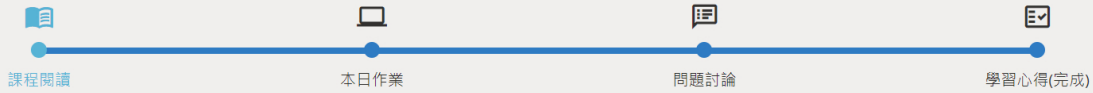


D3：機器學習 - 流程與步驟



1.2 如何開啟、處理檔案? >

2. 定義目標 >

2. 設定評估準則 >

3. 建立模型並調整參數 >

4. 導入 >

常見問題 >

解題時間 >

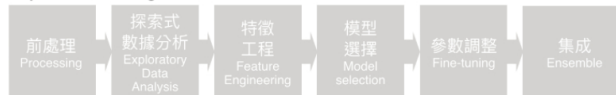
機器學習-流程與步驟



知識地圖

機器學習概論 Introduction of Machine Learning

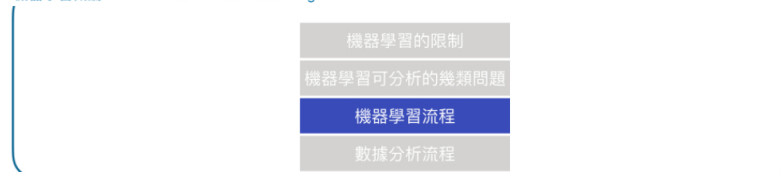
監督式學習 Supervised Learning



非監督式學習 Unsupervised Learning



機器學習概論 Introduction of Machine learning



課程講解

eupoy AI共學社群 我的

機器學習-流程與步驟 >

知識地圖 >

課程講解 >

本日知識點目標 >

機器學習專案開發流程 >

1.1 資料在哪? 結構為何? >

1.2 如何開啟、處理檔案? >

2. 定義目標 >

A: 當我們訓練好一個機器學習模型，為了驗證其可行性，多半會讓模型正式上線，觀察其在實際資料進來的結果；有時也會讓模型跟專家進行 PK，挑一些真實資料讓模型與專家分別測試，評估其準確率。

解題時間

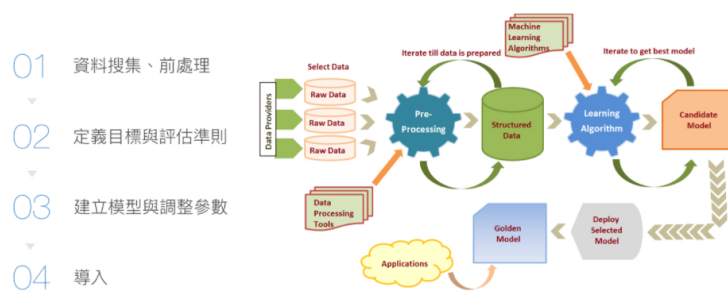
8:30 / 8:30

本日知識點目標



- 了解一個完整機器學習專案的細節
- 機器學習專案的開發流程步驟
- 每個步驟的意義及該如何進行

機器學習專案開發流程



- 01 資料搜集、前處理
- 02 定義目標與評估準則
- 03 建立模型與調整參數
- 04 導入

機器學習啟動的必要流程與準備事項

做機器學習(machine learning)應該要準備哪些東西呢？絕對不是單單一個人就可以完成，而是整個團隊一同建構。但流程應該是如何？從最深的坑：資料搜集、處理，就開始發問：搜集哪些資料？整理成什麼樣子？瞭解自己的資料：資料標記、特徵工程，反而模型調整、優化、訓練是最簡單快速的地方。有過無數辛酸血淚史的講師，告訴我們每一個階段都需要思考的問題。



1.1 資料在哪？結構為何？

政府公開資料集、Kaggle 資料集

- 結構化資料
 - Excel 檔 (.xlsx)
 - CSV 檔 (.csv, 逗號分隔)
- 非結構化資料
 - 圖片
 - 影音
 - 文字

1.2 如何開啟、處理檔案?

 Python



- 多數檔案都能使用 Python 的套件開啟
 - 開啟圖片: PIL, skimage, open-cv...
 - 開啟文件: pandas
- 資料前處理
 - 缺失值填補
 - 離群值處理
 - 標準化

2. 定義目標

- 回歸問題? 分類問題?
- 要預測的目標是甚麼? (target 或 y)
- 要用甚麼資料來進行預測? (predictor 或 x)
- 將資料分為
 - 訓練集 training set
 - 驗證集 validation set
 - 測試集 test set

2. 設定評估準則

- 不同問題有不同的評估指標
- 回歸問題 (預測值為實數)
 - a. RMSE, Root Mean Square Error
 - b. Mean Absolute Error
 - c. R-Square
- 分類問題 (預測值為類別)
 - a. Accuracy
 - b. F1-score
 - c. AUC, Area Under Curve

3. 建立模型並調整參數

根據設定目標建立機器學習模型

- Regression · 回歸模型
- Tree-based model · 樹模型
- Neural network · 神經網路
- 各模型都有其超參數需調整，根據經驗與對模型了解、訓練情形等進行調參

4. 導入

- 建立資料搜集、前處理等流程
- 送進模型進行預測
- 輸出預測結果
- 視專案需求整合前後端
 - 建議統一資料格式，方便讀寫 (.json, .csv)

常見問題



Q: 如何確立一個機器學習模型的可用性？

A: 當我們訓練好一個機器學習模型，為了驗證其可行性，多半會讓模型正式上線，觀察其在實際資料進來時的結果；有時也會讓模型跟專家進行 PK，挑一些真實資料讓模型與專家分別測試，評估其準確率。

解題時間



Sample Code & 作業
開始解題



[下一步：完成作業](#)

