



D22：類別型特徵 - 基礎處理



PDF 下載

全螢幕

Sample Code & 作業內容

作業1：觀察範例Day_022_LabelEncoder_and_OneHotEncoder，在房價預測中調整標籤編碼(Label Encoder) / 獨熱編碼(One Hot Encoder) 方式，對於線性迴歸以及梯度提升樹兩種模型，何者影響比較大？

作業2：鐵達尼號例題中，標籤編碼 / 獨熱編碼又分別對預測結果有何影響？

作業請繳交：Day_022_HW.ipynb

檢視範例

參考資料

數據預處理：獨熱編碼（One-Hot Encoding）和 LabelEncoder標籤編碼

神馬文庫 [網頁連結](#)

其實 One Hot Encoding 與 Label Encoder 是類別型資料最常見的編碼方式，因此實現的程式碼也頗為常用，其中 One Hot Encoding 常見的兩種做法：pandas.get_dummies 與 sklearn 的 OneHotEncoder 在這網頁中都有清楚的展示，本課程今日範例中會用到前者，在之後的葉編碼中則會用到後者，所以同學不妨先了解一下寫法。

```
In [10]: s1 = ['a', 'b', np.nan]
pd.get_dummies(s1, dummy_na=True)

Out[10]:
   a  b  nan
0  1  0   0
1  0  1   0
2  0  0   1

In [12]: df2 = pd.DataFrame({'A':['a','b','a'],'B':['b','a','c'],'C':[1,2,3]})
pd.get_dummies(df2, prefix=['col1','col2'])

Out[12]:
   C  col1_a  col1_b  col2_a  col2_b  col2_c
0  1     1     1     0     0     1     0
1  2     0     1     1     0     0     0
2  3     1     0     0     0     0     1
```

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

<https://github.com/>

確定提交

[如何提交](#) ▼

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](#) ▼