

D22：特徵工程簡介



簡報閱讀



範例與作業



問題討論

特徵工程簡介

知識地圖

本日知識點目標

什麼是特徵工程?

特徵工程舉例

前述流程 / python程式
對照(範例：房價預測)

重要知識點複習

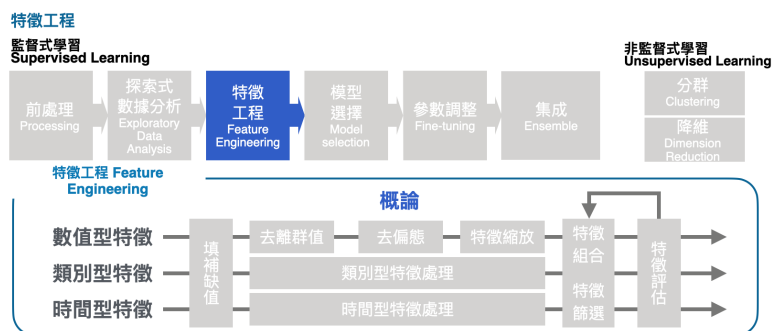
延伸閱讀

推薦延伸閱讀

特徵工程簡介



知識地圖



本日知識點目標

今日知識點目標

- 初步理解特徵工程的概念
- 能從程式中辨識特徵工程的區塊與意義
- 知道特徵工程至少需要那些部分

什麼是特徵工程?

通常我們會這樣考慮事情...

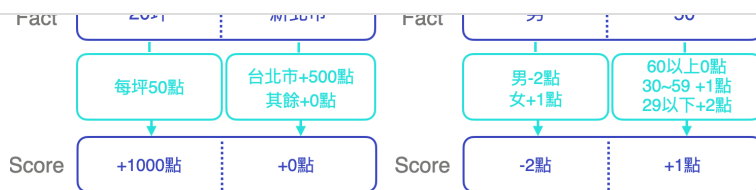
| | 重要性 | 緊急度 | 所需時間 | |
|-----------|-----|-----|------|-------|
| Question1 | 5★ | 1★ | 5★ | → 11★ |
| Question2 | 2★ | 5★ | 1★ | → 8★ |
| Question3 | 3★ | 4★ | 3★ | → 10★ |

但其實考慮的關鍵是...

| | 重要性 | 緊急度 | 所需時間 | |
|-------|------|------|-------|------|
| Fact | 非常重要 | 3週以內 | 約36小時 | |
| | 如何轉換 | | | |
| Score | ?★ | ?★ | ?★ | → ?★ |

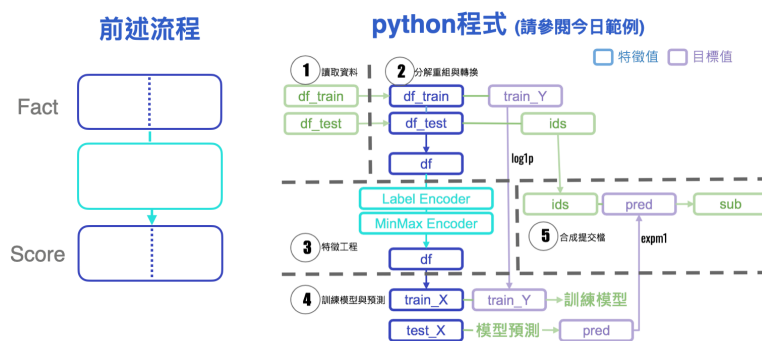
從事實到對應分數的轉換，我們稱為特徵工程

特徵工程舉例



點數未必直接對應到總價或機率，但會是後續評估的依據

前述流程 / python程式 對照(範例：房價預測)



重要知識點複習

- 特徵工程是事實對應到後續評估分數的轉換
- 在程式語法中，特徵工程位於資料彙整之後，以及擬合模型之前
- 由於資料包含類別型(文字型)特徵以及數值型特徵，所以最小的特徵工程至少要包含一種類別編碼(範例使用標籤編碼)，以及一種特徵縮放方法(範例使用最小最大化)

延伸閱讀

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

知乎-特徵工程到底是什麼

特徵工程到底是什麼？

大概知道一些，但是不是很清楚領域內的知識是怎麼幫助特徵工程的呢？

www.zhihu.com

本文重點主要是希望同學大致知道特徵工程大致包含哪些部分，若對細節有興趣，還可以從這篇中了解一些概念其中一部分的內容，會在後面的課程中說明並練習，詳情請參閱百日馬拉松課綱

痞客邦-IT邦2019鐵人賽：為什麼特徵工程很重要

為什麼特徵工程很重要 - iT

數據科學家和機器學習工程師經常收集數據以解決問題。因為他們試圖解

ithelp.ithome.com.tw

本文主要在描述現實中資料科學工作的時間比重(右圖)，其中大部分的時間在於資料清理，少部分為特徵探勘，雖然這兩部份都是特徵工程，但在學習階段與實務階段，比重卻有著天壤之別



Sample Code & 作業

開始解題



[下一步：閱讀範例與完成作業](#)