

D5 : 如何新建一個 dataframe? 如何讀取其他資料? (非 csv 的資料)



課程閱讀



本日作業



問題討論



學習心得(完成)

如何新建一個 dataframe ?

知識地圖

課程講解

本日知識點目標

為什麼新建一個
dataframe 重要 ?

如何新建一個 dataframe ?



知識地圖

機器學習概論 Introduction of Machine Learning

監督式學習 Supervised Learning



前處理 Processing



課程講解



The screenshot shows a horizontal progress bar at the top with four segments: '課程閱讀' (blue), '本日作業' (green), '問題討論' (yellow), and '學習心得(完成)' (red). Below the bar is a sidebar with links: '如何新建一個 dataframe?' (highlighted in blue), '知識地圖' (grey), and '本日知識點目標' (grey). The main content area displays a video thumbnail for 'Day 5-1 資料清理數據前處理' with the title '如何新建一個 dataframe?'.

本日知識點目標



- 了解如何快速驗證 dataframe 操作的程式碼

為什麼新建一個 dataframe 重要？

需要把分析過程中所產生的數據或者結果儲存為結構化的資料

- 例子1：將每筆交易資料匯總計算平均值、標準差等統計數值
- 例子2：Kaggle 比賽要上傳的結果

測試程式碼

- 有時候原始資料太大了，有些資料的操作很費時，先在具有同樣結構的資料上測試程式碼是否能夠得到理想中的結果。
- 不確定視覺化程式碼中所需要的資料結構，用新建立的 dataframe 結構來去了解，而不是急著在原始資料上操作。

重要知識點複習

- 在資料量很大時，可以先在和資料具有同樣結構的小樣本驗證程式碼執行的結果是否符合預期
- 用 pd.DataFrame 來創建一個 dataframe
- 用 np.random.randint 來產生隨機數值

5-2如何讀取其他資料？(非CSV的資料)





本日知識點目標



- 學會如何讀取其他資料格式 : txt / jpg / png / json / mat / npy / pkl ...

讀取其他非csv資料格式 ?

檔案格式

- 圖像檔 (PNG / JPG ...)

讀取範例 :

```
Import cv2  
image = cv2.imread(...) # 注意 cv2 會以 BGR 讀入  
image = cv2.cvtColor(image, cv2.COLOR_BGR2RGB)
```

```
from PIL import Image  
image = Image.read(...)  
import skimage.io as skio  
image = skio.imread(...)
```

- Python npy

讀取範例 :

```
import numpy as np  
arr = np.load(example.npy)
```

- Pickle (pkl)

讀取範例 :

```
import pickle  
with open('example.pkl', 'rb') as f:  
arr = pickle.load(f)
```

重要知識點複習

不同的資料有不同讀取方式

- 文字格式 : 通常可以用 with open()
- 圖像格式 : 可以使用 PIL, Skimage 或是 CV2 (CV2 的速度較快 , 但須注意讀入的格式為 BGR)
- 其他形式如 npy / pickle 可以儲存經過處理後的資料

延伸閱讀

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

[Pandas Foundations](#)

The screenshot shows a DataCamp video player. The title is "pandas DataFrames". Below the title is a bulleted list: "• Example: DataFrame of Apple Stock data". A table representing the DataFrame is shown:

Date	Open	High	Low	Close	Volume	Adj Close
2014-09-16	99.80	101.26	98.89	100.86	66818200	100.86
2014-09-15	102.81	103.05	101.44	101.63	61216500	101.63
2014-09-12	101.21	102.19	101.08	101.66	62626100	101.66
...

A video player interface at the bottom shows a progress bar, a play button, and a timestamp of 3:47. To the right is a portrait of a man in a blue sweater.

第一個 chapter 是免費的，建議可用來預習 pandas，如果覺得英文聽不懂也沒關係，可以按部就班跟著我們後面的課程，也可以學到相關的內容。

Github repo

機器學習百日馬拉松之後也會有 pandas 操作相關的練習，但若你迫不及待想要更精進自己 pandas 技能，可以到 [github repo](#) 挑戰！

The screenshot shows a GitHub repository page for "guipsamora / pandas_exercises". The repository has 128 stars and 1,695 forks. The main page displays a Jupyter notebook file named "Exercise_with_Solutions.ipynb". The notebook content includes a section titled "Ex2 - Getting and Knowing your Data" with a note about pulling data from the internet. It also shows a code cell starting with "In [1]: import pandas as pd".

解題時間



Sample Code & 作業
開始解題



範例影片

應用程式 Gmail YouTube Python電子書 TensorFlow 閱讀清單

jupyter Day_005-1_build_dataframe_from_scratch Last Checkpoint: 3 小時前 (autosaved)

Logout Not Trusted Python 3

File Edit View Insert Cell Kernel Widgets Help

In [33]: `import pandas as pd`

方法一

In [20]: `data = {'weekday': ['Sun', 'Sun', 'Mon', 'Mon'], 'city': ['Austin', 'Dallas', 'Austin', 'Dallas'], 'visitor': [139, 237, 326, 456]}`

In [34]: `visitors_1 = pd.DataFrame(data)`
`print(visitors_1)`

0:00 / 19:39

下一步：完成作業

音量 全屏 ⋮

