# 📘 第十一章：決策樹與集成學習（Decision Trees and Ensemble Learning）

## 📖 章節簡介與補充重點

本章探討決策樹（Decision Trees）與集成學習（Ensemble Learning）模型的運作機制與應用。  
  
重點主題包括：  
1. 決策樹的建構方式（ID3、CART）、分裂準則（如 Information Gain、Gini Index）。  
2. 過擬合與剪枝技術（Pre-pruning、Post-pruning）。  
3. 集成學習的兩大類：Bagging（例如 Random Forest）與 Boosting（例如 AdaBoost）。  
4. 集成模型如何提升穩定性與準確率，以及其原理與比較。

## 🧠 名詞解釋（Term Explanation）

決策樹（Decision Tree）  
中文：一種透過條件分裂建立預測路徑的模型。  
英文：A model that uses conditional branching to make predictions.

ID3 演算法  
中文：根據資訊增益（Information Gain）選擇最佳分裂屬性。  
英文：Uses Information Gain to choose the best attribute to split.

CART 演算法  
中文：採用基尼指數（Gini Index）為分裂準則，可進行分類與回歸。  
英文：Uses Gini Index; supports both classification and regression.

資訊增益（Information Gain）  
中文：衡量某特徵分裂後對不純度的減少量。  
英文：Measures impurity reduction from a feature split.

基尼指數（Gini Index）  
中文：衡量資料集中類別的不純度程度。  
英文：Measures impurity level within a dataset.

剪枝（Pruning）  
中文：移除過度擬合的子樹以提升泛化能力。  
英文：Removes overfitted subtrees to improve generalization.

集成學習（Ensemble Learning）  
中文：結合多個模型提升穩定性與準確率。  
英文：Combines multiple models to improve stability and accuracy.

Bagging  
中文：隨機取樣子集後訓練多個模型並進行平均或投票。  
英文：Trains multiple models on random subsets and averages/votes results.

Boosting  
中文：連續訓練模型，每次針對前一次錯誤進行修正。  
英文：Trains sequential models correcting previous errors.

隨機森林（Random Forest）  
中文：以 Bagging 為基礎的多棵決策樹組合模型。  
英文：A Bagging-based ensemble of multiple decision trees.

## 📚 名詞比較（Term Comparison）

📌 ID3 vs. CART  
中文：ID3 使用資訊增益進行分類；CART 使用基尼指數，支援分類與回歸。  
英文：ID3 uses Information Gain for classification; CART uses Gini Index and supports regression.  
  
📌 Bagging vs. Boosting  
中文：Bagging 並行訓練模型減少變異；Boosting 序列訓練模型降低偏差。  
英文：Bagging trains in parallel to reduce variance; Boosting trains sequentially to reduce bias.  
  
📌 決策樹 vs. 隨機森林  
中文：決策樹單一模型，易過擬合；隨機森林為多樹結合，效果更穩定。  
英文：Decision Tree is a single model prone to overfitting; Random Forest is an ensemble, more stable.

## 📝 模擬出題與中英文詳解（包含解釋性問題）

Q1: Q1. 請比較 Bagging 與 Boosting 的訓練方式與偏差、變異之處理。  
答：Bagging 並行訓練，降低變異（variance）；Boosting 逐步修正錯誤，降低偏差（bias）。  
Bagging trains in parallel to reduce variance; Boosting trains sequentially to reduce bias.

Q2: Q2. 為何決策樹容易過擬合？有何因應方式？  
答：因為可完美擬合訓練資料，解法包括剪枝、限制深度與使用集成學習。  
It fits training data too well; solved by pruning, depth limiting, or using ensemble methods.

Q3: Q3. 隨機森林如何改善單一決策樹的不穩定性？  
答：透過多棵樹的平均結果減少對資料變異的敏感度。  
Averaging multiple trees reduces sensitivity to data fluctuations.

Q4: Q4. 為何 Gini Index 可用於分類任務？其數學意義為何？  
答：Gini 衡量資料的不純度，值越小表示分類越集中。  
Gini measures impurity; smaller values mean more concentrated classification.