodes.
3. **Leaf node (end node) :** It represents the final result of a combination of decisions or events.

4.6. List out the application areas of machine learning.

AKTU 2018-19, Marks 02

Ans: Application areas of machine learning are :

1. Financial services
2. Marketing and sales
3. Government
4. Healthcare
5. Transportation

4.7. Define supervised and unsupervised learning in machine learning.

AKTU 2018-19, Marks 02

Ans: Supervised learning : Supervised learning is also known as associative learning, in which the network is trained by providing it with input and matching output patterns.

Unsupervised learning : Unsupervised learning is also known as self-organization, in which an output unit is trained to respond to clusters of pattern within the input.

4.8. What is decision tree ?

AKTU 2018-19, Marks 02

Ans: A decision tree is a decision support tool that uses a tree-like graph or model of decisions and their possible consequences, including chance event outcomes, resource costs, and utility.

4.9. Write down the advantages and disadvantages of reinforcement learning.

Ans: Advantages of reinforcement learning are :

1. Maximizes performance.
2. Sustain change for a long period of time.
3. Increases behaviour.

Disadvantages of reinforcement learning :

1. Too much reinforcement can lead to overload of states which can diminish the results.
2. It only provides enough to meet up the minimum behavior.

4.10. Define the types of reinforcement learning.

Ans. Types of reinforcement :

1. Positive reinforcement learning.
2. Negative reinforcement learning.

5 UNIT

Pattern Recognition (2 Marks Questions)

5.1. What are the features of pattern recognition ?

Ans: Features of pattern recognition :

1. Pattern recognition system should recognize familiar pattern quickly and accurate.
2. It recognizes and classifies unfamiliar objects.
3. It accurately recognizes shapes and objects from different angles.
4. It identifies patterns and objects even when partly hidden.
5. It recognizes patterns quickly with ease and with automaticity.

5.2. Define support vector machine. **AKTU 2015-16, Marks 02**

Ans: A Support Vector Machine (SVM) is machine learning algorithm that analyzes data for classification and regression analysis. SVM is a supervised learning method that looks at data and sorts it into one of the two categories.

5.3. Discuss the various types of model of parallel algorithm with example. **AKTU 2015-16, Marks 02**

Ans: Various types of models of parallel algorithm :

1. **Data parallel model :** In data parallel model, tasks are assigned to processes and each task performs similar types of operations on different data. For example, dense matrix multiplication.
2. **Task graph model :** In the task graph model, parallelism is expressed by a task graph. A task graph can be either trivial or non-trivial. For example, parallel quick sort.
3. **Work pool model :** In work pool model, tasks are dynamically assigned to the processes for balancing the load. Therefore, any process may potentially execute any task. For example, parallel tree search.
4. **Pipeline model :** It is also known as the producer-consumer model. Here a set of data is passed through a series of processes, each of which performs some task on it. For example, parallel LU factorization algorithm.

5. **Hybrid model :** A hybrid algorithm model is required when more than one model may be needed to solve a problem. A hybrid model is composed of either multiple models applied hierarchically or sequentially to different phases of a parallel algorithm. For example, parallel quick sort

5.4. Write down some applications of pattern recognition.

AKTU 2017-18, Marks 02

Ans: Applications of pattern recognition :

i. Image processing
ii. Computer vision
iii. Speech recognition
iv. Finger print identification

5.5. What are the types of neural networks ?

AKTU 2017-18, Marks 02

Ans: Types of neural network are :

i. Feed forward neural network
ii. Single layer perceptron
iii. Multi layer perceptron
iv. Adaline
v. Kohonen self-organizing network

5.6. Differentiate between classification and regression ?

AKTU 2018-19, Marks 02

Ans:

S. No.	Classification	Regression
1.	Classification is the task of predicting a discrete class label.	Regression is the task of predicting a continuous quantity.
2.	A classification algorithm may predict a continuous value, but the continuous value is in the form of a probability for a class label.	A regression algorithm may predict a discrete value, but the discrete value is in the form of an integer quantity.
3.	Classification predictions can be evaluated using accuracy.	Regression predictions can be evaluated using root mean squared error.

5.7. Discuss the features of support vector machine.

AKTU 2018-19, Marks 02

Ans: Features of support vector machine are :

i. Flexibility in choosing a similarity function.
ii. Sparseness of solution when dealing with large data sets.
iii. Ability to handle large feature spaces.

5.8. What are the design principles of pattern recognition system ?

Ans: Design principles of pattern recognition system are :

1. Statistical approach.
2. Syntactic approach.

5.9. Write down the activities used to design a pattern recognition system.

Ans: Activities used to design a pattern recognition system are :

i. Data collection
ii. Feature choice
iii. Model choice
iv. Training
v. Evaluation

5.10. What are the components used to design a pattern recognition system ?

Ans: Components used to design a pattern recognition system :

i. Sensing
ii. Segmentation and grouping
iii. Feature extraction
iv. Classification
v. Post processor

B. Tech.
(SEM. VII) ODD SEMESTER THEORY EXAMINATION, 2019-20
ARTIFICIAL INTELLIGENCE

Time : 3 Hours **Max. Marks : 70**

Note : Attempt **all** sections. If require any missing data; then choose suitably.

SECTION-A

1. Attempt **all** questions in brief. **(2 × 7 = 14)**

a. Write the history of artificial intelligence.

Ans. Refer Q. 1.6, Page 1–5A, Unit-1.

b. Describe optimal problem with suitable example.

Ans.

i. An optimization problem is the problem of finding the best solution from all feasible solutions. An optimization problem is given as :

1. A set of variables, each with an associated domain.
2. An objective function that maps total assignments to real numbers.
3. An optimality criterion is to find a total assignment that minimizes or maximizes the objective function.

ii. The aim is to find a total assignment that is optimal according to the optimality criterion.

Example : TSP (Travelling Salesman Problem).

c. Define utility theory.

Ans. Refer Q. 3.25, Page 3–21A, Unit-3.

d. What are statistical learning models ?

Ans. Refer Q. 4.12, Page 4–13A, Unit-4.

e. Define Bayes classifier.

Ans: Refer Q. 5.15(ii), Page 5–13A, Unit-5.

f. Justify the use of searching in game.

Ans:

1. In game theory, a game tree is a directed graph whose nodes are positions in a game and whose edges are moves.
2. The complete game tree for a game is the game tree starting at the initial position and containing all possible moves from each position, the complete tree is the same tree as that obtained from the extensive-form game representation.
3. Game trees are important in artificial intelligence because one way to pick the best move in a game is to search the game tree using any of numerous tree search algorithms, combined with minimax-like rules to prune the tree.
4. The game tree for tic-tac-toe is easily searchable, but the complete game trees for larger games like chess are much too large to search.

g. Write the difference between the propositional and predicate logic.

Ans:

S. No.	Propositional logic	Predicate logic
1.	Area of logic deals with proposition.	Area of logic deals with both logic relation and internal structure of subject.
2.	It uses logical operator to join two proposition.	It uses logical operators and concept of variable and quantifier.
3.	Propositional logic deals with finite models.	Predicate logic deals with infinite structures.

SECTION-B

2. Attempt any **three** of the following : **(7 × 3 = 21)**

a. Define Principle Component Analysis (PCA). Determine the 2 PCA of the following set of observations of 2-dimensional data having 5 examples.

S. No.	X	Y
1.	–1.4	–1.9
2.	–0.5	–0.8
3.	0.1	0.1
4.	0.8	1.1
5.	1.4	1.8

Ans: **PCA :** Refer Q. 5.8, Page 5–8A, Unit-5.
Numerical :

X	Y
–1.4	–1.9
–0.5	–0.8
0.1	0.1
0.8	1.1
1.4	1.8

Covariance matrix :

$$\text{Let } A = \begin{bmatrix} -1.4 & -1.8 \\ -0.5 & -0.9 \\ 0.1 & 0.1 \\ 0.8 & 1.1 \\ 1.4 & 1.8 \end{bmatrix}$$

Mean of matrix $A = \bar{X} = [0, 0]$

Matrix of deviation from the mean :

$$\begin{bmatrix} -1.4 & -1.9 \\ -0.5 & -0.8 \\ 0.1 & 0.1 \\ 0.8 & 1.1 \\ 1.4 & 1.8 \end{bmatrix} - \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} -1.4 & -1.9 \\ -0.5 & -0.8 \\ 0.1 & 0.1 \\ 0.8 & 1.1 \\ 1.4 & 1.8 \end{bmatrix}$$

$$\text{Standard deviation} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})(X_i - \bar{X})^T$$

$$\frac{1}{(5-1)}\begin{bmatrix} -1.4 & -0.5 & 0.1 & 0.8 & 1.4 \\ -1.9 & -0.8 & 0.1 & 1.1 & 1.8 \end{bmatrix}\begin{bmatrix} -1.4 & -1.9 \\ -0.5 & -0.8 \\ 0.1 & 0.1 \\ 0.8 & 1.1 \\ 1.4 & 1.8 \end{bmatrix}$$

$$= \begin{bmatrix} 1.205 & 1.617 \\ 1.797 & 2.177 \end{bmatrix}$$

$$\text{Now, eigen values} = \begin{bmatrix} 1.205-\lambda & 1.617 \\ 1.797 & 2.177-\lambda \end{bmatrix} = 0$$

$$= [(1.205-\lambda)(2.177-\lambda) - 1.617 \times 1.797] = 0$$

$$= 2.623 - 1.205\lambda - 2.177\lambda + \lambda^2 - 2.905 = 0$$

$$= \lambda^2 - 3.382\lambda - 0.282 = 0$$

$$= 1000\lambda^2 - 3382\lambda - 282 = 0$$

$$\lambda_1 = 3.463 \qquad \lambda_2 = -0.081$$

$$\text{Eigen values} = \begin{bmatrix} 3.463 \\ -0.081 \end{bmatrix}$$

λ_2 is very small. Hence neglect.

$$\text{Eigen vector} = \begin{bmatrix} 1.205-3.463 & 1.617 \\ 1.797 & 2.177-3.463 \end{bmatrix}$$

$$= \begin{bmatrix} -2.256 & 1.617 \\ 1.794 & -1.286 \end{bmatrix}$$

$$\text{Final data} = \begin{bmatrix} -2.256 & 1.617 \\ 1.797 & -1.286 \end{bmatrix}\begin{bmatrix} -1.4 & -0.5 & 0.1 & 0.8 & 1.4 \\ -1.9 & -0.8 & 0.1 & 1.1 & 1.8 \end{bmatrix}$$

$$= \begin{bmatrix} 0.0889 & -0.1646 & -0.0641 & -0.0277 & -0.2506 \\ -0.0724 & 0.1303 & 0.0511 & 0.023 & 0.201 \end{bmatrix}$$

b. Explain about the Hill climbing algorithm with its drawback and how it can be overcome ?

Ans: Refer Q. 2.12, Page 2–10A, Unit-2.

c. Describe the rules of inference in first order predicate logic with suitable example.

Ans: Refer Q. 3.10, Page 3–10A, Unit-3.

d. Define reinforcement learning. Differentiate between the passive and active reinforcement learning. Is evolution reinforcement learning an appropriate abstract model for human learning ?

Ans: Refer Q. 4.14, Page 4–14A, Unit-4.

Yes, evolution reinforcement learning is an appropriate abstract model for human learning.

e. Explain the role of artificial intelligence in natural language processing.

Ans: Refer Q. 1.16, Page 1–15A, Unit-1.

SECTION-C

3. Attempt any **one** part of the following : **(7 × 1 = 7)**

a. Define intelligent agent. Explain various types of agent program with suitable example.

Ans: Refer Q. 1.8, Page 1–7A, Unit-1.

b. Explain computer vision in parlance to the artificial intelligence.

Ans: Refer Q. 1.14, Page 1–14A, Unit-1.

4. Attempt any **one** part of the following : **(7 × 1 = 7)**

a. What is heuristic function ? Differentiate between blind search and heuristic search strategies.

Ans: Refer Q. 2.10, Page 2–8A, Unit-2.

b. What is adversarial search ? Write the steps for game problem formulation. State and explain minimax algorithm with tic-tac-toe game.

Ans: **Adversarial search :** Refer Q. 2.17, Page 2–14A, Unit-2.

Steps for game problem formulation : Problem formulation involves deciding what actions and states to consider for the given goal. A problem can be defined formally by five components :

1. The initial state of the agent.
2. The possible actions available to the agent, corresponding to each of the state the agent resides in.
3. The transition model describing what each action does.
4. The goal test, determining whether the current state is a goal state.
5. The path cost function, which determines the cost of each path, which is reflecting in the performance measure.

Minimax algorithm : Refer Q. 2.20, Page 2–17A, Unit-2.

5. Attempt any **one** part of the following : **(7 × 1 = 7)**

a. Differentiate between forward and backward chaining of inference with the help of example.

Ans: Refer Q. 3.20, Page 3–17A, Unit-3.

b. Translate the following sentences in formulas in predicate logic and casual form :

i. John likes all kind of food.

ii. Apples are food.

iii. Chicken is food.

iv. Anything anyone eats and is not killed by is food.

v. Bill eats peanuts and is still alive.

vi. Sue eats everything Bill eats.

Ans: Refer Q. 3.11, Page 3–11A, Unit-3.

6. Attempt any **one** part of the following : **(7 × 1 = 7)**

a. Define machine learning. Explain supervised and unsupervised learning with suitable example.

Ans: **Machine learning :** Refer Q. 4.1, Page 4–2A, Unit-4.

Supervised and unsupervised learning : Refer Q. 4.3, Page 4–3A, Unit-4.

b. Explain the following in detail :

i. Naive Bayes model

ii. Learning with hidden data-EM algorithm

Ans:

i. Refer Q. 4.10, Page 4–9A, Unit-4.

ii. Refer Q. 4.11, Page 4–12A, Unit-4.

7. Attempt any **one** part of the following : **(7 × 1 = 7)**

a. How Linear Discriminant Analysis is different from logistics regression ? Explain Linear Discriminant Analysis (LDA) with suitable example.

Ans: **Difference :**

1. Logistics Regression (LR) makes no assumptions on the distribution of the explanatory data, LDA has been developed for normally distributed explanatory variables.
2. LDA to give better results in the case when the normality assumptions are fulfilled, but in all other situations LR should be more appropriate.
3. The goal of LR is to find the best fitting and to describe the relationship between the outcome (dependent or response variable) and a set of independent (predictor or explanatory) variables. In LR, no assumptions are made regarding the distribution of the explanatory variables.

LDA : Refer Q. 5.11, Page 5–11A, Unit-5.

Example :

1. In computerised face recognition, each face is represented by a large number of pixel values.
2. Linear discriminant analysis is used to reduce the number of features to a more manageable number before classification.

3. Each of the new dimensions is a linear combination of pixel values, which form a template.
4. The linear combinations obtained using Fisher's linear discriminant are called Fisher faces, while those obtained using the related principal component analysis are called eigenfaces.

b. What is clustering ? Describe *k*-mean clustering technique.

Ans. Refer Q. 5.17, Page 5–18A, Unit-5.