

Day 18

特徵工程

特徵類型



常見特徵類型：數值型特徵 / 類別型特徵

常見特徵有兩大類 (舉Day017的範例)



- 數值型特徵：有不同轉換方式，函數 / 條件式都可以
- 類別型特徵：通常一種類別對應一種分數

其他特徵類型 (1 / 2)

二元特徵

True

False

- 只有 True / False 兩種數值的特徵
- 可以當作類別型，也可當作數值型特徵 (True:1 / False: 0)

排序型特徵

0

1

.....

99

- 例如名次 / 百分等級，有大小關係，但並非連續數字
- 通常當作數值型特徵處理，因為當作類別型會失去排序資訊

其他特徵類型 (2 / 2)

時間型特徵

2018/12/15 09:00:00

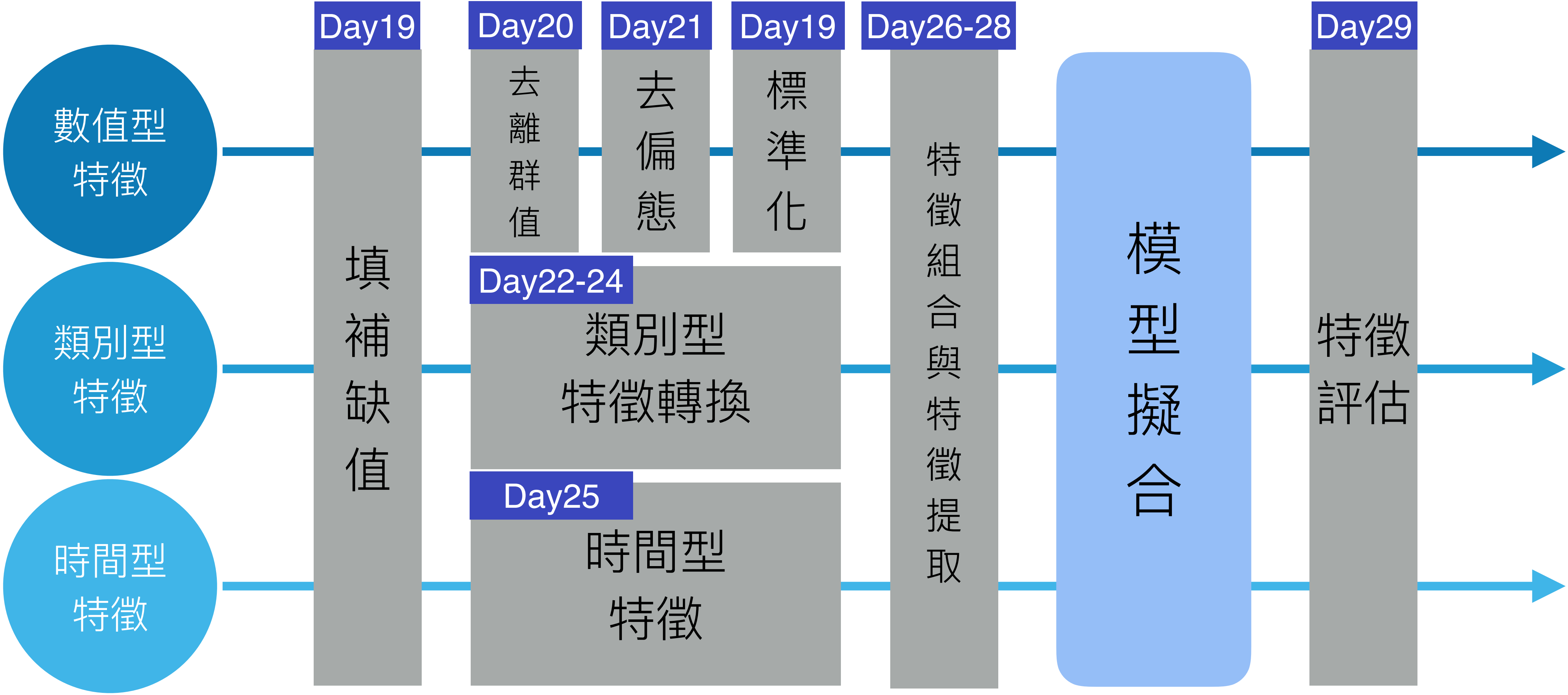
- 雖然時間型特徵可當作數值型特徵或類別型特徵，但都不適合
 - 取總秒數雖可變為數值，但會失去週期性 (ex 月 / 星期)
 - 使用本身可以當作類別，但會失去排序資訊，類別數量也過大
- 因此時間型特徵我們會個別於 Day 25 的課程中講解

補充說明

因為程式講解需要，會以 `cross_val_score` 顯示改善效果：分數越高表示效果越好，但不會在現階段講解這部分原理，有興趣提前了解的同學請研讀延伸閱讀內容：[k- fold cross validation](#)

特徵工程 - 學習地圖

接下來會以此圖為主，逐步為各位解說知識點



解題時間 It's Your Turn

請跳出PDF至官網Sample Code & 作業
開始解題

