



人工智慧 Artificial Intelligence

國立東華大學電機工程學系 楊哲旻

Outline

01 人工智慧發展史



02 人工智慧的任務



03 人工智慧的學習方法



04 資料格式

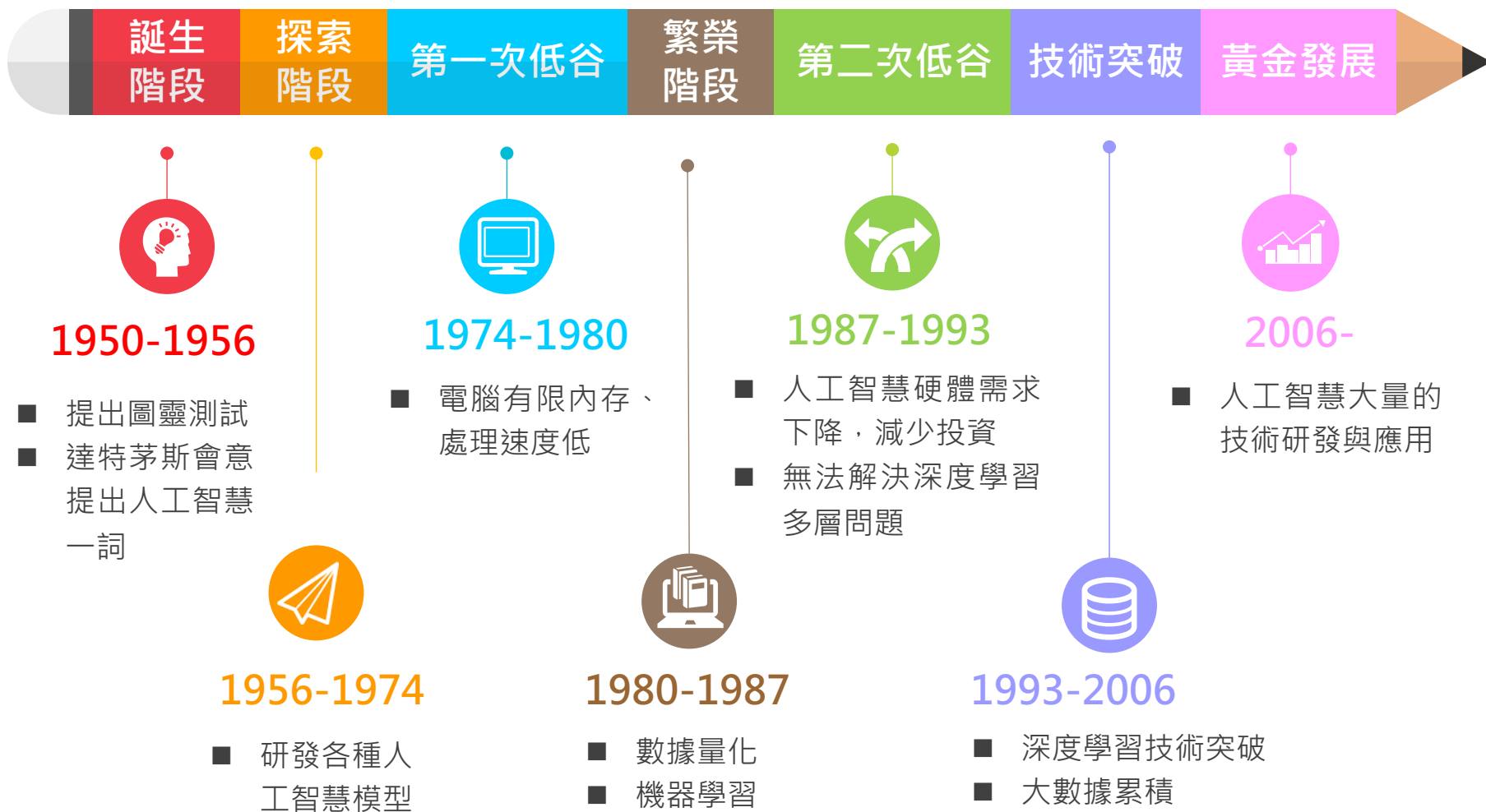




01 人工智慧發展史

人工智慧發展史

1



人工智慧發展史

2

■ 人工智慧 (Artificial Intelligence)

- 強人工智慧
- 弱人工智慧

大數據時代來臨

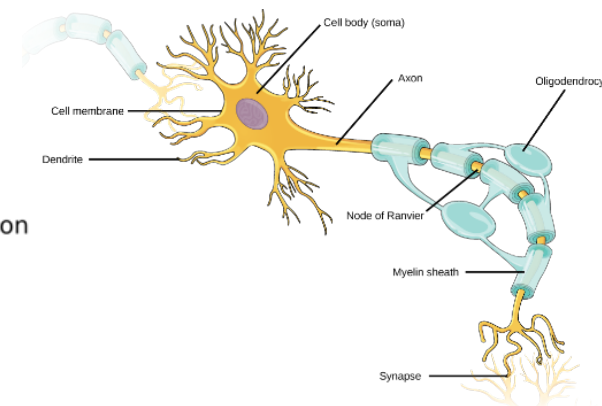
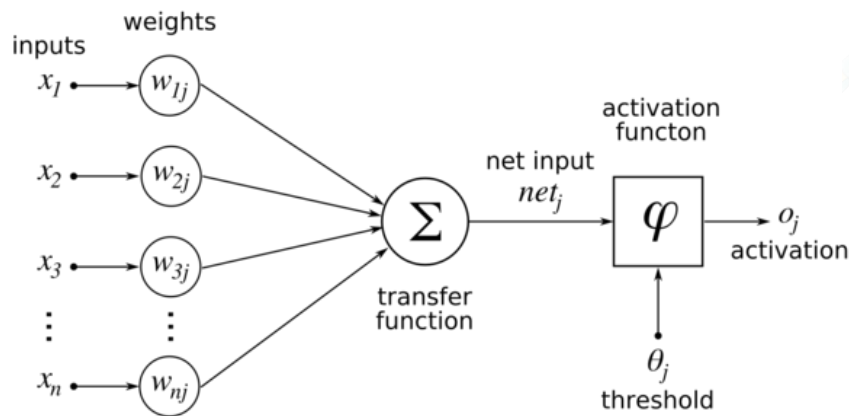
■ 機器學習 (Machine Learning)

- 非監督式學習
- 監督式學習
- 增強式學習

仿生物神經元架構

■ 深度學習 (Deep Learning)

- 深度神經網路
- 卷積神經網路
- 循環神經網路



中央處理器 Central Processing Unit, CPU

圖形處理器 Graphics Processing Unit, GPU

張量處理器 Tensor Processing Unit, TPU





分類 (Classification)

輸出為離散數值，例：疾病有病無病，影像是貓是狗等



分群 (Cluster)

又稱聚類，輸出為離散數值，藉由距離或數學公式進行類似的分類任務



回歸 (Regression)

輸出為連續數值，例：天氣溫度、PM2.5濃度等



其它複雜性任務

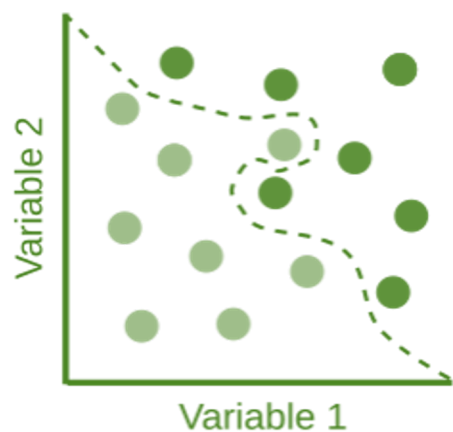
輸出為複雜性結構，例：輸出為圖，影像生成畢卡索畫風，人像照片生成老化人像；輸出為語音，模仿某人人聲等

根據已標註的資料集進行訓練，自動學習出一個函數，此函數是所有權重與輸入之組合



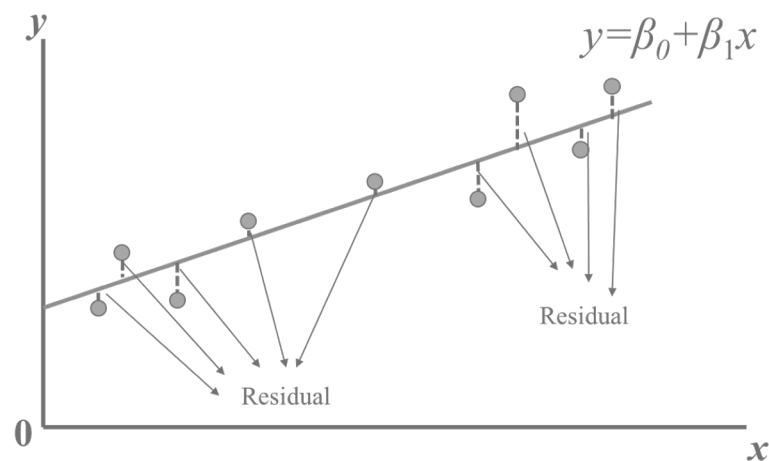
分類

設計一函數可將類別的資料點分開



回歸

設計一函數的預測輸出與實際值(標籤)誤差最小





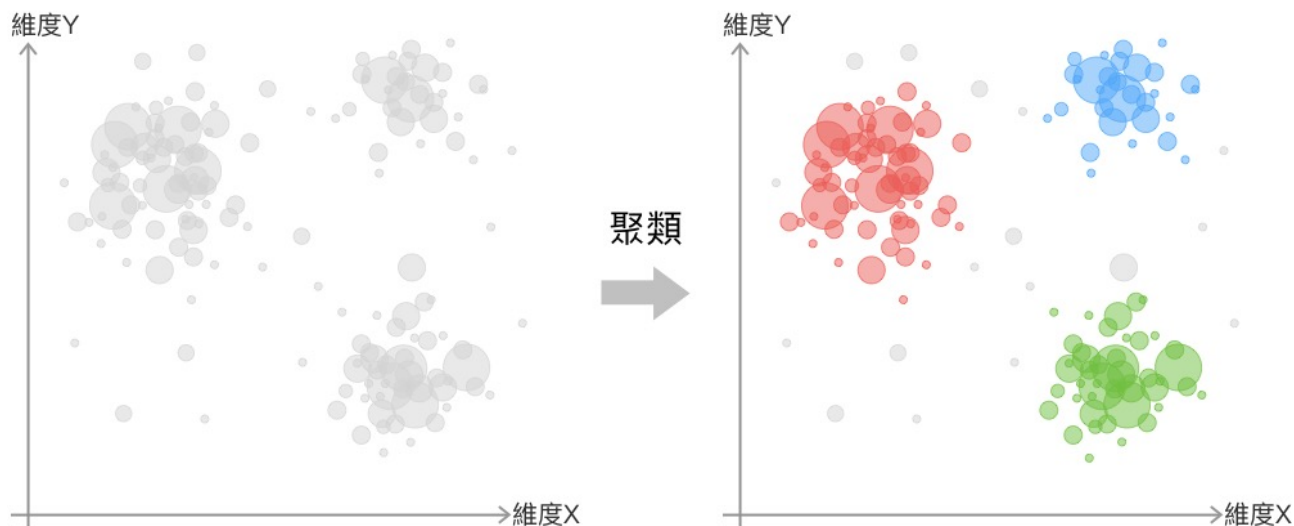
分群

透過群的代表點，針對它們
近距離的資料點歸於同類



其它複雜性任務

使用生成對抗網路生成不同
畫風的圖像



人工智慧的任務 – 影像相關

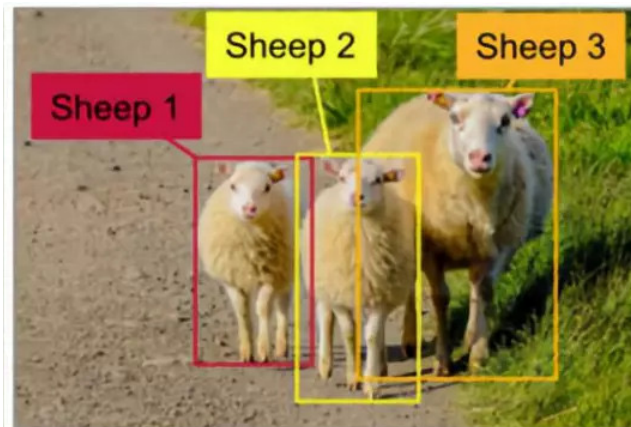
6

分類



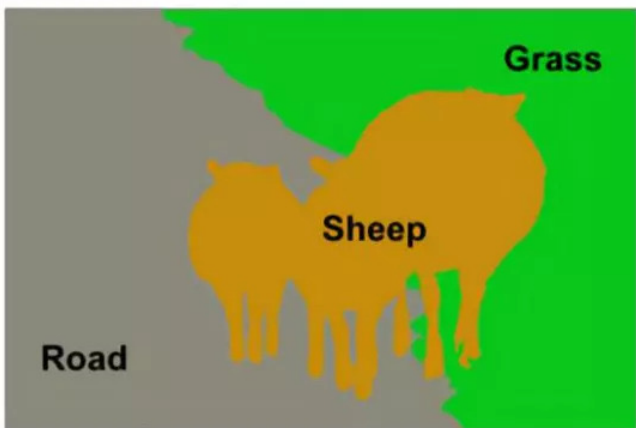
Classification

目標檢測



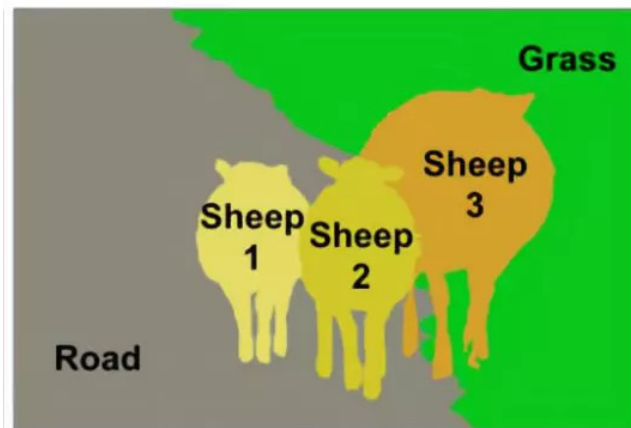
Object Detection

語義分割



Semantic Segmentation

實例分割



Instance Segmentation

人工智慧的學習方法

7

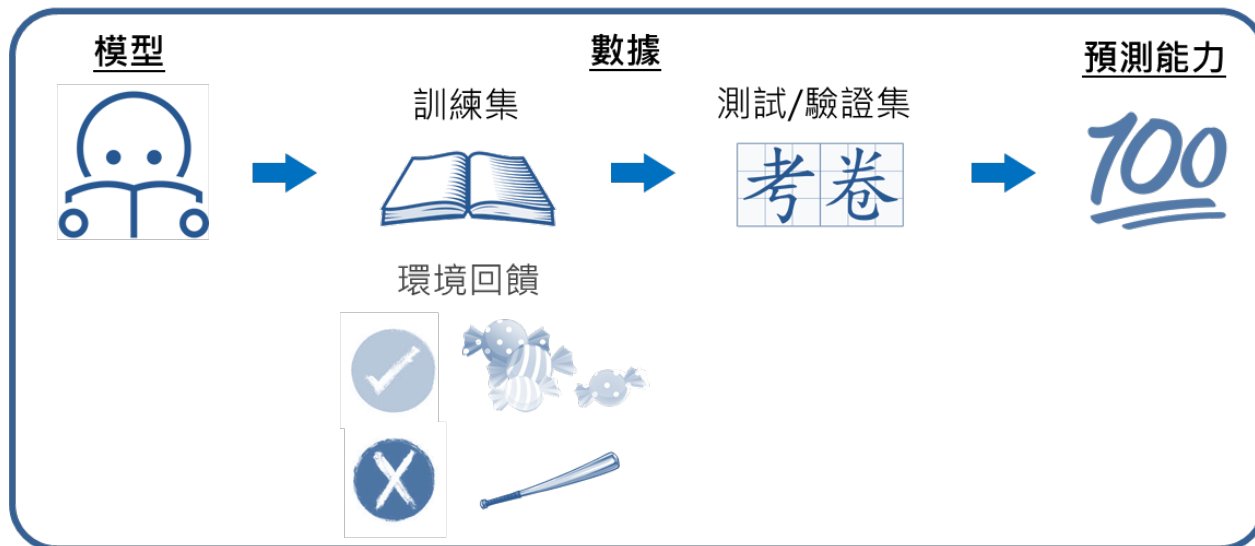
監督式學習(Supervised learning)



非監督式學習(Unsupervised Learning)



強化學習(Reinforcement Learning)



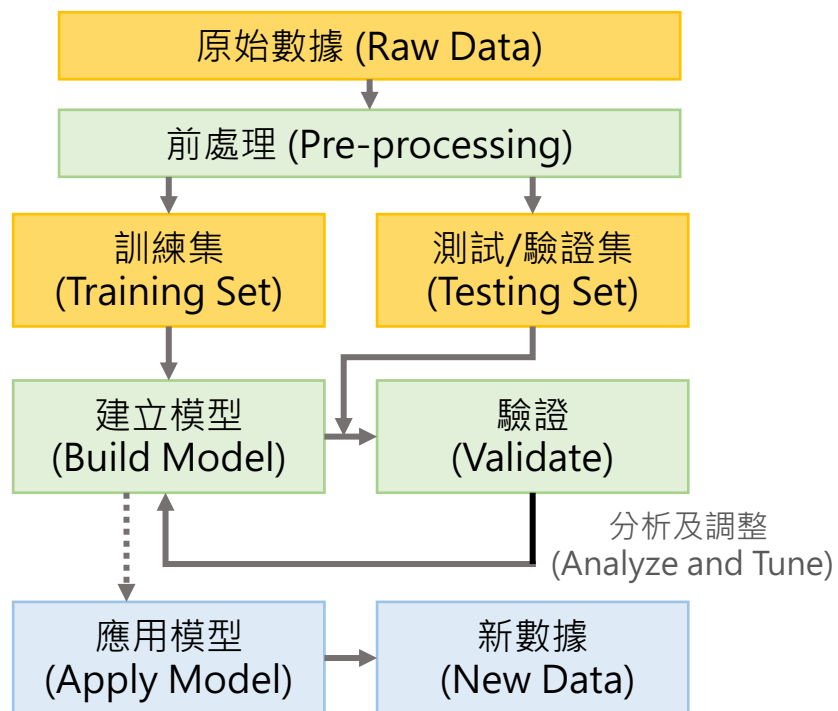


監督式學習(Supervised learning)

監督式學習是電腦從標籤化(Labeled) 的資訊中分析模式後做出預測的學習方式

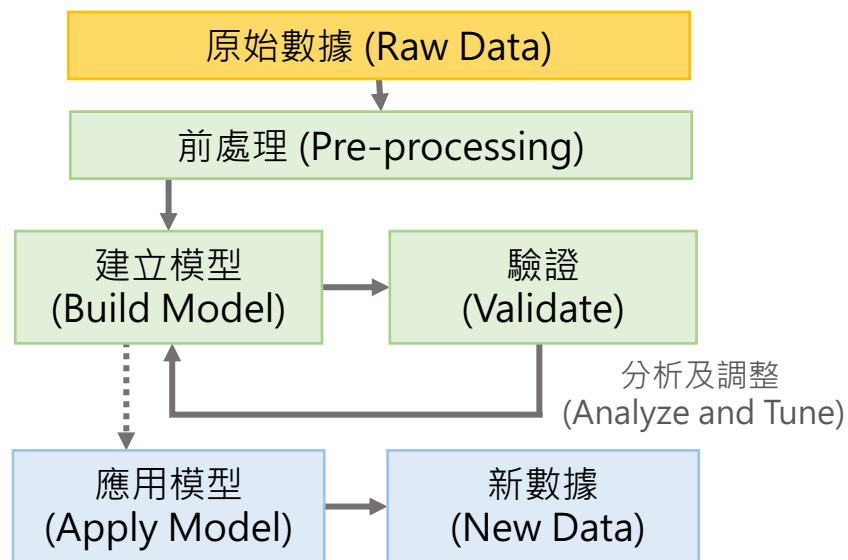
常見的監督式學習模型為：

1. 線性回歸
2. 邏輯回歸
3. 決策樹
4. 隨機森林
5. 支持向量機
6. 多層感知器
7. 卷積神經網路





非監督式學習(Unsupervised Learning)



非監督式學習不需事先以人力處理標籤，機器面對資料時，做的處理是依照關聯性去歸類、找出潛在規則與套、形成集群。

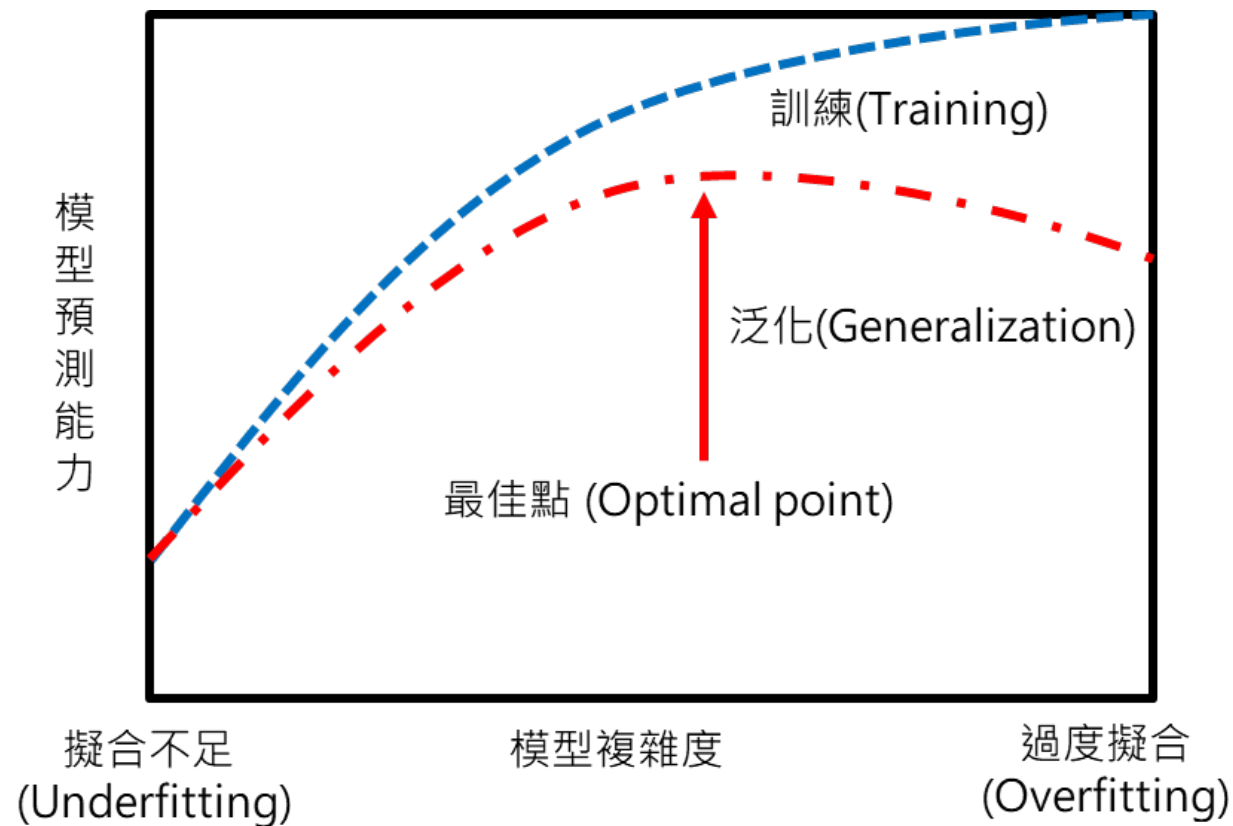
常見的非監督式學習模型為：

1. K平均演算法
2. 主成分分析
3. 孤獨森林
4. 單類別支持向量機
5. 自編碼器
6. 生成對抗網路

泛化能力(Generalization Ability)

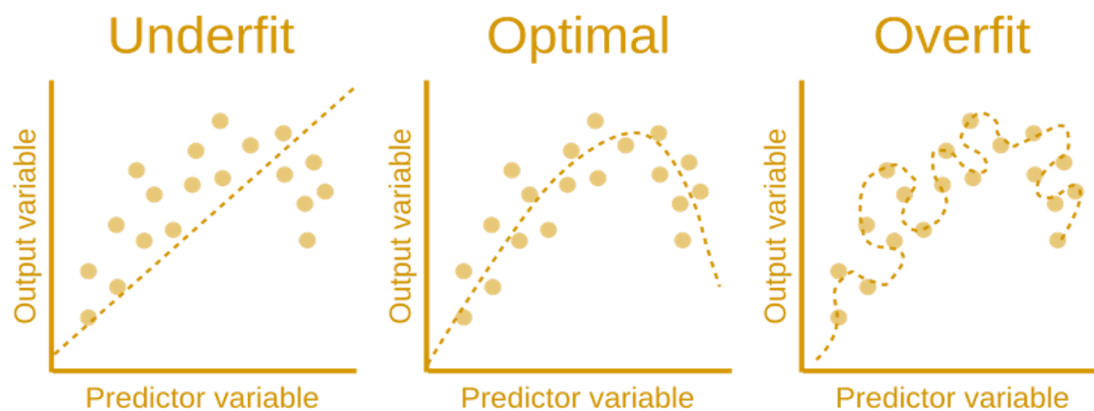
泛化能力是指模型適應新數據的能力。以下方法可以提高泛化能力：

1. 資料樣本增大
2. 資料前處理
3. 調降低整模型**超參數**來模型複雜度

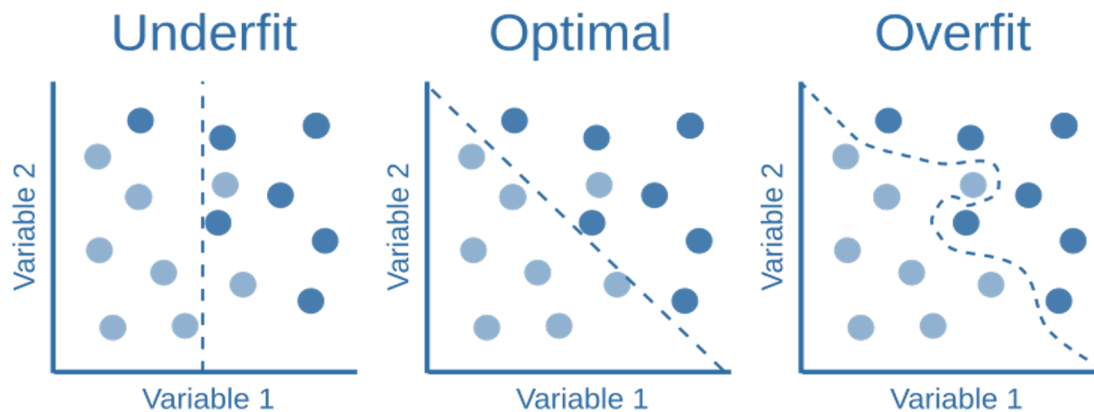



泛化能力(Generalization Ability)

回歸



分類



 在數據分析與人工智慧模型訓練前的第一步，就是要了解資料與蒐集資料，目前常見的資料格式分為**結構化**、**半結構化**與**非結構化**資料：

| | 結構化資料 (Sturctured Data) | 半結構化資料 (Semi-Structured Data) | 非結構化資料 (Unstructured Data) |
|-----|---|--|-------------------------------------|
| 定義 | 嚴謹定義為資料可以被呈現在資料庫表格的行與欄，即已被整理過的資料 | 便於資料交換，其特性同時具備欄位概念與欄位可拓展性，可透過欄位查詢資料，並可根據使用者需求來增減欄位 | 形式自由且不遵循標準的格式規範，一團沒有組織的數據，即未經整理過的資料 |
| 優缺點 | 查詢資料快速，佔用存儲空間少；缺點是拓展新的欄位比較麻煩，在資料交換上的規定也比較嚴格 | 利於資料交換與傳輸，並可以增減欄位；缺點每筆資料的結構可能會不一致 | 佔用更多存儲空間，無法直接用於數據分析、未規則性的資料很難處理與整理 |
| 範例 | 關聯式資料庫(MySQL, Oracle等)的資料、Excel | CSV、JSON與XML | 文字、圖片、音樂、影片、PDF、網頁等 |

※ 先有結構，再有資料

半結構化資料

➤ XML

```
<?xml version="1.0" encoding="UTF-8"?>
<root>
  <row>
    <order>A001</order>
    <item>紅茶</item>
    <sugar_level>25%</sugar_level> <!-- 只增加糖度欄位 -->
    <unit_price>30</unit_price>
    <quantity>5</quantity>
    <total_amount>150</total_amount>
  </row>
  <row>
    <order>A002</order>
    <item>綠茶</item>
    <ice_level>25%</ice_level> <!-- 只增加冰塊欄位 -->
    <unit_price>30</unit_price>
    <quantity>3</quantity>
    <total_amount>90</total_amount>
  </row>
</root>
```

➤ JSON

```
{
  "table": {
    "row": [
      {
        "order": "A001",
        "item": "紅茶",
        "sugar_level": "25%", <-- 只增加糖度欄位
        "unit_price": "30",
        "quantity": "5",
        "total_amount": "150"
      },
      {
        "order": "A002",
        "item": "綠茶",
        "ice_level": "25%", <-- 只增加冰塊欄位
        "unit_price": "30",
        "quantity": "3",
        "total_amount": "90"
      }
    ]
  }
}
```

非結構化資料



批踢踢實業坊 > 看板 Gossiping

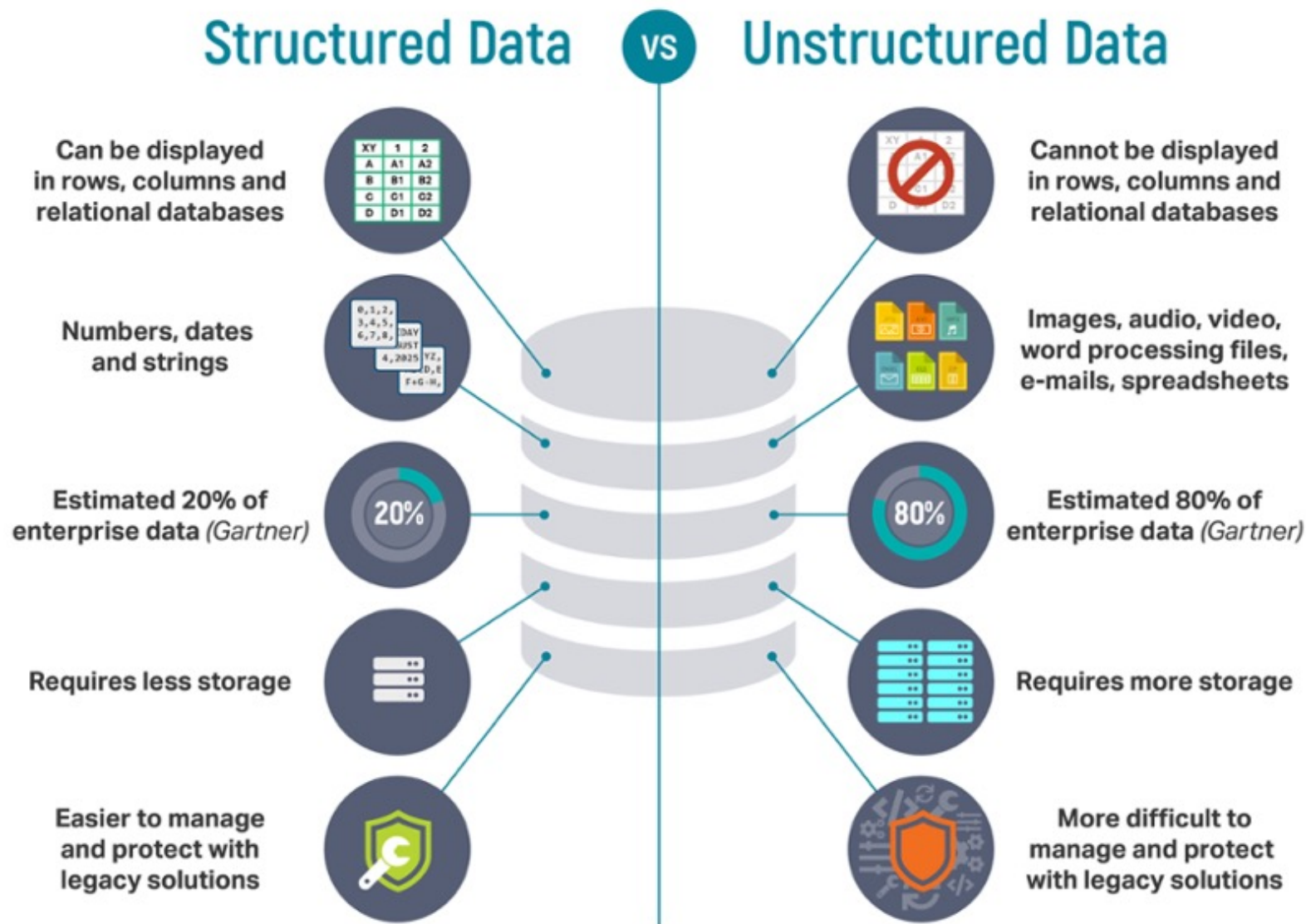
13 Re: [問卦] 登入破三千的都給我進來
shotholisi

Re: [新聞] 拚基本工資28K 五一勞工大遊行
trylin

結構化資料

網路爬蟲

| 發文日期 | 作者 | 主題 | 回復人數 |
|------------|------------|------------------|-------|
| 2018/05/01 | shotholisi | 登入破三千的都給我進來 | 13 |
| 2018/05/01 | trylin | 拚基本工資28K 五一勞工大遊行 | 0 |
| | | | |





Google Colab



基本功能

- 程式檔預設附檔名為.ipynb，可以透過「File」中的功能鍵另存為.py檔案
- File中的許多功能可以另存到本機、雲端硬碟與Github
- 文字編輯則是以「Markdown」純文字語法來撰寫

The screenshot shows the Google Colab web interface. Above the interface, six colored boxes with arrows point to specific features:

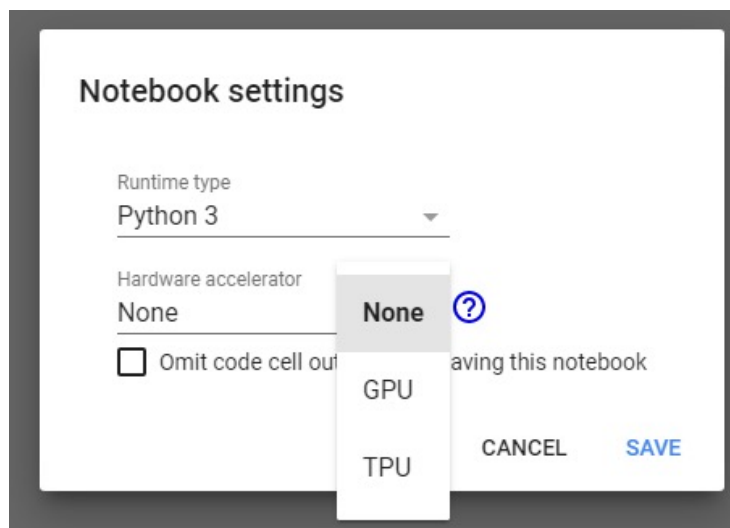
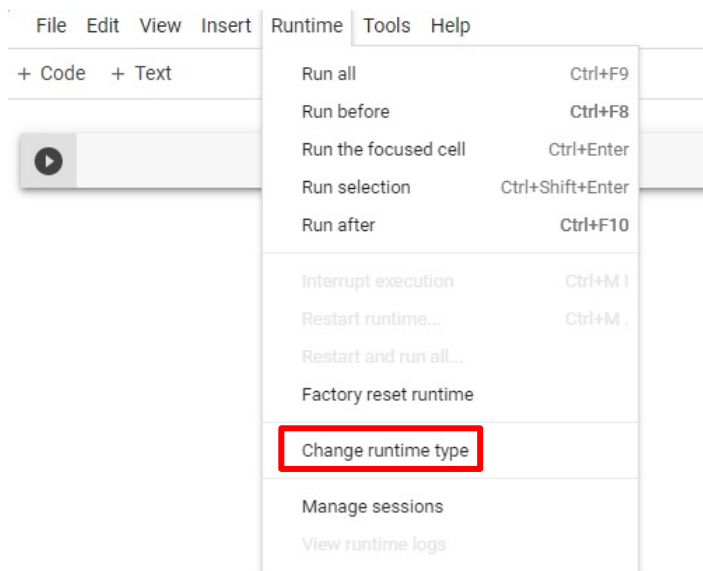
- 程式碼名稱** (Code Name): Points to the "My Colab.ipynb" title.
- 新增程式碼** (Add Code): Points to the "+ Code" button in the top toolbar.
- 移動單元格** (Move Cell): Points to the up/down arrow icons in the cell toolbar.
- 新增文字** (Add Text): Points to the "+ Text" button in the top toolbar.
- 文字編輯** (Text Editing): Points to the "Editing" button in the cell toolbar.
- 雲端帳號登入** (Cloud Account Login): Points to the user profile icon in the top right corner.

The interface itself shows a "Table of contents" on the left, a "My Colab" section with a code cell containing `[] print("hello world!")`, and a bottom toolbar with "Code" and "Text" buttons.



基本功能

選擇「Runtime」→「Change runtime type」來更改設定，可以選擇程式語言(但目前只支援Python 2與3)，另外也可指定筆記本是否要使用硬體加速器GPU/TPU。



基本功能

- 執行「!pip list」可以查看所有安裝套件與其版本
- 若需要使用尚未安裝套件則可以用「! pip install 套件名稱」來安裝

```
!pip list
```

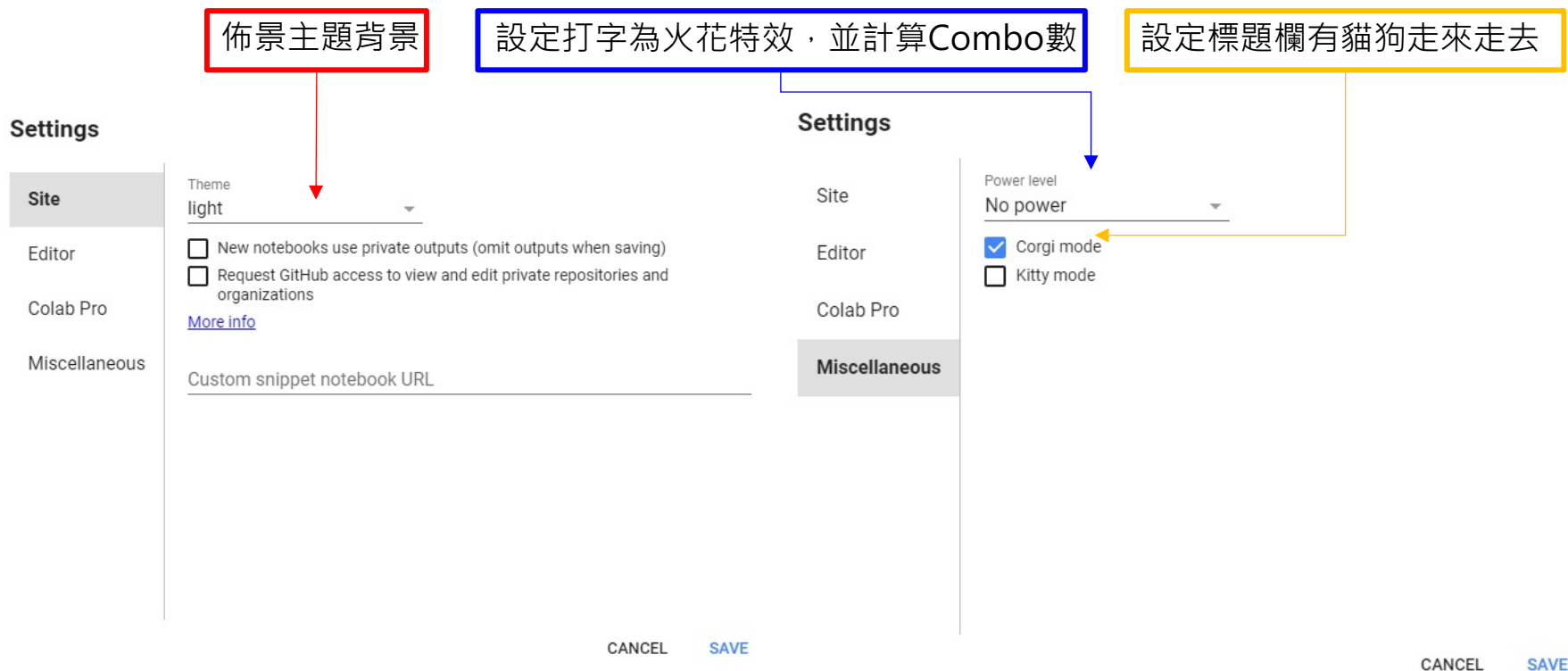
| | |
|-------------------------------|--------|
| tensor2tensor | 1.14.1 |
| tensorboard | 1.15.0 |
| tensorboardcolab | 0.0.22 |
| tensorflow | 1.15.0 |
| tensorflow-addons | 0.8.3 |
| tensorflow-datasets | 2.1.0 |
| tensorflow-estimator | 1.15.1 |
| tensorflow-federated | 0.12.0 |
| tensorflow-gan | 2.0.0 |
| tensorflow-gcs-config | 2.1.8 |
| tensorflow-hub | 0.7.0 |
| tensorflow-metadata | 0.21.1 |
| tensorflow-model-optimization | 0.2.1 |
| tensorflow-privacy | 0.2.2 |
| tensorflow-probability | 0.7.0 |
| termcolor | 1.1.0 |
| terminado | 0.8.3 |
| testpath | 0.4.4 |
| text-unidecode | 1.3 |
| textblob | 0.15.3 |
| textgenrnn | 1.4.1 |
| tflearn | 0.3.2 |

- 執行「! pip install 套件名稱==版本編號」則可安裝指定的版本
- 執行「! pip install -upgrade 套件名稱」則可以安裝指定的版本
(注意：自行安裝或更新的套驗在虛擬機器關閉後即失效，建議安裝指令保存在筆記本的單元格中)

在指令前面加 **!** 就表示要執行命令列的執行檔或指令(而非執行Python程式碼)

其餘功能

- 可以設定佈景主題背景、編輯器功能及一些有趣的特效



佈景主題背景

設定打字為火花特效，並計算Combo數

設定標題欄有貓狗走來走去

Settings

Site

Theme
light

☐ New notebooks use private outputs (omit outputs when saving)

☐ Request GitHub access to view and edit private repositories and organizations

[More info](#)

Custom snippet notebook URL

CANCEL SAVE

Settings

Site

Editor

Colab Pro

Miscellaneous

Power level
No power

☒ Corgi mode

☐ Kitty mode

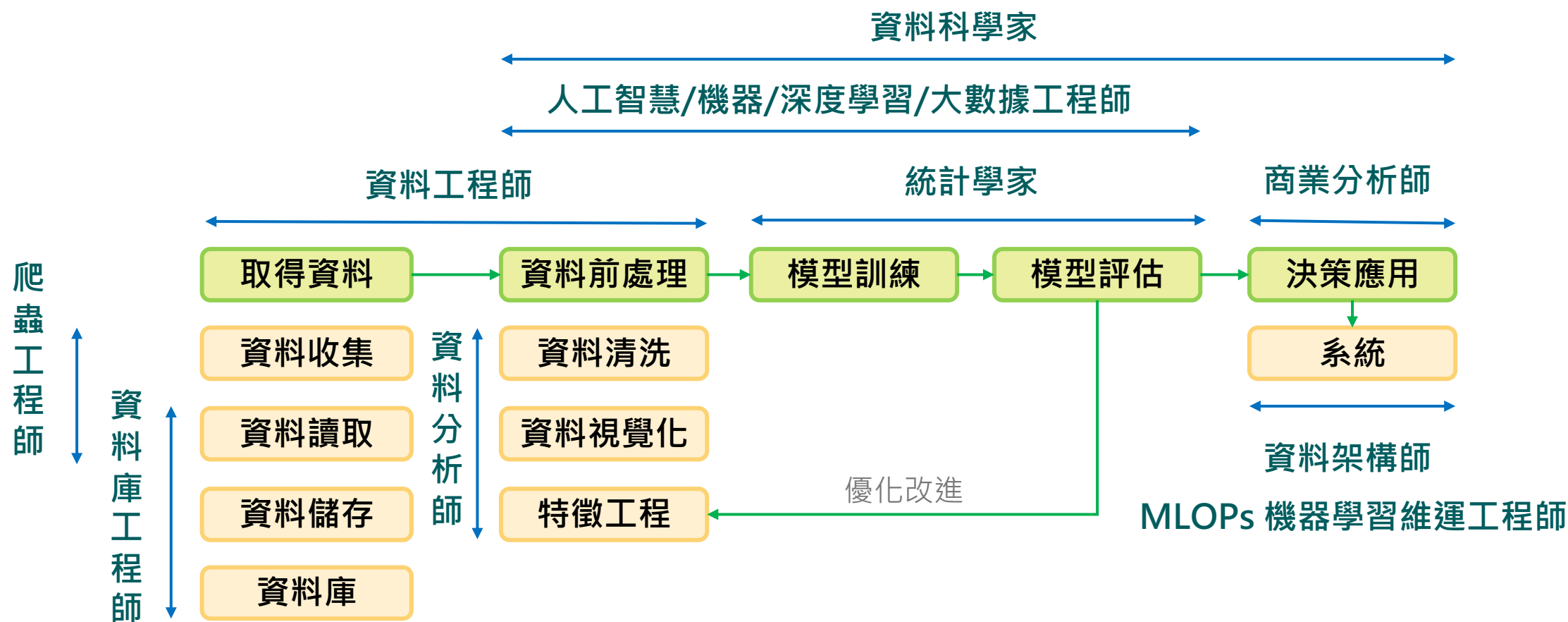
CANCEL SAVE



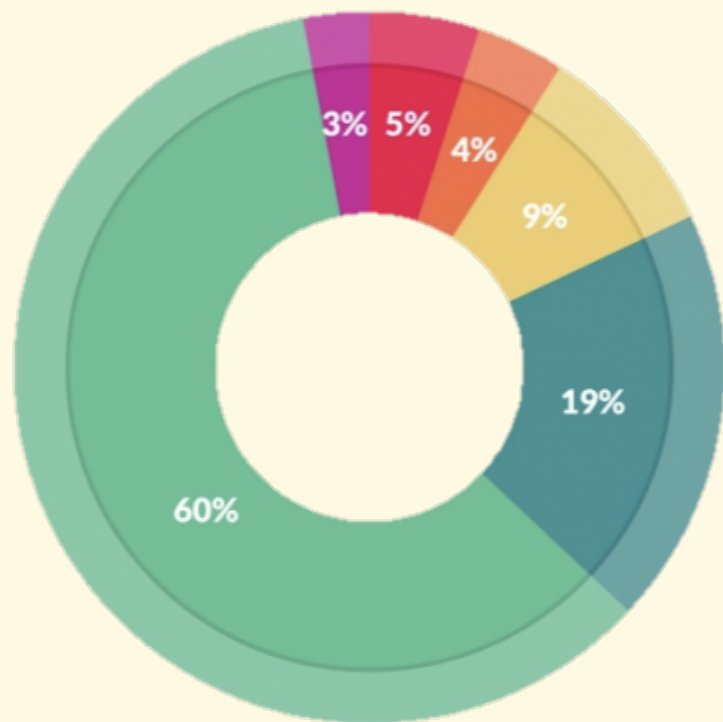
AI相關的軟體工程師

AI相關的軟體工程師

19

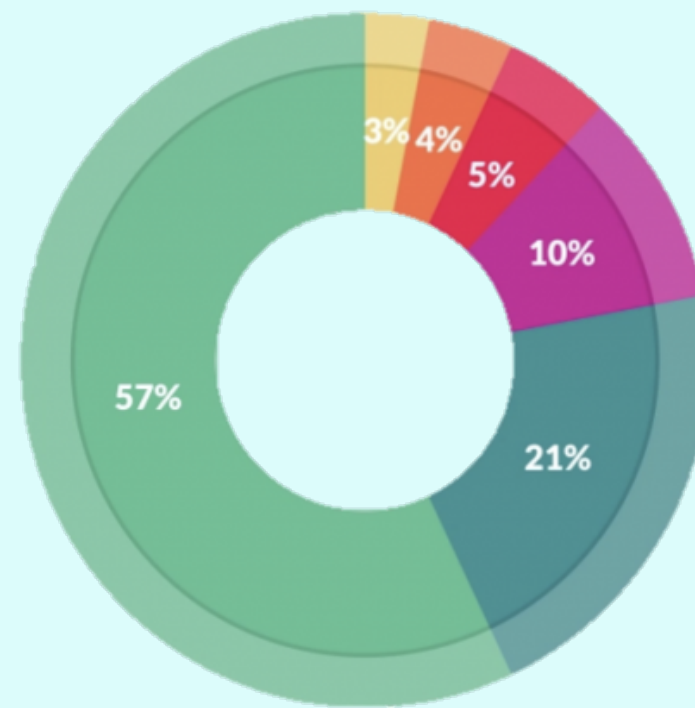


所花費工作項目時間佔比



- 清洗與組織資料: 60%
- 資料收集: 19%
- 特徵探勘與工程: 9%
- 創造訓練資料集: 3%
- 其他: 4%
- 優化機器學習演算法: 5%

最不喜歡的工作項目佔比



- 清洗與組織資料: 57%
- 資料收集: 21%
- 創造訓練資料集: 10%
- 其他: 5%
- 優化機器學習演算法: 4%
- 特徵探勘與工程: 3%