

## D19：數值型特徵-補缺失值與標準化



PDF 下載

全螢幕

### Sample Code & 作業內容

請參考範例程式碼 Day\_019\_Fill\_NaN\_and\_Scalers.ipynb: 房價預測

執行作業範例Day\_019\_HW.ipynb：鐵達尼生存預測

作業1：試著在補空值區塊, 替換並執行兩種以上填補的缺值, 看看何者比較好?

作業2：使用不同的標準化方式 ( 原值 / 最小最大化 / 標準化 )，搭配羅吉斯迴歸模型，何者效果最好?

作業請提交Day\_019\_HW.ipynb

檢視範例

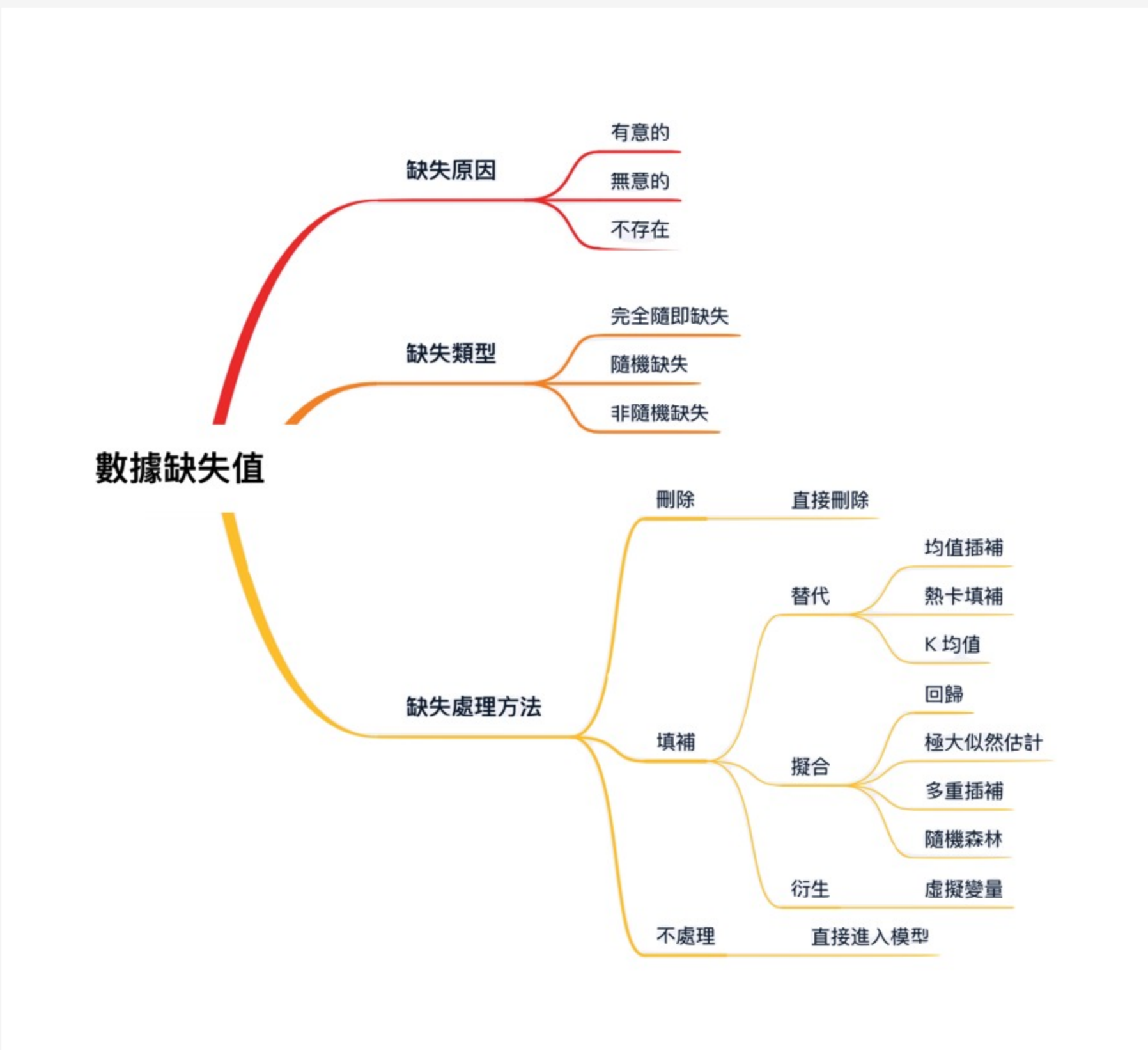
### 參考資料

課後補充:

### 掘金：Python數據分析基礎：數據缺失值處理

網頁連結

這篇文章更詳細地介紹了各種缺失值的種類，以及處理的各種方式優缺點，如果要徹底搞懂缺失值的話，這是一份不錯的補充資料。但是我們還是要強調：補缺是因資料而異，所以熟悉方法與觀察資料本身都是同樣重要的，因此這在實務上會是一個辛苦的環境。



### 數據標準化 / 歸一化normalization

網頁連結

本文重點如下圖，介紹了標準化 / 最大最小化以外的一些特徵縮放方式，雖然文中也提到這兩種就是最常見的方式了，但是其他幾種方式也是不錯的參考，提供同學查閱。

#### 常見的數據歸一化方法

最常用的是 min-max 標準化 和 z-score 標準化。

**min-max 標準化(Min-max normalization)/0-1 標準化(0-1 normalization)/線性函數歸一化/離差標準化**

### 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](#) ▼