

機器學習演算法十大經典與應用

什麼是機器學習演算法？

核心概念

演算法是讓電腦從資料中自動學習、預測與決策的數學模型。

學習方式分類

依據資料標籤與學習方式，分為監督式、非監督式與強化學習三大類。

選擇關鍵

選擇合適演算法需考慮資料特性、問題類型與計算資源等因素，以達到最佳效果。



監督式學習的基石：線性迴歸與邏輯迴歸

線性迴歸

使用直線模型預測連續數值型資料，如精準預估房價或銷售額走勢。



邏輯迴歸

運用S型函數進行二元分類預測，常見於疾病診斷（例如：有/無）或信用評估。



- 兩者皆需去除高度相關變數與異常值，以提升模型的穩定性與預測能力。



強大且直觀的分類工具：決策樹與隨機森林



決策樹

以「if-then」規則對資料進行分割，模型易於解釋，但存在過度擬合的風險。



隨機森林

由多棵決策樹集成，透過隨機抽樣降低過度擬合，顯著提升模型的準確度與穩定性。

廣泛應用於：信用風險評估、醫療診斷輔助系統、客戶行為分群等複雜場景。

距離與鄰近的智慧：K-近鄰與學習向量量化

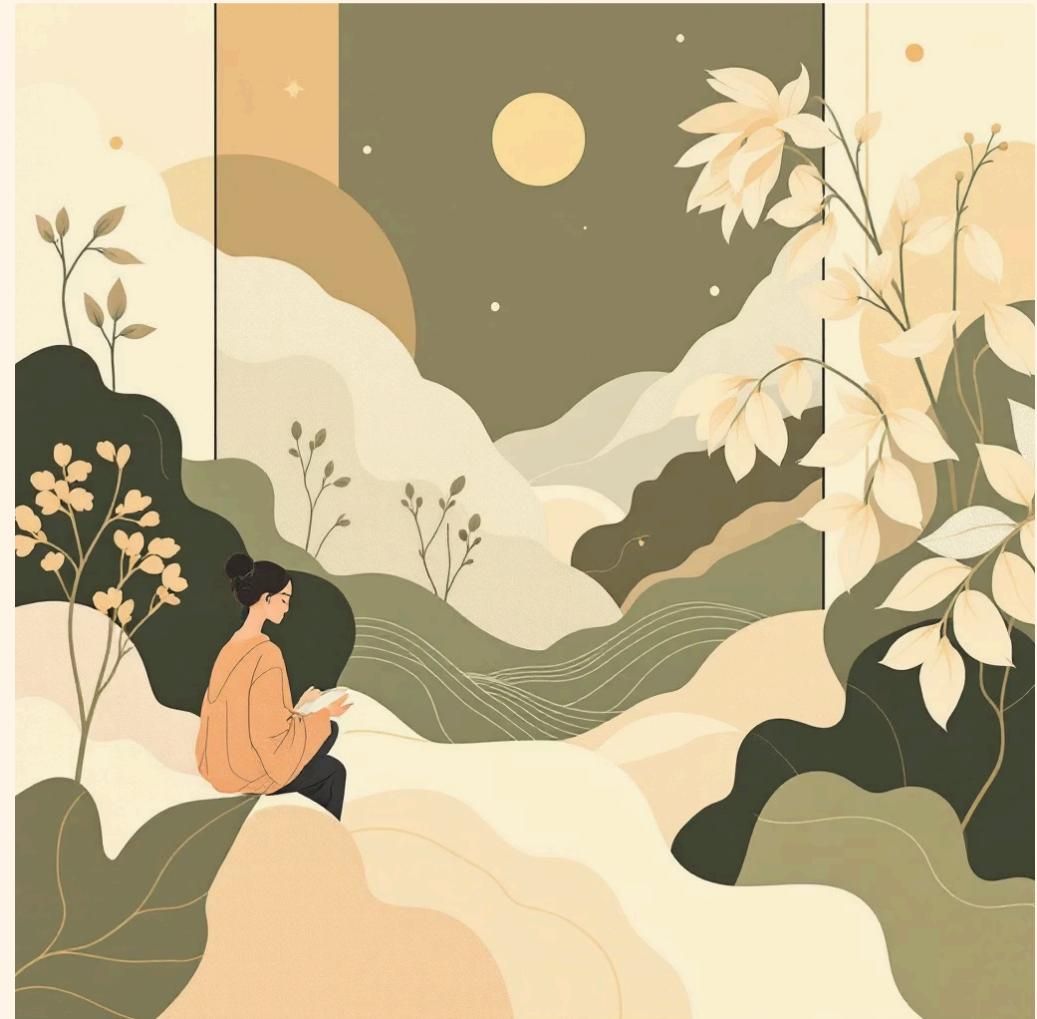
K-近鄰 (KNN)

依據鄰近資料點的「投票」進行分類或回歸，概念簡單直觀，但面對大型資料集時記憶體需求較高。



學習向量量化 (LVQ)

透過少量代表向量取代KNN的資料點，有效減少儲存負擔，提升效率。



這兩種方法特別適合處理小至中型規模的資料集，解決分類與回歸問題。

支援向量機 (SVM)：最大邊界的分類藝術

尋找最佳超平面

SVM 的核心目標是找出一個最佳的超平面，以最大化不同類別間的邊界距離。

提升分類穩健性

透過最大化邊界，模型能夠更穩健地對未見資料進行分類，減少誤判。

核函數處理非線性

利用核函數 (Kernel Function)，SVM 能將非線性可分的資料映射到高維空間，使其在高維度下變得線性可分。

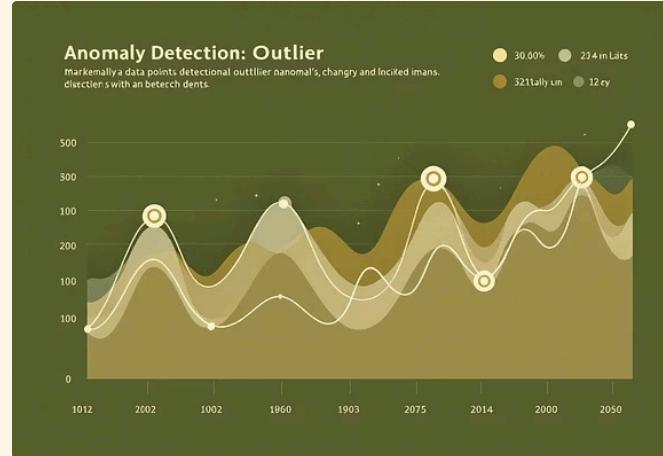
廣泛應用

在文字分類、圖像識別、生物資訊等領域有廣泛應用，表現出色。

參數調整與資料標準化

為確保模型效能，需特別注意參數調整與資料標準化處理，以發揮SVM的最大潛力。

非監督式學習：K-均值聚類與異常偵測



K-均值聚類

將資料分組到預設數量的簇中，揭示潛在的資料結構，廣泛應用於客戶分群和市場細分分析。

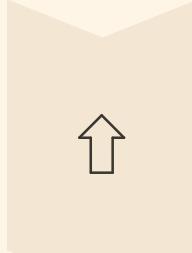
異常偵測

識別偏離正常模式的資料點，對於金融詐騙偵測、設備故障預警和網路安全監控至關重要。

非監督式學習的獨特之處在於無需預先標籤的資料，使其成為探索性分析的強大工具。



進階集成技術：Boosting與梯度提升機 (GBDT)



Boosting

串聯多個弱分類器，每個後續模型都會修正前一個模型的錯誤，逐步提升整體性能。



梯度提升機 (GBDT)

以決策樹為基礎，利用梯度下降優化損失函數，每次迭代都使模型更精確，達到極高的準確度。

這些強大的集成方法在金融風險評估、貸款違約預測以及許多預測性建模任務中被視為不可或缺的利器。



如何選擇適合的機器學習演算法？

1

評估資料與問題

根據資料量、特徵數量以及問題的複雜度，確定最適合的演算法類別。

2

從簡入繁原則

對於簡單問題，優先嘗試線性模型；面對複雜任務，可考慮集成學習或深度學習方法。

3

持續優化與評估

不斷評估模型效能，並透過調整參數與精進特徵工程來達到最佳結果。

結語：機器學習演算法的未來與挑戰

AI核心
演算法是人工智慧的核心驅動力，推動自動化與智慧決策的發展。



技術創新

持續的技術創新與資料品質提升，將帶來機器學習更廣泛的應用前景。

無限可能

深入掌握機器學習演算法，將為我們開啟智慧時代的無限可能。