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.

一張含有 文字, 螢幕擷取畫面, 圖表, 字型 的圖片

自動產生的描述

一張含有 文字, 螢幕擷取畫面, 圖表, 行 的圖片

自動產生的描述

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:

講義上的Perceptron

A perceptron is a single neuron that classifies a set of inputs into one of two categories.

 If the inputs are in the form of a grid, a perceptron can be used to recognize visual images of shapes.

 The perceptron usually uses a step function, which returns 1 if the weighted sum of inputs exceeds a threshold, and –1 otherwise

感知器是單一神經元，它將一組輸入分類為兩個類別之一。  如果輸入採用網格形式，則感知器可用於辨識形狀的視覺影像。  感知器通常使用階躍函數，如果輸入的加權和超過閾值，則傳回 1，否則回傳 –1

一張含有 圖表, 行, 文字, 字型 的圖片

自動產生的描述

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 n

eural 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)

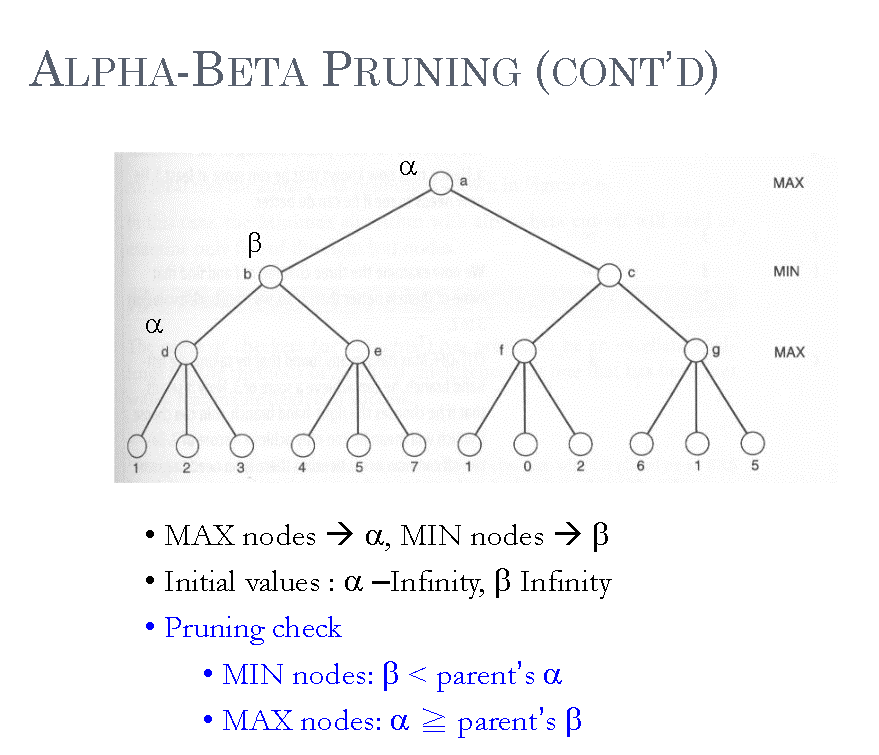
MAX nodes → a, MIN nodes → b

• Initial values : a –Infinity, b Infinity

• Pruning check

• MIN nodes: b < parent’s a

• MAX nodes: a ≧ parent’s b

 一張含有 圖表, 行, 樣式 的圖片

自動產生的描述

1. 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

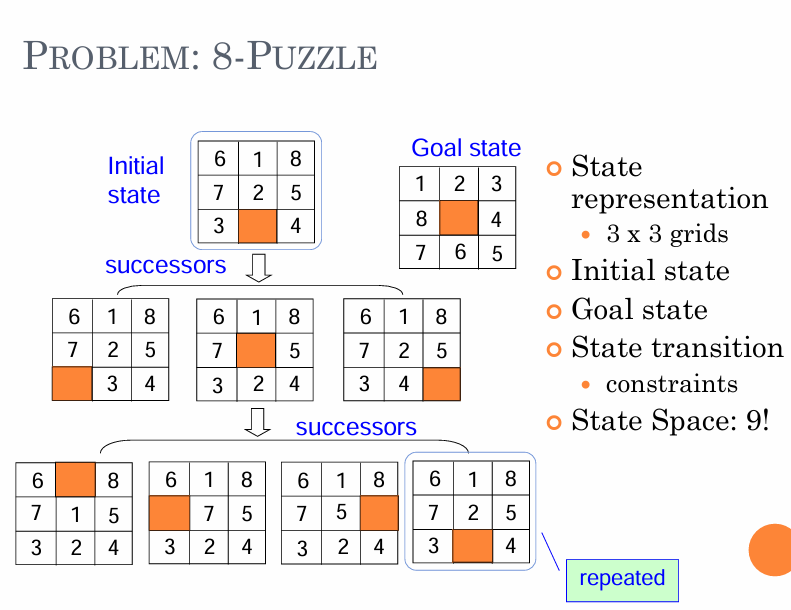
MAX

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。

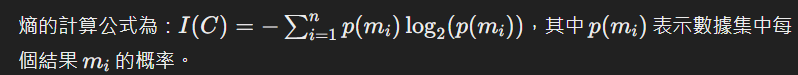
1. Please describe how inductive learning algorithm (ID3) for decision trees is conducted請描述決策樹的歸納學習演算法（ID3）是如何進行的. (20).

1. ID3: Iterative Dichotomiser 3, an algorithm based on Information Gain, which is based on the concept of entropy to measure the information after classification of each feature The higher the amount of information, the more priority it will take as a decision-making condition.

1. ID3：疊代二元樹3代（Iterative Dichotomiser 3），一種基於資訊獲利（Information Gain）的演算法，其基礎是借助熵（Entropy）的概念來衡量每種特徵分類後的資訊量，資訊量越高就越優先做為決策條件。

**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.請描述聚集聚類演算法的過程，包括樹狀圖是如何生成的，以及聚類是如何確定的 (20)

Agglomerative algorithm (aggregation)

 A bottom-up process

 Rain drops (small → big)

 Regards each object as a cluster initially

 The clusters are merged into larger clusters

 A dendrogram is constructed

凝聚演算法（聚合） ❖ 自下而上的過程  雨滴（小→大） ❖ 最初將每個物件視為一個簇 ❖ 簇被合併成更大的簇 ❖ 建構樹狀圖

1.Place each sample in its own cluster. Construct a list of inter-cluster distances for all pairs of samples, and sort this list in ascending order. 將每個樣本放入自己的簇中。構造一個所有對的簇間距離列表樣本，並按升序排列此清單。

2. Step through the sorted list of distances, forming for each distinct threshold value dk a graph of the samples where pairs of samples closer than dk are connected into a new cluster by a graph edge. If all the samples are members of a connected graph, stop. Otherwise, repeat this step. 2. 遍歷已排序的距離列表，形成對於每個不同的閾值 dk 的圖比 dk 更近的樣本對是透過圖邊連接成一個新的簇。如果所有樣本都是連通圖的成員，停止。否則，請重複此步驟。

3. The output of the algorithm is a nested hierarchy of graphs, which can be cut at the desired dissimilarity level forming a partition (clusters) identified by simple connected components in the corresponding subgraph. 3.演算法的輸出是一個嵌套的層次結構圖表，可以根據需要進行切割形成分區（簇）的相異程度

由簡單的連接組件來識別對應的子圖。

1. Please explain the meaning of each term (in *italic*) and its usage in the algorithm.
2. *Temperature* in simulated annealing. (8)
3. *Perceptron* in artificial neural network. (8)
4. *Pheromone* in ant colony optimization. (8)
5. *Gradient Descent* for optimization. (8)
6. *Schema* for genetic algorithm. (8)

(i) \*Temperature in simulated annealing\*:

- \*\*Meaning\*\*: 在模擬退火算法中，溫度代表著系統的隨機性程度和接受更差解的機率。在算法開始時，溫度較高，系統更容易接受較差的解答；隨著時間的推移，溫度逐漸降低，系統更傾向於接受更好的解答。

- \*\*Usage\*\*: 模擬退火算法中的溫度參數影響著解答的品質和搜索的效率，可以根據問題的特性調整溫度的變化方式。

\*Temperature in simulated annealing\*:

- \*\*Meaning\*\*: In simulated annealing, temperature represents the level of randomness in the system and the probability of accepting worse solutions. At the beginning of the algorithm, the temperature is high, making it easier for the system to accept worse solutions; as time progresses, the temperature decreases, and the system becomes more inclined to accept better solutions.

- \*\*Usage\*\*: The temperature parameter in simulated annealing affects the quality of solutions and the efficiency of search, and it can be adjusted based on the characteristics of the problem to control the variation of temperature.

(j) \*Perceptron in artificial neural network\*:

- \*\*Meaning\*\*: 在人工神經網絡中，感知器是一種簡單的人工神經元，用於二元分類任務。它通過加權和激活函數來計算輸入的加權和，並根據結果進行二元分類。

- \*\*Usage\*\*: 感知器是人工神經網絡中的基本組成單元，通常用於解決簡單的分類問題，但對於複雜的非線性問題效果有限。

\*Perceptron in artificial neural network\*:

- \*\*Meaning\*\*: In artificial neural networks, a perceptron is a simple artificial neuron used for binary classification tasks. It computes a weighted sum of inputs using weights and an activation function, then performs binary classification based on the result.

- \*\*Usage\*\*: Perceptrons are basic building blocks in artificial neural networks, typically used for solving simple classification problems, but they have limitations in handling complex nonlinear problems.

(k) \*Pheromone in ant colony optimization\*:

- \*\*Meaning\*\*: 在螞蟻族群優化中，信息素是一種螞蟻之間通信的化學物質。螞蟻在搜索過程中會釋放信息素，其他螞蟻會根據信息素的濃度選擇路徑，從而實現全局搜索和最優化。

- \*\*Usage\*\*: 信息素的蒸發和更新機制影響螞蟻族群的搜索效率和收斂速度，可以通過調整信息素的分泌和感知規則來優化算法的性能。

\*Pheromone in ant colony optimization\*:

- \*\*Meaning\*\*: In ant colony optimization, pheromones are chemical substances used by ants to communicate with each other. Ants release pheromones during the search process, and other ants choose paths based on the concentration of pheromones, enabling global search and optimization.

- \*\*Usage\*\*: The evaporation and updating mechanism of pheromones affect the search efficiency and convergence speed of ant colonies. The performance of the algorithm can be optimized by adjusting the secretion and perception rules of pheromones.

(l) Gradient Descent for optimization:

Meaning: 梯度下降是一種優化算法，用於尋找函數的局部最小值或最大值。它通過計算函數的梯度（斜率）方向來更新參數，使得函數值不斷減小或增大，直到達到最優解。

Usage: 梯度下降廣泛應用於機器學習和深度學習中，用於訓練模型的參數，例如線性回歸、神經網絡等。

Gradient Descent for optimization:

Meaning: Gradient descent is an optimization algorithm used to find the local minimum or maximum of a function. It updates parameters by computing the gradient (slope) direction of the function, reducing or increasing the function value until reaching the optimal solution.

Usage: Gradient descent is widely used in machine learning and deep learning for training model parameters, such as in linear regression, neural networks, etc.

(m) Schema for genetic algorithm:

Meaning: 在遺傳算法中，基因型中的一部分稱為基因型模式（schema），它描述了可能繁衍和遺傳的潛在結構。基因型模式通常由一組位元表示，代表基因中的特定位置和組合。

Usage: 基因型模式在遺傳算法中用於形成新個體和保留優良特徵，透過交叉和突變操作來探索解空間並找到最優解。

Schema for genetic algorithm:

Meaning: In genetic algorithms, a part of the genotype is called a schema, which describes potential structures for reproduction and inheritance. Schemas are typically represented by a set of bits, representing specific positions and combinations within genes.

Usage: Schemas are used in genetic algorithms to form new individuals and preserve favorable features, exploring the solution space and finding the optimal solution through crossover and mutation operations.

1. 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對角) (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)

(a) 在決策樹歸納學習算法中，決定每個節點跨越時的最佳問題的標準是什麼？ (5)

(b) 應用於分類問題時，分類和回歸樹的局限性是什麼？ (5)

(c) 在具有39維度點的10個混合的高斯混合模型中有多少參數？（假設其協方差矩陣都是對角的）(5)

(d) 假設籃子中有4個紅球，2個白球，1個藍球和1個綠球。請計算籃中球的分佈的熵值。 (10)

(e) 給定兩個向量 X1=[0 1 1 0 1 0 0 1 0 1] 和 X2=[1 0 1 1 0 0 1 1 0 0]，請計算簡單匹配係數、Jaccard係數和Rao's係數。 (10)

(A)確定決策樹歷遍每個節點時的最佳問題標準是基於熵的減少。目標是選擇一個能夠最大程度降低數據集熵的問題，從而創建更加同質的子集。

The criteria for determining the optimal question when spanning each node in the inductive learning algorithm of a decision tree are based on entropy reduction. The goal is to select a question that maximally reduces the entropy of the dataset, thereby creating more homogeneous subsets.

(b) 應用於分類問題時，分類和回歸樹的局限性包括：

- 對離散和連續特徵的處理不一致。

- 小數據集容易過擬合。

- 對異常值和缺失值敏感。

- 遞歸分割導致過於複雜的模型。

決策樹中產生的最大的限制就是輸入輸出的資料類型要是離散型的，因為模型要變成分枝，輸出入都不能是連續型資料，決策樹用於回歸、分類分群，不過分類的類別要是離散的，若資料的類型是數字型也可以透過類別來將資料變成離散的，比如資料是速限，那可以把速限變成等級，如高速、中速、低速，若資料型態是連續數字是就要用回歸，不然你的決策樹就會難以收斂

When applied to classification problems, limitations of classification and regression trees include:

- Inconsistent handling of discrete and continuous features.

- Small data sets are prone to overfitting.

- Sensitive to outliers and missing values.

- Recursive segmentation leads to overly complex models.

The biggest limitation in decision trees is that the input and output data types must be discrete, because the model needs to become branched, and the input and output cannot be continuous data. Decision trees are used for regression and classification, but the classification categories must be Discrete. If the data type is numeric, the data can also be made discrete through categories. For example, if the data is a speed limit, then the speed limit can be changed into a level, such as high speed, medium speed, low speed. If the data type is If it is a continuous number, you must use regression, otherwise your decision tree will have difficulty converging.

(c) **在具有 10 個混合成分**和 39 維度點的高斯混合模型中，**每個混合成分包含 39 個均值參數和 39 個對角協方差矩陣參數**（假設協方差矩陣是對角的）。因此，總參數數量為 \(10 \times (39 + 39) = 780\)。

In a Gaussian mixture model with 10 mixtures and 39-dimensional points, each mixture has parameters including 39 mean parameters and 39 parameters for diagonal covariance matrices (assuming diagonal covariance). Therefore, the total number of parameters is \(10 \times (39 + 39) = 780\).

(d) 假設籃子裡有 4 個紅球、2 個白球、1 個藍球和 1 個綠球（共 8 個球），根據顏色比例計算球分佈的熵：

- 紅球：4/8

- 白球：2/8

- 藍球：1/8

- 綠球：1/8

計算熵 \(H\)：



1. 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)

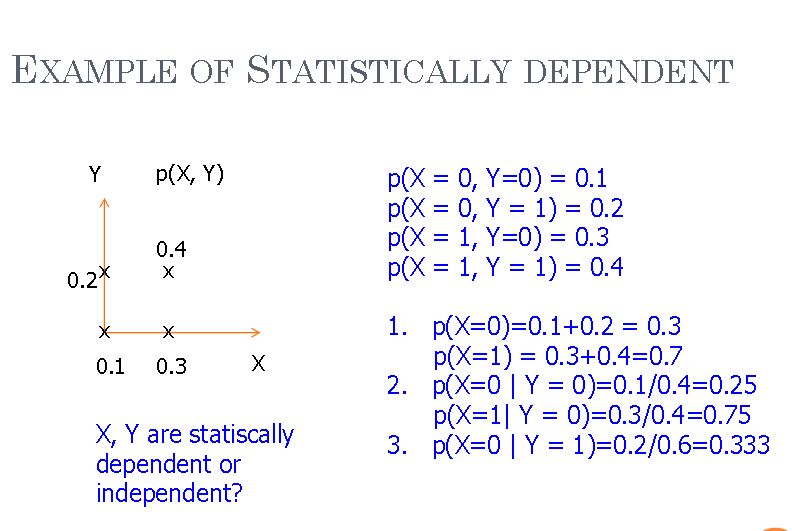
c) Assume we observe a fact that Y is impossible to be C. What could we make decision for X (X=1 or X=2) under this condition? What decision should we make if no obsertation is available? (10)

a） 請計算條件概率 P（X=1|Y=A） （5）

b） 這兩個變數在統計上是獨立的嗎？為什麼？ (5)

Two variables A and B are statistically independent if the likelihood that A will occur has nothing to do with whether B occurs

c） 假設我們觀察到一個事實，即 Y 不可能是 C。在這種情況下，我們可以如果沒有可用的觀察論文，我們應該做出什麼決定？ (10)



一張含有 文字, 螢幕擷取畫面, 字型 的圖片

自動產生的描述

If we observe that Y is impossible to be C, we need to consider the probabilities of X given Y=A and Y=B:

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)

一張含有 文字, 螢幕擷取畫面, 字型 的圖片

自動產生的描述

一張含有 文字, 螢幕擷取畫面, 字型, 數字 的圖片

自動產生的描述

(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. 假設我們有一組資料 {(xi, yi)}，其中 xi 和 yi 都是實數。說明如何將非線性迴歸問題表示為具有變數變換的線性迴歸問題，分別用於以下迴歸函數族。

(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)

Clustering 一張含有 文字, 螢幕擷取畫面, 字型 的圖片

自動產生的描述 一張含有 文字, 字型, 螢幕擷取畫面, 數字 的圖片

自動產生的描述