Assignment – 4

1. Explain alpha-beta cut off search with an example. State a case when to do alpha pruning.

- Alpha-beta pruning is a modified version of the minimax algorithm. It is an optimization technique for the minimax algorithm.
- ALPHA-BETA cutoff is a method for reducing the number of nodes explored in the Minimaxstrategy.
- For the nodes it explores it computes, in addition to the score, an alpha value anda beta value.
- As we have seen in the minimax search algorithm that the number of game states it has to examine are exponential in depth of the tree. Since we cannot eliminate the exponent, but we can cut it to half. Hence there is a technique by which without checking each node of the game tree we can compute the correct minimax decision, and this technique is called pruning. This involves two threshold parameter Alpha and beta for future expansion, so it is called alpha-beta pruning.
- It is also called as **Alpha-Beta Algorithm**.
- Alpha-beta pruning can be applied at any depth of a tree, and sometimes it not only prune the tree leaves but also entire sub-tree.

- The two-parameter can be defined as:

  (1) Alpha: The best (highest-value) choice we have found so far at any point along the path of Maximizer. The initial value of alpha is -∞.

  (2) Beta: The best (lowest-value) choice we have found so far at any point along the path of Minimizer. The initial value of beta is +∞.

- Alpha-Beta has two types:
  (1) Worst ordering
  (2) Ideal ordering

- Key points about alpha-beta pruning:
- The Max player will only update the value of alpha.
- The Min player will only update the value of beta.
- While backtracking the tree, the node values will be passed to upper nodes instead of values of
- alpha and beta.
- We will only pass the alpha, beta values to the child nodes.

The Alpha-beta pruning to a standard minimax algorithm returns the same move as the standard algorithm does, but it removes all the nodes which are not really affecting the final decision but making algorithm slow. Hence by pruning these nodes, it makes the algorithm fast.

2. Discuss Min-Max search method.

- Mini-max algorithm is a recursive or backtracking algorithm which is used in decision-making and game theory. It provides an optimal move for the player assuming that opponent is also playing optimally.
- Mini-Max algorithm uses recursion to search through the game-tree.
- Min-Max algorithm is mostly used for game playing in AI. Such as Chess, Checkers, tic-tac-toe,go, and various tow-players game. This Algorithm computes the minimax decision for the current state.
- In this algorithm two players play the game, one is called MAX and other is called MIN.
- Both the players fight it as the opponent player gets the minimum benefit while they get themaximum benefit.
- Both Players of the game are opponent of each other, where MAX will select the maximizedvalue and MIN will select the minimized value.
- The minimax algorithm performs a depth-first search algorithm for the exploration of thecomplete game tree.
- The minimax algorithm proceeds all the way down to the terminal node of the tree, thenbacktrack the tree as the recursion.

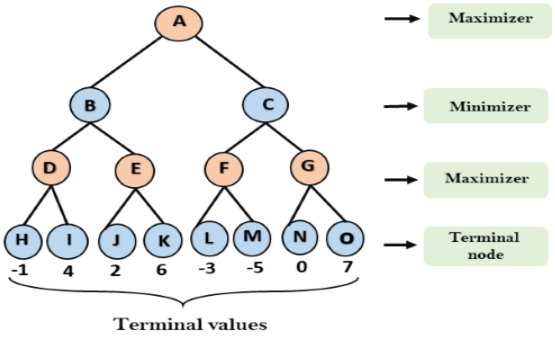
Properties of Mini-Max algorithm:

- **Complete**- Min-Max algorithm is Complete. It will definitely find a solution (if exist), in the finite search tree.
- **Optimal**- Min-Max algorithm is optimal if both opponents are playing optimally.
- **Time complexity**- As it performs DFS for the game-tree, so the time complexity of Min-Max algorithm is O(bm), where b is branching factor of the game-tree, and m is the maximum depth of the tree.
- **Space Complexity-**Space complexity of Mini-max algorithm is also similar to DFS which is O(bm).

Working of Min-Max Algorithm:

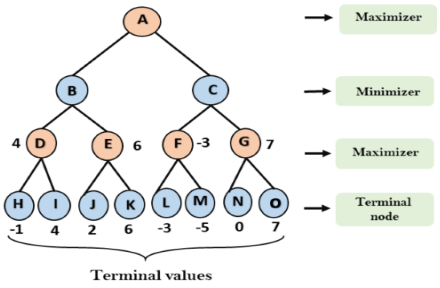
- The working of the minimax algorithm can be easily described using an example. Below we have taken an example of game-tree which is representing the two-player game.
- In this example, there are two players one is called Maximizer and other is called Minimizer.
- Maximizer will try to get the Maximum possible score, and Minimizer will try to get the minimum possible score.
- This algorithm applies DFS, so in this game-tree, we have to go all the way through the leaves to reach the terminal nodes.
- At the terminal node, the terminal values are given so we will compare those value and backtrack the tree until the initial state occurs. Following are the main steps involved in solving the two-player game tree:

Step-1: In the first step, the algorithm generates the entire game-tree and apply the utility function to get the utility values for the terminal states. In the below tree diagram, let's take A is the initial state of the tree. Suppose maximizer takes first turn which has worst-case initial value =- infinity, and minimizer will take next turn which has worst-case initial value = +infinity.



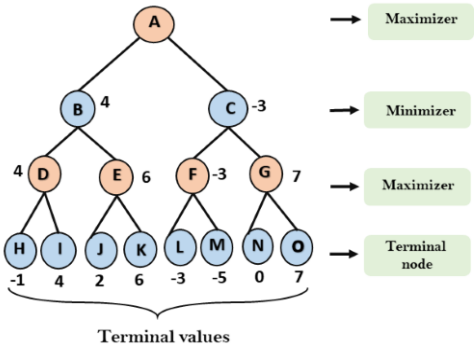
Step 2: Now, first we find the utilities value for the Maximizer, its initial value is -∞, so we will compare each value in terminal state with initial value of Maximizer and determines the higher nodes values. It will find the maximum among the all.

For node D max(-1,- -∞) => max(-1,4)= 4
For Node E max(2, -∞) => max(2, 6)= 6
For Node F max(-3, -∞) => max(-3,-5) = -3
For node G max(0, -∞) = max(0, 7) = 7

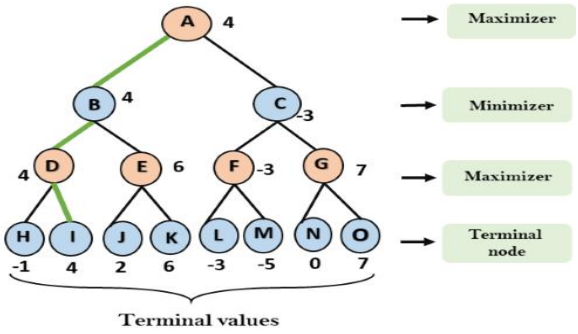


Step 3: In the next step, it's a turn for minimizer, so it will compare all nodes value with +∞, and will find the 3rd layer node values.

For node B= min(4,6) = 4
For node C= min (-3, 7) = -3



Step 4: Now it's a turn for Maximizer, and it will again choose the maximum of all nodes value and find the maximum value for the root node. In this game tree, there are only 4 layers, hence we reach immediately to the root node, but in real games, there will be more than 4 layers.
For node A max(4, -3)= 4



3. Discuss Turing Test.

- In 1950, Alan Turing introduced a test to check whether a machine can think like a human ornot, this test is known as the Turing Test. In this test, Turing proposed that the computer can besaid to be an intelligent if it can mimic human response under specific conditions.
- Turing Test was introduced by Turing in his 1950 paper, "Computing Machinery and Intelligence," which considered the question, "Can Machine think?"
- The Turing test is based on a party game "Imitation game," with some modifications.
- This gameinvolves three players in which one player is Computer, another player is human responder, andthe third player is a human Interrogator, who is isolated from other two players and his job is tofind that which player is machine among two of them.
- Consider, Player A is a computer, Player B is human, and Player C is an interrogator. Interrogatoris aware that one of them is machine, but he needs to identify this on the basis of questions and their responses.

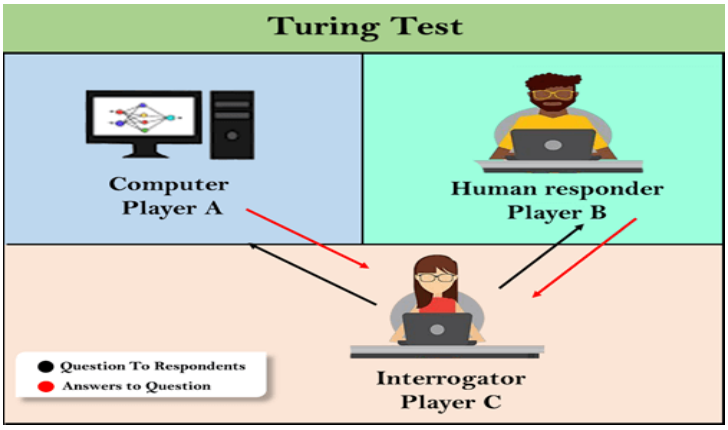
The questions and answers can be like:

- **Interrogator**: Are you a computer?
- **PlayerA** (**Computer**): No
- **Interrogator**: Multiply two large numbers such as (256896489*456725896)
- **PlayerA**: Long pause and give the wrong answer.

Features required for a machine to pass the Turing test:

- **Naturallanguageprocessing**: NLP is required to communicate with Interrogator in generalhuman language like English.
- **Knowledgerepresentation**: To store and retrieve information during the test.
- **Automatedreasoning**: To use the previously stored information for answering the questions.
- **Machinelearning**: To adapt new changes and can detect generalized patterns.

- **Vision (For total Turing test):**To recognize the interrogator actions and other objects during atest.
- **Motor Control (For total Turing test):**To act upon objects if requested.



4. Discuss Simulated Annealing search method. Compare it with hill climbing method.

- A hill-climbing algorithm which never makes a move towards a lower value guaranteed to beincomplete because it can get stuck on a local maximum. And if algorithm applies a randomwalk, by moving a successor, then it may complete but not efficient.
- Simulated Annealing is analgorithm which yields both efficiency and completeness.
- In mechanical term Annealing is a process of hardening a metal or glass to a high temperaturethen cooling gradually, so this allows the metal to reach a low-energy crystalline state.
- The sameprocess is used in simulated annealing in which the algorithm picks a random move, instead ofpicking the best move.
- If the random move improves the state, then it follows the same path.Otherwise, the algorithm follows the path which has probability of less than 1 or it movesdownhill and chooses another path.
- **Simulated annealing** (**SA**) is a probabilistic technique for approximating the global optimum of a given function.

Pseudo code:

- Let $s = s_0$
- For $k = 0$ through k_{max} (exclusive):
  - $T \leftarrow$ temperature($(k+1)/k_{max}$)
  - Pick a random neighbour, $s_{new} \leftarrow$ neighbour(s)
  - If $P(E(s), E(s_{new}), T) \geq$ random(0, 1):
  - $s \leftarrow s_{new}$
- Output: the final state s

Hill Climbing algorithm	Simulated annealing algorithm
It is always gets stuck in a local maxima because downward moves are not allowed.	It is technique that allows downward steps in order to escape from a local maxima.
It is easy to implementation.	It is also easy to implementation.
Hill Climbing is a heuristic search used for mathematical optimization problems in the field of Artificial Intelligence.	Simulated annealing is an algorithm which yields both efficiency and completeness.
Hill climbing algorithm is a local search algorithm which continuously moves in the direction of increasing elevation/value to find the peak of the mountain or best solution to the problem. It terminates when it reaches a peak value where no neighbor has a higher value.	Simulated annealing is based on metallurgical practices by which a material is heated to a high temperature and cooled. At high temperatures, atoms may shift unpredictably, often eliminating impurities as the material cools into a pure crystal.
It is also called greedy local search as it only looks to its good immediate neighbor state and not beyond that.	
Problems in hill climbing are local maxima, plateau, ridges .	Problems are shift input elements randomly, permute input sequence, move all points 0 to 1 units in random direction.

5. Discuss Nonlinear Planning using Constraint Posting with example.

- Let us reconsider the SUSSMAN ANOMALY
- Problems such as this one require subproblems to be worked on simultaneously.
- Thus a nonlinear plan using heuristics such as:
- Try to achieve ON(A,B) clearing block A putting block C on the table.
- Achieve ON(B,C) by stacking block B on block C.
- Complete ON(A,B) by stacking block A on block B.
- Constraint posting has emerged as a central technique in recent planning systems(E.g. MOLGEN and TWEAK)
- Constraint posting builds up a plan by:suggesting operators, trying to order them, and produce bindings between variables in the operators and actual blocks.
- The initial plan consists of no steps and by studying the goal state ideas for the possible stepsare generated.
- There is no order or detail at this stage.
- Gradually more detail is introduced and constraints about the order of subsets of the steps areintroduced until a completely ordered sequence is created.
- In this problem means-end analysis suggests two steps with end conditions ON(A,B) and
- ON(B,C) which indicates the operator STACK giving the layout shown below where the operatoris preceded by its preconditions and followed by its post conditions:

CLEAR(B) CLEAR(C)
*HOLDING(A) *HOLDING(B)

STACK(A,B) STACK(B,C)
ARMEMPTY ARMEMPTY
ON(A,B) ON(B,C)
CLEAR(B)
CLEAR(C)
HOLDING(A)
HOLDING(B)

- A plan that consists of sub-problems, which are solved simultaneously is said to be a non-linearplan.

6. Explain various steps of Natural Language Processing.

- NLP stands for Natural Language Processing, which is a part of Computer Science, Humanlanguage, and Artificial Intelligence. It is the technology that is used by machines to understand,analyse, manipulate, and interpret human's languages.
- It helps developers to organize knowledgefor performing tasks such as translation, automatic summarization, Named Entity Recognition(NER), speech recognition, relationship extraction, and topic segmentation.
- Natural Language Processing (NLP) refers to AI method of communicating with an intelligentsystems using a natural language such as English.

Components of NLP

- There are two components of NLP as given:

1. Natural Language Understanding (NLU)

- Natural Language Understanding (NLU) helps the machine to understand and analyse humanlanguage by extracting the metadata from content such as concepts, entities, keywords,emotion, relations, and semantic roles.
- NLU mainly used in Business applications to understand the customer's problem in both spokenand written language.

NLU involves the following tasks -

- It is used to map the given input into useful representation.
- It is used to analyze different aspects of the language.

2. Natural Language Generation (NLG)

- Natural Language Generation (NLG) acts as a translator that converts the computerized datainto natural language representation. It mainly involves Text planning, Sentence planning, andText Realization.

NLP Libraries:

1. Scikit learn
2. Natural language Toolkit (NLTK)
3. Pattern
4. Textblob
5. Quepy
6. Spacy
7. Gensim

Steps in NLP:

There are general five steps –

1. Lexical Analysis
2. Syntactic Analysis (Parsing)
3. Semantic Analysis
4. Disclosure Integration
5. Pragmatic Analysis

NLP Terminology:

- **Phonology**− It is study of organizing sound systematically.
- **Morphology**− It is a study of construction of words from primitive meaningful units.
- **Morpheme**− It is primitive unit of meaning in a language.
- **Syntax**− It refers to arranging words to make a sentence. It also involves determining thestructural role of words in the sentence and in phrases.
- **Semantics**− It is concerned with the meaning of words and how to combine words intomeaningful phrases and sentences.
- **Pragmatics**− It deals with using and understanding sentences in different situations and howthe interpretation of the sentence is affected.
- **Discourse**− It deals with how the immediately preceding sentence can affect the interpretationof the next sentence.
- **WorldKnowledge**− It includes the general knowledge about the world.

Application:

- Question answering
- Chatbot
- Spam detection
- Sentiment analysis
- Machine translation
- Spelling correction
- Speech recognition
- Information extraction
- Natural language understanding (NLU)

AdvantagesofNLP:

- NLP helps users to ask questions about any subject and get a direct response within seconds.
- NLP offers exact answers to the question means it does not offer unnecessary and unwantedinformation.
- NLP helps computers to communicate with humans in their languages.
- It is very time efficient.

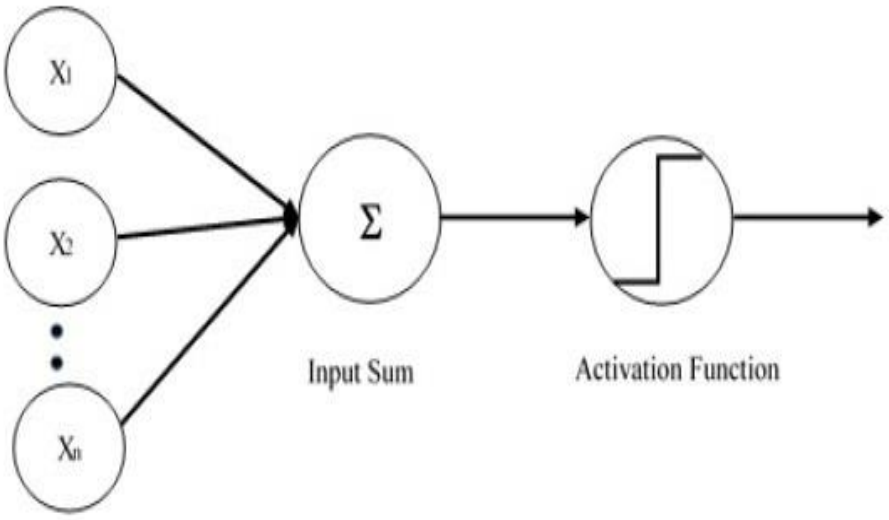
DisadvantagesofNLP:

- NLP may not show context.
- NLP is unpredictable
- NLP may require more keystrokes.

- NLP is unable to adapt to the new domain, and it has a limited function that's why NLP is built fora single and specific task only.

7. Discus perceptron.

- Perceptron is a single layer neural network, or we can say a neural network is a multi-layerperceptron. Perceptron is a binary classifier, and it is used in supervised learning. A simplemodel of a biological neuron in an artificial neural network is known as **Perceptron**.
- A function that can decide whether or not an input which is represented by a vector of numberbelongs to some specific class is known as binary classifiers.
- The binary classifier is a type oflinear classifier. A linear classifier is a classification algorithm which makes its predictions basedon a linear predictor function combining a set of weight with the feature vector.
- The perceptron algorithm was designed to categorizing subjects into one of two types,classify visual input and separating groups with a line. Classification is a key part of image processing and machine learning.
- The perceptron algorithm classifies patterns, i.e., find and classify by many different means using machine learning algorithm, and groups by finding the linear separation between different objects and patterns which are received through numeric or visual input.



- Single layer perceptron is the first proposed neural model created. The content of the localmemory of the neuron consists of a vector of weights.
- The computation of a single layerperceptron is performed over the calculation of sum of the input vector each with the valuemultiplied by corresponding element of vector of the weights.
- The value which is displayed in theoutput will be the input of an activation function.

Perceptron consist of four parts and which are required to understand for the implementation of the perceptron model in PyTorch.

- **Input values or one input layer**: The input layer of a perceptron is made of artificial input neurons and brings the initial data intothe system for further processing.

- **Weights and bias:**Weight represents the strength or dimension of the connection between units. If the weight fromnode 1 to node 2 has the greater quantity, then neuron 1 has greater influence over neuron 2.How much influence of the input will have on the output, is determined by weight.Bias is similar to the intercept added in a linear equation. It is an additional parameter which task is to adjust the output along with the weighted sum of the inputs to the neuron.

- **ActivationFunction:**A neuron should be activated or not, is determined by an activation function. Activation functioncalculates a weighted sum and further adding bias with it to give the result.

- **NeuralNetwork:** Itis based on the Perceptron, so if we want to know the working of the neuralnetwork, learn how perceptron work.

Limitations of Perceptron:

- (1) The output values of a perceptron can take on only one of two values (0 or 1) due to thehard-limit transfer function.
- (2) Perceptron can only classify linearly separable sets of vectors. If a straight line or a planecan be drawn to separate the input vectors into their correct categories, the input vectors arelinearly separable. If the vectors are not linearly separable, learning will never reach a point whereall vectors are classified properly
- The Boolean function XOR is not linearly separable (Its positive and negative instances cannotbe separated by a line or hyperplane). Hence a single layer perceptron can never compute theXOR function. This is a big drawback which once resulted in the stagnation of the field of neuralnetworks. But this has been solved by multi-layer.

8. Discuss various defuzzification methods.

Defuzzification:

- It is the inversion of fuzzification, there the mapping is done to convert the crisp results intofuzzy results but here the mapping is done to convert the fuzzy results into crisp results.
- This process is capable of generating a nonfuzzy control action which illustrates the possibilitydistribution of an inferred fuzzy control action.

- **Defuzzification**is a module or component, which takes the fuzzy set inputs generated bythe Inference Engine, and then transforms them into a crisp value.

- It is the last step in theprocess of a fuzzy logic system. The crisp value is a type of value which is acceptable by theuser. Various techniques are present to do this, but the user has to select the best one forreducing the errors.
- **Defuzzification**is the process of producing a quantifiable result in Crisp logic, given fuzzy setsand corresponding membership degrees. It is the process that maps a fuzzy set to a crisp set.
- It is typically needed in fuzzy control systems. These will have a number of rules that transforma number of variables into a fuzzy result, that is, the result is described in terms of membershipin fuzzy sets.
- For example, rules designed to decide how much pressure to apply might result in"Decrease Pressure (15%), Maintain Pressure (34%), increase Pressure (72%)". defuzzification isinterpreting the membership degrees of the fuzzy sets into a specific decision or real value.
- The simplest but least useful defuzzification method is to choose the set with the highestmembership, in this case, "Increase Pressure" since it has a 72% membership, and ignore the others, and convert this 72% to some number.
- The problem with this approach is that it loses information. The rules that called for decreasing or maintaining pressure might as well have not been there in this case.

Defuzzification methods:

1. Max-Membership Principle: This method is also known as height method and is limited to peak output functions. This method is given by the algebraic expression:
μ(z*) >= **μ**(z) for all z ∈ Z.

2. Centroid Method: This method is also known as the centre of mass, centre of area or centre of gravity. It is the most commonly used defuzzification method. The defuzzified output z* is given by:
z* = ∫μ(z).zdz / ∫μ(z)dz

3. Weighted Average Method: This method is valid for symmetrical output membership functions only. Each membership function is weighted by its maximum membership value. The output in the case is given by
z* = ∑μ(z').z' / ∑μ(z') ; where z' is the maximum value of the membership function.

4. Mean-Max Membership: This method is also known as the middle of the maxima. This is closely related to the max-membership method, except that the locations of the maximum membership can be nonunique. The output here is given by:
z* = ∑z' / n ; where z' is the maximum value of the membership function.

5. Centre of Sums: This method employs the algebraic sum of the individual fuzzy subsets instead of their union. The calculations here are very fast, but the main drawback is that the intersecting areas are added twice. The defuzzified value z* is given by
z* = ∫ z*∑μ(z).zdz / ∫ ∑μ(z)dz

6. Centre of Largest Area: This method can be adopted when the output of at least two convex fuzzy subsets which are not overlapping. The output, in this case, is biased towards a side of one membership function. When output fuzzy st has at least two convex regions, then the centre of gravity of the convex fuzzy subregion having the largest are is used to obtain the defuzzified valuez*. The value is given by

$$z^* = \int \mu c(z).zdz / \int \sum \mu c(z)dz$$

9. Define: Frames. Draw Semantic Net for following statements. a) Every kid likes candy. b) Every school going kid likes candy.

- Marvin Minsky proposed 1975 frames are as a means of common sense knowledge.
- It proposed that knowledge is organized into small packets called frames. The content of frames are certain slots which have values.
- Frames are record like structure that have slots and slot values for an entity.
- A frame can be defined as static data structure that has slots for various objects and a collection of frames consists of expectation for a given situation.
- All frames of a given situation constitute the system, whenever one encounter a situation, a series of related frames are activated and reasoning is done.
- A frame is a digital data transmission unit in computer networking and telecommunication.
- A frame typically include frame synchronization features consisting of a sequence of bits or symbols that indicate to the receiver the beginning and end of the payload data within the stream of symbols or bits it receives.

Types of frames:

Frame is used to represent two types of knowledge

1. Declarative / Factual /Situational/ frame
2. Procedural/ Action frame

Three types of frames can be found in a frame system

1. Class frame
2. Sub class frame
3. An instance frame

Frame relationship:

1. Is-A relationship
2. A-part-of relationship
3. Semantic relationship

Advantages:

- The knowledge can be structured.
- Flexible inference by using procedural knowledge.
- Layered representation and inheritance is possible.

Disadvantages:

- The design of the interpreter is not easy.
- The validity of the inferences is not guaranteed.
- Hard to maintain consistency between the knowledge.

Examples:

(a) Every kid likes candy. (b) Every school going kid likes candy.