

Day 47

機器學習

超參數調整



# 機器學習模型中的超參數

---

- 之前接觸到的所有模型都有超參數需要設置
  - LASSO，Ridge:  $\alpha$  的大小
  - 決策樹：樹的深度、節點最小樣本數
  - 隨機森林：樹的數量
- 這些超參數都會影響模型訓練的結果，建議先使用預設值，再慢慢進行調整
- 超參數會影響結果，但提升的效果有限，資料清理與特徵工程才能最有效的提升準確率，調整參數只是一個加分的工具。

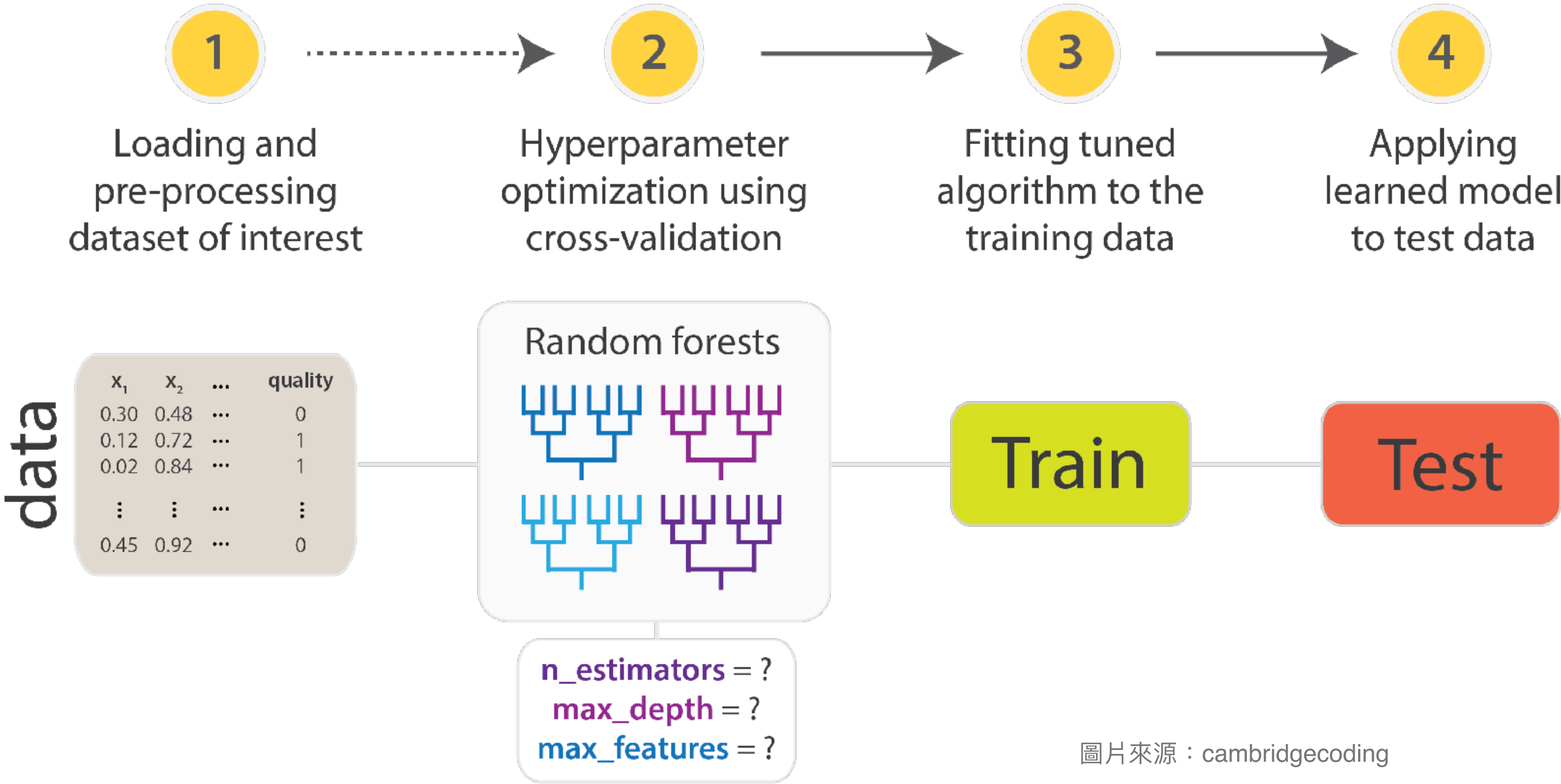
# 超參數調整方法

---

- 窮舉法 (Grid Search)：直接指定超參數的組合範圍，每一組參數都訓練完成，再根據驗證集 (validation) 的結果選擇最佳參數
- 隨機搜尋 (Random Search)：指定超參數的範圍，用均勻分布進行參數抽樣，用抽到的參數進行訓練，再根據驗證集的結果選擇最佳參數
- 隨機搜尋通常都能獲得更佳的结果，[詳見此](#)



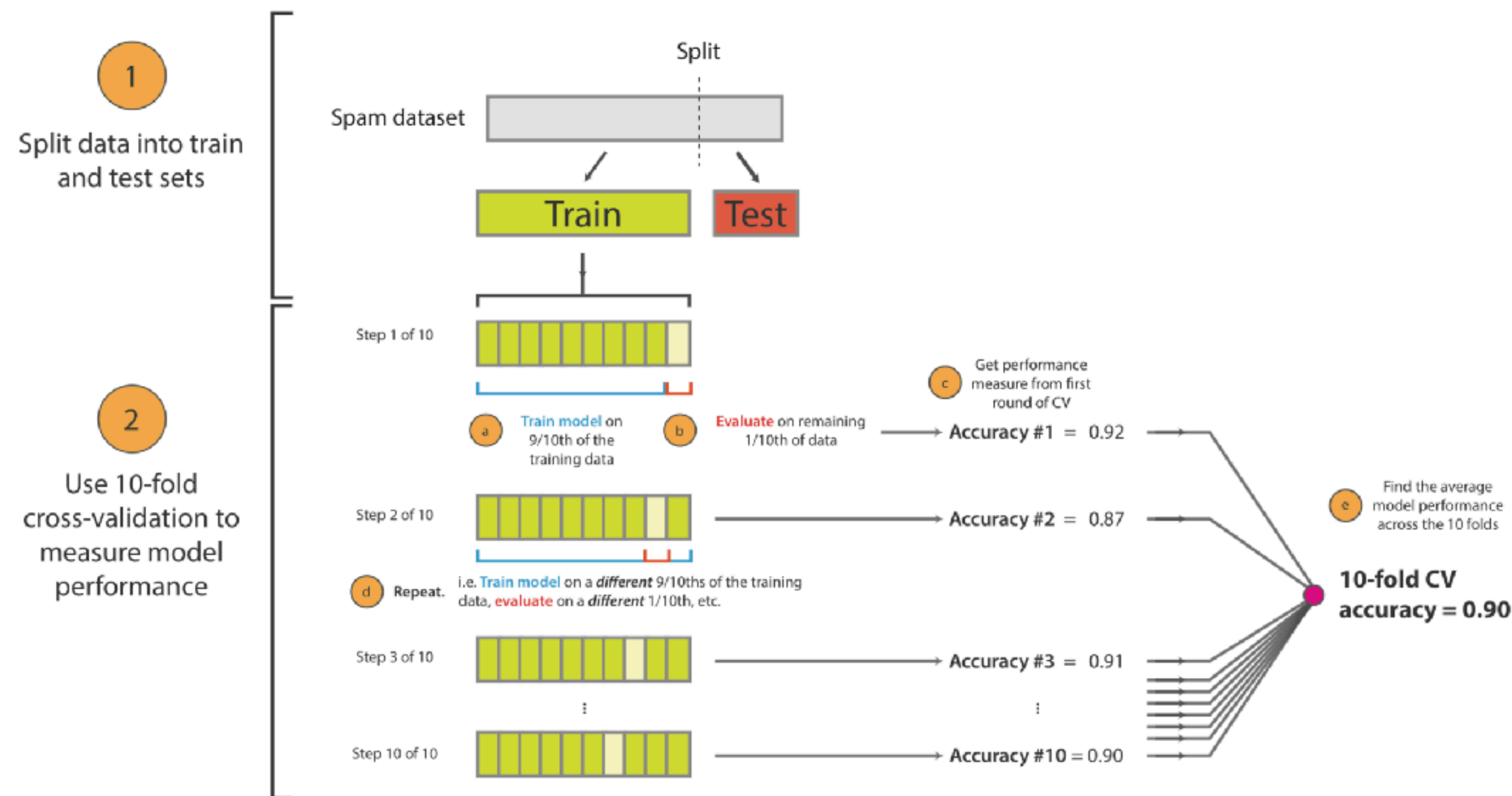
# 機器學習模型訓練步驟



圖片來源：cambridgecoding

# 正確的超參數調整步驟

- 若持續使用同一份驗證集 (validation) 來調參，可能讓模型的參數過於擬合該驗證集，正確的步驟是使用 Cross-validation 確保模型泛化性
- 先將資料切分為訓練/測試集，測試集保留不使用
- 將剛切分好的訓練集，再使用 Cross-validation 切分 K 份訓練/驗證集
- 用 grid/random search 的超參數進行訓練與評估
- 選出最佳的參數，用該參數與全部訓練集建模
- 最後使用測試集評估結果



# 解題時間 It's Your Turn

請跳出PDF至官網Sample Code & 作業  
開始解題

