內容

[FNN（Feedforward Neural Network） 2](#_Toc166456121)

[RNN（Recurrent Neural Network） 3](#_Toc166456122)

[CNN（Convolutional Neural Network） 3](#_Toc166456123)

[FNN RNN CNN三者之間的差別 3](#_Toc166456124)

[激活函數（Activation Function） 5](#_Toc166456125)

[Entropy 與 Cross Entropy 的差別 6](#_Toc166456126)

[Entropy（熵） 6](#_Toc166456127)

[Cross Entropy（交叉熵） 6](#_Toc166456128)

[DFS 8](#_Toc166456129)

[BFS 9](#_Toc166456130)

[DFS VS. BFS 9](#_Toc166456131)

[CHARACTERISTICS OF SEARCH ALGORITHMS搜尋演算法的特性 10](#_Toc166456132)

[Minimax 11](#_Toc166456133)

[🕰️ 最小最大演算法的時間複雜度 12](#_Toc166456134)

[⚖️ 最小最大演算法的局限性 12](#_Toc166456135)

[✂️ α-β剪枝法中 12](#_Toc166456136)

[Alpha是甚麼? 12](#_Toc166456137)

[beta 是甚麼? 13](#_Toc166456138)

[ALPHA-BETA PRUNING 13](#_Toc166456139)

[🔄 最佳移動策略 14](#_Toc166456140)

[➕ 以极大和极小的策略进行移動 15](#_Toc166456141)

[🌳 遊戲樹的分析 15](#_Toc166456142)

[決策樹 (Decision tree) 16](#_Toc166456143)

[今日學習目標 16](#_Toc166456144)

[決策樹 16](#_Toc166456145)

[決策樹如何生成？ 17](#_Toc166456146)

[決策樹的混亂評估指標 18](#_Toc166456147)

[評估分割資訊量 19](#_Toc166456148)

[熵 (Entropy) 20](#_Toc166456149)

[Gini 不純度 (Gini Impurity) 21](#_Toc166456150)

[迴歸樹 22](#_Toc166456151)

[樹越深模型越複雜 23](#_Toc166456152)

[迴歸樹該如何選擇切割點? 24](#_Toc166456153)

[CART 決策樹 25](#_Toc166456154)

[決策樹模型的優缺點 26](#_Toc166456155)

[決策樹總結 26](#_Toc166456156)

## FNN（Feedforward Neural Network）

- 中文描述：FNN（前饋神經網路）是一種**最基本的神經網路模型**，其中**訊息只會向前流動，不會形成循環或回饋。**它由多個層組成，每一層包含多個神經元（節點），**相鄰層之間的神經元之間存在權重**。每個神經元的輸出是前一層神經元的加權和，**經過激活函數處理後得到**。

- 英文描述：A Feedforward Neural Network (FNN) is a basic neural network model where information flows only in one direction, without cycles or feedback loops. It consists of multiple layers, each containing multiple neurons (nodes), with weighted connections between adjacent layers. The output of each neuron is the weighted sum of inputs from the previous layer, followed by an activation function.

## RNN（Recurrent Neural Network）

- 中文描述：RNN（循環神經網路）是一種**具有循環結構的神經網路**，能夠處理序列數據或時間序列數據。**RNN的特點是在每個時間步都會將當前的輸入與上一個時間步的隱藏狀態進行計算**，**產生當前時間步的輸出**。RNN**具有記憶功能**，可以**捕捉序列數據中的時間依賴性**。

- 英文描述：A Recurrent Neural Network (RNN) is a type of neural network with a loop structure that allows it to handle sequential or time-series data. In RNN, the current input at each time step is processed along with the hidden state from the previous time step to produce the output at the current time step. RNNs have the ability to capture temporal dependencies in sequential data.

## CNN（Convolutional Neural Network）

- 中文描述：CNN（卷積神經網路）是一種**專門用於處理圖像數據的神經網路模型**。它主要**包含卷積層（Convolutional Layer）、池化層（Pooling Layer）和全連接層（Fully Connected Layer）**。CNN通過使用卷積操作來提取圖像的特徵，並通過池化操作進行特徵降維，最終通過全連接層進行分類或回歸。

- 英文描述：A Convolutional Neural Network (CNN) is a type of neural network designed specifically for processing image data. It typically consists of convolutional layers, pooling layers, and fully connected layers. CNNs use convolution operations to extract features from images and pooling operations to reduce the dimensionality of the features, ultimately performing classification or regression tasks using fully connected layers.

## FNN RNN CNN三者之間的差別

**結構差異：**

FNN：訊息只向前傳播，沒有循環結構。

RNN：具有循環結構，能夠處理序列或時間序列數據。

CNN：主要用於處理圖像數據，包含卷積、池化和全連接層。

**Structural Differences:**

**Structural Differences:**

* **FNN (Feedforward Neural Network):** Information flows only forward without any cyclic structure.
* **RNN (Recurrent Neural Network):** Contains cyclic structure and is capable of handling sequential or time-series data.
* **CNN (Convolutional Neural Network):** Primarily used for processing image data and includes convolutional layers, pooling layers, and fully connected layers.

**應用領域差異：**

FNN：適用於一般的分類或回歸任務。

RNN：適用於序列建模，如自然語言處理（NLP）、時間序列預測等。

CNN：主要用於計算機視覺任務，如圖像分類、物體檢測等。

**Application Domain Differences:**

* **FNN:** Suitable for general classification or regression tasks.
* **RNN:** Suitable for sequence modeling tasks such as natural language processing (NLP) and time-series prediction.
* **CNN:** Mainly used for computer vision tasks like image classification and object detection.

**特點差異：**

FNN：較簡單，無法處理序列數據的時間依賴性。

RNN：具有記憶能力，能夠捕捉序列數據中的時間依賴性。

CNN：專門用於處理圖像數據，具有平移不變性（Translation Invariance）。

**Characteristic Differences:**

* **FNN:** Relatively simple and unable to handle time dependencies in sequential data.
* **RNN:** Possesses memory capabilities and can capture temporal dependencies in sequential data.
* **CNN:** Specifically designed for processing image data with translation invariance properties.

## 激活函數（Activation Function）

神經網路中的一種重要組件，它通常**被應用於神經元的輸出，用於引入非線性特性**，並決定神經元是否應該被激活（fired）。

在神經網路中，**每個神經元接收來自前一層的加權輸入（包括偏差項），然後通過激活函數計算出神經元的輸出**。激活函數**將線性組合的輸入轉換成非線性輸出**，這對於神經網路的學習和表示能力至關重要。

An activation function is a critical component in neural networks applied to determine the output of a neuron based on its input. It introduces non-linearity into the network, allowing it to model complex relationships and make decisions about whether a neuron should be activated (fired).

In neural networks, each neuron receives weighted inputs from the previous layer, including a bias term, and computes its output using an activation function. The activation function transforms the linear combination of inputs into a non-linear output, which is essential for the learning and representational capabilities of neural networks.

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**Sigmoid Function** (Logistic Function): It compresses the input into the range of (0,1)(0,1) and is often used in the final layer of binary classification tasks.

**Tanh Function** (Hyperbolic Tangent Function): It compresses the input into the range of (−1,1)(−1,1) with a centering property.

**ReLU Function** (Rectified Linear Unit Function): It outputs the input if it's positive, otherwise outputs zero, commonly used in hidden layers.

**Softmax Function**:It converts a vector into a probability distribution, often used in the final layer of multi-class classification tasks

## Entropy 與 Cross Entropy 的差別

中文描述：

### Entropy（熵）

是訊息理論中衡量**隨機變量不確定度或混亂程度的度量**。在機器學習中，熵用於**衡量一個概率分佈的不確定性**，計算方式為將每個事件的概率乘以其自然對數，然後**求和取負值。熵越高，概率分佈的不確定性越大**。

1. Entropy:

- Definition: Entropy is a measure of uncertainty or disorder of a random variable in information theory. In machine learning, entropy is used to quantify the uncertainty of a probability distribution. It is calculated by multiplying each event's probability by its natural logarithm, then summing the results and taking the negative value. Higher entropy indicates greater uncertainty in the probability distribution.

### Cross Entropy（交叉熵）

是一種**衡量兩個概率分佈之間的差異或不匹配程度的指標**。在機器學習中，交叉熵**常用於衡量模型預測的概率分佈與真實概率分佈之間的距離**。交叉熵的計算方式是**將真實概率分佈的每個事件的概率乘以其自然對數，然後與模型預測的概率分佈相乘，最後求和取負值**。

英文描述：

2. Cross Entropy:

- Definition: Cross Entropy is a metric that measures the difference or mismatch between two probability distributions. In machine learning, cross entropy is commonly used to quantify the distance between the predicted probability distribution of a model and the true probability distribution. Cross entropy is calculated by multiplying each event's probability from the true distribution by its natural logarithm, then multiplying it with the corresponding probability from the model's predicted distribution, and finally summing the results and taking the negative value.

總之，熵衡量機率分佈內的不確定性，而交叉熵衡量兩個機率分佈之間的差異。交叉熵通常用作機器學習任務中的損失函數，例如分類，以根據預測分佈和真實分佈之間的差異來優化模型參數

In summary, entropy measures uncertainty within a probability distribution, while cross entropy measures the divergence between two probability distributions. Cross entropy is often used as a loss function in machine learning tasks such as classification to optimize model parameters based on the difference between predicted and true distributions.

Formula

## DFS

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

It follows each path to its greatest depth before

moving on to the next path

 When leaf node is reached, trace it back 到達葉子節點時回溯

A stack is maintained for controlling the search 控制堆疊來控制搜尋

 The first entry is popped 第一個入的已經彈出

 Spanned nodes are inserted at head of the list 新加入的節點家在list的前面

## BFS

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

Examine all nodes one level down from the root node

More memory usage more than depth-first

 Depth increases gradually

 Goals in lower depth will be found first

A queue佇列is maintained for controlling the search, Newly spanned nodes新跨入的節點 are inserted at back of the queue插入到後面的佇列

## DFS VS. BFS

Ways of managing open list

DFS: inserted at front (as stack)

BFS: inserted at back (as queue)  
 It is often not easy to consider optimality, time complexity and space complexity simultaneously

DFS: lower memory requirement

BFS: minimum-length path  
 Which is better?

It depends on the problems

What is the shape of state space

Where are the goal states located

DFS VS. BFS

 DFS is more often used: Less memory usage

 Search tree may have very deep path, and goal node is in shallower part

Breadth-first good, but depth-first poor

8-Puzzle: very deep path, DFS not good

 All paths are of similar length

DFS good (do not try lower level)

 Search tree has high branching factor

DFS good (BFS consumes too many memories)

## CHARACTERISTICS OF SEARCH ALGORITHMS搜尋演算法的特性

Completeness完整性：如果存在目標狀態，則如果保證找到目標狀態，則搜尋方法是完整的。例如拼圖問題的 DFS 不完整。

A search method is complete if it is guaranteed to find a goal state provided there exists goal state.

e.g. DFS for puzzle problem is not complete.

Optimality 最適性 如果保證找到現有的最佳解，則搜尋方法是最優的。「最好」與「善良」的定義相關

A search method is optimal if it is guaranteed to find the best solution that exists“Best” is relevant to the definition of “Goodness” .

Irrevocability (no back tracing) 不使用回溯的方法是不可撤銷的。不可撤銷的方法，例如爬山，往往會被局部最優所愚弄

Methods that do not use backtracking are irrevocable. Irrevocable methods, such as hill-climbing, tend to be fooled by local optima.

Admissibility可接受性（Admissibility） 如果搜尋演算法能夠保證在任何時候存在一條最優路徑時都能找到解決方案，那麼該搜尋演算法就是可接受的。  任何演算法 A\* 都是可接受的。

A search algorithm is admissible if it is guaranteed to find a optimal path to a solution whenever such a path exists.

Any algorithm A\* is admissible.

## Minimax

一張含有 文字, 螢幕擷取畫面, 設計 的圖片

自動產生的描述

一張含有 行, 圖表, 設計, 三腳架 的圖片

自動產生的描述

## 🕰️ 最小最大演算法的時間複雜度

The time complexity of the minmax algorithm depends on the path factor and depth of the game tree. The degree is the depth power of the channel factor, which is O(B^D). The complexity increases with the branching factor and depth.

最小最大演算法的時間複雜度取決於遊戲樹的分支因子和深度。分支因子表示每個遊戲狀態下可能的移動數量，而深度表示遊戲樹中的層數或回合數。最小最大演算法的時間複雜度是分支因子的深度次方，即O(B^D)。複雜度的增長隨著分支因子和深度的增加而增加。

## ⚖️ 最小最大演算法的局限性

The main limitation of the minmax algorithm is that **it is not efficient when dealing with certain games**. **When the branching factor and depth of the game are very large, the game tree will become very large, causing the game tree search to be very time-consuming and computationally resource-consuming**.

For example, in chess, the branching factor is on average 35 and the depth is on average 100, which results in a very large number of nodes in the game tree, making it difficult to perform a complete search. In order to solve this problem, we need to use other algorithms, such as α-β pruning.

最小最大演算法的**主要局限性在於它在處理某些遊戲時的效率不高**。當遊戲的**分支因子和深度非常大時，遊戲樹會變得非常龐大，導致遊戲樹搜索的時間和計算資源非常消耗**。

例如在國際象棋中，分支因子平均為35，深度平均為100，這導致遊戲樹的節點數量非常龐大，難以進行完整的搜索。為了解決這個問題，我们需要使用其他的演算法，如α-β剪枝法。

## ✂️ α-β剪枝法中

### Alpha是甚麼?

alpha is the temporary maximum during the execution.

alpha 是執行期間的暫時最大值。

alpha can increase only (could be replaced by a larger entry) during the the maximization process.

**alpha 只能在最大化過程中增加**（可以用更大的條目取代）

alpha is the maximum finally after all elements are visited

**alpha 是所有元素被存取後最終的最大值**

### beta 是甚麼?

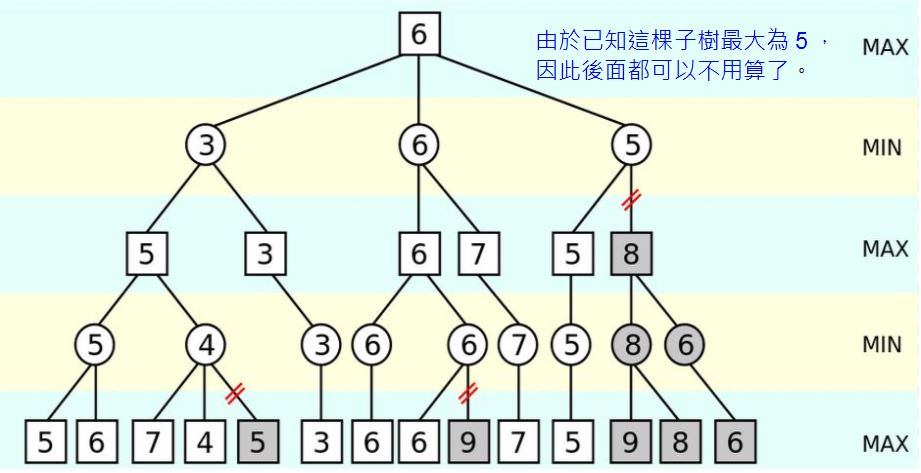
**beta is the temporary minimum during the execution.**

**beta 是執行期間的暫時最小值。**

**beta can decrease only beta 只能減少**

**beta is the maximum finally after all elements are visited** **beta 是所有元素訪問後最終的最大值**

### ALPHA-BETA PRUNING



A method that can often cut off a large part of the game tree. Based on the idea that if a move is clearly bad, there is no need to waste time to search the consequences of it completely

ALPHA-BETA 剪枝 

一種通常可以剪掉賽局樹大部分的方法。基於這樣的想法：如果一個舉動明顯是不好的，那麼就沒有必要浪費時間去徹底尋找它的後果

**The α-β pruning method is an optimized min-max algorithm used to reduce the amount of calculations in game tree search**. **It improves search efficiency by limiting the search scope and avoiding unnecessary parts of the game tree**.

**When a node is searched, the α-β pruning method will determine the nodes that do not need to be searched based on the known best movement strategy, and then prune.**

This method can greatly reduce the time and computing resources of game tree search and improve the efficiency of the game algorithm.

**α-β剪枝法是一種優化的最小最大演算法，用於减少遊戲樹搜索中的計算量。它通過限制搜索範圍，避免遊戲樹搜索不必要的部分，從而提高搜索效率。**

**當搜索到某個節點時，α-β剪枝法會根據已知的最佳移動策略來確定不必搜索的節點，從而剪枝。**

這種方法可以大大減少遊戲樹搜索的時間和計算資源，提高遊戲演算法的效率。

## 🔄 最佳移動策略

In game theory, **the optimal movement strategy refers to the strategy developed by players in the game that can not only maximize their own benefits, but also inhibit their opponents' maximum benefits.** The player's goal is **to gain maximum benefit through optimal movement strategies, while opponents work hard to prevent you from winning**. The optimal movement strategy can be formulated based on game rules and game status, and can be adjusted accordingly in each turn based on the opponent's movements.

遊戲理論中，**最佳移動策略表示玩家在遊戲中制定的既能使自己獲得最大利益，又能抑制對手獲得最大利益的策略**。玩家的目標**是通過最佳移動策略來獲得最大利益，同時對手也在努力阻止你獲勝。**最佳移動策略可以根據遊戲規則和遊戲狀態來制定，並且可以在每一個回合中根據對手的移動作出相應的調整。

## ➕ 以极大和极小的策略进行移動

**The "maximum" and "minimum" in the name of the minmax algorithm represent the two players in the game, namely the maximum player and the minimum player. The largest player tries to maximize his own interests, while the smallest player tries to minimize the largest player's interests.**

When it is the turn of the largest player to move, **he will choose the move that maximizes his own interests**; and when it is the turn of the smallest player to move, **he will choose the move that maximizes the interests of the largest player.** Such strategies are carried out alternately until the conditions for the end of the game are reached.

**最小最大演算法的名稱中的"最大"和"最小"表示的是遊戲中的兩個玩家，即最大玩家和最小玩家。最大玩家試圖最大化自己的利益，而最小玩家則試圖最小化最大玩家的利益。**

當輪到最大玩家移動時，他會**選擇使自己利益最大化的移動**；而當輪到最小玩家移動時，**他會選擇使最大玩家得到最小利益的移動**。這樣的策略交替進行，直到達到遊戲結束的條件。

## 🌳 遊戲樹的分析

The minmax algorithm uses game trees to analyze various movement strategies of the game. **A game tree is a tree-like structure consisting of all possible moves and game states of the game.**

By recursively computing each node of the game tree, **the minmax algorithm can determine the benefit generated by each move and select the next move based on the best benefit.**

The depth of the game tree depends on the complexity of the game and the length of the game.

最小最大演算法使用遊戲樹來分析遊戲的各種移動策略。**遊戲樹是由遊戲的所有可能移動和遊戲狀態構成的樹狀結構**。

通過遞歸運算遊戲樹的每個節點，**最小最大演算法可以確定每個移動所產生的利益，並根據最佳的利益來選擇下一步的移動**。

遊戲樹的深度取決於遊戲的複雜性和棋局的長度。

# 決策樹 (Decision tree)

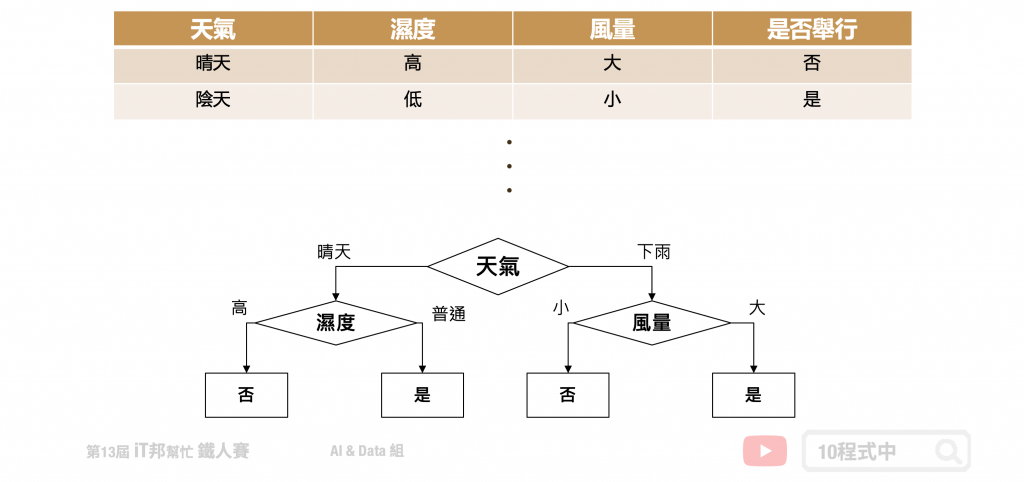
## 今日學習目標

* 決策樹演算法介紹
  + 決策樹如何生成？
  + 如何處理分類問題？
  + 如何處理迴歸問題？
* 實作決策樹分類器
  + 觀察決策樹是如何生成的。
* 實作決策樹迴歸器
  + 查看決策樹方法在簡單線性迴歸和非線性迴歸表現。

## 決策樹

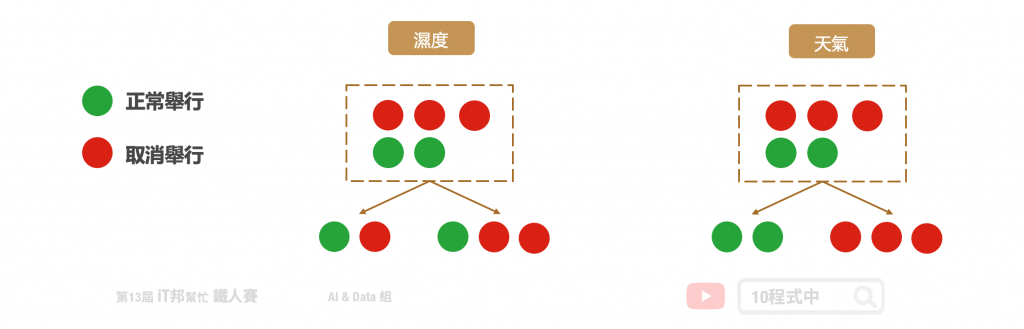
決策樹會根據訓練資料產生一棵樹，依據訓練出來的規則來對新樣本進行預測。決策樹演算法可以使用不同的方式來評估分枝的好壞(亂度)，例如像是 Information gain、Gain ratio、Gini index。依據訓練資料找出合適的規則，最終生成一個規則樹來決策所有事情，其目的使每一個決策能夠使訊息增益最大化。就好比我們評估今天比賽是否舉行，天氣因子可能站比較大的因素，而 Co2 的濃度高低可能佔的因子程度較低。因此在第一層的決策中以天氣的特徵先進行第一次的決策判斷。接著第二層再從所有特徵中尋找最適合的決策因子，直到設定的最大樹的深度即停止樹的生長。

The decision tree generates a tree based on the training data, and the new sample is predicted according to the trained rules. Decision tree algorithms can be used in different ways to evaluate the quality (disorder) of branches, such as information gain, gain ratio, and gini index. Based on the training data, the appropriate rules are found, and finally a rule tree is generated to make decisions about everything, with the aim of maximizing the message gain for each decision. For example, if we evaluate whether or not a race will be held today, the weather factor may be a relatively large factor, and the concentration of CO2 may account for a lower factor. Therefore, in the first level of decision-making, the first decision-making judgment is made based on the characteristics of the weather. The second layer then looks for the most suitable decision-making factor from all the characteristics until the maximum depth of the tree is set, which stops the tree growth.



## 決策樹如何生成？

**決策樹是以一個貪婪法則來決定每一層要問什麼問題，目標是分類過後每一群能夠很明顯的知道是屬於哪一種類別。**延續上面的例子，以分類問題來說假設要評估明天比賽是否舉行。在樹的第一層節點中我們要從已知的兩個特徵分別是溫度與特徵選一個作為該層的決策因子。假設目前訓練集有五筆資料，其中正常舉行的有兩筆資料，取消舉行的有三筆資料。在樹的結構中左子樹為決策正常取行，而右子樹是決策取消舉行。我們可以發現當特徵為天氣的時候可以一很清楚的將這兩類別完整分開，因此我們會將天氣作為這一層判斷的因子。這就是決策樹在生成中的貪婪機制**。然而要如何去判斷每次決策的好壞，就必須依靠亂度的評估指標。The decision tree uses a law of greed to determine what questions to ask at each level, with the goal of making it clear to each group which category it belongs to. Continuing with the above example, let's assume that the game will be held tomorrow in the case of a classification problem. In the first layer of the tree, we want to choose one of the two known features, temperature and features, as the decision factor of the layer. Suppose that there are five data in the current training set, of which two data are held normally and three data are cancelled. In the structure of the tree, the left subtree is the normal row for the decision, while the right subtree is the decision cancellation held. We can find that when the feature is weather, the two categories can be clearly separated, so we will use weather as a factor in this layer of judgment. This is the mechanism of greed in the generation of decision trees. However, how to judge the quality of each decision, we must rely on the evaluation index of chaos.**



## 決策樹的混亂評估指標

我們需要客觀的標準來決定決策樹的每個分支，因此我們需要有一個評斷的指標來協助我們決策。決**策樹演算法可以使用不同的指標來評估分枝的好壞，常見的決策亂度評估指標有 Information gain、Gain ratio、Gini index。**我們目標是從訓練資料中找出一套決策規則，讓每一個決策能夠使訊息增益最大化。**以上的指標都是在衡量一個序列中的混亂程度，其數值越高代表越混亂。然而在 Sklearn 套件中預設使用 Gini。**

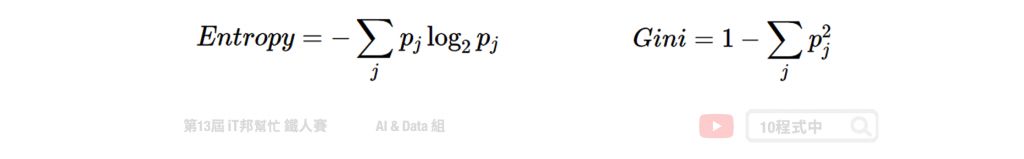
**We need objective criteria to decide on each branch of the decision tree, so we need to have a judgmental metric to assist us in our decision-making. The decision tree algorithm can use different indicators to evaluate the quality of branches, and the common decision chaos evaluation indicators are Information gain, Gain ratio, and Gini index. Our goal is to find a set of decision rules from the training data so that each decision maximizes the message gain. All of the above metrics measure the degree of chaos in a series, with higher values representing more confusion. However, Gini is used by default in the Sklearn suite.**

* Information gain (資訊獲利)
* Gain ratio (吉尼獲利)
* Gini index (吉尼係數) = Gini Impurity (吉尼不純度)

## 評估分割資訊量

Information Gain 透過從訓練資料找出規則，讓每一個決策能夠使訊息增益最大化。其算法主要是計算熵，因此經由決策樹分割後的資訊量要越小越好。**而 Gini 的數值越大代表序列中的資料亂，數值皆為 0~1 之間，其中 0 代表該特徵在序列中是完美的分類。**常見的資訊量評估方法有兩種：資訊獲利 (Information Gain) 以及 Gini 不純度 (Gini Impurity)。

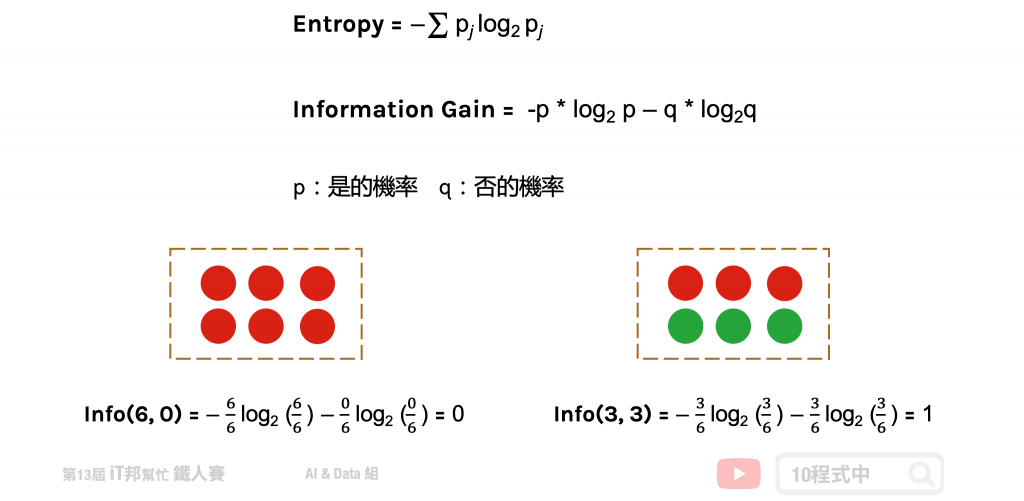
Information Gain maximizes message gain by identifying rules from training data to make every decision. The algorithm mainly calculates entropy, so the amount of information segmented by the decision tree should be as small as possible. The larger the value of Gini, the larger the data in the sequence, the value is between 0~1, where 0 means that the feature is perfectly classified in the sequence. There are two common methods of information assessment: Information Gain and Gini Impurity.



## 熵 (Entropy)

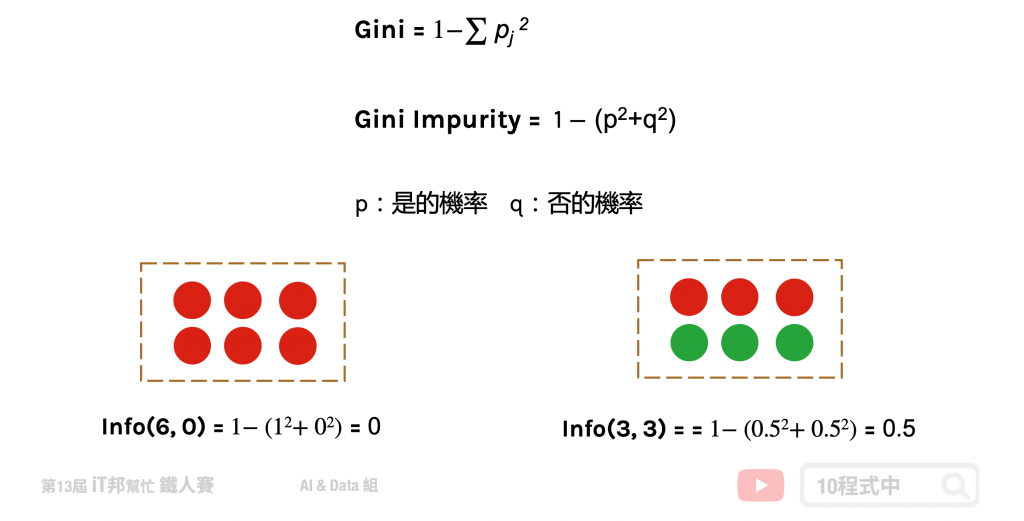
熵 (Entropy) 是計算 Information Gain 的一種方法。在了解 Information Gain 之前要先了解熵是如何被計算出來的。**其中在下圖公式中 p 代表是的機率、q 代表否的機率。我們可以從圖中範例很清楚地知道當所有的資料都被分類一致的時候 Entropy 即為 0，當資料各有一半不同時 Entropy 即為 1。**

**Entropy is a method of calculating Information Gain. Before we get to know Information Gain, it's important to understand how entropy is calculated. In the formula below, p represents the probability of being yes and q represents the probability of no. As we can clearly see from the example in the diagram, Entropy is 0 when all data are classified consistently, and 1 when half of the data is different.**



## Gini 不純度 (Gini Impurity)

Gini 不純度是另一種亂度的衡量方式，**它的數字越大代表序列中的資料越混亂**。公式如下所示，其中 p 代表是的機率、q 為代表否的機率。我們可以從圖中範例很清楚地知道當所有的資料都被分類一致的時候混亂程度即為 0，當資料各有一半不同時混亂程度即為 0.5。



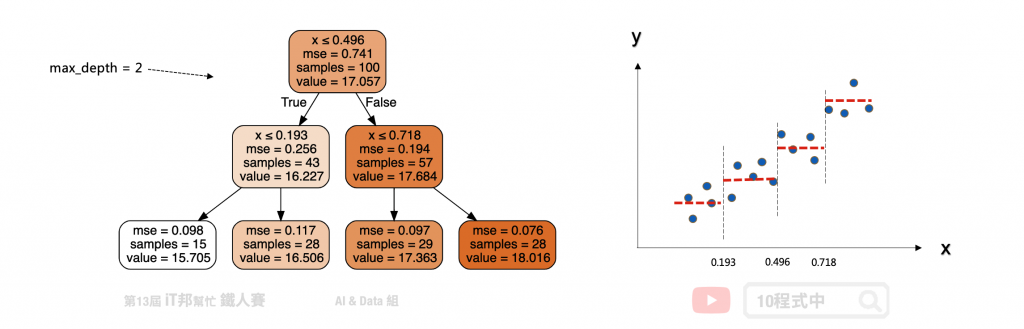
## 迴歸樹

**決策樹迴歸方法與分類有點類似差別僅在於評估分枝好壞的方式不同，我們又可以稱作迴歸樹**。當數據集的輸出爲連續性數值時，該樹算法就是一個迴歸樹。透過樹的展開，並用葉節點的均值作爲預測值。從根節點開始，對樣本的某一特徵進行測試。經過評估後，將樣本分配到其子結點。此時每一個子節點對應著該特徵的一個值。依照這樣方式進行，直至到達葉結點。此時誤差值要最小化，並且越接近零越好。

The decision tree regression method is somewhat similar to classification, except that the way of evaluating the quality of the branches is different, which we can also call a regression tree. When the output of the dataset is a continuous value, the tree algorithm is a regression tree. Through the unfolding of the tree, the mean of the leaf nodes is used as the prediction value. Starting from the root node, a feature of the sample is tested. After evaluation, the samples are assigned to their child nodes. Each child node corresponds to a value for the feature. This is done until the leaf node is reached. In this case, the error value should be minimized, and the closer to zero the better.

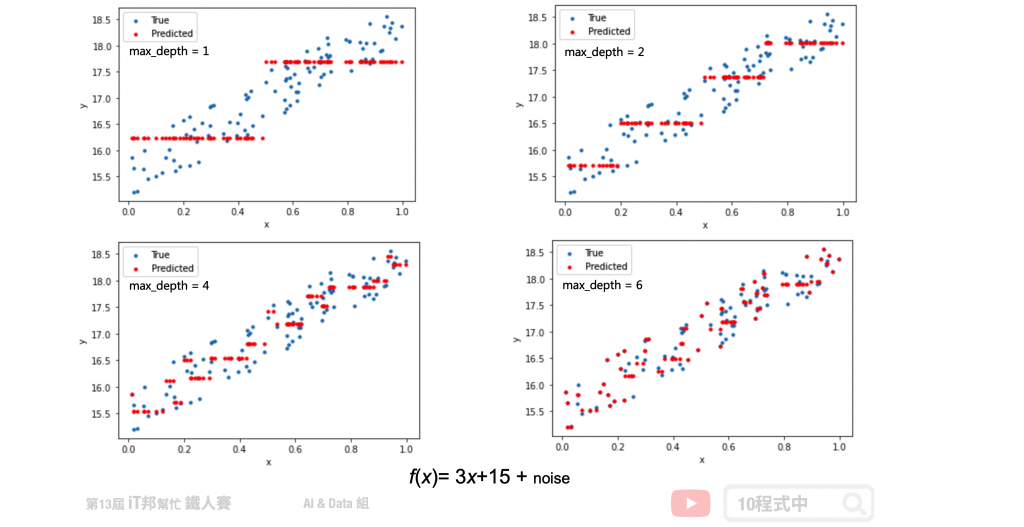
迴歸樹的生長過程很推薦看[這篇](https://zhuanlan.zhihu.com/p/82054400)文章

以下舉一個例子假設 x 是輸入 y 是輸出，我們可以在一個平面上繪製出資料與正確答案間的分佈。假設迴歸樹的最大深度設定兩層。首先在第一層中會將所有的資料從中間切一刀此斷點為 x=0.496 當大於設定的值的數據點會繼續往右子樹下去延伸，反之小於 0.496 的資料點會往左子樹走。此時將會切出一個分支出來並往下擴展並形成第二層的決策分支。一直不斷持續拓展直到設定的最大深度終止，此時的節點即為葉節點也就是最終的模型輸出值。As an example, assuming x is the input and y is the output, we can plot the distribution between the data and the correct answer on a plane. Suppose the maximum depth of the regression tree is set to two levels. First of all, in the first layer, all the data will be cut from the middle, this breakpoint is x=0.496, when the data point greater than the set value will continue to extend to the right subtree, otherwise, the data point less than 0.496 will go to the left subtree. A branch will be cut out and expanded down to form a second level of decision branch. It continues to expand until the set maximum depth is terminated, at which point the node is the leaf node, which is the final model output.



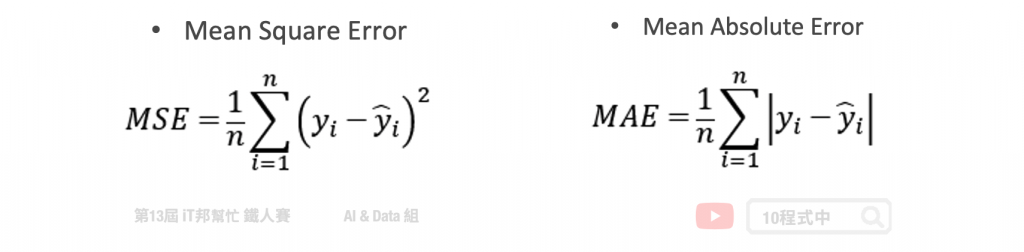
## 樹越深模型越複雜

假設我們生成一個 f(x) = 3x+15 + noise 的資料，其中 noise 為一個 0~1 之間的隨機數。從以下的測試可以看出隨著決策樹深度的增加，決策樹的擬合能力不斷上升。決策樹已經不僅僅擬合了我們的線性函式 3x+15，同時也擬合了我們添加的噪音(noise)。The deeper the tree, the more complex the model Let's say we generate a data with f(x) = 3x+15 + noise, where noise is a random number between 0~1. From the following tests, it can be seen that the fitting ability of the decision tree increases as the depth of the decision tree increases. The decision tree already fits not only our linear function 3x+15, but also the noise we add.



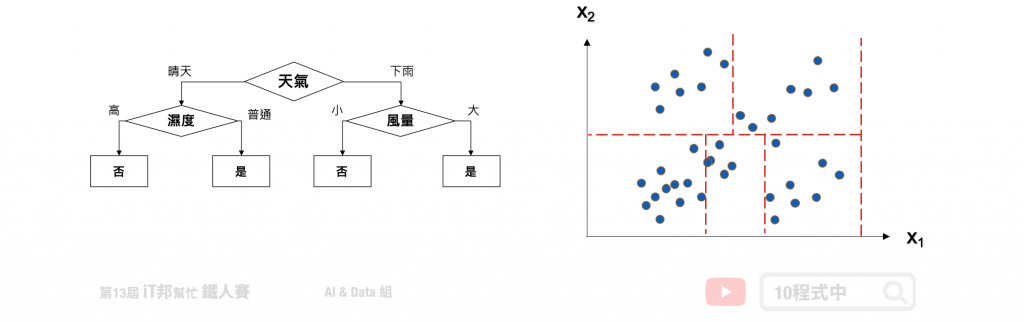
## 迴歸樹該如何選擇切割點?

在分類模型中決策樹是以亂度作為決策樹生成時候的評估指標。但是迴歸樹透過是 MSE 或 MAE 來評估模型，並找出誤差最小的值作為樹的特徵選擇與切割點。其中前者是均方差，後者是和均值之差的絕對值之和。How to choose a cutting point for a regression tree? In the classification model, the decision tree uses chaos as the evaluation index when the decision tree is generated. However, the regression tree evaluates the model by being an MSE or MAE, and finds the value with the least error as the feature selection and cutting point of the tree. where the former is the mean square deviation and the latter is the sum of the absolute values of the difference from the mean.



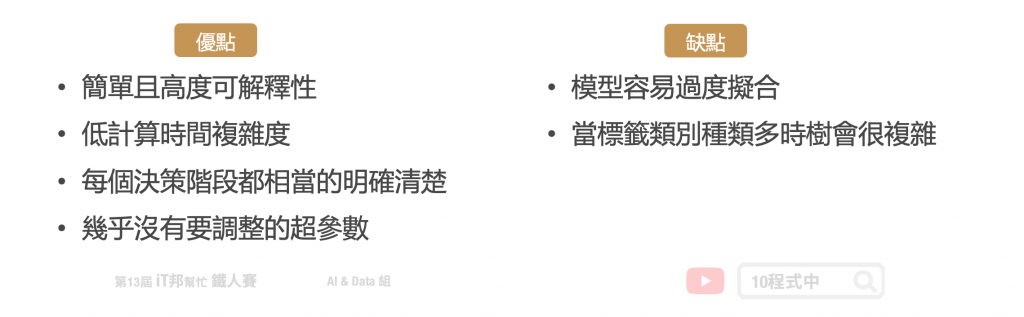
## CART 決策樹

在 Sklearn 套件中決策樹演算法是採用 CART (Classification and Regression Tree) 演算法，並且可以被拿來做分類和迴歸的預測。在決策樹的每一個節點上都是採用二分法，也就是每一個決策節點只分枝出兩個子節點。並且不斷地往下拓展，直到設定的最大深度為止，此時最大深度的節點稱為葉節點即為模型的預測輸出。In the Sklearn suite, the decision tree algorithm uses the CART (Classification and Regression Tree) algorithm and can be used to make predictions for classification and regression. The dichotomy is used at each node of the decision tree, that is, each decision node branches out only two subnodes. And continue to expand down until the maximum depth is set, at this time the node with the maximum depth is called the leaf node, which is the predicted output of the model.



## 決策樹模型的優缺點

建立決策樹的過程就是不斷的尋找特徵進行決策，透過這些決策盡量的使這些資料被分為同一個類別，且試著讓混亂程度越小越好。切記樹的深度越深不一定越好，他可能會造成過度擬合的問題。訓練好的模型我們能夠視覺化決策樹的結構，相對的可解釋性就變高。此外與其它的ML模型比較起來，決策樹執行速度是它的一大優勢。因為是樹狀結構，因此在進行機器學習的時候每個決策階段都相當的明確清楚，不是 0 就是 1。Advantages and disadvantages of the decision tree model The process of building a decision tree is to constantly look for features to make decisions, through which the data can be divided into the same category as much as possible, and try to make the confusion as small as possible. Keep in mind that the deeper the tree is, the better, and it can cause problems with overfitting. After the trained model is trained, we can visualize the structure of the decision tree, and the relative interpretability will be higher. In addition, the speed of decision tree execution is a major advantage compared to other ML models. Because it is a tree-like structure, each decision stage is quite clear when machine learning is carried out, and it is either 0 or 1.



## 決策樹總結

決策樹透過所有特徵與對應的值將資料切分，來找出最適合的分枝並繼續往下拓展。若決策樹深度越深則決策的規則將越複雜，模型預測也會越接近真實答案。但若訓練集中含有過多的雜訊，太深的樹就有可能產生過擬合的情形。因此單一的決策樹肯定是不夠用的，我們可以利用集成學習中的 Boosting 架構，對迴歸樹進行改良升級。

Decision Tree Summary The decision tree slices the data through all the features and corresponding values to find the most suitable branch and continue to expand downward. If the depth of the decision tree is deeper, the more complex the rules of the decision will be, and the closer the model prediction will be to the real answer. However, if the training set contains too much noise, trees that are too deep may be overfitted. Therefore, a single decision tree is definitely not enough, and we can use the boosting architecture in ensemble learning to improve and upgrade the regression tree.