BACS 2003
BACS 3074
BMCS 2003

Artificial Intelligence

Tutorial Book Sample Answer

Tutorial 1

1. Who is the father of A.I.? Describe the reason why he was recognized so. ()

Answer: John McCarthy

1. Due to his astounding contributions in the field of computer science and AI. By the mid-1950s,
2. John McCarthy created the term "artificial intelligence" (1956), which he defined as "the science and engineering of making intelligent machines. Dartmouth conference in 1956
3. The objective of the conference was to explore ways to invent a machine that could reason like a human, has problem solving skills and self-improvement.

Student to find the answer from any reference. Reference source must be indicated

2. Identify one key event or major achievement of A.I. development in the year of
   a) 1960-1969 ()
   b) 1970-1979 ()
   c) 1980-1989 ()
   d) 1990-1999 ()
   e) 2000-2009 ()
   f) 2010-2019 ()
   g) 2020-now

Answer:

1960-1969

1. Introduction of Unimate (1961): Unimate, the first industrial robot, was developed by George Devol and Joseph Engelberger. It was installed at General Motors' plant to automate tasks such as lifting and welding car parts.
2. Development of ELIZA (1966): ELIZA, created by Joseph Weizenbaum at MIT, was one of the first examples of a chatbot. It simulated conversation by using pattern-matching techniques and was programmed to mimic the behavior of a Rogerian psychotherapist.
3. Development of MacHack (1966): MacHack was one of the earliest chess-playing programs, developed by Richard Greenblatt at MIT. It was notable for being able to play chess at a reasonable level for its time, and it was the first computer program to compete in a formal chess tournament.

1970-1979

1. Introduction of the Expert System MYCIN (1972): MYCIN, developed at Stanford University, was an expert system designed to diagnose blood infections. It demonstrated the potential of AI in specialized fields like medicine.
2. Development of Wabot-1 (1973): Wabot-1 was the first human-sized robot developed at Wako University in Japan. It was designed to simulate human conversation and perform basic tasks like walking, moving its arms, and speaking. Wabot-1 used early AI techniques for its interactions, such as speech recognition and basic natural language processing.

1980-1989

1. Commercial expert system: Systems like Diagnostic Pathfinder were used in automotive industries to diagnose car engine problems, improving efficiency and reducing costs. Expert systems like XCON (used by Digital Equipment Corporation) provided financial advice and optimized computer system configurations
2. Rise of Neural Networks (1986): Geoffrey Hinton, David Rumelhart, and Ronald J. Williams introduced the backpropagation algorithm, making neural networks more practical and efficient, leading to advances in machine learning.
3. Development of Polly (1980s): Polly was an early example of behavior-based robotics, developed by Dr. Rodney Brooks and his team at the Massachusetts Institute of Technology (MIT). Polly was designed to demonstrate how robots could exhibit complex behaviors without relying on detailed programming or pre-defined instructions.

1990-1999

1. Development of ERS-110 Robot Pet Dog (1999): The ERS-110 was a robotic pet dog created by Sony, part of the AIBO (Artificial Intelligence Robot) series. It featured

advanced sensors, motors, and artificial intelligence to mimic the behavior of a real dog. The ERS-110 could recognize voice commands, respond to touch, and express emotions through movement and sound. It marked a significant advancement in consumer robotics, demonstrating the potential for robots to interact with humans in a more personal and emotional way. Deep Blue Defeats Kasparov (1997): IBM's chess-playing computer, Deep Blue, defeated world champion Garry Kasparov, marking a significant milestone in AI's capability to outperform humans in complex problem-solving tasks.

2. Development of ALICE (1995): The Artificial Linguistic Internet Computer Entity (ALICE), created by Richard Wallace, was a natural language processing chatbot. ALICE utilized a scripting language called AIML (Artificial Intelligence Markup Language) to simulate conversation. While not as advanced as modern AI models, it contributed significantly to data collection and research in natural language processing (NLP).

2000-2009

1. Advancement in Recommendation Technology: During this decade, recommendation systems saw rapid growth and development, powered by machine learning algorithms and increased computational capabilities. Companies like Amazon, Netflix, and Spotify popularized recommendation technology by leveraging algorithms to suggest products, movies, and music to users based on their preferences, past behaviors, and the behavior of similar users.
2. Development of Roomba (2002): Roomba, developed by iRobot, was one of the first commercially successful autonomous robotic vacuum cleaners. Roomba used a combination of sensors, motors, and algorithms to navigate and clean floors without human intervention. It could detect dirt and adjust its cleaning pattern accordingly, making it an early example of practical, autonomous robotics in the home.
3. Development of Watson (2007): IBM's Watson was a significant leap in the field of artificial intelligence, aimed at advancing natural language processing (NLP) and machine learning. Watson was designed to understand and process human language, allowing it to answer complex questions posed in natural language. It gained worldwide recognition in 2011 when it competed on the popular quiz show Jeopardy! against two of the show's greatest champions, Ken Jennings and Brad Rutter, and won.

2010-2019

1. Launch of Siri (2011): Siri, Apple's virtual assistant, was released in 2011 as part of the iPhone 4S. Siri is a voice-activated assistant that uses natural language processing and machine learning to interpret and respond to user commands. It can perform a variety of tasks, such as sending messages, setting reminders, providing weather updates, and answering general knowledge questions. Development of Never Ending Image Learner (NEIL) (2012): NEIL, developed by researchers at the University of Washington, was an AI system designed to learn to recognize objects in images by continuously processing and interpreting data from the internet. NEIL was significant because it demonstrated the ability to learn from raw, unstructured visual data and improve its recognition capabilities without human intervention or labeled datasets.
2. Launch of Cortana (2014): Cortana, developed by Microsoft, was one of the leading virtual assistants introduced during this decade. Named after an AI character in the Halo video game series, Cortana was designed to assist users with tasks such as setting reminders, managing calendars, answering questions, and providing information based on voice commands. Cortana integrated with Microsoft services and apps, including Outlook and Microsoft Office, to streamline user productivity.
3. AlphaGo's Victory (2016): DeepMind's AlphaGo defeated world champion Go player Lee Sedol. Go was considered a game of high complexity and intuition, and this achievement showcased AI's advancements in reinforcement learning and strategy.

2020-now

1. Rise of Generative AI Tools:
    a) Text-Based Tools (e.g., GPT-3, ChatGPT): These models excel in generating human-like text for various applications such as content creation, coding assistance, customer service automation, and academic research. GPT-3 and GPT-4, developed by OpenAI, are some of the most prominent models used to power generative AI chatbots like ChatGPT.

b) Image Generation (e.g., DALL·E, Stable Diffusion, MidJourney): These AI models can generate realistic images from textual descriptions. For example, OpenAI's DALL·E can create unique visuals from simple text prompts, while MidJourney and Stable Diffusion also enable users to produce high-quality artistic visuals. These tools are increasingly used in design, advertising, and art creation.
c) Video Generation (e.g., Runway ML, Synthesia): Video generation tools are becoming more sophisticated, allowing users to create AI-generated videos based on text input or pre-existing media. For example, Synthesia can create AI-generated avatars that speak in different languages or narrate text, often used for training and marketing videos.
d) Music Composition (e.g., OpenAI's Jukedeck, Amper Music): AI-generated music tools like Amper Music and Jukedeck allow users to create original music compositions based on mood, genre, or style, using machine learning models trained on vast music datasets.
e) Code Generation (e.g., GitHub Copilot): AI tools like GitHub Copilot, powered by OpenAI's Codex, help developers by suggesting code snippets or completing programming tasks based on natural language input, speeding up the software development process.

Each event of a decade to be answered by different students. They are to find the answers from any references. Reference source must be indicated.

3. Name an artificial intelligence application that is created by a Malaysian company. Briefly describe its A.I. functions. ()

Answer:

1. AskAILA is an artificial intelligence (AI) legal assistant developed by the Malaysian law firm Shang & Co. It is designed to provide accessible legal advice at a lower cost than traditional law firms, particularly in the area of labor law. The AI is trained and equipped with knowledge of Malaysian labor laws and regulations, and it can provide assistance 24/7
2. Aerodyne is a global enterprise solutions provider specializing in drone technology, data technology, and digital transformation (DT3). It is ranked the world's #1 drone service provider in 2021 and 2022 by Drone Industry Insights (DII). The company helps integrate digital technology into various areas of businesses, changing how they operate and deliver value to their customers. Aerodyne's services are designed to be faster, better, cheaper, and safer. The company has been providing global DT3 solutions since 2014. It helps organizations leverage drone data and AI to resolve industrial challenges, enabling digital transformation, optimal operations, productivity gains. Its solutions cover infrastructure asset management, precision agriculture, oil & gas, telecoms, construction, utilities, and more.
3. Billion Prima Sdn Bhd is a Malaysian company established in 2007. Its office is located at No. 2C-3, Jalan Jubli Perak 22/1 Section 22, 40400 Shah Alam, Selangor. The company specializes in the assembly, supply, and maintenance of scanning machines. With over 15 years of experience, Billion Prima has developed and deployed cargo scanners and has transformed from a service maintenance-oriented firm to a world-class cargo x-ray scanner R&D and manufacturing company. The company's cargo inspection system is designed for manifest verification, homeland security, and contraband interdiction. Billion Prima won the Security Technology - Cargo Handling category award at the Malaysia Technology Excellence Awards in 2023
4. MeLAB Global Sdn Bhd is a tech company established in 2016 in Malaysia (registration number 1233436H/201701019271) by Bong Kum Tim, Bong Mei Lai, and Alexander Henderson. Their business includes mobile apps, web application programming, website development, blockchain, machine learning, AI, and the education industry. MeLAB has developed facial recognition and license plate recognition (LPR) technologies using AI for security in residential buildings. Their system offers automatic license plate detection and recognition. Their facial recognition system allows door access control, offering convenience and security without need for passcards. It facilitates hands-free operation.

MeLAB claims their system can work offline and online, providing a complete hardware and software solution. This allows operation even during server downtimes. They offer rental, leasing or purchase options for their systems. Their customers range from low cost flats to luxury high rise buildings. MeLAB Global office is located at: B1-1-W9, Pusat Perniagaan Sri Ehsan, Jalan Sri Ehsan 15, Taman Sri Ehsan, Kepong 52100 Kuala Lumpur, Malaysia. In summary, MeLAB Global is a Malaysian tech startup focused on AI, facial recognition and license plate recognition systems for security in residential buildings. Their systems offer convenience, security and savings.

Student to find the answer from any reference. Reference source must be indicated

4. Differentiate between the following, then provide an example for each type of system.
   a) Systems that think like humans. ()
   b) Systems that act like humans. ()
   c) Systems that think rationally. ()
   d) Systems that act rationally. ()

Answer:

System that thinks like humans - The cognitive modelling approach

System that thinks like humans is that system that has the capability of self-learning, pattern recognition and logical thinking. This requires "getting inside" of the human mind to see how it works and then comparing our computer programs to this. This is what cognitive science attempts to do. The system is designed to solve problems by thinking, reasoning, and remembering, to mimic the way the human brain works Machine Learning and vision system (pattern recognition, object detection)

Example: IBM's Watson—Watson uses machine learning and natural language processing to simulate human-like thinking. It can understand, process, and respond to human language, often in complex scenarios like answering questions on Jeopardy! or providing diagnostic support in healthcare.

System that acts like humans - The Turing Test approach

Systems that act like humans are systems that have human characteristics and reflect the human condition/ has the capability and characteristics of acting like a person or doing human work. It can perform human's level action or interaction with minimal human assistance like communicating with humans in human language without understanding the meaning of the action. These programs must behave according to certain normal conventions of human interaction in order to make themselves understood. The underlying representation and reasoning in such a system may or may not be based on a human model.

Example 1: Siri or ChatGPT—These virtual assistants can understand and respond to natural language queries, providing information, setting reminders, and assisting with tasks, much like how a human would interact with someone in a conversation.

Example 2: Roomba—The autonomous robotic vacuum cleaner mimics human actions by navigating the home and cleaning the floor, using sensors to avoid obstacles, much like how a human would move around while vacuuming.

System that acts rationally – Agent-based system

A system that can act autonomously and independently in accordance with goals that are either created or predefined. The system is sensitive to its environment and can react to changes. Acting rationally means acting so as to achieve one's goals, given one's beliefs. An agent is just something that perceives and acts.

Example: Self-driving cars—Autonomous vehicles make decisions based on data from their environment, using algorithms to act in ways that optimize for safety, speed, and efficiency, without mimicking human-like behavior.

System that thinks rationally - Logic-based system

Thinking rationally means making inferences based on rules. Systems that think rationally use symbolic logic to represent the laws of rational thought as symbols that can be manipulated. Reasoning involves manipulating these symbols according to well-defined rules, similar to how algebra works. The outcome is an idealized model of human reasoning. Thinking rationally involves the use of logic, where reasoning processes are applied to derive new representations of the world and determine how to respond to them.

Example: MYCIN—An expert system for diagnosing bacterial infections and recommending antibiotics. It used a series of logical rules to evaluate symptoms and test results to make medical decisions, based on a rational, rule-based approach rather than human intuition.

5. Describe input (conditions), processing layer, output and how the agent react for each of the system below:
    i. Agent of an automated class scheduling system. ()
    ii. Agent of an intelligent air-conditioning system. ()
    iii. Agent of autonomous driving car. ()

Answer:

Agent of an automated class scheduling system

Input Layer:
a) Course requirements (e.g., subjects, timings, prerequisites).
b) Faculty availability.
c) Classroom resources and capacity.
d) Student preferences and availability.

Processing Layer:
a) Knowledge Base: Stores rules for scheduling, such as no overlapping classes and prioritizing required courses.
b) Inference Engine: Matches input data with rules to create a feasible schedule.
c) Conflict Resolution Module: Handles overlapping requests or resource conflicts.

Output Layer:
a) Generates a complete class schedule.
b) Provides feedback or options for adjustments (e.g., alternate slots for conflicting courses).

Reaction:
The agent analyzes inputs, resolves conflicts, and outputs a schedule optimized for available resources and constraints.

Agent of an intelligent air-conditioning system

Input Layer:
a) Room temperature (via sensors).
b) User preferences (e.g., desired temperature and humidity).
c) Environmental data (e.g., outdoor temperature).

Processing Layer:
a) Control Algorithm: Determines the required cooling/heating adjustments.
b) Learning Module: Adapts to user behavior over time to optimize settings.

Output Layer:
a) Adjusts AC temperature, fan speed, and mode (cooling, heating, etc.).
b) Sends notifications (e.g., "Temperature set to 22°C").

Reaction:

The agent monitors the room environment, processes user input and external conditions, and adjusts settings in real-time for optimal comfort and energy efficiency.

Agent of autonomous driving car

Input Layer:

a) Sensors: Cameras, LiDAR, radar, GPS, and ultrasonic sensors to detect the environment.
b) User Input: Destination and driving preferences.
c) External Data: Traffic updates, weather conditions, and road maps.

Processing Layer:

a) Perception Module: Analyzes the environment, identifying lanes, obstacles, and traffic signs.
b) Decision Module: Plans the route, determines the next actions (e.g., braking, turning, accelerating).
c) Control Module: Executes decisions by controlling steering, acceleration, and braking.

Output Layer:

a) Safe navigation and real-time adjustments based on inputs.
b) Notifications to the user (e.g., "Approaching destination").
c)

Reaction:
The agent continuously processes sensory inputs and external data to make decisions and control the car, ensuring safe and efficient navigation.

Students can provide any other relevant answers

Tutorial 2

Instead of asking, "Can machines think?", Alan Turing said we should ask, "Can machines pass a behavior test for intelligence?". Turing predicted that by the year 2000, a computer could be programmed to have a conversation with a human interrogator for five minutes and would have a 30% chance of deceiving the interrogator that it was a human. (Negnevitsky, 2002).

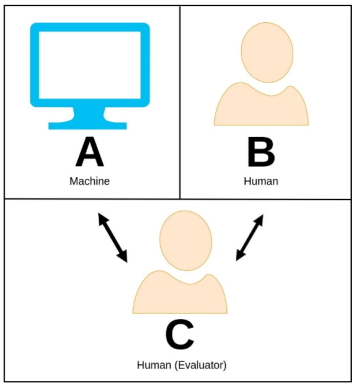
1. Explain Turing Test. Illustrate your answer with appropriate diagram.

Answer:

The Turing Test typically involves three participants:

1. A human judge (interrogator): This person conducts the test by engaging in a conversation with both a human and a machine.
2. A machine: The AI being tested for its ability to exhibit intelligent behavior.
3. A human: To provide a baseline for comparison with the machine's responses.

The judge communicates with both the human and the machine via a text-based interface, such as a computer keyboard and screen, so the judge cannot see or hear them directly. This ensures that the evaluation is based solely on the responses and not on physical appearance or voice. During the test, the judge asks questions and engages in a conversation with both the human and the machine. The machine's goal is to respond in a manner that is indistinguishable from the human participant. If the judge cannot consistently tell which participant is the machine, the machine is considered to have passed the Turing Test (Intelligent)



2. Criticize Turing's criteria for judging a computer's intelligence.

Answer:

Lack of Contextual Understanding: The Turing Test emphasizes conversational mimicry rather than true understanding or problem-solving. A machine might generate convincing responses without comprehending the context, meaning, or implications of the conversation.

Limited Scopes (Multi-Modal Intelligence): In the Turing Test, the interrogator can only communicate with the machine through language, without any visual or physical cues. Having language skills alone is not enough to prove that a machine has human-level intelligence.

Ignores Non-Verbal Intelligence: The Turing Test primarily focuses on linguistic and conversational intelligence. It doesn't account for other forms of intelligence, such as problem-solving, creativity, or emotional intelligence.

Students can provide any other relevant answers

3. Suggest how could this test be used (or modified) to assess other kind of artificial intelligence besides a chatbot. Provide an example to elaborate your answer.

Answer:

Without modified: CAPTCHA, (Completely Automated Public Turing test to tell Computers and Humans Apart) is inspired by the principles of the Turing Test, but it operates in a reverse manner. While the Turing Test evaluates whether a machine can act like a human, CAPTCHA determines whether a user is a human or a machine.

Purpose:

a) Turing Test: Designed to evaluate a machine's ability to mimic human intelligence.
b) CAPTCHA: A test used to distinguish humans from machines, specifically bots, by presenting challenges that are easy for humans but difficult for AI to solve.

Example:

a) Text-Based CAPTCHA:

Distorted or obscured text is presented, requiring the user to type what they see. Humans can usually recognize patterns despite distortions, while many machines struggle.

b) Image-Based CAPTCHA:

Users are asked to select all images containing specific objects (e.g., traffic lights, cars). This tests visual recognition, which is generally easier for humans than for machines, though AI has made significant progress in this area.

Modified

Task-Oriented Evaluation:

Instead of testing conversational ability, the test could be applied to measure how well an AI performs tasks requiring intelligence within a specific domain, such as decision-making, creativity, or problem-solving.

For instance, AI systems could be assessed on their ability to make medical diagnoses, recommend financial investments, or design solutions to engineering problems.

Incorporation of Contextual Understanding:

The test could require the AI to demonstrate its understanding of context, cultural nuances, or real-world scenarios. This would ensure the AI's responses are not only accurate but contextually appropriate.

Example: An AI tasked with responding to a patient's medical history would need to tailor its recommendations to the patient's specific conditions rather than offering generic advice.

Multi-Modal Interaction:

Beyond text-based communication, AI could be tested on its ability to process and respond to inputs from multiple modalities, such as images, videos, or sensor data.

Example: A self-driving car could be evaluated based on its ability to interpret visual data from its cameras and make safe driving decisions in diverse environments.

Students can provide any other relevant answers

4. The Loebner Prize is an annual competition in artificial intelligence that awards those computer programs considered by the judges to be the most human-like, using format of a standard Turing Test. The conversation scope between the programs and the judges has been unrestricted since 1995, and the duration of the conversation has been increased from 5 minutes to 25 minutes since 2010 (http://www.loebner.net/).
   a) Discuss **TWO** (**2**) reasons why Turing Test is considered **not effective enough** in assessing machine intelligence.
   b) Discuss **TWO** (**2**) challenges to build a computer program that can win the Grand Loebner Prize, in which judges totally cannot distinguish it from a real human.

Answer:

a)

1. Emphasis on Imitation, not Understanding:

The Turing Test primarily focuses on a machine's ability to imitate human conversational behavior, rather than genuinely understanding the content or meaning of the conversation. A machine can pass the test by producing responses that mimic human language and behavior without possessing true comprehension or intelligence. This means that the test may not effectively differentiate between machines that simulate understanding and those that actually understand the information they are processing. In essence, it places more importance on surface-level mimicry rather than deep comprehension or intelligence.

2. Lack of Objective Measurement:

The Turing Test relies on human judges to determine whether a machine's responses are indistinguishable from those of a human. This introduces subjectivity into the assessment, as the judgment may vary from one human evaluator to another. Additionally, there are no clear and objective criteria for passing the test. The lack of a quantifiable measure makes it challenging to establish a standardized benchmark for machine intelligence. As a result, the test's outcomes can be inconsistent, and it may not provide a robust and reliable way to assess and compare the intelligence of different machines.

Any other acceptable answer

b)

1. Natural Language Understanding and Contextual Awareness:

One of the foremost challenges is developing natural language understanding capabilities that allow a computer program to comprehend and respond to a wide range of conversational topics, nuances, and context. Human language is incredibly complex, often involving idiomatic expressions, cultural references, sarcasm, humor, and ambiguity. To pass as a human convincingly, a program must possess the ability to grasp not only the surface-level meaning of words but also the deeper contextual implications. Achieving this level of language understanding requires advanced natural language processing (NLP) techniques, large datasets, and machine learning algorithms. Even with these tools, building a program that can consistently handle a diverse array of conversational scenarios and maintain contextual coherence is a substantial challenge.

2. Emotional and Empathetic Responses:

Humans often convey emotions and empathy in their conversations, which can be crucial for creating a lifelike conversational AI. Developing a computer program that can generate authentic emotional responses is a complex task. It involves not only recognizing emotional cues in the input but also generating empathetic and contextually appropriate responses. While sentiment analysis techniques can help identify emotions in text, conveying empathy and emotions in a way that resonates with human judges is a formidable challenge. It requires understanding the emotional context of a conversation and crafting responses that reflect genuine empathy, care, or other emotional states. Achieving this level of emotional authenticity without overstepping ethical boundaries or appearing insincere is a delicate balancing act.

3. Passing the Total Turing Test:

The Grand Loebner Prize requires a computer program to pass not just a text-based Turing Test, but a Total Turing Test. This means the program must be able to interact with the world in the same way a human does, including understanding and responding to visual and auditory input. This adds an additional layer of complexity to the challenge, as it requires the program to have capabilities in areas like computer vision and speech recognition, in addition to natural language processing and understanding

4. Complexity of human language:

The complexity of human language presents a multifaceted challenge. Different languages necessitate distinct grammatical structures, making language understanding and generation a demanding task. Moreover, humans continually invent new words and expressions, exemplified by recent additions like "Covid-19" and "yyds."

Any other acceptable answer

5. The Chinese room argument by John Searle is one of the best known and widely credited criticism of Turing Test. Explain John Searle's Chinese room concept.

Answer:
The Chinese room argument holds that a digital computer executing a program cannot be shown to have a "mind", "understanding" or "consciousness", regardless of how intelligently or human-like the program may make the computer behave.

Searle imagines himself alone in a room following a computer program for responding to Chinese characters. Searle understands nothing of Chinese, and yet, by following the program for manipulating symbols and numerals just as a computer does, he sends appropriate strings of Chinese characters back out under the door, and this leads those outside to mistakenly suppose there is a Chinese speaker in the room. John Searle describes that although a system is clearly running a program and passes the Turing Test, it does not equally understand anything of its inputs and outputs. In conclusion, running the right program does not necessarily generate understanding.

6. Try to chat with the following chatbots within a few minutes. Then discuss what are the characteristics/behaviors of a chatbot should have in order to deceive any human.
   a) **Mitsuku**, the 5-time Loebner Prize winner - https://www.pandorabots.com/mitsuku/
   b) **Eliza**, the first chatbot - https://web.njit.edu/~ronkowit/eliza.html (not the original Eliza website)

Answer:

Mitsuku is an award-winning chatbot developed by Steve Worswick. Known for its advanced conversational abilities, Mitsuku possesses several key characteristics that contribute to its popularity and success:

1. Natural Language Understanding: Mitsuku demonstrates a strong ability to comprehend and interpret human language. It can understand context, extract meaning from complex sentences, and handle conversational nuances, allowing for more engaging and natural interactions.
2. Contextual Memory: Mitsuku has a robust memory that enables it to remember and recall previous conversations. This contextual memory allows it to maintain continuity in dialogues and provide personalized responses based on past interactions, creating a more personalized user experience.
3. Emotional Intelligence: One notable characteristic of Mitsuku is its emotional intelligence. It can recognize and respond to human emotions, showing empathy and understanding in its conversational style. This ability to empathize with users contributes to a more engaging and authentic interaction.
4. Conversational Depth: Mitsuku is capable of engaging in extended and meaningful conversations on a wide range of topics. It displays a substantial knowledge base, covering subjects such as general knowledge, trivia, news, and even personal questions, making it versatile and capable of providing informative responses.
5. Personality: Mitsuku exhibits a distinct personality, characterized by a friendly and conversational tone. It incorporates humor, wit, and occasionally playful responses, making interactions with the chatbot more enjoyable and engaging for users.
6. Continuous Learning: Mitsuku leverages machine learning techniques to continually improve its conversational abilities. It learns from user interactions, allowing it to adapt and refine its responses over time. This enables Mitsuku to grow and evolve, providing users with an increasingly refined conversational experience.
7. Multiple Language Support: Mitsuku supports multiple languages, including English, Spanish, French, German, and Italian. This multi-language capability extends its accessibility to a broader range of users, facilitating interactions with individuals from different linguistic backgrounds.

Overall, Mitsuku's key characteristics, including its natural language understanding, contextual memory, emotional intelligence, conversational depth, personality, continuous learning, and multi-language support, contribute to its reputation as a highly sophisticated and engaging chatbot.

Eliza is a classic chatbot developed in the 1960s by Joseph Weizenbaum. Although considered simplistic by today's standards, Eliza played a significant role in the history of chatbots and psychotherapy. Here are some of the key characteristics of Eliza:

1. Rule-Based Approach: Eliza operates on a rule-based approach, utilizing a set of predefined patterns and responses. It uses pattern matching techniques to identify keywords or phrases in user input and generates pre-programmed responses based on those patterns.
2. Reflection of User Input: One notable characteristic of Eliza is its ability to reflect user input. It achieves this by rephrasing or repeating certain parts of the user's statements as questions. This technique, known as "Rogerian psychotherapy," helps simulate a conversation by encouraging users to explore their thoughts and feelings further.
3. Lack of Genuine Understanding: Eliza does not possess genuine understanding or semantic comprehension. It primarily focuses on mimicking human-like conversation by utilizing generic and context-independent responses. It does not process or analyze the meaning of user input beyond simple pattern matching.
4. Minimal Context Awareness: Eliza lacks deep context awareness and does not maintain a memory of previous interactions. It treats each user input independently, without considering the larger conversation history. As a result, Eliza's responses do not evolve or adapt based on the ongoing conversation.
5. Limited Scope: Eliza's capabilities are confined to basic psychotherapy interactions. It aims to engage users in open-ended discussions about their feelings and concerns, employing techniques such as active listening, reflection, and probing questions. However, it lacks the ability to engage in conversations on diverse topics outside the scope of psychotherapy.
6. Text-Based Interface: Eliza operates through a text-based interface, where users input text messages that the chatbot responds to. It does not incorporate any visual or multimedia elements and solely relies on textual interactions to simulate a conversation.
7. Influence on Chatbot Development: Despite its limitations, Eliza has had a significant impact on the development of chatbot technology. Its rule-based approach and the concept of using reflective techniques to simulate conversation laid the foundation for subsequent chatbot iterations, inspiring further advancements in natural language processing and dialogue systems.

In summary, Eliza's key characteristics include its rule-based approach, reflection of user input, lack of genuine understanding, minimal context awareness, limited scope, text-based interface, and its historical significance in shaping the field of chatbot development.

Tutorial 3

1. Discuss the advantages and disadvantages of **breadth-first search** and **depth-first search**.

Answer:

	Breadth-first search	Depth-first search
Advantages	1) Guarantee a result if it exists 2) Guarantee the shortest path 3) If the goal is at higher level, then the search is fast, hence consumption of memory could be less	1) Guarantee a result if it exists 2) Usually requires lesser memory as it only stores the nodes in the search path 3) Usually the search time is faster
Disadvantage	1) Needs more memory to store all branches 2) Large branching factors will lead to longer search time	1) Cannot guarantee shortest path 2) It might stuck at infinite loop 3) The search path can be very long if the goal is very deep down, so may also consume more memory

2. **Figure 1** shows a puzzle problem that requires rearrangement of the tiles to transform the order from start to goal state. One is only permitted to slide the empty tile follow the sequence of **left**, **right**, **up** or **down.**
Remark: Avoid repeated state

Figure 1: The Puzzle Problem

a) Provide the goal formulation of the puzzle problem above.
b) Formulate the puzzle problem above by specifying the initial state, successor functions, goal test, step cost, and path cost.
c) Perform **breadth-first search** and **depth-first search** on the puzzle problem above. Draw the resulting search trees for both.

Answer:

a)

Goal formulation

Goal: to reach the goal state as shown in Figure 1

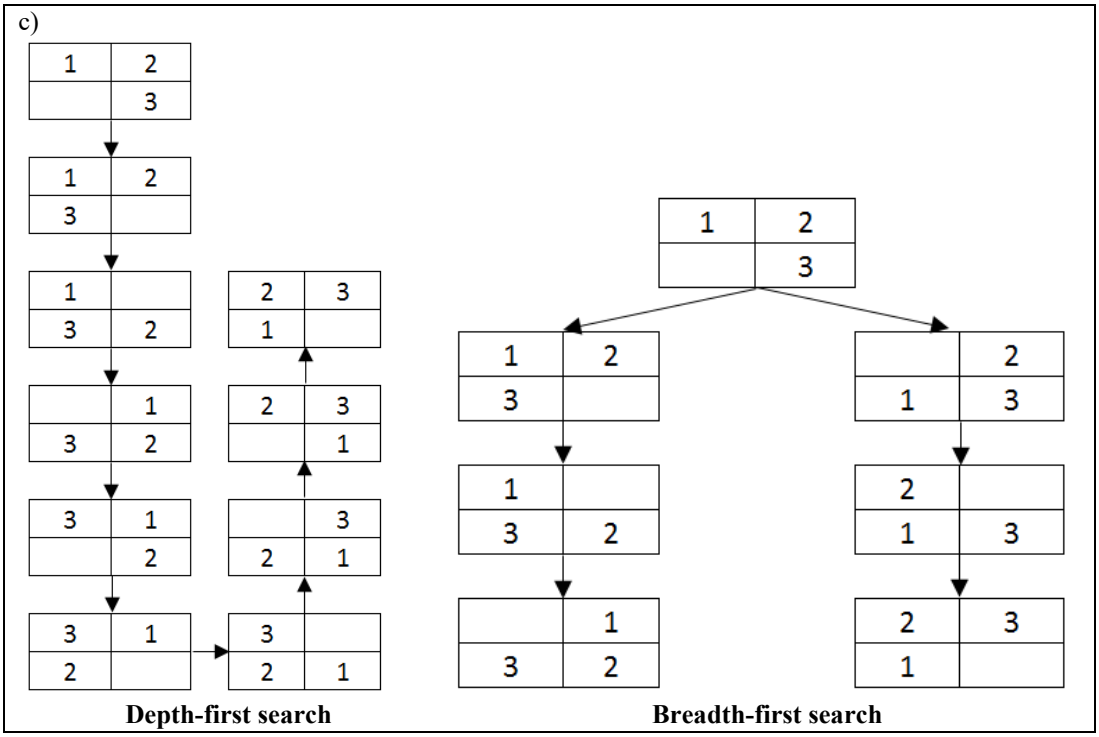
2	3
1	

Optimal solution: find the shortest path to reach the goal
Abstraction: time, material

b)

Problem formulation

Initial state: Start state
Successor function: All possible movements to successor state from current state
Goal test – to check if the goal state has been reached, otherwise continue to search
Step cost – the distance between one node to the other, e.g. each step is 1.
Path cost – the total step cost between start state and the goal state along a path



3. **Figure 2** shows a puzzle problem that requires rearrangement of the tiles to transform the order from start to goal state. One is only permitted to slide the empty tile follow the sequence of **down, right, up or left**
Remark: Avoid repeated state

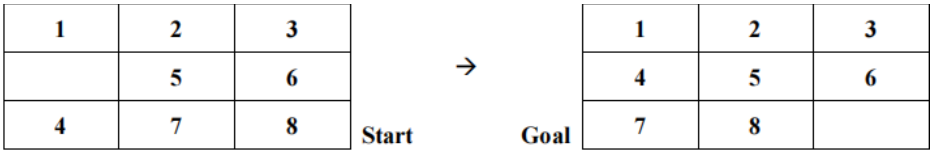
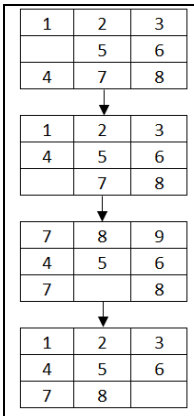


Figure 2: The 8-Puzzle Problem

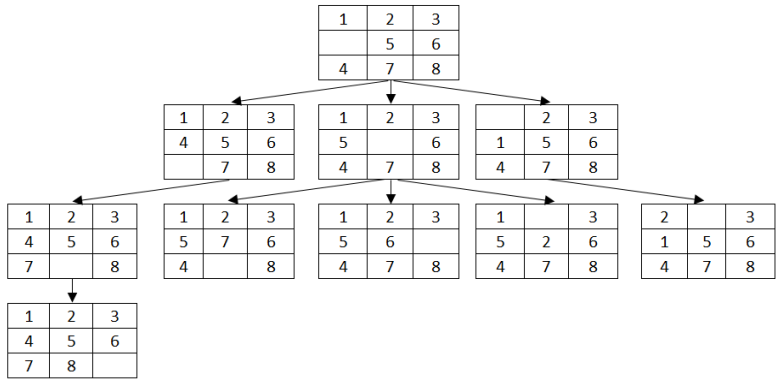
a) Perform **breadth-first search** and **depth-first search** on the 8-puzzle problem above. Draw the resulting trees for both. (Remark: you may stop the search at level 4)

b) Evaluate the efficiency of **breadth-first search** and **depth-first search** in terms of completeness, optimality, time efficiency and space efficiency in solving the problem above.

Answer:
a)
Successor function: Slide the empty tile follows the sequence of down, right, up and left



Depth-first search



Breadth-first search

b)

Efficiency	Breadth-first search	Depth-first search
Completeness	Yes	Yes
Optimality	Yes	Yes
Time-complexity	Consume more time	Consume less time
Space complexity	Consume more memory	Consume less memory

4. In the family river-crossing problem, two parents are with their two children - a son and a daughter - came to a wide river. The only way to get to the other side was to ask a fisherman if he could lend them his boat. However, the boat could carry only two persons. For safety reason, no child should be left alone without the supervision of at least one parent. The family must get to the other side and finally returns the boat to the fisherman, assuming only the fisherman and the two parents know how to row the boat. Based on this family river-crossing problem, answer the following questions.
   Remark: Avoid repeated state
   a) Suggest a simple representation of the initial state. You must briefly explain the representation.
   b) Describe the goal formulation and problem formulation.
   c) Draw the depth-first search tree to show how all the states are being traversed.

Answer:

a)

Representation:

[2, 2, 1, L, 0, 0, 0] == ([LP LC LF L RP RC RF])

[no. of parents at left side, no. of children at left side, no. of fisherman at left side, position of boat, no. of parents at right side, no. of children at left side, no. of fisherman at right side])

b)

<u>Goal formulation:</u>

Goal- to reach [0 0 0 R 2 2 1] (The family and fisherman at the right hand side)

Optimal solution- least number of step (trip) to reach the goal

Abstraction- no time limit in reaching the goal (eg. Do not care about the time used to reach the other side of the river, which can be caused by unexpected factors such as weather)

<u>Problem Formulation</u>

Initial state - [2, 2, 1, left, 0, 0, 0]

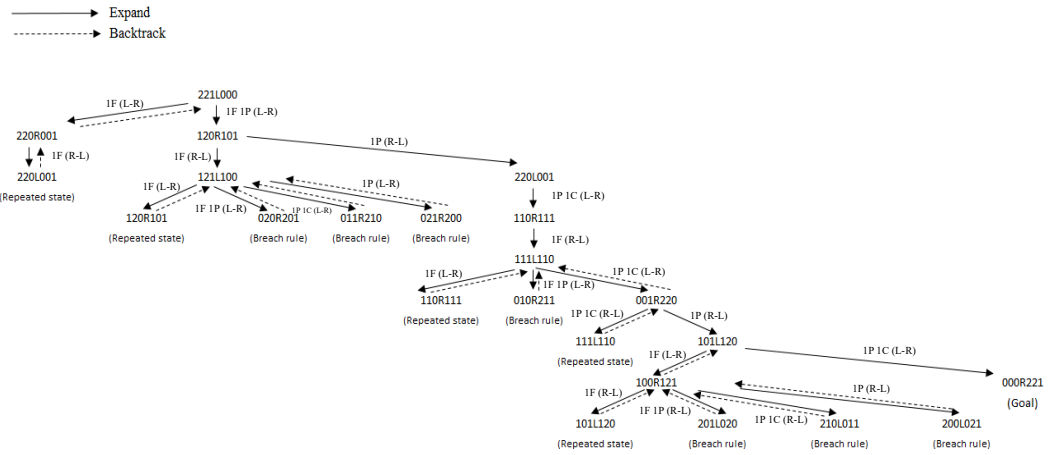
Successor Function - The possible actions available to the agent that will change the state from the current state.

	If boat at left side	If boat at right side
1F	L to R	R to L
1F+1P		
1P+1C		
1P		

Therefore, the successor functions are shown as below:
If boat at left:
[LP LC LF-1 R RP RC RF+1]
[LP-1 LC LF-1 R RP+1 RC RF+1]
[LP-1 LC-1 LF R RP+1 RC+1 RF]
[LP-1 LC LFR RP+1 RC RF]

If boat at right:
[LP LC LF-1 L RP RC RF+1]
[LP-1 LC LF-1 L RP+1 RC RF+1]
[LP-1 LC-1 LF L RP+1 RC+1 RF]
[LP-1 LC LF L RP+1 RC RF]
F: Fisherman; P:Parent; C:Child; L:Left; R:Right
2P and 1C are not feasible as it breach the rule of "no child should be left alone without the supervision of at least one parent"
Goal Test - To check whether the given state is a goal state. (state([0,0,0,R,2,2,1]))
Step Cost - The route distance from one state to another state. (1)
Path Cost - The sum of costs of the individual actions along the path.

c)



5. **Figure 3** below shows a directed graph. Assume that the traversal would start from **Vertex 0** to **Vertex 7**. All vertices to be visited in **ascending order** (i.e. from smaller number to bigger number). **Remark: Avoid repeated state**

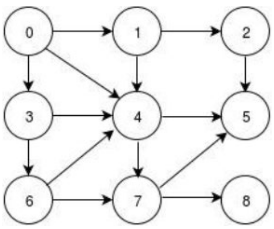
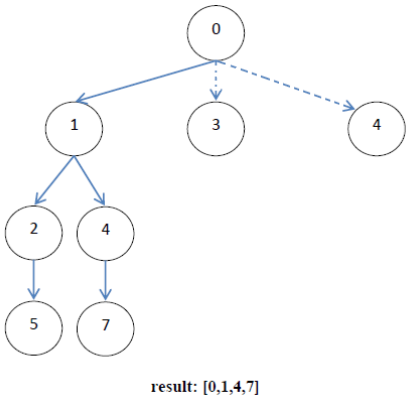


Figure 3: The Directed Graph

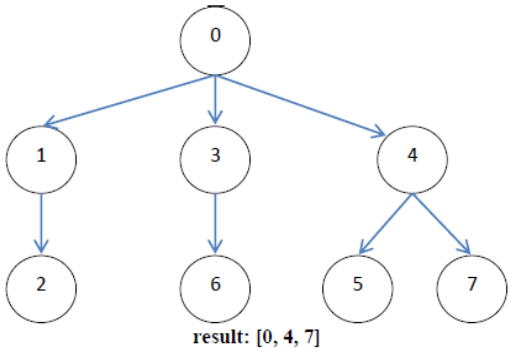
a) Perform a **depth-first search** on the directed graph to traverse from Vertex 0 to Vertex 7. Draw the resulting search ree and list the returned path.
b) Perform a **breadth-first search** on the directed graph to traverse from Vertex 0 to Vertex 7. Draw the resulting search tree and list the returned path. Avoid repeated state.
c) Evaluate the efficiency of **breadth-first search** and **depth-first search** in terms of completeness, optimality, time efficiency and space efficiency in solving the problem above.

Answer:

a) b)



Depth-first search



Breadth-first search

c)

Efficiency	**Breadth-first search**	**Depth-first search**
Completeness	Yes	Yes
Optimality	Yes	No
Time-complexity	Consume more time	Consume less time
Space complexity	Consume more memory	Consume less memory

Tutorial 4

1. **Figure 1.1** below shows a Block World Problem. A robot will move the blocks one by one from initial state S1 to reach the goal state S19.

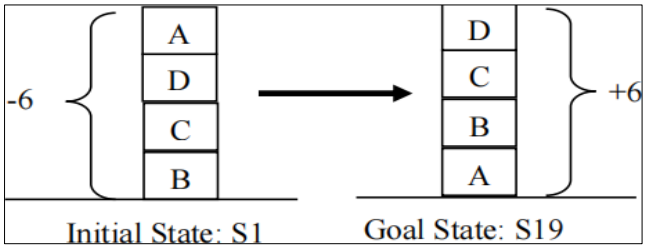


Figure 1.1

Figure 1.2 shows the state space of the Block World Problem and the heuristic costs for each state are shown in parentheses next to their respective nodes.

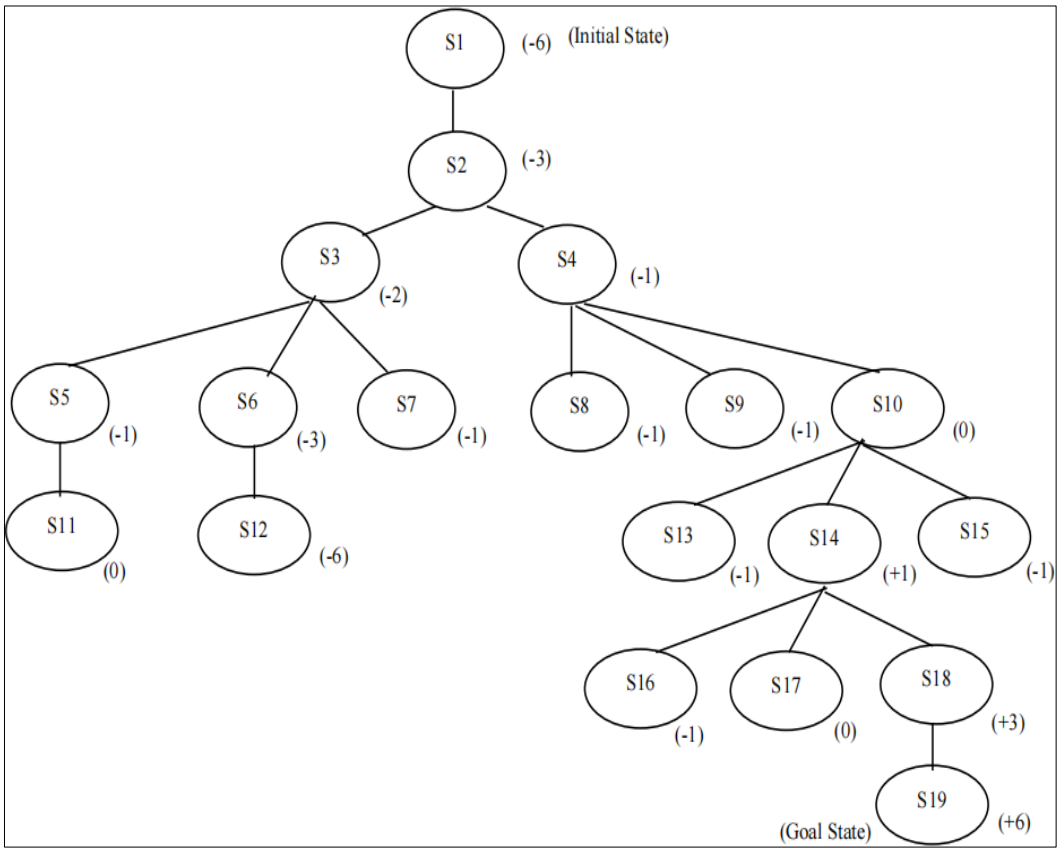


Figure 1.2

a) Explain step cost used in problem formulation. Specify the value of the step cost for the problem above.

b) Hill climbing search is unable to guarantee completeness and optimality as it may be trapped into local maximum.
   i. Explain local maximum.
   ii. Discuss why hill climbing search always lead to a local maximum.
   iii. Use simple hill-climbing and steepest-ascent hill-climbing to search for the best path from S1 to S19 on the state space shown in **Figure 1.2**. Then for each search technique,

draw the resulting search tree that shows the visited nodes. Show that hill-climbing technique can be trapped into a local maximum. (**Remark: Avoid repeated state)**

c) A search technique can be evaluated based on four criteria: completeness, optimality, time complexity and space complexity. Evaluate the efficiency of breadth-first search and steepest-ascent hill-climbing. Conclude which technique is better to solve the Block World Problem mentioned in **Figure 1.1**.

Answer:

a)

Step cost refers to the cost of an action. In other words, it is the cost of an action that transforms a state into another state. For instance, in the case of Block World problem, the step cost for all actions can be any but the same constant numeric value. However, in the case of path finding problem, step cost of an action might refer to the distance (km) or traveling time (seconds) between two states. Thus, most probably every step cost will have a different numeric value.
The value of step cost is 1.

b)(i)

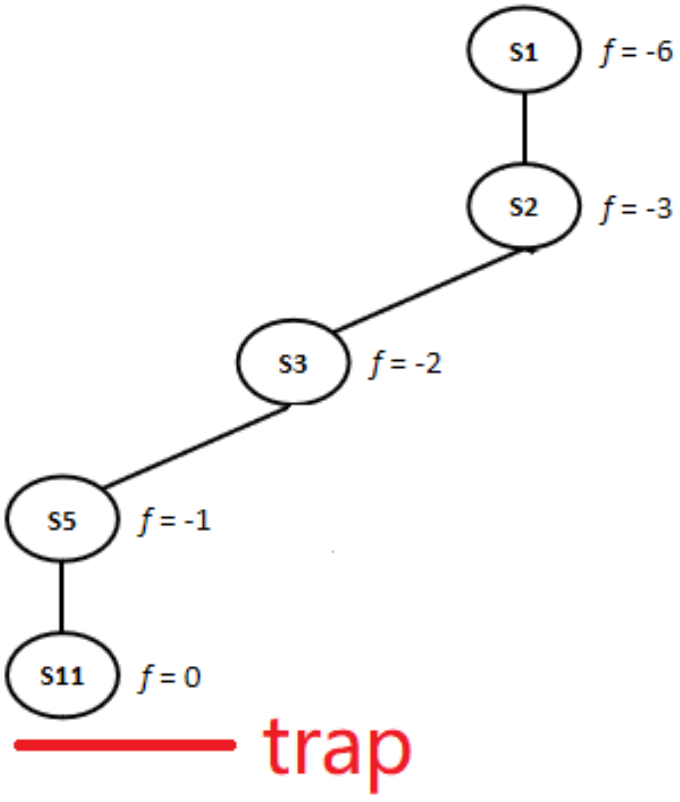
Local maximum refers to a state with no better child states or successors. The presence of local maxima makes a search fails to find a solution (e.g., 8-queen problem) or an optimal solution (e.g., travelling salesman problem), depending on the types of problem to be solved. For instance, in the case of 8-queen problem,
Foothill, plateau and ridge are 3 types of local maximum.

b)(ii)

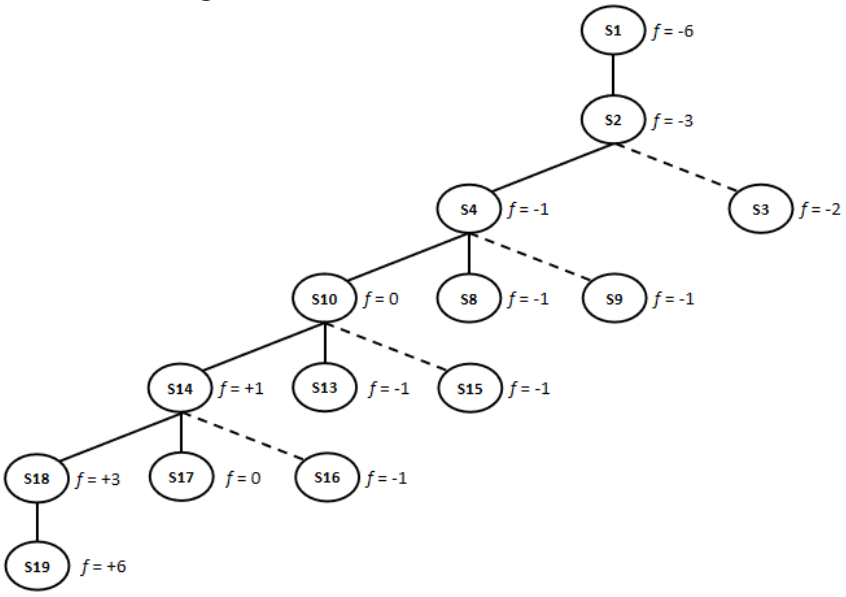
The nature of its uphill moves that keep looking for a better successor but unable to backtrack to any least promising unexplored path increases its tendency to reach and terminate at a peak (or local maximum) where no neighbor has a higher value.

b)(iii)

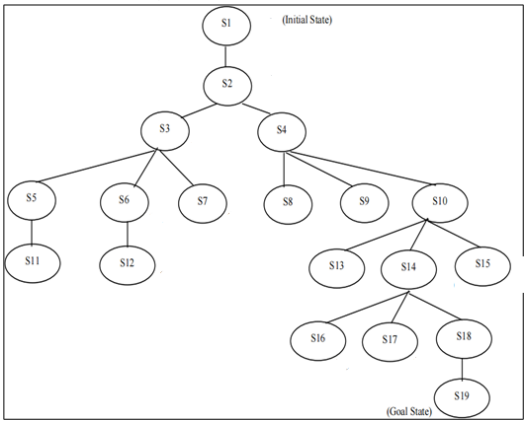
Simple hill climbing



Steepest ascent hill climbing



c) To answer this, student will need to draw the resulting search tree for BFS before comparing



Search tree for BFS

	Breadth-First Search	**Steepest Ascent Hill Climbing**
Completeness	Complete	Complete
Optimality	Optimal (if compare no. of station)	Optimal
Time complexity	Has a higher time consumption because it explores all paths by expanding all successors.	Has a lower time consumption because it explores only the most promising path by expanding only the best successor.
Space complexity	Has a higher memory consumption because it keeps track of all expanded nodes	Has a lower memory consumption because it doesn't keep track of all expanded nodes

Steepest ascent hill climbing is a better approach because it has a better performance in terms of time and space efficiency.

2. **Figure 2** below shows an 8-puzzle problem, which requires rearrangement of the tiles to transform the order from start state to goal state. One is only permitted to slide the empty tile in sequence of **up**, **down**, **right or left**. (**Remark: Avoid repeated state)**

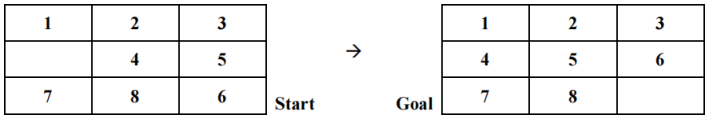


Figure 2: The 8-Puzzle Problem

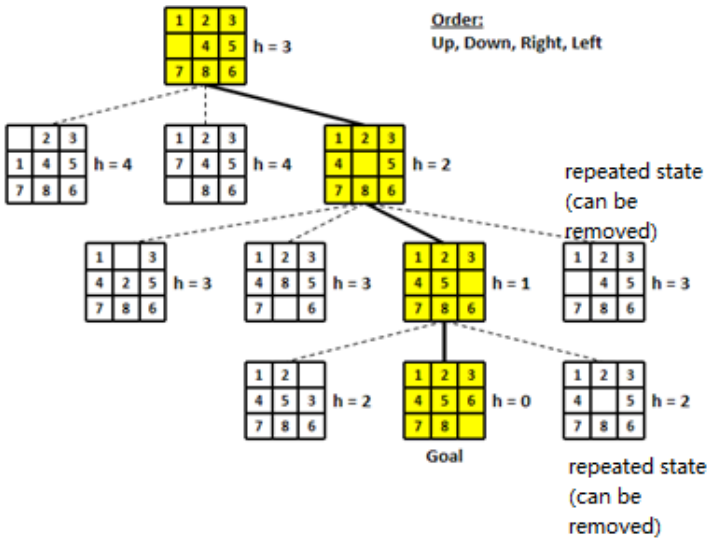
a) Suggest a heuristic function to produce a heuristic cost for a state. Demonstrate how such heuristic cost can be computed on the **start state**. Then perform best-first search.

b) Evaluate the efficiency of **breadth-first search** and **best-first search** in terms of completeness, optimality, time efficiency and space efficiency in solving the problem above.

c) Can hill-climbing find a solution for this problem? Draw a resulting search tree to support your answer.

Answer:

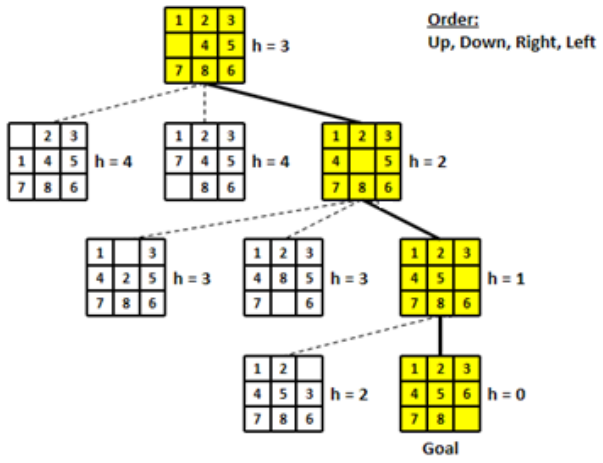
a)

Heuristic	Description
Hamming distance OR	The total number of misplaced tiles $f = 0 + 0 + 0 + 1 + 1 + 1 + 0 + 0 = 3$
Manhattan distance	The sum of the distances of each tile from its goal position $f = 0 + 0 + 0 + 1 + 1 + 1 + 0 + 0 = 3$
Best-First Search (assuming hamming distance is used in this example)	

b) To compare BFS and Best-first search, student will have to complete the search tree for BFS

	Breadth-First Search	Best-First Search
Completeness	Complete	Complete
Optimality	Optimal	Optimal
Time complexity	Has a higher time consumption because it explores all paths by expanding all successors.	Has a lower time consumption because it explores only the most promising path by expanding only the best successor.
Space complexity	Has a higher memory consumption because it keeps track of all expanded nodes	Has a lower memory consumption because it doesn't keep track of all expanded nodes

c)

Simple hill climbing	 Yes.

3. The following graph in Figure 1.1 shows all the nodes in a telecommunication network. The distance (in km) from one node to another is shown on the arc.

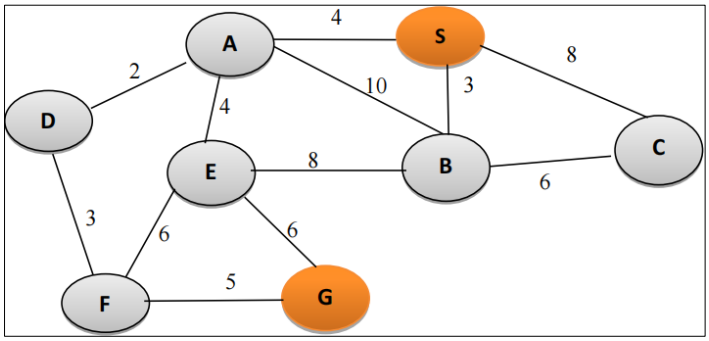


Figure 3: A search graph of a new LRT network

The Euclidean distance (in km), which is used as the heuristic cost (h) for different node, is provided in Table 1 below.

S	A	B	C	D	E	F	G
14	10	13	8	8	11	5	0

a) Assume that some data are to be sent from node S to node G using the shortest route. Describe the goal formulation and problem formulation.

b) Show the resulting search tree of A* search to find the shortest path from S to G. State the shortest path. (**Remark: Avoid repeated state)**

c) Evaluate the efficiency of A* search in solving the path-finding problem above

Answer:

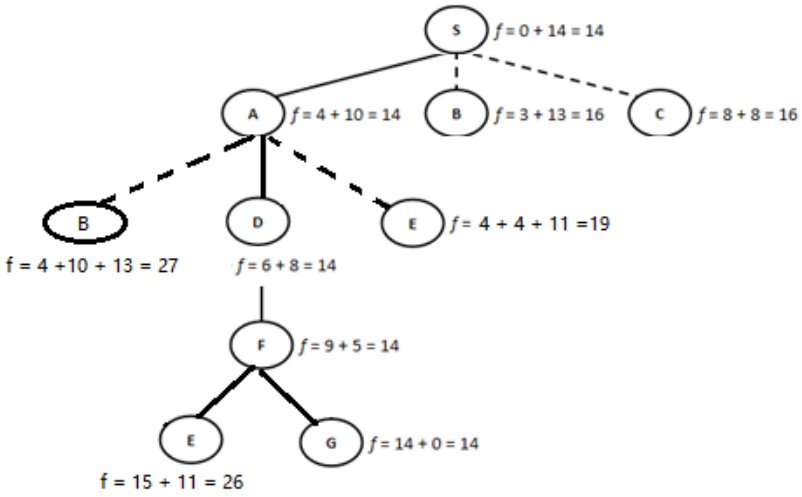
a)

Goal	Node G
Optimal Solution	The shortest path used to send data from node S to node G.
Abstraction	Ignore the chance that some data might be lost during transmission.

Initial state	Node S
Successor function	The possible actions available to the agent that will change the state from the current state OR a function to generate a successor state (output) from a current state (input).e.g From A generates S etc. OR all possible movement available to the agent.
Goal test	A test to check if a node is node G.

Step cost	The cost of transforming a state to another state, indicated by the value of an edge connecting any two nodes. E.g., the step cost of transforming S to A is 4.
Path cost	The total step costs along a path. E.g., the path cost for S-A-D is 4 + 2 = 6.

b)



The shortest path discovered b A* is S-A-D-F-G

c)

	A*
Completeness	Complete
Optimality	Optimal
Time complexity	Time efficiency is high because it explores only the most promising path by expanding only the best successor.
Space complexity	Space efficiency is high because it keeps track only expanded nodes.

4. Consider 2 heuristic h_1 and h_2 of A* for the puzzle problem are defined as:
   $h_1(n)$ = number of misplaced tiles
   $h_2(n)$ = total Manhattan distance

3	1
2	

Start State

Goal State

1	2
3	

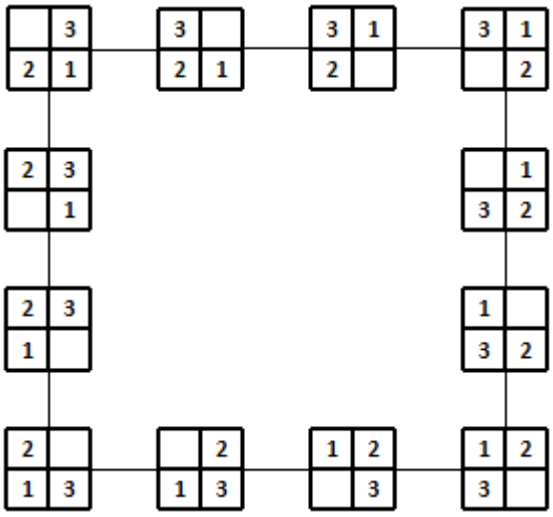
   a) Illustrate the **state space** of the puzzle to reach the goal state based on:
      - $h_1(n)$
      - $h_2(n)$

   b) Show the resulting search trees of A* search to find the shortest path using the heuristic functions of:
      - $h_1(n)$
      - $h_2(n)$

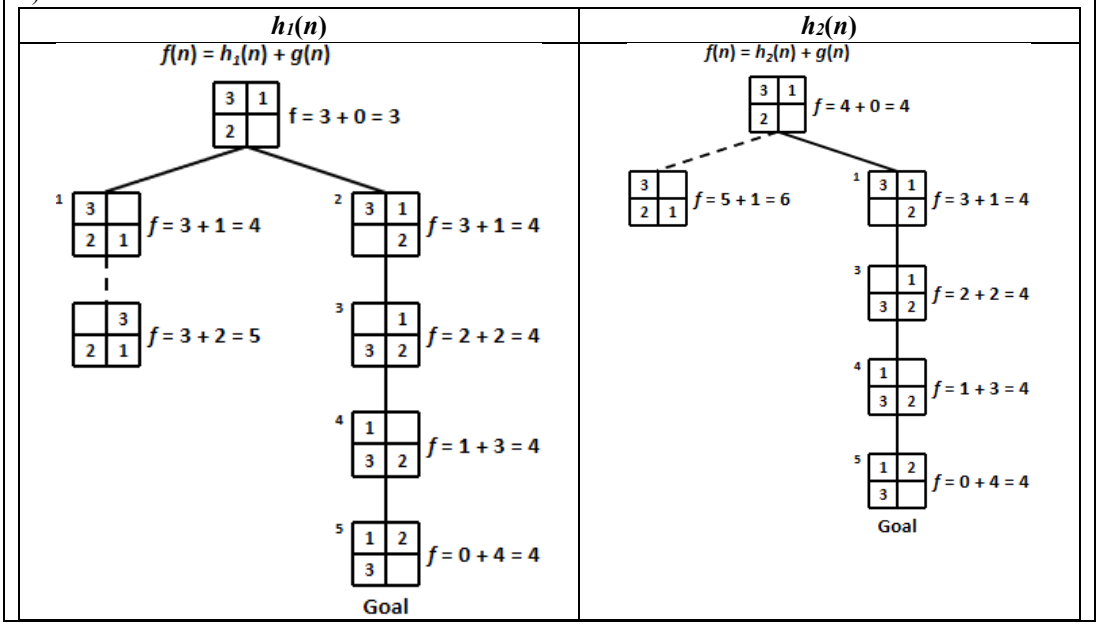
   (Remark: Avoid repeated state)
   You must clearly show the function cost, given that: $f(n) = h(n) + g(n)$, where $g(n)$ is the path cost.

Answer:
a)

A **state space** is a set of all possible states that it can reach from the start state.



b)

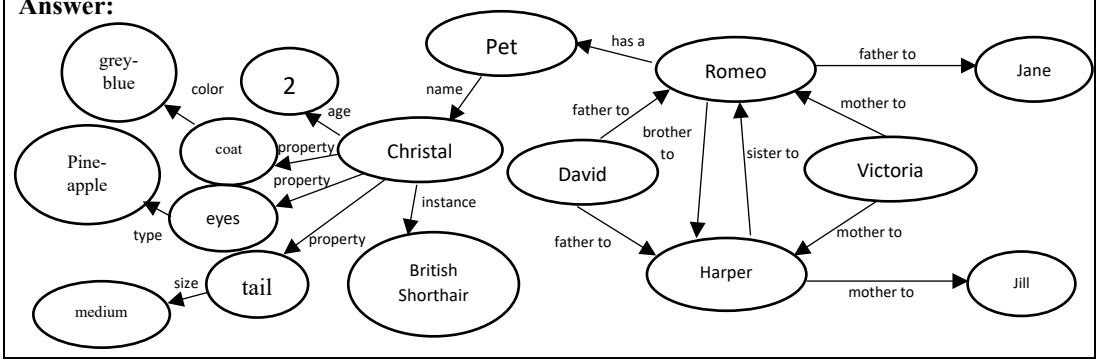


Tutorial 5

1. Semantic network is a knowledge base that represents semantic relations between concepts in a network. Draw a semantic network for the following facts and relations.

"David and Victoria are the parents of Romeo and Harper. Romeo and Harper are brother and sister. Jane is the child of Romeo, and Jill is the child of Harper. Romeo has a pet called 'Christal,' a British Shorthair. Christal is 2 years old now and has a grey-blue coat, pineapple eyes, and a medium-sized tail."

Answer:

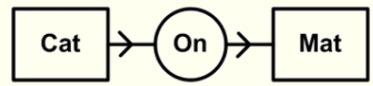


2. How are semantic networks different from Conceptual Graph? Explain your answer with the help of **ONE** (**1**) diagram for each of the methods based on the following statement:

"A cat is on a mat."

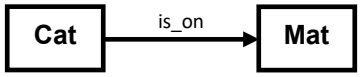
Answer:

Conceptual Graph



- The nodes of the graph are either concepts or conceptual relations.
- Do not use labeled arcs

Semantic Networks



- The nodes of the graph are either concepts, events, or action.
- Labeled arcs to show the relationships

3. The following statements are given.

Albert is a human.
Pepper is a robot.
The construction of human is biological, while robot is mechanical.
Both human and robot are autonomous system.
The behaviours of autonomous system are mobile and adaptive.

a) Between semantic network and frames, justify which knowledge representation tool you would use to represent the information above.
b) Discuss **ONE** (**1**) limitation of the knowledge representation tool that is selected in Question 3(a) above.

c) Illustrate the representation of the given statements above using the knowledge representation tool that you selected in Question 3(a) above.

Answer:

a)

Semantic network as it is more suitable to describe the relationship between objects and overall concepts. Or

Frames as frames are suitable to show the details of objects and their relationship.

Not description on syntax/semantic of a sentence, etc "bank"

b)

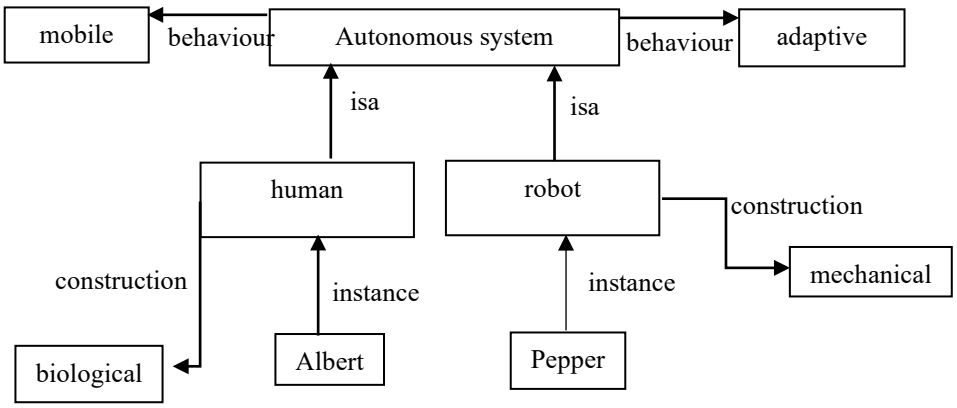
Semantic network is less organized and difficult to view the relationship between a rule and another.

Frames are only able to represent stereotyped objects, but are not able to represent events

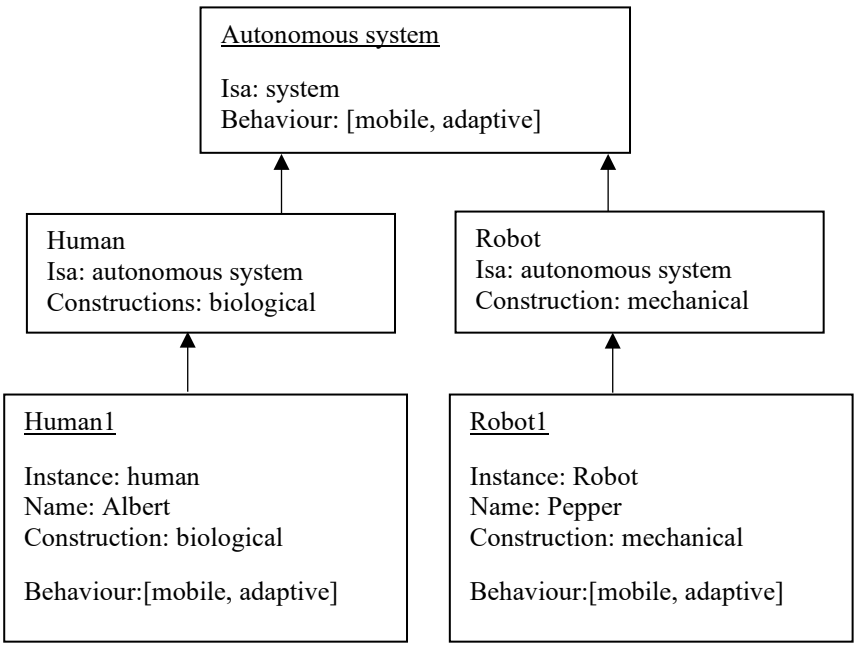
A generalized picture of a person, created without taking the whole person into account; to make such a generalization. Context: When we stereotype a group of people, we depict all of the individuals within that group as having the same characteristics.

Semantic shows height/weight/carry heavy load/size but frame is not able to represent it

Semantic Network



Frames



4. The following statements are given.

Bird is a living being that builds nest. Insect is an invertebrate living being. Ant is a kind of insects that builds nest. Both dragonfly and fly are insects that have wings.

a) Between semantic network and frames, justify which knowledge representation tool you would use to represent the information above.

b) Discuss **ONE** (**1**) limitation of the knowledge representation tool that is selected in Question 4(a) above.

c) Illustrate the representation of the given statements above using the knowledge representation tool that you selected in Question 4(a) above.

Answer:

a)

Semantic network as it is more suitable to describe the relationship between objects and overall concepts Semantic network is not intelligent, it depends on the creator. If the creator is not intelligent enough, the connection will become choosy. Or

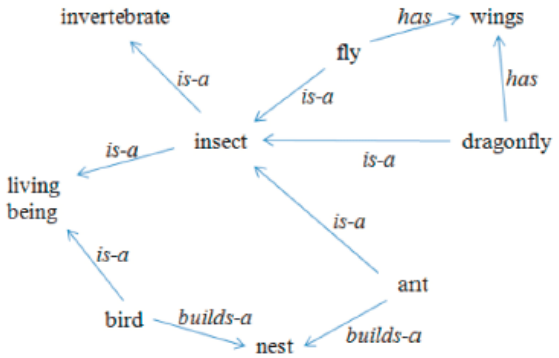
Frames as frames are suitable to show the details of objects and their relationship.

b)

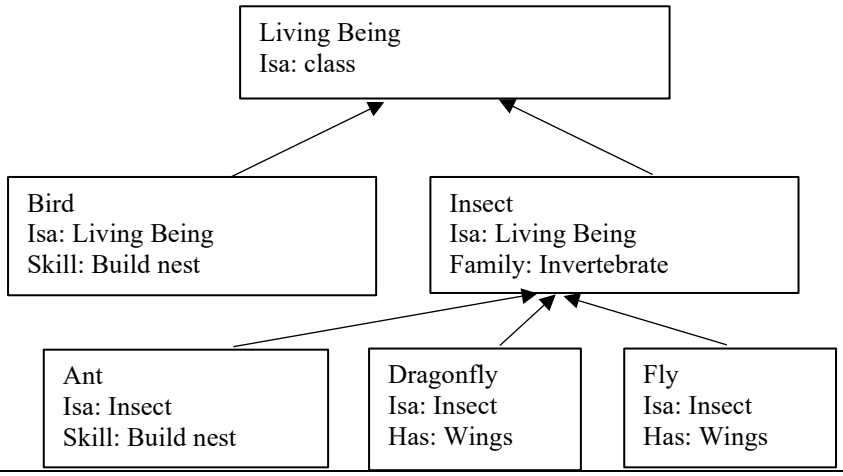
Semantic network is less organized and difficult to view the relationship between a rule and another. In semantic network representations, there is no formal semantics, no agreed-upon notion of what a given representational structure means. For example, the system is limited by the user's understanding of the meanings of the links in a semantic network.Or

Frames are only able to represent stereotyped objects, but are not able to represent events

Semantic Network



Frames



5. Suggest and describe a real-life application where semantic network is applied.

Answer:
Open-ended question. Students find sample applications online (please remind them to place the reference). Sample can obtain from Wikipedia.

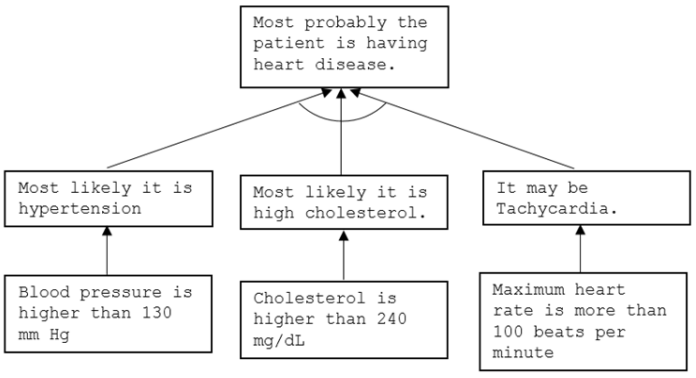
1) The real-life application is WordNet, a lexical database of English that has been used by major search engines and IR research projects for many years. It groups English words into sets of synonyms called synsets, provides short, general definitions, and records the various semantic relations between these synonym sets. Some of the most common semantic relations defined are meronymy (A is a meronym of B if A is part of B), holonymy (B is a holonym of A if B contains A), hyponymy or troponymy (A is subordinate of B; A is kind of B), hypernymy (A is superordinate of B), synonymy (A denotes the same as B) and antonymy (A denotes the opposite of B). WordNet properties have been studied from a network theory perspective and compared to other semantic networks created from Roget's Thesaurus and word association tasks. From this perspective the three of them are a small world structure.
2) Google search engine

6. A rules based heart failure rule-based system is shown below.

R1:	IF	Blood pressure is higher than 130 mm Hg
	THEN	Most likely it is hypertension.
R2:	IF	Cholesterol is higher than 240 mg/dL
	THEN	Most likely it is high cholesterol.
R3:	IF	Maximum heart rate is more than 100 beats per minute
	THEN	It may be Tachycardia.
R4:	IF	There is hypertension
	AND	High cholesterol
	AND	Tachycardia
	THEN	Most probably the patient is having heart disease.

Proposed **ONE (1)** technique to represent the knowledge above. Then, represent the knowledge by using the proposed technique.

Answer:
And/OR graph (1m)

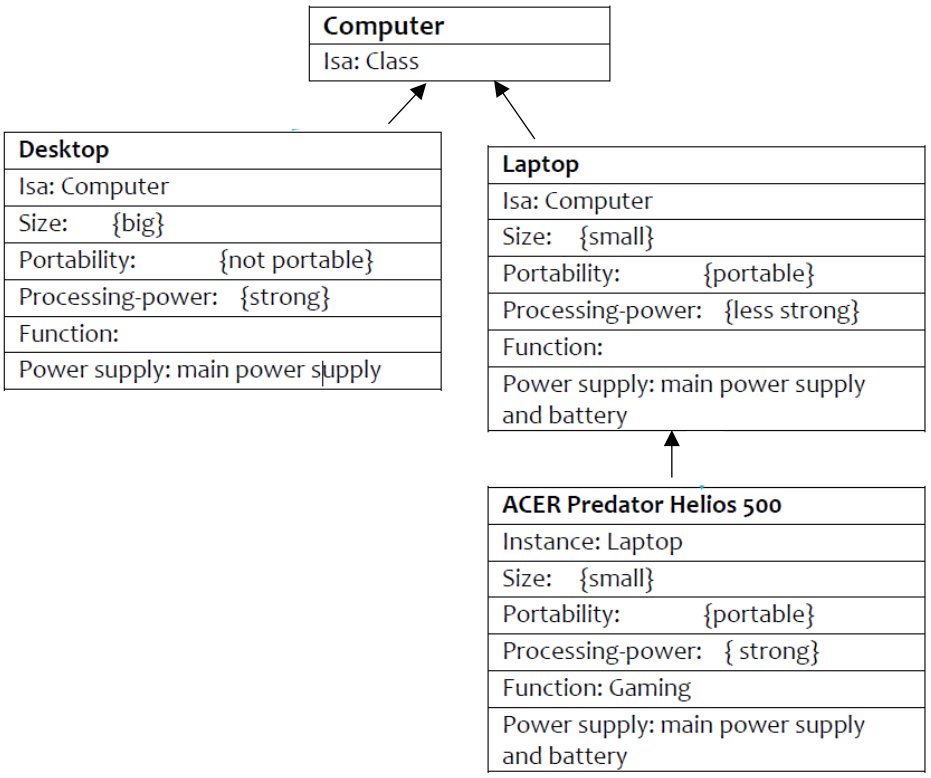


7. The information below describes the differences between a desktop and a laptop. Illustrate the information below with frames.

Laptop and desktop are the commonly used computer among students. Laptop is generally small and easily portable. It can run either on battery or main power supply. Desktop, on the other hand, is large and not portable. It only runs on main power supply. Besides that, desktop normally has more powerful processor as compared to laptop. ACER Predator Helios 500, which is a gaming laptop, however, is equipped with a powerful Intel i9 processor.

Answer:

Student's answer may be varied



Tutorial 6

1. Discuss the **THREE** (**3**) major obstacles involved in Natural Language Understanding (NLU).

Answer:

1. A large amount of human knowledge is assumed. Language acts describe relationships is an often complex world. Knowledge of these relationships must be part of any understanding system. Too much of assumptions make the understanding not "real".
2. Language is pattern based: phonemes are components of words and worls make phrases and sentences. Phoneme, word, and sentence orders are not random. Communication is impossible without a rather constrained use of these components. However, it is rather difficult to set the constraint of the relationships most of the time.
3. Limitations of programming tools make the coding becomes so difficult.
4. Ambiguity and Polysemy: Natural language is inherently ambiguous due to factors like polysemy (words having multiple meanings) and context-dependent interpretations. NLU systems must contend with identifying the correct meaning based on the surrounding context. Resolving ambiguity requires sophisticated techniques to disambiguate words and phrases to ensure accurate comprehension.
5. Communication with natural language heavily depends on our knowledge within the domain of disclosure, which involves transmission of words, knowledge, assumption and more.
6. Sentences formed with natural language cannot be understand through a simplistic, literal treatment of meaning as we need to consider the ambiguity or sarcasm meaning of a word or sentence.
7. Contextual Dependency and Inferences: Language meaning heavily relies on context, making contextual analysis a crucial aspect of NLU. Words or phrases may have different meanings depending on the surrounding text. Successful NLU involves not only understanding the immediate context but also making inferences about implied information and connections between different parts of the conversation.
8. Misspelled or misused words can create problems for text analysis. Autocorrect and grammar correction applications can handle common mistakes, but don't always understand the writer's intention.

Any other acceptable answers

2. What are the three typical stages involved in a Natural Language Processing (NLP) application development? Elaborate these three stages with the aid of appropriate diagrams.

Answer:

<u>Parsing:</u>

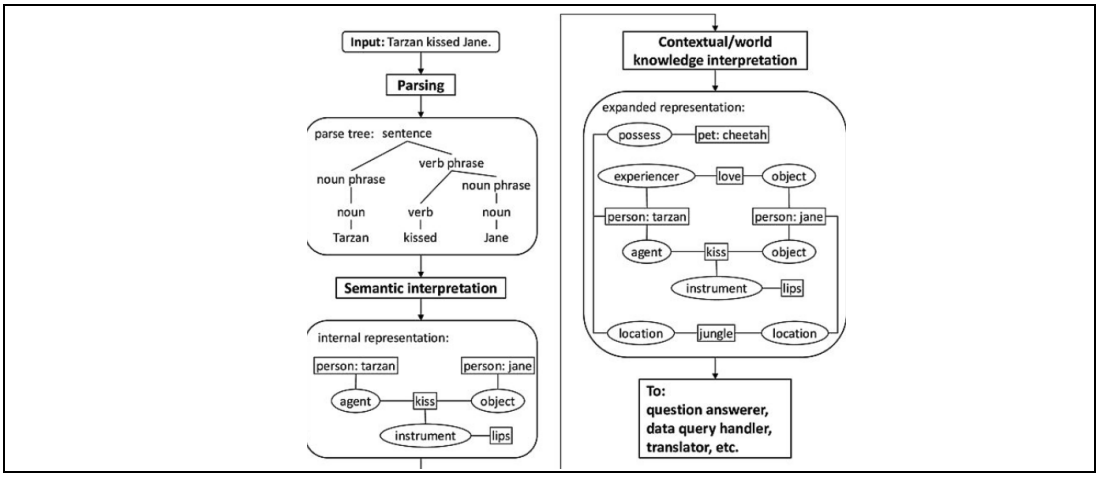
- Analyze the syntactic structure of sentences, by identifying the major relations such as subject-verb, verb-object, noun-modifier, e.g. Tarzan and Jane are nouns, and kiss is a verb.
- Often represented as parse tree
- Employs knowledge of language syntax, morphology, and some semantics.

<u>Semantic interpretation:</u>

- Produce a representation of the meaning of the text, such as conceptual graph, conceptual dependencies, frames, etc.
- Uses knowledge about the meaning of words and linguistic, e.g. the meaning of kiss is added "lips" as the instrument.
- Perform consistency checks, such as verb "kiss" may include constraints that only be performed by person to another person, but not tree.

<u>World knowledge interpretation:</u>

- Produce expanded representation of the sentence's meaning and add to knowledge base.
- Necessary for complete understanding such as Tarzan loves Jane, they live in jungle, Tarzan has pet Cheetah.
- The resulting representation is used for further processing such as handling database query, answering questions, translating the meaning, etc.



3. Besides the three main phases involved in NLP as discussed in Question 2, describe three other significant analyses that can be done to improve NLP applications.

Answer:

Refer lecture notes

1. Prosody – rhythm and intonation of language (mood, emotion, get out from this room)
2. Phonology – combination of sounds to form language (doubt)…focused on how speech sounds change and behave when in a syllable, word, or sentence
3. Morphology – components that make up words, such as prefixes (un-, non-, anti-) and suffixes (-ing, -ly) that modifying the meaning of root words
4. Syntax – rules for combining words into legal phrases and sentences
5. Semantics – the meaning of words, phrases, and sentences
6. Pragmatics – the study of the ways in which language is used and its effects on the listener. In contrast to semantic is more emphasis to the context/meaning of words, pragmatic refers to how words are being used in practical sense, words have different meaning at different situation, we will apply our understanding of symbols as we read or listen to others.

4. NLP application requires the use of knowledge about human languages. Suggest an example of NLP application and describe the difficulties that a researcher would face during the development of the application that you have suggested.

Answer:

Chatbot:

Spelling Variations:

Example: "color" (American English) vs. "colour" (British English)

Difficulty: Recognizing regional spelling preferences and suggesting the correct version based on the user's context or preference.

New Words and Slang:

Example: "LOL, that's so lit! "

Difficulty: Recognizing slang terms like "lit" and ensuring they are not flagged as errors.

Typographical Errors:

Example: "I am goiing to the party."

Difficulty: Detecting repeated letters ("goiing") and suggesting the correct form, "going."

Proper Nouns:

Example: "John Smith lives in New York."

Difficulty: Avoiding corrections to proper nouns like "John Smith" and "New York."

User Intent:

Example: "Their going to the park."

Difficulty: Understanding that "Their" should be corrected to "They're" based on the user's intent, despite the grammatical error.

False Positives and Negatives:

Example (False Positive): Correcting "I saw their cat" to "I saw there cat" (unnecessary correction).

Example (False Negative): Not correcting "I wont go" to "I won't go" (missed correction).

Machine Learning Models:

Difficulty: Developing and training machine learning models to balance correction accuracy without over-correcting or introducing errors in the text.

Any other acceptable answers: Idioms etc

5. In NLP, representation is important as it can solve issues like canonical form of sentences and syntactic problem of a sentence.
   a) Explain the meaning of canonical form of sentences. Provide examples to elaborate your answer.
   b) By referring to the below statement, there is a syntactic problem. Identify the problem and then draw **TWO** (**2**) different representation (by selecting either the semantic network or the conceptual graph) to solve the syntactic problem.

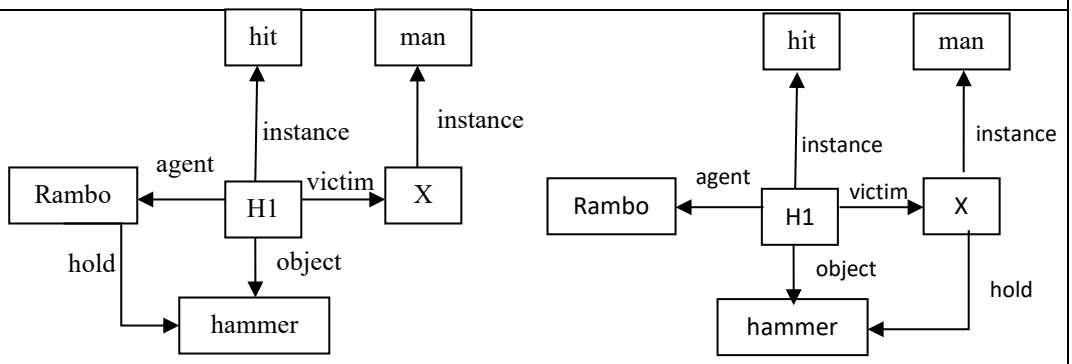
"Rambo hit the man with a hammer"

Answer:

a)

Canonical form of sentences means two different sentences have the same meaning. For example the sentences “Jen was given a gift by Sam” and “Sam gave Jen a gift” are having the same meaning, but written in different ways.

b)



1st diagram shows Rambo is the one who holds the hammer and hit the man

2nd diagram shows Rambo was hit by a man who holds a hammer

6. Parse tree is a popular tool used in one of the phases of NLP called parsing.
   a) Explain the importance of parsing in NLP.
   b) With the aid from the simple English grammar for simple transitive sentence as shown next page, draw the respective **parse trees** to verify the sentences "**the boy likes the girl**" and "**time flies like an arrow**"

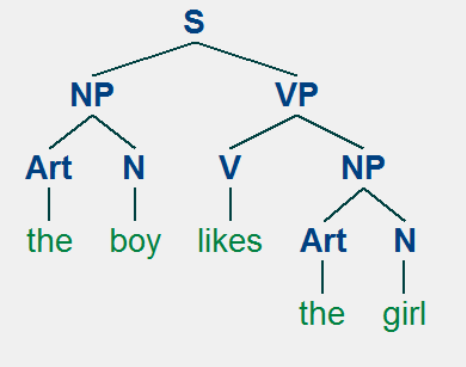
Sentence	-> noun_phrase verb_phrase
Noun phrase	-> noun
Noun-phrase	-> article noun
Verb phrase	-> verb noun_phrase
Verb phrase	-> verb preposition noun_phrase
preposition	[like]
article	[a, an, the]
noun	[flies, time, arrow, boy, girl]
verb	[like, flies]

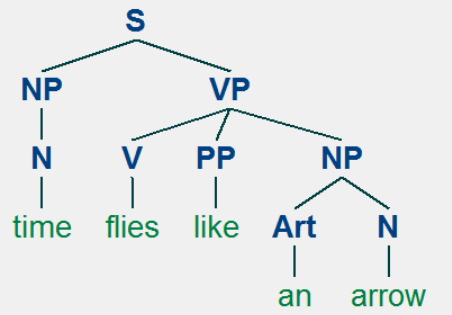
Answer:

a)

1. To analyze the syntactic structure of sentences (present, object can be used as noun or verb, diff meaning)
2. To identifying the major relations such as subject-verb, verb-object, noun-modifier

b)





7. Considering the sentence S = "**She beats George with one hand at the bank**".
   a) The sentence S consists of semantic ambiguity and syntactic ambiguity. Identify both of the ambiguities found from the sentence above.
   b) Given the grammar below, construct **ONE** (**1**) parse tree for the sentence S. (**Remark**: grammar in the parentheses **()** means it is optional.)

S -> NP VP (PP)
NP -> (DET) (ADJ) N (PP)
VP -> V NP (PP)
PP -> P NP
ADJ -> [one]
DET -> [a, the]
N -> [She, George, hand, bank]
V -> [beats]
P -> [with, at]

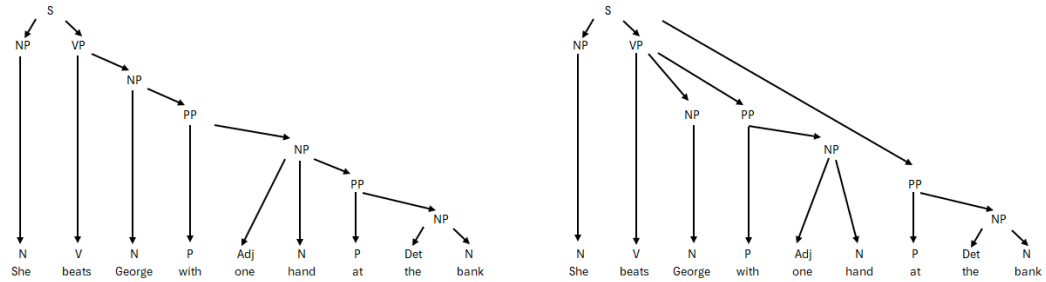
Answer:

a)

Semantic ambiguity: bank means different things, e.g. a river bank or a financial institution

Syntactic ambiguity: It is either she uses one hand to beat George, or George has only 1 hand

b)



8. Consider a sentence, *S = Ethan invited the person with microphone.*
   a) The sentence *S* consists of semantic ambiguity and syntactic ambiguity. Identify both of the ambiguities found from the sentence above.
   b) Given the grammar below, construct **ONE** (**1**) parse tree for the sentence *S*.

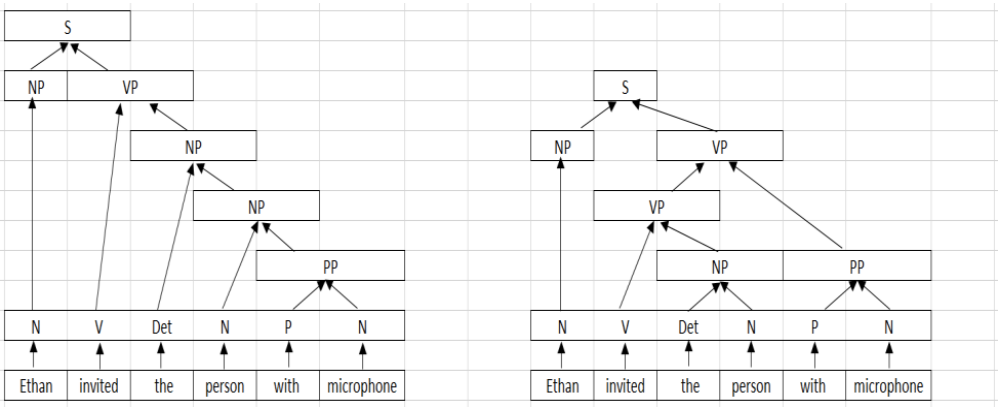
```
noun_phrase(NP)
verb_phrase(VP)
preposition_phrase(PP)
determiner(D)
noun(N)
verb(V)
S -> NP VP
NP -> N | N PP | D NP | D N
VP -> VP PP | V NP
PP -> P N
N -> [Ethan, person, microphone]
V -> [invited]
P -> [with]
D -> [the]
```

Answer:

a)

Syntactic ambiguity
Meaning 1: Ethan with microphone.
Meaning 2: The person with microphone

b)



Tutorial 7

1. Differentiate between artificial intelligence (A.I.) and machine learning (ML).

Answer:

Aspect	Artificial Intelligence (AI)	Machine Learning (ML)
Definition	AI refers to the simulation of human intelligence in machines. It involves the ability of machines to perform tasks that normally require human intelligence.	ML is a subset of AI that involves the development of algorithms and models that allow computers to learn patterns from data and make decisions based on that learning.
Scope	AI is a broader concept that aims to create machines capable of mimicking human intelligence across a wide range of tasks.	ML focuses specifically on the development of algorithms that allow computers to learn from and make predictions or decisions based on data.
Dependency on Data	AI systems can operate with or without large amounts of data. They might rely on pre-defined rules and logic.	ML heavily depends on data to train models. The more relevant and diverse the data, the better the model's performance.
Learning	AI systems may or may not involve learning from data. They can use predefined rules and logic to perform tasks.	ML is all about learning from data. Algorithms adjust their parameters automatically to improve performance as more data is processed.
Human Intervention	AI systems might require frequent human intervention and programming to adapt to new tasks or situations.	ML systems become more autonomous over time, as they learn and adapt from the data they process, requiring less human intervention for specific tasks.
Types of Tasks	AI can perform tasks that include reasoning, problem-solving, decision-making, language understanding, and more.	ML is commonly used for tasks like classification, regression, clustering, and pattern recognition.
Examples	Siri, self-driving cars, game-playing AI (like AlphaGo), and chatbots are examples of AI systems.	Recommender systems, image recognition, natural language processing (NLP), and fraud detection are examples of ML applications.
Goal	The goal of AI is to create machines that can replicate human-like intelligence and decision-making.	The goal of ML is to enable computers to learn from data, improve their performance over time, and make predictions or decisions based on that learning.
Relation	AI encompasses a broader concept that includes ML as a subset.	ML is a subset of AI, focusing specifically on the learning aspect.

Any other acceptable answer.

2. There are **TWO** (**2**) main types of ML which are Supervised and Unsupervised Learning.
   a) Discuss the differences between them.
   b) Provide **TWO** (**2**) algorithms for each type ML.

Answer:

Aspect	Supervised Learning	Unsupervised Learning
Definition	Supervised learning involves training a model using labeled data, where the input data is paired with corresponding correct output or target values. The goal is to learn a mapping from inputs to outputs.	Unsupervised learning involves training a model on unlabeled data, where the algorithm tries to find patterns or structure within the data without specific target values.
Learning Process	The model learns from examples provided in the form of input-output pairs. It learns to generalize and make predictions on new, unseen data.	The model learns to identify patterns, clusters, or relationships in the data without being explicitly given the correct answers.
Goal	The goal is to learn a mapping or relationship between inputs and outputs, enabling the model to predict the output for new, unseen inputs accurately.	The goal is to discover hidden patterns, groupings, or structure within the data, often for purposes like segmentation or anomaly detection.
Examples	Classification and regression are common tasks in supervised learning. Image classification, spam email detection, and predicting stock prices are examples.	Clustering, dimensionality reduction, and anomaly detection are typical tasks in unsupervised learning. Customer segmentation and topic modeling are examples.
Training Data	Labeled training data with input-output pairs is required for supervised learning.	Unlabeled or partially labeled data is used for unsupervised learning.
Evaluation	Models are evaluated using metrics like accuracy, precision, recall, and F1-score, which compare predicted outputs to actual target values.	Evaluation can be more challenging in unsupervised learning. It might involve assessing the quality of patterns, clusters, or other structures discovered in the data.
Human Intervention	Requires human effort to label the training data with correct output values.	Generally requires less human effort for data labeling, but might involve human interpretation of results for validation and understanding.
Use Cases	Supervised learning is used for tasks where the correct output is known, and the model needs to generalize from the training data to make predictions.	Unsupervised learning is used for tasks where the goal is to explore and uncover insights from the data, often in cases where labeled data is scarce or unavailable.
Example Algorithms	Decision trees, neural networks, support vector machines (SVM), and linear regression are examples of supervised learning algorithms.	K-means clustering, hierarchical clustering, and principal component analysis (PCA) are examples of unsupervised learning algorithms.

Any other acceptable answer.

3. Assume that the Sea Fisheries Department has approached you to analyze abalone from physical measurements. A set of data that contains 7 attributes and 4177 rows of records is given. Part of the data are shown in **Table 1** below:

Table 1: Partial Abalone Data*

Sex	Length (mm)	Diameter (mm)	Height (mm)	Whole weight (g)	Shell weight (g)	No. of rings
F	0.59	0.455	0.145	1.063	0.25	8
F	0.595	0.47	0.155	1.121	0.155	11
F	0.595	0.45	0.15	1.114	0.25	11
F	0.595	0.46	0.14	1.0045	0.2515	9
F	0.605	0.49	0.15	1.1345	0.295	9
M	0.595	0.475	0.165	1.213	0.274	9
M	0.595	0.455	0.15	1.044	0.27	9
M	0.605	0.475	0.155	1.161	0.275	9
M	0.605	0.47	0.165	1.2315	0.2925	11
M	0.61	0.47	0.15	1.1625	0.3085	11

Data Source: http://mlr.cs.umass.edu/ml/datasets/Abalone
Reference: Asuncion, A. & Newman, D.J. (2007). UCI Machine Learning Repository [http://www.ics.uci.edu/~mlearn/MLRepository.html]. Irvine, CA: University of California, School of Information and Computer Science.

Between supervised learning or unsupervised learning, which type of ML is suitable to perform the following data analytics? Justify your answer. Then suggest **ONE** (**1**) suitable ML algorithm for each case.

a) To determine 3 different grades of abalone based on diameter and whole weight.
b) To predict the age of an abalone. (To determine the age of an abalone, we can check the number of rings, which is between 1 to 29)

Answer:
a)
Unsupervised learning –To cluster based on the similar features
K-means or any other acceptable answer
b)
Supervised learning – the output / lable, ie. Number of rings is known.
Logistic regression or any other acceptable answer

4. Table 2 shows the partial data collection for stroke patients. The data is classified into two different classes which are Yes (1), and No (0). Identify the class that a new patient belongs to by using K-Nearest Neighbour (K-NN) algorithm given that the new patient's avg_glucose_level and bmi are 200.7 and 30.5, respectively (**Note:** K = 5).

Table 2: Attributes of the stroke dataset

patient_id	avg_glucose_level	bmi	stroke
9046	228.69	36.6	1
38047	100.98	28.2	1
56669	186.21	29.0	1
2374	213.37	36.0	0
15528	223.36	41.5	0
70374	122.41	40.3	0
1665	174.12	24.0	1
29908	103.26	25.4	0
11091	75.39	37.8	0
53882	70.09	27.4	1

Source: https://www.kaggle.com/datasets/fedesoriano/stroke-prediction-dataset?resource=download

Answer:

id	avg_glucose_level	bmi	stroke	distance
2374	213.37	36.00	0	13.81 (1 mark)
56669	186.21	29.00	1	14.57 (1 mark)
15528	223.36	41.50	0	25.19 (1 mark)
1665	174.12	24.00	1	27.36 (1 mark)
9046	228.69	36.60	1	28.65 (1 mark)
70374	122.41	40.30	0	78.90 (1 mark)
29908	103.26	25.40	0	97.57 (1 mark)
38047	100.98	28.20	1	99.75 (1 mark)
11091	75.39	37.80	0	125.52 (1 mark)
53882	70.09	27.40	1	130.65 (1 mark)

P(Yes| new patient) = 3/5 = 0.6 (1 mark)

P(No| new patient) = 2/5 = 0.4 (1 mark)
Conclusion, the patient most probably belong to class "Yes" (1 mark)

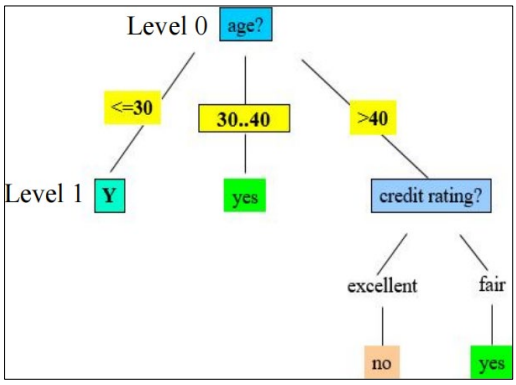


Figure 3: Partial decision tree of buys_computer

Table 2: The decision table of buys_computer

age	income	student	credit_rating	buys_computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
31…40	high	no	fair	yes
>40	medium	no	fair	yes
>40	low	yes	fair	yes
>40	low	yes	excellent	no
31…40	low	yes	excellent	yes
<=30	medium	no	fair	no
<=30	low	yes	fair	yes
>40	medium	yes	fair	yes
<=30	medium	yes	excellent	yes
31…40	medium	no	excellent	yes
31…40	high	yes	fair	yes
>40	medium	no	excellent	no

5. Given the table and the corresponding decision tree as above.
   a) Determine the appropriate attribute Y (income, student or credit_rating) at level 1 (Figure 3) by using Information Gain. Justify your answer.
   b) Produce the complete decision tree based on the answer from Question 6(a).
   c) Given a new sample, X = < <=30, low, no, excellent >, predict the class it belongs to.
   d) Refer to **Table 2**, given a new sample, X = < <=30, low, no, excellent >, predict the class it belongs to, P(C|X), using Naïve Bayes classification.

To calculate the information gain:

$$I(p,n) = -\frac{p}{p+n}\log_2\frac{p}{p+n} - \frac{n}{p+n}\log_2\frac{n}{p+n}$$

$$E(A) = \sum_{i=1}^{v}\frac{p_i+n_i}{p+n}I(p_i,n_i)$$

$$Gain(A) = I(p,n) - E(A)$$

Answer:
a)
Overall I(P, N) = I(2, 3) = **0.9710**
For income attribute:

Income Level	P (Result with Yes)	N (Result with No)	I(P, N)
Low	1	0	0
Medium	1	1	1
High	0	2	0

E(Income_Level) = (1/5)*0 + (2/5)*1 + (2/5)*0 = 0.4
Gain(Income_Level) = 0.9710 – 0.4 = 0.5710
For student attribute:

Income Level	P (Result with Yes)	N (Result with No)	I(P, N)
No	0	3	0
Yes	2	0	0

E(Student) = (3/5)*0 + (2/5)*0 = 0
Gain(Student) = 0.9710 – 0 = 0.9710

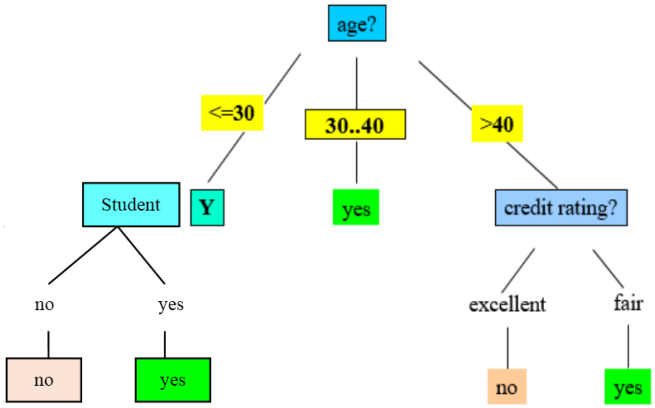
For credit rating attribute:

Income Level	P (Result with Yes)	N (Result with No)	I(P, N)
Fair	1	2	0.9183
Excellent	1	1	1

E(Credit_Rating) = (3/5)*0.9183 + (2/5)*1 = 0.9510
Gain(Credit_Rating) = 0.9710 – 0.9510 = 0.02

The attribute "student" is selected as Y attribute as it has the highest Information Gain.

b)



c) Buys_computer = no
d) P(P|X) = P(<=30|P) * P(low|P) * P(no|P) * P(excellent|P)* P(P)
= 2/9 * 3/9 * 3/9 * 3/9 * 9/14 = 0.00529

P(N|X) = P (<=30|N) * P(low|N) * P(no|N) * P(excellent|N)* P(N)
= 3/5 * 1/5 * 4/5 * 3/5 * 5/14 = 0.02057
The conclusion of buys_computer = no

6. Assess the following performance of the classifier as shown in Figure 4:
   a) Overall accuracy
   b) Precision for cat classification
   c) Recall for dog classification

		Predicted		
		Cat	Dog	Rabbit
Actual class	Cat	5	3	0
	Dog	2	3	1
	Rabbit	0	2	11

Figure 4: Confusion Matrix

Answer:
(i)Overall Accuracy

= (TP of Cat + TP of Dog + TP of Rabbit) / Total
= (5+3+11) / 27
= 0.7037

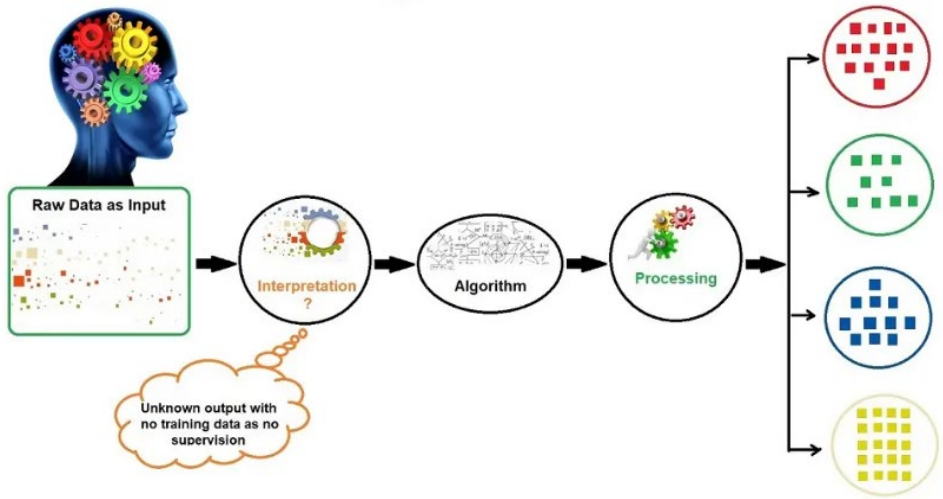
(ii)Precision for cat classification
= TP of Cat / Predicted Yes
= 5/7
= 0.7143

(iii)Recall for dog classification
= TP of Dog / Actual Yes
= 3/6
= 0.5

Tutorial 8

1. Draw a diagram to depict unsupervised learning model.

Answer:



1. Raw Data as Input: The system begins with raw data, which could be any unprocessed data. It's represented as a scatter of various colored squares, indicating diversity in the data points.

2. Interpretation: This step involves understanding or interpreting the data.

3. Algorithm: The data is then processed through an algorithm. Algorithms are sets of rules or instructions given to a computer to help it perform a specific task or solve problems. This step likely involves the transformation, analysis, or clustering of the data based on the algorithm's design.

4. Processing: This is the execution phase where the algorithm processes the data.

5. Output: Finally, we have the outputs, which are shown as colored grids. Each grid has a uniform color and order, suggesting that the algorithm has processed the diverse raw data into distinct categories or results.

2. Considering a partial dataset as shown in Table 1:

a) Explain how regression model is different from clustering.
b) Explain how regression model and clustering can be applied on the dataset below

Table 1: Sample of data

Age	Education	Gender	Income
31	Doctoral	Male	150k
45	Bachelor's	Female	78k
26	Master's	Male	76k
…	…	…	…

Answer:
a)
It predicts continuous valued output. The Regression analysis is the statistical model which is used to predict the numeric data instead of labels. It can also identify the distribution trends based on the available data or historic data. Clustering is the task of partitioning the dataset into groups, called clusters. The goal is to split up the data in such a way that points within single cluster are very similar and points in different clusters are different. It determines grouping among unlabeled data.
b)
Regression: Predicting a person's income from their age, education is example of regression task.
Clustering: Example is to group the users based on their age, income and education.

3. Clustering is the grouping of unlabeled data, it can be used for knowledge discovery. Discuss about this and provide **TWO (2)** real world applications of clustering.

Answer:
Clustering enable us to discover knowledge from the unlabeled data in finding the association (in features) and group a set of objects in such a way that objects in the same group (called a cluster) are more similar (in some sense or another) to each other than to those in other groups (clusters).

Two real life applications:
1) Cluster on the T-Shirt size of human – This is very expensive and tedious job to custom made T-Shirt for everyone in the world. So clustering helps in generalize the size into a few categories which can fit majority of the human size.
2) Image Segmentation – Segment the image into regions that represent Sky, Grass, Building, etc..
3) marketing in credit card /loan
Other appropriate answer can be accepted.

4. K-means is used to cluster the data in Table 2. Randomly we choose two cluster centers (**k = 2**) for two clusters; **C1 = (1.0, 1.0)** and **C2 = (5.0, 7.0)**.

Table 2: Sample of data

Individual	Variable 1	Variable 2
1	1.0	1.0
2	1.5	2.0
3	3.0	4.0
4	5.0	7.0
5	3.5	5.0
6	4.5	5.0
7	3.5	4.5

a) List the clusters that each data will be assigned to.
b) Identify the new centers for the first iteration.
c) Suggest **TWO (2)** stopping conditions for K-means.
d) Discuss on the limitation of K-means.

a)
Divide these data into 2 groups, By using Euclidean distance measure, calculate the distance of each data towards center 1 (1.0, 1.0) and center 2 (5.0, 7.0)

Individual	Center 1	Center 2
1	0	7.21
2	1.12	6.10
3	3.61	3.61
4	7.21	0
5	4.72	2.5
6	5.31	2.06
7	4.30	2.92

we obtain two clusters containing: {1,2,3} and {4,5,6,7}

b)
New cluster centers:
C1_new = (1/3 * (1.0+1.5+3.0), 1/3 * (1.0+2.0+4.0)) = (1.83, 2.33)
C2_new = (1/4 * (5.0+3.5+4.5+3.5), 1/4 * (7.0+5.0+4.5)) = (4.12, 5.38)

c)
2 stopping conditions:
- no (or minimum) re-assignments of data points to different clusters, or
- no (or minimum) change of centroids, or
- minimum decrease in the **sum of squared error**

d)
Limitation of K-means:
- Must manually choose K
- K-means has problems when clusters are of differing Sizes, Densities, and Non-globular shapes (http://faculty.juniata.edu/rhodes/ml/clusterAn.htm)
- K-means has problems when the data contains outliers.

5. Mean Shift is one of the popular clustering algorithms:
   a) Demonstrate the algorithm for Mean Shift
   b) Discuss the advantages and disadvantages of Mean Shift.

Answer:
a)
Algorithm of Meanshift (Refer to the animation in Lecture note)
1. Choose a search window (width and location)
2. Compute the mean of the data in the search window
3. Center the search window at the new mean location
4. Repeat until convergence

b)
Advantages
- Does not assume number of clusters
- Just a single parameter (window size)
- Finds variable number of modes (Local Maxima)
- Robust to outliers

Disadvantages
- Output depends on window size
- Computationally expensive
- Does not scale well with dimension of feature space

6. Explain why cluster evaluation is a hard problem. Identify **TWO** (**2**) methods based on internal information that we can use to evaluate a clustering algorithm.

Answer:
The quality of a clustering is very hard to evaluate because
- We do not know the correct clusters.

- Cluster centers are arbitrary (unclear/randomly)

Methods based on internal information that can be used to evaluate a clustering algorithm.
a) Intra-cluster cohesion (compactness):
Cohesion measures how near the data points in a cluster are to the cluster centroid. Sum of squared error (SSE) is a commonly used measure.
b) Inter-cluster separation (isolation):
Separation means that different cluster centroids should be far away from one another.

Tutorial 9

1. Answer the following questions:
   a) Research in Artificial Neural Networks (ANNs) started with a view of mimicking the functioning of Biological Neural Networks (BNNs). Illustrate with diagrams, how the BNNs and ANNs are similar to each other.
   b) Figure 1 below shows the simplest neural network model that is known as perceptron, which can be used for classification tasks. Briefly explain the functions of each of the layers A, B and C as shown in the diagram.

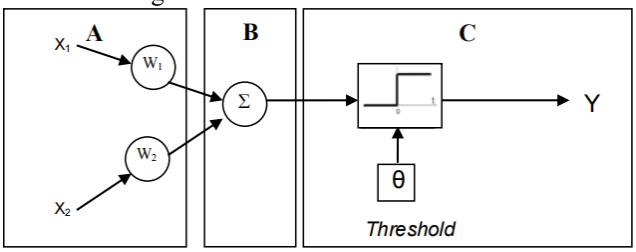


Figure 1: A sample neural network model

   c) Back propagation Multilayer Artificial Neural Network can be applied into a system that is able to recognize and interpret two different types of flowers. Discuss the prerequisite for the Neural Network system and design **ONE** (**1**) Neural Network structure for the purpose. State the advantage of back propagation network.

Answer:

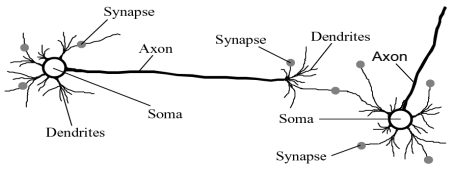
a)

A neuron consists of a cell body, soma, a number of fibers called dendrites, and a single long fiber called the axon. The connection between neurons is called biological neural network.

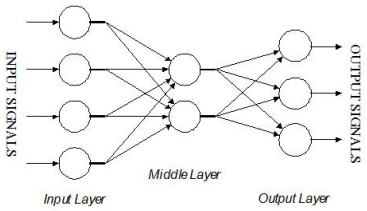
An artificial neuron consists of a neuron, a number of inputs with their correspondent adjusted weights, and an output. The connection between artificial neurons is called artificial neural network.

The similarities between the two include:

- The processing element receives many signals (signal can be features, for e.g. identify is a girl or a man, features= body shape, hair, eyes).
- Signals may be modified by a weight at the receiving synapse.
- The processing element sums the weighted inputs.
- With sufficient input, a neuron transmits a single output.
- The output from a particular neuron may go to many other neurons.
- Fault tolerance capacity.



Biological Neural Network



Artificial Neural Network

Biological Neural Network	Artificial Neural Network
Soma	Neuron
Dendrite	Input
Axon	Output
Synapse	Weight

b)

Layers	Functions	Elaborations
A	Input	Received data ($x1$ and $x2$) and the weights ($w1$ and $w2$) are used to express the strength or the importance of neuron input
B	Linear Combiner	To combine the input with the weight, $\sum x.w$
C	Activation function	Step function is used to classify the inputs into two classes based on the threshold θ. The output Y is to determine the result of the classification; or to pass output to other neuron

c)

Student's answers may vary.

First they have to explain the training data for the flowers (with labels) should be provided as the prerequisite to train the neural network.

Secondly, they have to explain how a neural network is integrated in the methodology to recognize two types of flowers. The answers must include the design of inputs (e.g. the size of petals and sepals), the multilayer network structure (input layer, hidden layer and output layer). They must include the explanation of how the weights are adjusted in back propagation network (E.g. When there is error in the output, based on the learning rate, number of epoch, and so on..).

Lastly, the student must explain the advantage of back propagation network as it can perform error correction and fine tune the network structure. If there is an error - or in other words a difference between actual and desired output patterns - the weights are adjusted to reduce this error.

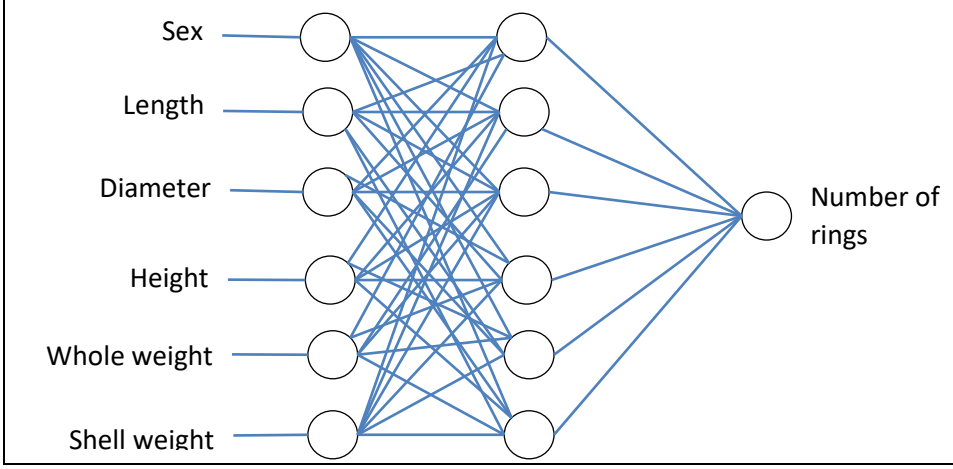
2. Assume that the Sea Fisheries Department has approached you to analyze abalone from physical measurements. A set of data that contains 9 attributes and 4177 rows of records is given. Part of the data are shown in Table 1 below:

Table 1: Partial Abalone Data

Sex	Length (mm)	Diameter (mm)	Height (mm)	Whole weight (g)	Shell weight (g)	No. of rings
F	0.59	0.455	0.145	1.063	0.25	8
F	0.595	0.47	0.155	1.121	0.155	11
F	0.595	0.45	0.15	1.114	0.25	11

Demonstrate how a **multi-layer feedforward back-propagation neural network** can be constructed to predict the number of rings of abalone. Illustrate the architecture of the neural network with appropriate labels of each layer in the diagram. In your diagram, also include all the attributes and output as shown in Table 1.

Answer:



3. A combinational logic circuit are made up from basic logic AND, OR, NAND, NOR or NOT gates that are "combined" or connected together to produce more complicated switching circuit. The output of the gate will be determined by combination of 1's and 0's which are present at the input terminals. The truth table of 3-input combinational logic circuit is given by Table 2 and Figure 2 where A, B and C are the inputs and Q is the output.

Table 2

A	B	C	Q
0	0	0	1
0	0	1	1
0	1	0	0
0	1	1	1
1	0	0	0
1	0	1	0
1	1	0	0
1	1	1	1

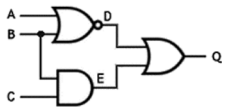


Figure 2

Train the perceptron to perform 3-input combinational logic circuit operation. The formula of output Q is given by

$$\mathbf{Q = w_1A + w_2B + w_3C}$$

where w_1, w_2 and w_3 are weight for A, B and C respectively. Assume that the threshold $\theta = 0.25$, learning rate $\alpha = 0.1$ and **step function** is used for activation function. Calculate the final weights of w_1, w_2 and w_3 for epoch 1 only. (Assume the initial weights $w_1 = 0.3$, $w_2 = 0.2$ and $w_3 = 0.1$)

Answer:

Epoch	Inputs			Desired Output	Initial Weights			Actual Output	Error	Final Weights		
	A	B	C	Q_D	w_1	w_2	w_3	Q	e	w_1	w_2	w_3
1	0	0	0	1	0.3	0.2	0.1	0	1	0.3	0.2	0.1
	0	0	1	1	0.3	0.2	0.1	0	1	0.3	0.2	0.2
	0	1	0	0	0.3	0.2	0.2	0	0	0.3	0.2	0.2
	0	1	1	1	0.3	0.2	0.2	1	0	0.3	0.2	0.2
	1	0	0	0	0.3	0.2	0.2	1	-1	0.2	0.2	0.2
	1	0	1	0	0.2	0.2	0.2	1	-1	0.1	0.2	0.1
	1	1	0	0	0.1	0.2	0.1	1	-1	0.0	0.1	0.1
	1	1	1	1	0.0	0.1	0.1	0	1	0.1	0.2	0.2

4. In a supervised learning session, a 2-input neuron and the respective truth table are given in Figure 3 and Table 3, where x_1 and x_2 are the inputs and y is the output.

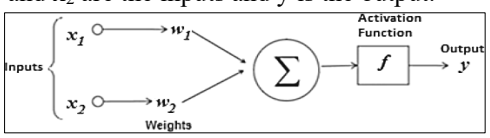


Figure 3: The 2-input neuron

Table 3

x_1	x_2	*Desired Ouput, y_d*
0	0	-1
0	1	1
1	0	1
1	1	1

The formula of output y is given by

$$y = \text{sign}(n) = \begin{cases} 1 \; if \; n \geq \theta \\ -1 \; if \; n < \theta \end{cases}$$

where $n = w_1x_1 + w_2x_2$, and w_1 and w_2 are the weights for x_1 and x_2 respectively. The **sign function** is used for activation function. The weights, $w_1 = 0.1$ and $w_2 = 0.2$, the threshold, $\theta = 0.2$ and the learning rate, $\alpha = 0.1$ are set initially for the training. Train the neuron and demonstrate the final weights of w_1 and w_2 **for epoch 1 only**.

Answer:

x_1	x_2	*Desired Ouput,* y_d	w_1	w_2	*Actual output,* y	e	w_1	w_2
0	0	-1	0.1	0.2	-1	0	0.1	0.2
0	1	1	0.1	0.2	1	0	0.1	0.2
1	0	1	0.1	0.2	-1	2	0.3	0.2
1	1	1	0.3	0.2	1	0	0.3	0.2

Tutorial 10

1. Noise reduction is very important in image processing.
   a) Discuss the above statement.
   b) Identify **TWO** (**2**) types of noises in an image.
   c) Name a filter to perform noise reduction.
   d) Demonstrate how image filtering works in reducing the noise by using median filter. Also, Perform median filtering onto the image in Figure below by using 3×3 kernel. (Note: Replace the pixel value to 0 for border problem due to not enough neighbouring pixels to perform the filtering).

Answer:

a)

Digital images are prone to a variety of types of noise. Noise is the result of errors in the image acquisition process that result in pixel values that do not reflect the true intensities of the real scene. Noise reduction is important in the context of:

- To recover from the Image noise that might caused by different intrinsic (i.e., sensor) and extrinsic (i.e., environment) conditions which are often not possible to avoid in practical situations.
- To ensure the smoothness and the best performance of the later processing steps.
- To eliminate unintended information during feature extraction.

b)

Salt and pepper noise

Contains random occurrences of black and white pixels

Gaussian Noise

Variations in intensity drawn from a Gaussian normal distribution

Impulse Noise

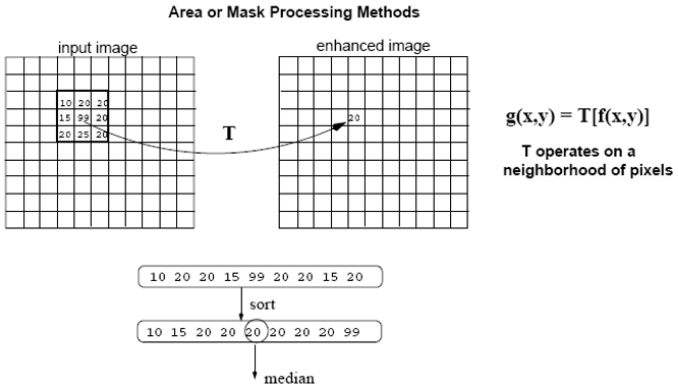
Contains random occurrences of white pixels

c)

Average Filter (impulse noise), Median Filter (salt and pepper noise), Gaussian Filter

d)

Image filtering is used to modify the spatial frequency characteristics of an image. It works by determining the value of a central pixel (in a fixed kernel size which 3x3 is often used) and supported by all its neighbours pixel values.



0	0	0	0	0	0
0	163	145	144	138	0
0	198	198	144	135	0
0	198	198	144	124	0
0	171	173	170	135	0
0	157	168	168	144	0
0	0	0	0	0	0

2. Morphological operation is important in image processing.
   a) Explain **FOUR** (**4**) morphological operations.
   b) Suggest a morphological operation that could be used to recover the image as shown in figure below (the objective is to close the gap between the pixels of the number). Explain your answer.

Answer:

a)

1. Erosion

Erosion shrinks the white regions in an image and enlarges the black regions. It works by moving a structuring element (a predefined shape) over the image and retaining a pixel in the output image only if all pixels under the structuring element are 1 (white). This operation is useful for removing small white noises, detach two connected objects, or to erode away the boundaries of an image region

2. Dilation

Dilation is the opposite of erosion. It adds pixels to the boundaries of objects in an image, expanding the white regions and shrinking the black regions. The structuring element is used to probe the image, and a pixel in the output image is set to 1 (white) if any pixel under the structuring element is 1.

Dilation is used to fill small holes, connect adjacent objects, and to recover the size of an object after erosion

3. Opening

Opening is a two-step process involving erosion followed by dilation, using the same structuring element for both operations. This operation is effective in removing small objects or extrusions from an image while preserving the shape and size of larger objects. It can also be used to smooth the contour of an image and to clear small holes

4. Closing

Closing is also a two-step process but in the reverse order of opening: it involves dilation followed by erosion. This operation is useful for closing small holes inside the foreground objects or small black points on the object. It can also be used to smooth the contour of an image while generally maintaining the shape and size of the larger objects in the image

b)Morphological operation: Dilation

Explanation: The dilation operation usually uses a structuring element for probing and expanding the shapes contained in the input image.

3. Explain **ONE** (**1**) importance of morphological image analysis in image processing.

Answer:

- useful to remove noise
- improve image texture / image enhancement

any other acceptable answer

4. In morphological image analysis, morphological operators often probe an image with a **structuring** element. A 6x6 image is given in **Figure 1** below.(Note: Replace the pixel value to 0 for border problem due to not enough neighbouring pixels to perform the filtering)

	0	1	2	3	4	5
0	0	0	0	0	0	1
1	0	1	1	1	1	1
2	0	1	1	1	1	0
3	0	1	1	1	1	0
4	0	1	1	1	0	0
5	1	1	1	0	0	0

Figure 1: 6x6 binary image

a) Based on the 3x3 structuring element with the origin in the middle as shown in Figure 2, perform an erosion effect on the 6x6 binary image as shown in Figure 1. Provide the resulting binary image.

1	1	1
1	1	1
1	1	1

Figure 2: 3x3 square structuring element

b) Based on the 2x2 structuring element with the origin in the middle as shown in Figure 3, perform an dilation effect on the 6x6 binary image as shown in Figure 1. Provide the resulting binary image.

0	1
1	0

Figure 3: 2x2 square structuring element

Answer

Erosion

	0	1	2	3	4	5
0	0	0	0	0	0	0

1	0	0	0	0	0	0
2	0	0	**1**	**1**	0	0
3	0	0	**1**	0	0	0
4	0	0	0	0	0	0
5	0	0	0	0	0	0

Dilation

	0	1	2	3	4	5
0	0	0	0	0	0	0
1	0	0	**1**	**1**	**1**	**1**
2	0	**1**	**1**	**1**	**1**	**1**
3	0	**1**	**1**	**1**	**1**	**1**
4	0	**1**	**1**	**1**	**1**	0
0	0	**1**	**1**	**1**	0	0

5. The scanned document in Figure 4 shows a missing portion of pixels (with an actual value of 2) resulting from a poor scanning process. Figure 5 represents the pixel value of Figure 4. Utilising the structuring element depicted in Figure 5, apply both dilation and erosion operations to Figure 5. Then, develop a resulting complete pixel for each operation. The origin of the structuring element is denoted by the bold box in Figure 6. Also, conclude which technique is more suitable to restore the missing part of the Figure 4. (Note: Replace the pixel value to 0 for border problem due to not enough neighbouring pixels to perform the filtering)

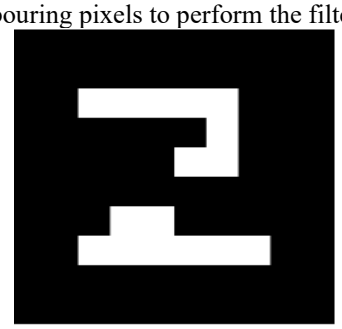


Figure 4: Scanned image

0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	1	1	1	1	1	0	0	0
0	0	0	0	0	0	1	0	0	0
0	0	0	0	0	1	1	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	1	1	0	0	0	0	0
0	0	1	1	1	1	1	1	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0

Figure 5: Pixel value

1	1
1	**1**

Figure 6: Structuring element

Answer:

0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	1	1	1	1	1	1	0	0
0	0	1	1	1	1	1	1	0	0
0	0	0	0	0	1	1	1	0	0
0	0	0	0	0	1	1	1	0	0
0	0	0	1	1	1	0	0	0	0
0	0	1	1	1	1	1	1	1	0
0	0	1	1	1	1	1	1	1	0
0	0	0	0	0	0	0	0	0	0

Dilation

0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	1	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0

Erosion

6. Pre-processing is one of the important steps in digital image analysis.
   a) Explain the importance of image pre-processing.
   b) Identify the suitable pre-processing operations based on the images in **Figure 7**. Provide brief description for each of the suggested operations.

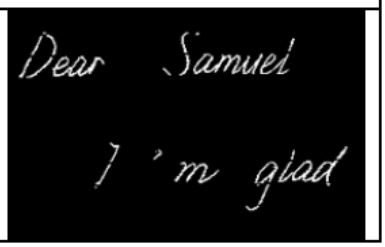
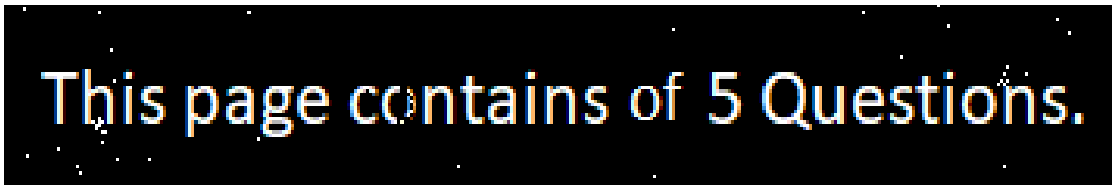
Image A (objects are less distinctive from its background)	**Image B** (salt and pepper in the image)	**Image C** (broken alphabet)
		

Figure 7: Sample images

Answer:
a)
To improve the image quality (to increase chance for success of the later on processes)
b)
1. Contrast enhancement - to enhance the contrast of the image so that the intensity values of the image are sharpen.
2. Noise removal – to remove noise (such as salt and pepper, Gaussian noise) from the image.
3. Morphological operation – may perform dilation to fill up the broken strokes.

7. Figure below shows a part of scanned image which will be recognized by optical character recognition (OCR) algorithm. Nevertheless, the detection performance is relatively poor compared to the actual sentence "Tbis page ccntains of 5 Questions." Identify TWO (2) potential output error produced by OCR algorithm. In an addition, provide ONE (1) preprocessing technique to increase the accuracy of OCR performance for each potential output error.



Answer:
Potential errors:
1) "This" is detected as Tbis
2) "contains" is detected as ccntains
3) "Questions" is detected as Questiohs

Solution:
1) Dilation or closing to enlarge the object to recover letter 'o' in "ccntains"
2) Erosion/Opening/noise remover to shrink the object to remove the "noise" for letter 'h' or 'n'

Tutorial 11

Part I: Expert Systems

1. Given a customer has problem deciding whether or not to purchase a car. He consulted an expert for advice and the consultant suggested some rules as follows.

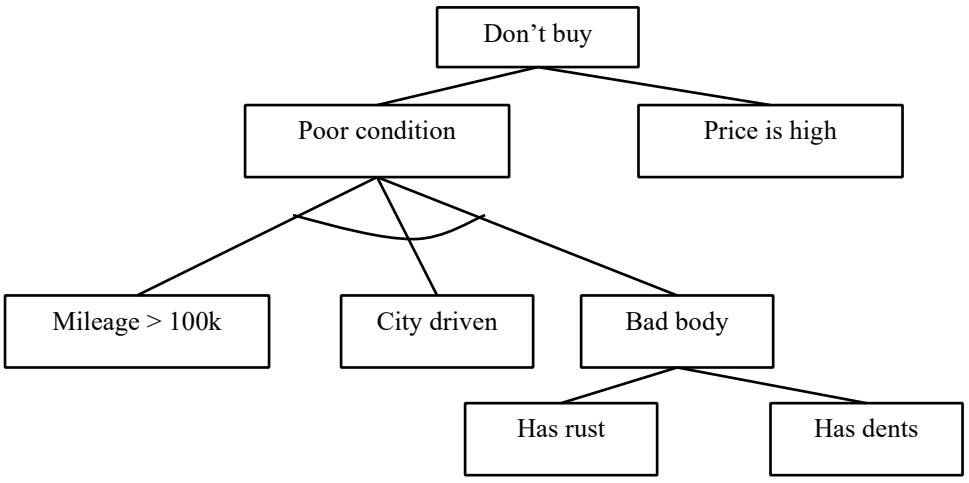
Rule 1:	IF	The condition of car is poor
	OR	The price of the car is high
	THEN	Don't buy the car
Rule 2:	IF	Mileage on the car exceeds 100,000
	AND	The car is city driven
	AND	The body of the car is bad
	THEN	The condition of the car is poor
Rule 3:	IF	The car has dents
	THEN	The body of the car is bad
Rule 4:	IF	The car has rust
	THEN	The body of the car is bad

a) Assume the hypothesis is "Don't buy the car", draw an inference network (and-or graph) that supports the problem above.
b) Given the observation by the expert as follows, demonstrate the inference process by using forward-chaining.

Mileage: 160,000 miles Location of car: KL city Dents on car: right rear door and back door Rust on car: not noticed Price of car: $1000 (remark: considered low)

c) If you are going to develop the above expert system, which inference mechanism are you going to implement? Forward or backward chaining? Justify your answer.

Answer:
Assume the hypothesis is "Don't buy the car", draw an inference network (and-or graph) that supports the problem above.

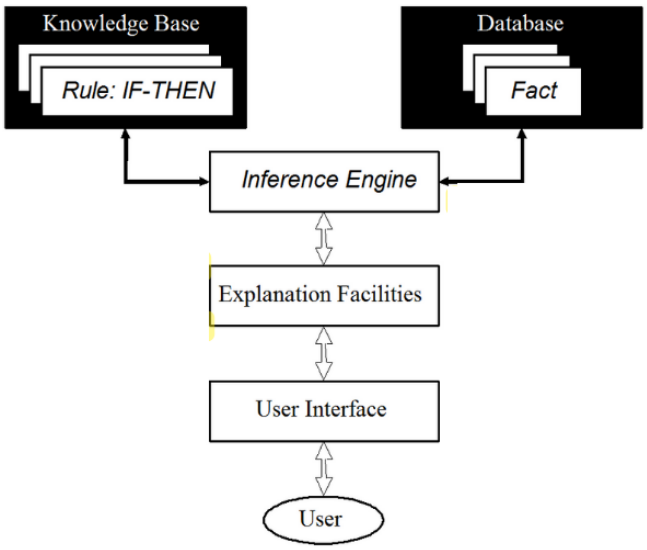


b)
If the user chose Forward Chaining, then the processes involved are:
Rule 3 is fired :
E6 – The car has dents is true, Then E5 – the car body is bad is true
Rule 2 is fired:
E3 – mileage > 100k is true,
E4 – car is driven in city is true,
E5 – the car body is bad is true, Then E1 – the car condition is poor is true
Rule 1 is fired:
E1 – Car condition is poor is true
E2 – Car price is high is false (but is OR so this rule is acceptable), Then H1 – Don't buy car is true

c)
Both forward and backward can be accepted based on student's preference. However, they must support their answer with justification.
Examples of answers:
1. In this case since the expert begins with a hypothetical solution ("don't buy") and then attempts to find facts to prove it, therefore backward chaining should be chosen.
2. If the customer first needs to gather some information, such as condition and price of car, then tries to infer it whatever can be inferred, then choose forward chaining.

2. A basic structure of a rule-based expert system involves the components such as knowledge base, database, inference engine, explanation facilities, user interface, and user.
   a) Draw a diagram to illustrate the relations between the components in an expert system as given above, and describe the functions of each component.
   b) Define the fundamental characteristics of an expert system.
   c) What is the reason that an expert-system separates its knowledge base from the inference mechanism in the system?

Answer:
a)



The basic structure of an expert system consists of several key components that work together to emulate the decision-making abilities of a human expert. These components are:
Knowledge Base: This contains the domain-specific knowledge, facts, and rules that the expert system uses to make decisions and solve problems. It is the core repository of the expert system's understanding of the subject matter

Inference Engine: This is the processing unit of the expert system that applies logical rules to the knowledge base to deduce new information or make decisions. It is responsible for the reasoning process and can work with both forward and backward chaining methods

User Interface: This component allows users to interact with the expert system, inputting data and questions and receiving advice or solutions. It is designed to be user-friendly to facilitate easy communication between the user and the system

Explanation Facility: This part of the expert system provides users with explanations of the reasoning process, detailing how it arrived at a particular conclusion or decision. This helps in understanding the logic behind the system's advice and increases the user's trust in the system

Knowledge Acquisition Facility: This is a tool or module that assists in the process of updating and expanding the knowledge base by capturing new knowledge from domain experts or other sources

b)

- Narrow but specialized domain.
- High-quality performance - correctness
- Speed - for instance, in an emergency, when a patient dies or a nuclear power plant explodes.
- Apply heuristics to guide reasoning and reduce search space
- Contains explanation capability
- Employ symbolic reasoning when solving a problem.

d)

- Easier to maintain: new knowledge can be easily added without affecting the inference mechanism. Vice versa, if we change the inference mechanism, we may not have to change the rules in the knowledge base.
- Easy to understand: we can read the read the rules in the knowledge base without need to understand how the inference mechanism works.

3. Given the rules below:

R1:	If X breathes using gills Then X lives in sea
R2:	If X has backbone and has fin and breathes using gills Then X is fish
R3:	If X lives in sea and has backbone and has fin Then X is seahorse
R4:	If X is fish and has fin and has two eyes at two sides Then X is shark
R5:	If X is fish and has two eyes at one side Then X is flounder

A subject with ID "X1001" was observed and some facts were gathered as listed below:

- X1001 has backbone
- X1001 has fins
- X1001 breathes through gills
- X1001's two eyes on two sides of its head

a) From the observation above, conflicts may occur when you infer the rules above. Analyze what is the potential conflict that might occur.
b) Suggest and discuss **THREE (3)** approaches of conflict resolution.

c) Derive a conclusion for the problem above for each of the conflict resolution approaches that you have suggested in Question 3(b) by using forward chaining.

Answer:
a) Initially R1 and R2 could be fired therefore conflict is generated.

b)
Any acceptable answer should be considered, as there are many types of conflict resolutions. Three approaches are required

Examples of answer:

1. Most Specific rules: Fire the most specific (longest matching) rule. If a rule's condition part is a superset of another, use that particular rule since it is more specialized for the current task. For example, R2 is the superset of R1, so we fire R2.
2. Highest Priority (Bottom most): Arrange the data in a priority list. Choose the rule that applies to data that have the highest priority. For example, R1 and R2 can be fired, but R2 has more priority than R1 so we fire R2.
3. FCFS: Fire the earliest rule. In this case, R1 will be fired first.

c)

1. Most Specific rules: R2 is fired → R4 is fired (Conclusion: X1001 is shark)
2. Highest Priority (Bottom most): R2 is fired → R4 is fired (Conclusion: X1001 is shark)
3. FCFS: R1 is fired → R2 is fired → R3 is fired → R4 is fired (Conclusion: X1001 is shark)

Part II: Recommender Systems

1. There are two common types of recommendation engine algorithms, i.e. collaborative filtering models and content-based models. Briefly explain the differences between them. Provide examples of applications to aid your answers.

Answer:
Recommendation engine algorithms usually categorize in two kinds: collaborative filtering models and content-based models. They differ by the type of data involved.
Collaborative filtering models compute their predictions using a dataset of feedback from users to items (typically star ratings or thumb-up/thumb-down), i.e. they are built on a dataset of user/item feedback. This may be either explicit feedback such as a star rating or thumb-up/thumb-down, or implicit feedback such as the number of episodes watched in a TV show.

Content-based models use only characteristic features of the items (like the price of a product or its color). The approach is identical to the previous User-User or Item-Item algorithms, except that the similarities are computed using only content-based features. To train a model solving the item cold-start (resp. user cold-start) you need a dataset including detailed descriptions of your items (resp. of your users), such as the genre of a movie, its budget, its duration, or any variable that may help the recommendation.

2. Discuss the limitations of collaborative filtering models.

Answer:
Collaborative filtering models suffer from what is called the cold-start problem. Because the recommendations are computed using a dataset of users feedback on items, they can't recommend items with no or only a few feedback, such as new items. Similarly they can't recommend anything to a new user before they started to bootstrap by giving some feedback on enough items (this explains the tedious on boarding steps you surely experienced online).

3. Discuss the limitations of content-based models.

Answer:

- Limited in scope, for example, it can only make recommendations that are similar to the original seed.
- For users with thousands of purchases, however, it's impractical to base a query on all the items.
- The algorithm must use a subset or summary of the data, reducing quality.

4. There are two types of collaborative filtering models as follows. Differentiate thesetwo types with the aid of suitable examples.
   a) user-user
   b) item-item

Answer:

User-User

The most commonly used recommendation algorithm follows the "people like you like that" logic. We call it a "user-user" algorithm because it recommends to a user an item that similar users liked before. The similarity between two users is computed from the amount of items they have in common in the dataset. This algorithm is very efficient when the number of users is way smaller than the number of items. You can think of a medium sized online shop with millions of products. The major drawback is that adding a new user is expensive since it requires to update all similarities between users.

Item-Item

The "item-item" algorithm uses the same approach but reverses the view between users and items. It follows the logic "if you like this you might also like that". It recommends items that are similar to the ones you previously liked. As before the similarity between two items is computed using the amount of users they have in common in the dataset. This algorithm is best when the number of items is way smaller than the number of users, such as large-scale online shops. It is well suited if your items don't change too much, since you can pre-compute the full table of item-item similarities and then serve recommendations in real-time. Updating this table for adding a new item is unfortunately hard.

Tutorial 12

Part I: Probability Theory

1. **FOUR (4)** of the rules in a rule-based system are given as follows:

R1:	If there was drizzle last night, then there is 60% chance that the grass in the backyard is wet the next morning.
R2:	If there was rain last night, then there is 80% chance that the grass in the backyard is wet the next morning.
R3:	Drizzle occurs in 160 days in a year.
R4:	Rain occurs in 120 days in a year.

Assume that the drizzle and rain occur exclusively, and there are 365 days in a year.

a) Specify the above rules in probability assertion.

b) If you see the grass in the backyard is wet in the morning, by using Bayes Theorem, what is the event that most probably occurred last night?

c) After a conclusion is drawn from Question 2(b), your grandfather asserts that sometimes your neighbour would spray water in the backyard at night.

   i. Do you think this assertion would affect the conditional probability?

   ii. Do you think this assertion would affect your initial conclusion as well?

Explain your answers.

Answer:

a)

P(wet E|drizzle H) = 0.6

P(wet E|rain H) = 0.8

P(drizzle) = 160/365 = 0.438

P(rain) = 120/365 = 0.329

b)

Single evidence with multiple hypotheses

$$P(drizzle\ H|wet\ E) = \frac{P(wet|drizzle) * P(drizzle)}{P(wet|drizzle) * P(drizzle) + P(wet|rain) * P(rain)}$$

$$P(drizzle|wet) = \frac{0.6 * 0.438}{0.6 * 0.438 + 0.8 * 0.329} = \frac{0.2628}{0.526} = 0.4996 \approx 0.5$$

$$P(rain|wet) = \frac{P(wet|rain) * P(rain)}{P(wet|drizzle) * P(drizzle) + P(wet|rain) * P(rain)}$$

$$P(rain|wet) = \frac{0.8 * 0.329}{0.6 * 0.438 + 0.8 * 0.329} = \frac{0.2632}{0.526} = 0.5$$

It can be either rain or drizzle. No significant difference between the two events

c)

The conditional probability *P(wet|spray water)* will be added, and *H=spray water* will be added as the additional hypothesis in the problem.

The conclusion (posterior probability) will be affected by the frequent of the neighbour spray water at the backyard at night.

For instance, if the neighbour just sprays water once a while in a year, the *P(spray water)* will be very low (almost 0). Hence, this will not give big difference to the posterior hypotheses of the other two events.

However, if the neighbour sprays water frequently, e.g. almost everyday, then the *P(spray water)* will be very high (almost 1). Hence, this will definitely reduce the posterior hypotheses of the other 2 events.

2. Part of the rules in a liver disease diagnosis system is shown below.

R1:	IF	patient has loss of appetite
	THEN	patient has liver disease
R2:	IF	patient is a alcoholic
	AND	patient is a smoker
	THEN	patient has liver disease

Considering only the following information is given to you.

i. P(patient has loss appetite | patient has liver disease) = 0.8
ii. P(patient has liver disease) = 0.00013

Assume that a patient rated that he has loss appetite 70% of the time, demonstrate how do you predict P(patient has liver disease | patient has loss appetite) using Naïve Bayes Algorithm.

P(has disease H| loss appetite E) =

$$\frac{P(\text{loss appetite E}|\text{has disease H}) * P(\text{has disease H})}{P(\text{loss apetite E})}$$

= 0.8 * 0.00013 / 0.7 = 0.0001486

3. A police officer sees you run a red light late one Saturday night. Before pulling you over, the officer tries to guess what caused you to run the red light and comes up with the following possibilities:
   a) You are drunk
   b) You are in a hurry
   c) You are not wearing your glasses
   d) You felt you could do it and get away with it

   The policeman assigns the following evidential probabilities to these hypotheses given that the incident happened late on a Saturday night
   - P(run light late Sat. night | drunk driver) = 0.45
   - P(run light late Sat. night | driver in a hurry) = 0.60
   - P(run light late Sat. night | driver didn't see light) = 0.15
   - P(run light late Sat. night | didn't see cop/thinks can get away with it) = 0.05

   He further assigns the following prior probabilities:
   - P(drunk driver) = 0.10
   - P(hurried driver) = 0.33
   - P(driver not wearing glasses) = 0.10
   - P(driver feels he can get away with things) = 0.25

   Compute the conditional probabilities for the four possibilities. What decision is the police officer most likely to make?

Answer:

Single evidence with multiple hypotheses

P(run light late Sat. night E| drunk driver H) = 0.45 (conditional probability)

P(run light late Sat. night | driver in a hurry) = 0.60
P(run light late Sat. night | driver didn't see light) = 0.15
P(run light late Sat. night | didn't see cop/thinks can get away with it) = 0.05

H_1= drunk driver
H_2= hurried driver
H_3= driver not wearing glasses
H_4= driver feels he can get away with things
E = run light late Sat. Night

Joint conditional probability

$$\sum_i^4 P(E|H_i) * P(H_i) = 0.45 * 0.10 + 0.60 * 0.33 + 0.15 * 0.10 + 0.05 * 0.25 = 0.27$$

$$P(H_1|E) = \frac{P(E|H_1) * P(H_1)}{\sum_i^4 P(E|H_i) * P(H_i)} = \frac{0.45 * 0.10}{0.27} = 0.166$$

$$P(H_2|E) = \frac{P(E|H_2) * P(H_2)}{\sum_i^4 P(E|H_i) * P(H_i)} = \frac{0.60 * 0.33}{0.27} = 0.733$$

$$P(H_3|E) = \frac{P(E|H_3) * P(H_3)}{\sum_i^4 P(E|H_i) * P(H_i)} = \frac{0.15 * 0.10}{0.27} = 0.056$$

$$P(H_4|E) = \frac{P(E|H_4) * P(H_4)}{\sum_i^4 P(E|H_i) * P(H_i)} = \frac{0.05 * 0.25}{0.27} = 0.046$$

Conclusion: The driver may be in hurry

Part II: Certainty Factor

4. You receive a phone call from your classmate, James, informing you that your name is in the pass list. Preparing to tell your parents the good news, you recall that James is a liar who loves to play jokes on others. Thus, you decide to call another classmate, Sean. Although Sean is known for being occasionally careless, you believe that Sean is far more reliable.

R1: If James' call is true, then there is a 0.5 certainty that your name is in pass list.
R2: If Sean's call is true and Sean is not careless is true, then there is a 0.9 certainty that your name is in the pass list.
R3: f your name is in the pass list, then there is a 0.99 certainty that you pass the examination.
Conditions:
The certainty of James' call is 1.0.
The certainty of Sean's call is 1.0.
The certainty of Sean is not careless is 0.5

Based on the information provided, would you pass the examination?

Answer:
CF(1) = CF (names in the pass list – james call true)
=0.5*1
=0.5
CF(2) = CF (name in the pass list – Sean call true & sean not careless)
= min[1,0.5]*0.9
=0.45

CF(overall- name in the pass list)=CF(1)+CF(2)*(1-CF1)=0.5+0.45*(1-0.5)=0.725

CF(Pass)=0.99*0.725=0.72

Conclusion: almost certain I would pass the exam

5. Assume that a simple inference system is built using certainty factors. Demonstrate how to compute the certainty factor that a patient has liver disease based on the following rules.

R1:	IF	patient has loss of appetite [CF 0.7]
	THEN	patient has liver disease [CF 0.5]
R2:	IF	patient is a alcoholic [CF 0.8]
	AND	patient is a smoker [CF 1.0]
	THEN	patient has liver disease [CF 0.9]

Combining Certainty Factors:

$$\begin{cases} CF1 + CF2 - CF1*CF2 & \text{if both are positive,} \\ CF1 + CF2 + CF1*CF2 & \text{if both are negative,} \\ \dfrac{CF1 + CF2}{1 - min(|CF1|, |CF2|)} & \text{otherwise} \end{cases}$$

Answer:
CF(R1) = 0.7 * 0.5 = 0.35
CF(R2) = min(0.8, 1) * 0.9 = 0.72
CF(R1, R2) = CF1+ CF2 – CF1*CF2
=0.35+0.72 – 0.35*0.72 = 0.818

6. The Ebola outbreak has raised concerns of international spread in 2014, similar to the SARS epidemic in more than 10 years ago. Assume that you are assisting a medical centre to develop an online expert system that consults people on recognizing the symptoms of these diseases. Part of the rules of the system is as follows:

R1:	IF	body temperature exceeds 37 °C
	AND	headache presents AND muscle aches presents
	THEN	fever presents (CF 0.8)
R2:	IF	body temperature exceeds 37 °C
	THEN	fever presents (CF 0.2)
R3:	IF	body temperature exceeds 38 °C
	THEN	fever presents (CF 1.0)
R4:	IF	cough presents AND fever presents
	THEN	symptom of SARS (CF 0.5)
R5:	IF	cough presents AND cough starts 2-3 days after other symptoms
	AND	fever presents
	THEN	symptom of SARS (CF 0.6)
R6:	IF	IF (cough presents OR nausea presents OR stomach pain presents)
	AND	fever presents
	THEN	symptom of Ebola (CF 0.5)

R7:	IF	IF cough presents AND nausea presents AND stomach pain presents
	AND	fever presents
	THEN	symptom of Ebola (CF 0.8)

John has recently returned from oversea and he started to show the following symptoms. Based on R1 to R7, demonstrate the inference steps and conclude what disease that John could be infected.

Body temperature exceeds 37 °C	CF = 1.0
Headache presents	CF = 1.0
Muscle aches presents	CF = 1.0
Body temperature exceeds 38 °C	CF = 1.0
Cough presents	CF = 1.0
Cough starts 2-3 days after other symptoms	CF = 0.8
Nausea presents	CF = 0
Stomach pain presents	CF = 0.9

Remark: To calculate a combined certainty factor, use the following equation:

$$\begin{cases} CF1 + CF2 - CF1*CF2 & \text{if both are positive,} \\ CF1 + CF2 + CF1*CF2 & \text{if both are negative,} \\ \dfrac{CF1 + CF2}{1 - min(|CF1|, |CF2|)} & \text{otherwise} \end{cases}$$

Answer:

R1: CF(fever presents) = min(1, 1, 1)* 0.8 = 0.8
R2: CF(fever presents) = 1*0.2 = 0.2
R3: CF(fever presents) = 1
CF(fever presents) = CF1+CF2 – CF1* CF2 = 0.8+0.2 – (0.8*0.2) = 0.84
CF(fever presents) = 0.84 + 1 – (0.84*1) = 1

R4: CF(symptoms of SARS) = min(1,1) * 0.5 = 0.5
R5: CF(symptoms of SARS) = min(1,0.8,1) = 0.8*0.6 = 0.48
CF(symptoms of SARS) = 0.5 + 0.48 – (0.5*0.48) = 0.74

R6: CF(symptoms of Ebola) = min(max(1,0,0.9),1) = min(1,1)*0.5 = 0.5
R7: CF(symptoms of Ebola) = min(1,0,0.9,1) * 0.8 = 0
CF(symptoms of Ebola) = 0.5

Conclusion: John could be infected by SARS (CF 0.74) than Ebola (CF 0.5)

7. Given the report of an observation at 7am as below, demonstrate the inference and conclude the weather of tomorrow based on certainty factors.

Place of observation: Miami, Florida (tropical climate) (CF=1)
Observation data:

- Cloud height > 18,000 feet (CF = 0.8)
- Clouds are small, rounded puffs, appeared in long rows. (CF=0.8)
- Cloud patterns are sheet-like. (CF = 0.2)
- Color of cloud is white. (CF=0.7)
- Sun can shine through. (CF=0.9)
- Humidity: warm and sticky (CF=0.7)

Remark: The certain factor (CF) for other unknown data is 0.

IF	Cloud_Height > 18,000 feet [P1.1]
THEN	Cloud_Group is Cirrus [H1] (CF= 0.8)
IF	Cloud_Height is between 6,500 feet to 18,000 feet [P2.1]
THEN	Cloud_Group is Alto [H2] (CF=0.8)
IF	Cloud_Group is Cirrus [P3.1]
AND	Cloud_Pattern is sheet-like [P3.2]
AND	Cloud_Color is white [P3.3]
AND	Moon or Sun can shine through [P3.4]
THEN	Cloud_Type is Cirrostratus [H3] (CF=0.8)
IF	Cloud_Group is Cirrus [P4.1]
AND	Cloud_Pattern is small, rounded puffs with long rows [P4.2]
AND	Cloud_Color is white or gray [P4.3]
THEN	Cloud_Type is Cirrocumulus [H4] (CF=0.8)
IF	Cloud_Group is Alto [P5.1]
AND	Cloud_Color is gray or blue gray [P5.2]
AND	Moon or Sun can shine through [P5.3]
THEN	Cloud_Type is Altostratus [H5] (CF=0.7)
IF	Cloud_Group is Alto [P6.1]
AND	Cloud_Color is gray [P6.2]
THEN	Cloud_Type is Altocumulus [H6] (CF=0.8)
IF	Cloud_Type is Cirrostratus [P7.1]
THEN	Rain or snow storm occurs within 12-24 hours [H7] (CF=0.7)
IF	Cloud_Type is Cirrocumulus [P8.1]
AND	Climate_Zone is NOT of Tropical [P8.2]
THEN	Weather is fair but cold [H8] (CF=0.7)
IF	Cloud_Type is Cirrocumulus [P9.1]
AND	Climate_Zone is of Tropical [P9.2]
THEN	Hurricane is approaching [H9] (CF=0.6)
IF	Cloud_Type is Altostratus [P10.1]
THEN	Storms with continuous rain or snow is reaching [H10] (CF=0.7)
IF	Cloud_Type is Altocumulus [P11.1]
AND	Morning is warm and sticky [P11.2]
THEN	Thunderstorms by late afternoon [H11] (CF=0.7)

Part of the rules in the expert system (Reference: boatsafe.com)

Answer:

H1: 0.8 * 0.8 = 0.64 (CF(Cirrus) = 0.64)
H2: CF = 0 (CF(Alto) = 0)

H3 = min(0.64, 0.2, 0.7, 0.9) * 0.8 = 0.16 (CF(cirrostratus) = 0.16)
H4: min(0.64, 0.8, 0.7) * 0.8 = 0.512 (CF(cirrocumulus) = 0.512)
H5: min(0, 0, 0.9) * 0.7 = 0 (CF(altostratus) = 0)-alto
H6: min(0, 0) * 0.8 = 0

H7: 0.16*0.7 = 0.112 (CF(rain or snow storm) = 0.112)
H8: min(0.512, -1) * 0.7 = -0.7 (CF(weather is fair) = -1)
H9: min(0.512, 1) * 0.6 = 0.3072 (CF(hurricane) = 0.3072)
H10:0 *0.7=0 Storms with continuous rain or snow is reaching
H11: min(0, 0.7) * 0.7 = 0 (CF(thunderstorm) = 0)

Conclusion: hurricane may be approaching (CF 0.3072)

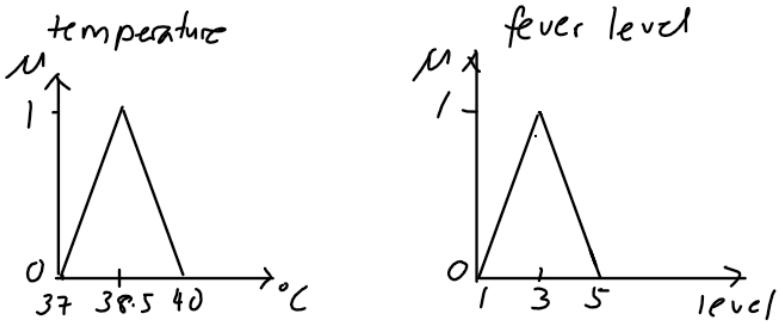
Tutorial 13

1. Following shows a sample crisp rule.

IF temperature > 37.5
THEN fever = TRUE

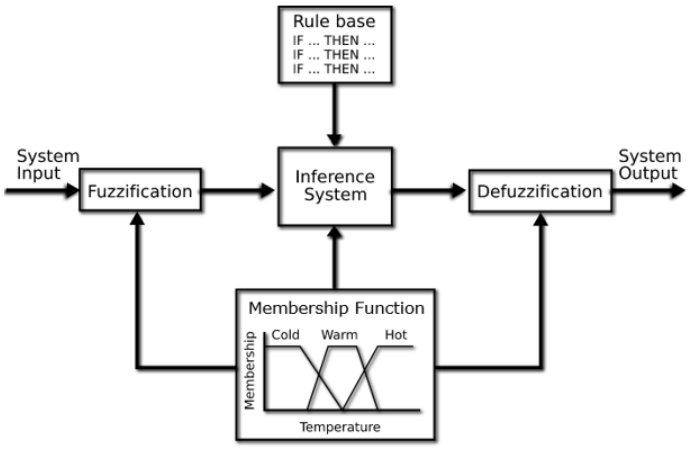
a) Explain the difference between crisp rule and fuzzy rule.
b) Suggest how do you convert the crisp rule above into a fuzzy rule.
c) Design the appropriate fuzzy sets to represent the input and output variables respectively.

Answer:
a) In crisp logic, the premise x is A can consequent y is B only can be true or false (is a binary). However, in a fuzzy rule, the premise x is A and the consequent y is B can be true to a degree, instead of entirely true or entirely false.> membership function, 0-1
b)
IF temperature is high
THEN fever is severe
c)



2. Illustrate the general structure of Fuzzy Inference System and explain each of the components.

Answer:



Fuzzification: is the first step in the fuzzy inferencing process. This involves a domain transformation where crisp inputs are transformed into fuzzy inputs by using the membership function. (E.g. fuzzify the crisp input into the degree of membership for each of the fuzzy set.)

Membership Function: A membership function is a curve that defines how each point in the input space is mapped to a membership value (or degree of membership) between 0 and 1.

Rule Base: Where fuzzy rule is defined as a conditional statement in the form: IF x is A. THEN y is B. where x and y are linguistic variables; A and B are linguistic values determined by fuzzy sets on the universe of discourse X and Y, respectively.

Inference System: A system that uses fuzzy set theory to map inputs (features in the case of fuzzy classification) to outputs (classes in the case of fuzzy classification).

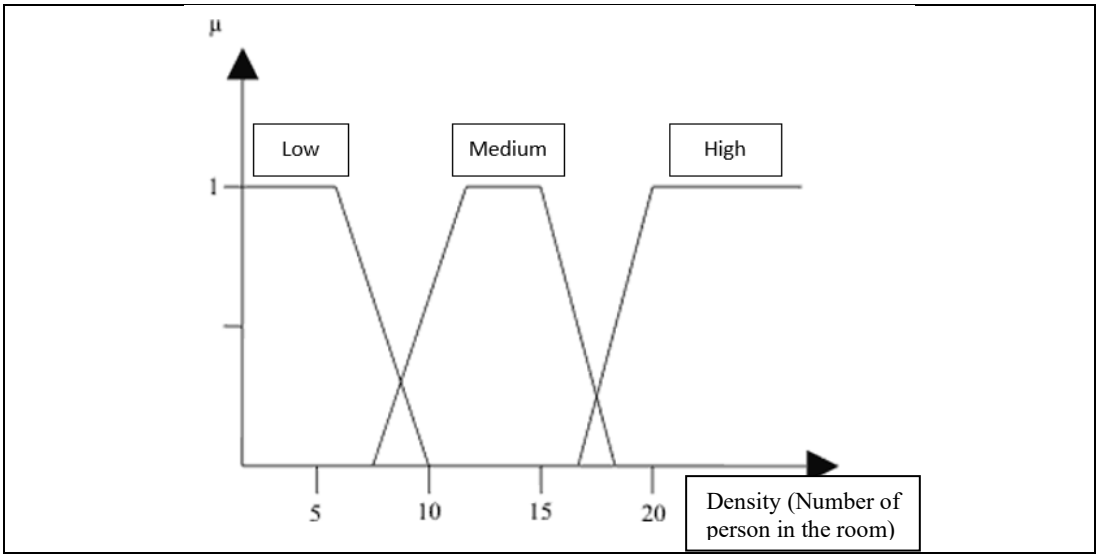
Defuzification: The process of producing a quantifiable result in Crisp logic, given fuzzy sets and corresponding membership degrees. For example, by using Centre of Gravity (COG).

3. Assume that you are designing an automated air conditioning system using Fuzzy Logic.
   a) Suggest an uncertainty in the above scenario with explanation.
   b) Generate the **TWO** (**2**) input and **TWO** (**2**) output fuzzy membership functions that are able to model the air conditioning system.
   c) Suggest **TWO** (**2**) fuzzy rules that associate to the Fuzzy membership functions generated in Question 3(b).

Answer:
a) The value of temperature to be consider as different level of coldness. For example is 21 degree Celsius consider cold or not?

b) Student can provide any appropriate membership function, example:

Input: Temperature of the room (cold, cool, comfortable, warm, hot), humidity (dry, moderate, humid), density, etc.
Output: Volume of the fans, cooling power (low, medium, high), fan speed (low, medium, high) etc.
Example of membership function:



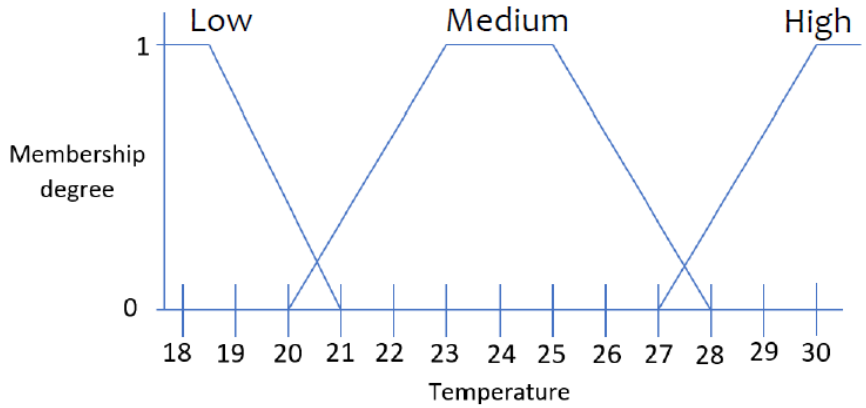
4. Suppose you are designing a fuzzy thermostat with the following membership function of the temperature.

```
Low(x)=     {0,             if temp (x) >= 21,
             f(temp(x)),    if 19 < temp(x) < 21,
             1,             if temp(x) <= 19}

Medium(x)=  {0,             if temp (x) <=20 or temp (x) >= 28,
             f(temp(x)),    if 20 < temp(x) < 23 or 25 < temp(x) <28
             1,             if 23 <= temp(x) <= 25}

High(x)=    {0,             if temp (x) <= 27,
             f(temp(x)),    if 27 < temp(x) < 30
             1,             if temp(x) >= 30}
```

(i) Illustrate the graph with labels to represent the above membership function.
(ii) Propose **ONE (1)** fuzzy rule that can be used in the fuzzy thermostat.

Answer:

i)



ii)

➢ If the room temperature is warm then the fan speed is high

- If the room density is low then fan speed is low.

any of the relevant answer is accepted