

D6: EDA: 欄位的資料類型介紹及處理

[PDF 下載](#)[全螢幕](#)

Sample Code & 作業內容

A、在初步 EDA 的過程，我們無可避免會想問的問題
不同資料類型各有多少個欄位？
類別型欄位 (pandas 中的 object) 的類別數量？
模型怎麼處理類別型的資料？有什麼表示方法？

B、這三個問題在參考程式碼範例(請點選下方檢視範例Day_006_column_data_type.ipynb)都會實現，第三個問題會更為複雜一些，簡單來說我們有兩種方法來處理類別型資料

- Label encoding: 把每個類別 mapping 到某個整數，不會增加新欄位
- One Hot encoding: 為每個類別新增一個欄位，用 0/1 表示是否

Label Encoding			One Hot Encoding			
Food Name	Categorical #	Calories				
Apple	1	95	Apple	Chicken	Broccoli	Calories
Apple	1	95	1	0	0	95
Chicken	2	231	0	1	0	231
Broccoli	3	50	0	0	1	50

C、作業1：請詳細閱讀[label-encoder-vs-one-hot-encoder](#)（英文）

D、作業2：請點選檢視範例，作業請提交Day_006_HW.ipynb檔

[檢視範例](#)

參考資料



Pandas所支持的數據類型

1. float
2. int
3. bool
4. datetime64 [ns]
5. datetime64 [ns,tz]
6. timedelta [ns]
7. category
8. object

默認的數據類型是int64,float64.

提交作業

請將你的作業上傳至 [Github](#)，並貼上該網網址，完成作業提交

[確定提交](#)[如何提交](#) ▾

熱門問答

[雷](#) [康勝歲](#) · 2019.09.10 · 2 回答

無法正常執行 [sample code](#)

上網找了很多方法去解決，但是還是不行我有試著用 [conda remove scikit-learn](#) 去移除再用 [conda install scikit-learn](#) 安裝回來，仍然也是不行

[吳明宗](#) · 2019.09.03 · 2 回答

矩陣列印的規則

`app_train = pd.read_csv(f_app_train)for col in app_train: if app_train[col].dtype == 'object':....etc`想請問以上的code在讀取csv資料後1.for迴圈中app_train[col]的col是指定循序讀取cloumn的意思嗎???2.有循序...

[林祐祥](#) · 2019.09.02 · 3 回答

[Day6_from sklearn.preprocessing import OneHotEncoder](#)使用問題

問題一：此二方程式分別為參考範例以及參考資料所寫，但得到的答案不相同，我要如何修改參考資料中的程式來使得答案與參考範例方程得到的答案相同？參考範例：[參考範例得到的結果](#)：參考資料：[參考...](#)

[查看更多](#)

到 Cupoy 問答社區提問，讓教練群回答你的疑難雜症

[向專家提問](#)[如何提問](#) ▾