Day 16 探索式數據分析

繪圖與樣式 &
Kernel Density Estimation (KDE)

出題教練 游為翔 / 杜靖愷

知識地圖 探索式數據分析 核密度函數

機器學習概論 Introduction of Machine Learning

監督式學習
Supervised Learning

前處理 Processing → **探索式數據分析 Exploratory Data Analysis** → 特徵工程 Feature Engineering → 模型選擇 Model selection → 參數調整 Fine-tuning → 集成 Ensemble

非監督式學習
Unsupervised Learning

- 分群 Clustering
- 降維 Dimension Reduction

探索式數據分析 Exploratory Data Analysis (EDA)

統計值的視覺化

- 相關係數
- **核密度函數**
- 離散化

- 繪圖排版
- 常用圖形
- 模型體驗

本日知識點目標

- 知道 matplotlib 的其他 theme
- 學會什麼是 Kernel Density Estimation (KDE) 與如何繪製

繪圖風格

```
plt.style.use('default') # 不需設定就會使用預設
plt.style.use('ggplot')
plt.style.use('seaborn') # 或採用 seaborn 套件繪圖
```

轉變繪圖風格的目的

用已經被設計過的風格，讓觀看者更清楚明瞭，包含色彩選擇、線條、樣式等。

Kernel Density Estimation (KDE)

不同 kernel function 的結果

1

採用無母數方法畫出一個觀察變數的機率密度函數

某個 X 出現的機率為何

2

Density plot 的特性

- 歸一：線下面積和為 1
- 對稱：K(-u) = K(u)

3

常用的 Kernel function

- Gaussian (Normal dist)
- Cosine

重要知識點複習

- KDE 的優點與缺點
  - 優：無母數方法，對分布沒有假設 (使用上不需擔心是否有一些常見的特定假設，如分布為常態)
  - 缺：計算量大，電腦不好可能跑不動
- 透過 KDE plot，我們可以較為清楚的看到不同組間的分布差異

延伸閱讀

除了每日知識點的基礎之外，推薦的延伸閱讀能補足學員們對該知識點的了解程度，建議您解完每日題目後，若有多餘時間，可再補充延伸閱讀文章內容。

推薦延伸閱讀

繪圖靈感資源參考

Python Graph Gallery (圖表參考)

網頁連結

- 這裡整合了 Python 許多繪圖函數的寫法，同學可以依據自己的喜好與資料形式，挑選適合的圖形寫作，並不需要全部看懂，只需要當成查詢用的工具手冊即可。

推薦延伸閱讀

R Graph Gallery

網頁連結

- 這裡整合了 R 許多繪圖函數的寫法，與上面的網站是相關網站，如果較擅長使用 R 做資料科學，可以先從這邊參考。

推薦延伸閱讀

R Graph Gallery (Interactive plot，互動圖)

網頁連結

- 可以由 R 語言繪製出的互動圖表，也是提供同學查詢之用。

解題時間

It’s Your Turn

請跳出PDF至官網Sample Code＆作業
開始解題