1. Assume in the search tree below, the open list is [ A ] initially. Please describe the traversal of the search tree

(a) by depth first search. (15)

(b) by breadth first search. (15)

Note you need to describe the states of the open list as the search progresses.

1. Please explain briefly the meaning of each term.
2. *Perceptron*  (10)
3. *Entropy*  (10)
4. *Dendrogram*  (10)
5. *Inverse document frequency (IDF)*  (10)
6. 感知器（Perceptron）：Perceptron:

A perceptron is a basic artificial neuron model used for binary classification tasks. It takes multiple inputs, each with an associated weight, and computes their weighted sum. This sum then passes through an activation function (like a step function) to produce an output. The goal of a perceptron is to adjust its weights to make its output match the expected output as closely as possible. Multiple perceptrons can be combined to form more complex neural networks.

感知器是一種簡單的人工神經元模型，用於二元分類問題。它接收多個輸入，每個輸入都有一個權重，然後計算加權和。這個加權和通過一個激活函數（例如符號函數）產生輸出。感知器的學習目標是通過調整權重來使其輸出盡可能接近預期輸出。多個感知器可以組成更複雜的神經網絡。

1. Entropy:

Entropy is a measure used in information theory to quantify the uncertainty of a random variable. In machine learning and data science, entropy is often used to measure the disorder or uncertainty of a system or random variable. Higher entropy indicates more disorder or uncertainty, while lower entropy indicates more order or certainty.

熵（Entropy）：

熵是信息理論中用來衡量隨機變量不確定性的度量。在機器學習和數據科學中，熵通常用於衡量一個系統或隨機變量的混亂程度或不確定性程度。熵的值越高，表示系統越混亂或不確定；反之，熵的值越低，表示系統越有序或確定。

1. Dendrogram:

A dendrogram is a tree-like diagram used to display the results of hierarchical clustering. It shows the similarity or distance between different data points (or clusters) during the clustering process. A dendrogram starts from a root node and branches downward, with each branch representing a clustering step. The horizontal axis typically represents samples or clusters, while the vertical axis represents similarity or distance.

分支圖（Dendrogram）：

分支圖是一種樹狀圖表，用於顯示層次聚類或階層性聚類的結果。它展示了聚類過程中不同數據點（或群集）之間的相似性或距離。分支圖從根節點開始，向下分支，直到最終的葉子節點，每個分支代表一個聚類步驟。分支圖的橫軸通常表示樣本或群集，縱軸表示相似性或距離。

1. Inverse Document Frequency (IDF):

Inverse Document Frequency (IDF) is a technique used in information retrieval and natural language processing to assess the importance of a term in a collection of documents. The IDF value is inversely proportional to the frequency of the term across the document collection – common terms have low IDF values, while uncommon terms have high IDF values. IDF is used to weight the importance of terms in retrieval results and is often multiplied with term frequency (TF) to form TF-IDF values used in document retrieval and information retrieval tasks.

逆文檔頻率（IDF）：

逆文檔頻率是一種用於信息檢索和自然語言處理的技術，用於評估一個詞彙對於一個文件集合的重要性。IDF的值與詞彙在文件集合中的出現頻率成反比，即常見詞彙的IDF值較低，不常見詞彙的IDF值較高。IDF用於加權詞彙在檢索結果中的重要性，通常與詞彙的詞頻（TF）相乘，形成TF-IDF（詞頻-逆文檔頻率）值，用於文檔檢索和信息檢索任務中。

1. Please answer the questions about search on game tree.
2. Please illustrate how min-max algorithm is performed for the following game tree. 為以下遊戲樹執行最小-最大演算法 (10)
3. Please describe the detailed process (e.g. alpha-pruning, beta-pruning) when alpha-beta pruning is applied to the min-max. 請描述將 alpha-beta 修剪應用於最小-最大值時的詳細過程 (10)
4. Please encircle those nodes that are visited in (b). (10)

圖一

3 2 9 6 0 8 2 4 1 5 7

MAX

MIN

MAX

MIN

圖二

4 6 9 3 5 2 3 4 7 1 8 3

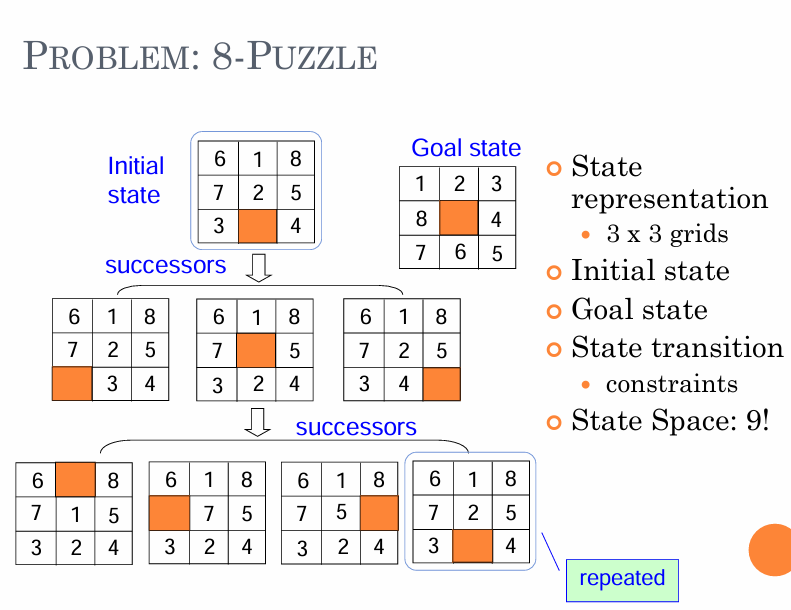
MAX

MIN

MAX

MIN

1. Please answer the following questions about search.
2. Use the problem of 3x3 puzzle to illustrate the basic factors for defining a state space search. 使用 3x3 謎題來說明定義狀態空間搜索的基本要素 (15)



State : Represent where we have reached, a "state" refers to a specific configuration or situation that the problem-solving agent can occupy.

「狀態」是指問題解決代理可以佔據的特定配置或情況。

State Space: Space consisting of all possible states , could be Huge or infinite, depend on the rules or description of question

In the context of a 3x3 puzzle problem, each arrangement and position of puzzle pieces represents a unique state. Therefore, the state space is the collection of all possible puzzle arrangements, and each state can be represented by a state vector.

狀態空間 (State Space)：在3x3拼圖問題中，每個拼圖的排列和位置都代表一種狀態。因此，狀態空間是所有可能的拼圖排列的集合，每個狀態可以用一個狀態向量表示。

Initial State: The initial state is the starting arrangement of the puzzle, which is the state at the beginning of the problem.

初始狀態 (Initial State)：初始狀態是拼圖的起始排列，即問題開始時的狀態。

Goal State: The goal state is the desired arrangement of the puzzle that we aim to achieve, which is the objective of solving the problem.

目標狀態 (Goal State)：目標狀態是我們希望達到的拼圖排列，也是解決問題的目標。

State transition: like Constraints, Conditions that limit the generation of successor states, such as blank cells that cannot move out of the boundary.

約束(Constraints):限制後繼狀態生成的條件,比如空白格不能移出邊界。

1. Explain briefly the constraints of algorithm A and algorithm A\*.簡要說明演算法 A 和A\* (10)

一張含有 文字, 圖表, 行, 折紙 的圖片

自動產生的描述一張含有 文字, 圖表, 行, 折紙 的圖片

自動產生的描述

Explain briefly the constraints of algorithm A and algorithm A\*.

簡要說明演**算法 A** 和**演算法 A\*** 的約束與限制 (10)

Algorithm A is suitable for **simple search problems** but may be limited by:

Potential inefficiency when dealing with large state spaces due to excessive search options.

No guarantee of finding the optimal solution during the search process, especially in larger search spaces.

**算法 A** 通常適用於簡單的搜索問題，但它可能會受到以下限制：

對於大型狀態空間，算法 A 可能會陷入過多的搜索選項，導致搜索效率低下。

算法 A 在搜索過程中不一定保證找到最佳解決方案，特別是在存在較大的搜索空間時。

Key constraints and conditions of Algorithm A\* include:

The effectiveness of the chosen heuristic function significantly impacts the performance of Algorithm A\*. **A poor heuristic may result in suboptimal solutions.**

It efficiency heavily relies on the **accuracy and applicability of the selected heuristic function**.

Algorithm A\* may face challenges in terms of **search efficiency when dealing with very large state spaces or complex search graphs**.

算法 A\* 的主要限制和條件包括：

啟發式函數的選擇可能會影響算法的效果**。一個不良的啟發式函數**可**能導致算法 A\* 僅產生次優解。**

算法 A\* 的效率高度**依賴於所選用的啟發式函數的準確性和適用性**。

算法 A\* **在狀態空間非常大或搜索圖複雜時**，仍可能面臨搜索效率的挑戰。

In summary, Algorithm A\* offers **higher search efficiency and better search accuracy** compared to Algorithm A but **requires appropriate heuristic functions and parameter adjustments** to address various search problems and environments. 總之，算法 A\* 相較於算法 A 具有更高的**搜索效率和更好的搜索準確性**，但也需要**合適的啟發式函數和適當的參數調整**，以應對不同的搜索問題和環境。

<https://www.almabetter.com/bytes/tutorials/artificial-intelligence/state-space-search-in-artificial-intelligence>

Algorithm A: Based on the best-first strategy

 Consider an evaluation function f(n) = g(n) + h(n) for each state n, where

 g(n) is the cost of n from the start state (have known)

 h(n) is the heuristic estimate of the cost of going from n to a goal node

If such evaluation function is used with the best-first search strategy, the algorithm is called Algorithm A.

演算法A：基於最佳優先策略

 考慮評估函數 f(n) = g(n) + h(n)，每個狀態 n，其中

g(n) 是從起始狀態開始的 n 的成本（已知）

h(n) 是從 n 開始的成本的啟發式估計到目標節點，如果這種評估函數與最佳優先順序一起使用搜尋策略，該演算法稱為演算法A。

Advantages: **Efficient** and often **optimal in finding** **the shortest path**.

Limitations: Heuristic quality greatly impacts performance. May not always be more efficient than BFS or DFS.

Algorithm A\*

search is an informed search algorithm that combines the principles of both BFS and DFS. It uses a heuristic function to estimate the cost of reaching the goal from each state. A\* considers both the cost of reaching a state and the estimated cost to the goal, making it a best-first search algorithm.

A\*搜尋是一種結合了BFS和DFS原理的知情搜尋演算法。它使用啟發式函數來估計從每個狀態達到目標的成本。 A\* 同時考慮達到某個狀態的成本和達到目標的估計成本，使其成為最佳優先搜尋演算法

If algorithm A is used with an evaluation function in which h(n) is less or equal to the cost of the minimum path from n to the goal, h\*(n), the resulting search algorithm is called Algorithm A\*. (Admissible) 如果演算法 A 與評估函數一起使用，其中 h(n) 小於或等於從 n 到目標的最小路徑的成本 h\*(n)，則產生的搜尋演算法稱為演算法 A\*。 （可接受）

Optimality: A\* is guaranteed to find an optimal solution, meaning it finds the shortest path to the goal in terms of total cost.

最優性：A\* 保證找到最優解決方案，這意味著它在總成本方面找到實現目標的最短路徑。

Completeness: meaning it will always find a solution if one exists, as long as the state space has a finite branching factor and the heuristic is admissible完備性：這意味著只要狀態空間具有有限的分支因子並且啟發式是可接受的，如果存在解，它總是會找到解

1. If there are two heuristics h1(n) and h2(n) for A\* search and h1(n) ≦ h2(n) for any state n. What is the difference of the two heuristics for search?(5)

For two A\* heuristics h1 and h2, if h1(n) ≦ h2(n) fir all states n in the search space, heuristic h2 is said to be more informed than h1. 對於兩個 A\* 啟發式 h1 和 h2，如果 h1(n) ≤ h2(n) 則所有在搜尋空間中狀態 n，h2 稱為比 h1 更了解狀況。

The key difference is in the accuracy of their estimates:

* h1(n) tends to underestimate the true cost, making it potentially more optimistic but less informative.
* h2(n) tends to overestimate the true cost, providing a more cautious and potentially more accurate estimate.
* h1(n) 傾向低估真實成本，其在搜索過程中可能更樂觀但信息量較少。
* ℎ2(𝑛) 傾向於高估真實成本，提供更謹慎和潛在更準確的估計。

h1(n) is more optimistic and may guide the search towards the goal more quickly but with potentially suboptimal paths.

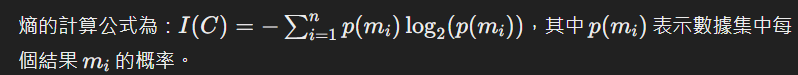
h2(n) is more conservative and provides safer estimates, potentially leading to more accurate and optimal paths at the cost of increased computational effort.

* h1(n) 更樂觀，會更快地引導搜索到達目標，但可能導致次優徑徑。
* ℎ2(𝑛)更保守，提供更安全的估計，可能會導致更準確和更優的路徑，但計算成本更高。

1. Please describe how inductive learning algorithm (ID3) for decision trees is conducted請描述決策樹的歸納學習演算法（ID3）是如何進行的. (20).

**1.Entropy Calculation**:

* Calculate the entropy (measure of uncertainty) of the current dataset C using the formula:
* Here, 𝑝(𝑚𝑖) represents the probability of each outcome 𝑚𝑖in the dataset.
* Entropy is used to determine the degree of chaos or uncertainty in the dataset. Lower entropy signifies higher homogeneity.



**2.Question Selection**:

* Choose a question (property) Q from a set of available questions that maximally reduces the entropy of the dataset. 從一組可用的問題中選擇一個能夠最大程度降低數據集熵值的問題
* Questions can partition the dataset 𝐶 into smaller subsets 𝐶1,𝐶2,...,𝐶𝑛based on specific criteria (e.g., income level, credit history).問題可以基於特定標準(如收入水平、信用紀錄)將數據集分成更小的子集

**3.** **遞歸切割Recursive Partitioning**:

* For each subset 𝐶𝑖 resulting from the question 𝑄:
* 把每個問題Q分割的子集ci :
* **If *Ci*​ has an entropy of 0** (indicating homogeneity), it becomes a leaf node in the decision tree. 如果商職為0則將其作為葉子節點
* Otherwise, recursively apply the ID3 algorithm to further partition 𝐶𝑖​ using the remaining set of questions. 否則利用ID3演算法對ci使用剩餘的問題進行再次切割

**4.建構決策數Tree Building**:

* As the algorithm progresses, a decision tree structure is built where each node represents a question and each leaf node represents a decision (output category). 隨著算法的進行，逐步構建決策樹結構，其中每個節點代表一個問題，每個葉子節點代表一個決策結果。

The goal of the ID3 algorithm is to efficiently partition the dataset into smaller subsets using the most informative questions (maximally reducing entropy) to construct a decision tree that accurately predicts outcomes for new, unseen data.

The effectiveness of the ID3 algorithm relies on the choice of questions and the ability to reduce entropy rapidly, leading to well-structured decision trees that generalize well to new data.

ID3 演算法的目標是使用資訊最豐富的問題（最大限度地減少熵）將資料集有效地劃分為更小的子集，以建立決策樹，從而準確預測新的、未見過的資料的結果。

Basic algorithm (a greedy algorithm)

* Tree is constructed in a top-down recursive divide-and-conquer manner
* At start, all the training examples are at the root
* Attributes are categorical (if continuous-valued, they are discretized in advance)
* Examples are partitioned recursively based on selected attributes
* Test attributes are selected based on a heuristic or statistical measure (e.g., Information gain、Gain ratio、Gini index

Conditions for stopping partitioning

* All samples for a given node belong to the same class
* There are no remaining attributes for further partitioning – majority voting is employed for classifying the leaf node
* There are no samples left

基本演算法（貪婪演算法）  
樹以自上而下的遞歸分而治之的方式構造  
一開始，所有的訓練示例都在根部  
屬性是分類的（如果連續值，則提前離散化）  
示例根據所選屬性遞歸分區  
根據啟發式或統計度量（例如，資訊增益）選擇測試屬性  
停止分區的條件  
給定節點的所有樣本都屬於同一類  
沒有剩餘的屬性用於進一步分區 - 多數投票用於對葉節點進行分類  
沒有樣品了

1. Please describe the process of agglomerative clustering algorithm, including how the dendrogram is generated and how the clusters are determined. (20)
2. Please explain the meaning of each term (in *italic*) and its usage in the algorithm.
3. *Temperature* in simulated annealing. (8)
4. *Perceptron* in artificial neural network. (8)
5. *Pheromone* in ant colony optimization. (8)
6. *Gradient Descent* for optimization. (8)
7. *Schema* for genetic algorithm. (8)
8. Please answer the questions.
   1. What are the criteria for determining optimal question when spanning each node in the inductive learning algorithm of decision tree? (5)
   2. What are the limitations of classification and regression tree when applied to classification problems? (5)
   3. How many parameters are there in a Gaussian mixture model of 10 mixtures have for 39 dimensional points? (assuming its covariance matrixes are all diagonal) (5)
   4. Assume there are 4 red balls, 2 white balls and 1 blue ball and 1 green ball in a basket. Please find the entropy for the distribution of the balls. (10)
   5. Given two vectors X1=[0 1 1 0 1 0 0 1 0 1] and X2=[1 0 1 1 0 0 1 1 0 0], please compute the simple matching coefficient, the Jaccard coefficient and the Rao’s coefficient. (10)
9. The joint distribution for two random variables, X and Y, is shown as below.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Y=A | Y=B | Y=C |
| X=1 | 0.06 | 0.12 | 0.30 |
| X=2 | 0.09 | 0.25 | 0.18 |

a) Please compute the conditional probability P(X=1|Y=A) (5)

b) Are the two variables statistically independent? Why? (5)

c) Assume we observe a fact that Y is impossible to be C. What could we make decision for X (X=1 or X=2) under this condition? What decision should we make if no obsertation is available? (10)

1. Assume the training data { (xi, yi) }, i = 1~n are given for regression analyses. All xi’s and yi’s are scalars. Describe how to find the regression coefficients for the following regression functions.

(a) Y = a + bX (10)

(b) Y = c + d·X + e·X2 + f·X3 (10)

(c) Y = ·eX (10)

1. Suppose we have a set of data, {(xi, yi)} where xi‘s and yi‘s are all real numbers. Illustrate how a nonlinear regression problem could be formulated as a linear regression problem with variable transformation for the following families of regression functions, respectively.

(a) . (10)

(b) (10)

1. Given two vectors X1=[0 1 1 0 1 0 0 1 0 1] and X2=[1 0 1 1 0 0 1 1 0 0], please compute the simple matching coefficient, the Jaccard coefficient and the Rao’s coefficient. (20)