

# D47：超參數調整與優化

[簡報閱讀](#)[範例與作業](#)[問題討論](#)[超參數調整](#)[知識地圖](#)[本日知識點目標](#)[機器學習模型中的超參數](#)[超參數調整方法](#)[機器學習模型訓練步驟](#)[正確的超參數調整步驟](#)[常見問題](#)[延伸閱讀](#)

## 超參數調整



## 知識地圖



## 本日知識點目標

- 了解何謂超參數，該如何調整
- 了解正確調整超參數的步驟
- 了解常用的調整超參數方法

## 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，詳見下方連結

Anyone that ever had to train a machine learning model had to go  
medium.com

## 機器學習模型訓練步驟

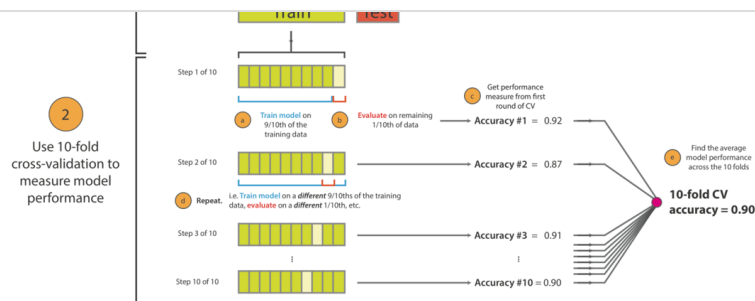


## 正確的超參數調整步驟

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

1. 先將資料切分為訓練/測試集，測試集保留不使用
2. 將剛切分好的訓練集，再使用 Cross-validation 切分 K 份訓練/驗證集
3. 用 grid/random search 的超參數進行訓練與評估
4. 選出最佳的參數，用該參數與全部訓練集建模
5. 最後使用測試集評估結果





## 常見問題

Q：超參數調整對最終結果影響很大嗎？

A：超參數調整通常都是機器學習專案的最後步驟，因為這對於最終的結果影響不會太多，多半是近一步提升 3-5 % 的準確率，但是好的特徵工程與資料清理是能夠一口氣提升 10-20 % 的準確率！因此建議專案一開始時，不需要花太多時間進行超參數的調整

## 延伸閱讀

- 劍橋實驗室教你如何調參數 - 英文

### Scanning hyperspace: how to tune

Introduction When doing machine learning using Python's Let's get cracking. Loading and exploring the [cambridgecoding.wordpress.com](https://cambridgecoding.wordpress.com)

- 教你使用 Python 調整隨機森林參數 - 英文

Improving the Random Forest Part  
Two

towardsdatascience.com

解題時間

[下一步：閱讀範例與完成作業](#)

