Day 25 特徵工程

時間型特徵

出題教練 陳明佑

知識地圖 特徵工程 時間型特徵

CUPOY

機器學習概論 Introduction of Machine Learning

監督式學習
Supervised Learning

前處理 Processing → 探索式數據分析 Exploratory Data Analysis → **特徵工程 Feature Engineering** → 模型選擇 Model selection → 參數調整 Fine-tuning → 集成 Ensemble

非監督式學習
Unsupervised Learning

- 分群 Clustering
- 降維 Dimension Reduction

特徵工程 Feature Engineering

概論

- **數值型特徵**
- **類別型特徵**
- **時間型特徵**

→ 填補缺值 →
- 去離群值 → 去偏態 → 特徵縮放
- 類別型特徵處理
- 時間型特徵處理

→ 特徵組合 / 特徵篩選 → 特徵評估

本日知識點目標

- 時間型特徵最常用的編碼方式
- 如何將時間最重要的特性 「循環」 改成特徵
- 時間常見的週期循環特徵有哪些，有什麼要注意的地方?

時間特徵分解 (1 / 2)

CUPOY

最常見的特殊欄位是時間欄位，想想看應該怎樣編碼?

時間戳記

2014-06-12 03:25:56

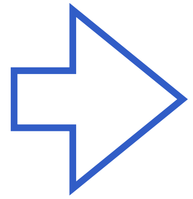
2011-07-16 01:19:59

2011-10-21 23:54:10

2015-02-03 10:42:03

2009-06-13 16:10:54

2010-08-09 14:46:03



?

時間特徵分解 (2 / 2)

CUPOY

最直覺的方式，就是依照原意義分欄處理，或加上第幾周或星期幾
但某些欄(例：分、秒)與目標值的關係很低，有沒有更有意義的特徵呢?

時間戳記	年	月	日	時	分	秒
2014-06-12 03:25:56	2014	6	12	3	25	56
2011-07-16 01:19:59	2011	7	16	1	19	59
2011-10-21 23:54:10	2011	10	21	23	54	10
2015-02-03 10:42:03	2015	2	3	10	42	3
2009-06-13 16:10:54	2009	6	13	16	10	54
2010-08-09 14:46:03	2010	8	9	14	46	3

週期循環特徵 (1 / 2)

時間也有週期的概念, 可以用週期合成一些重要的特徵
聯想看看：有哪幾種時間週期, 可串聯到一些可做特徵的性質?

- 年週期 與春夏秋冬季節溫度相關
- 月週期 與薪水、繳費相關
- 周週期 與周休、消費習慣相關
- 日週期 與生理時鐘相關

週期循環特徵 (2 / 2)

前述的週期所需數值都可由時間欄位組成, 但還首尾相接

因此週期特徵還需以正弦函數(sin)或餘弦函數(cos)加以組合

例如：年週期 (正：冷 / 負：熱)

cos((月/6 + 日/180)π)

周週期 (正：精神飽滿 / 負：疲倦)

sin((星期幾/3.5 + 小時/84)π)

日週期 (正：精神飽滿 / 負：疲倦)

sin((小時/12 + 分/720 + 秒/43200)π)

*註：此處小時是24小時制

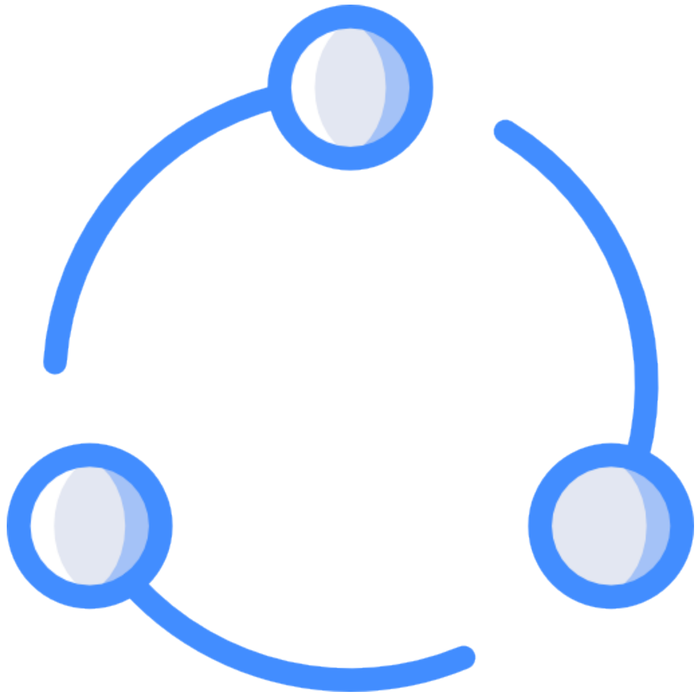
時段特徵

CUPOY

短暫時段內的事件計數，也可能影響事件發生的機率
如：網站銷售預測，點擊網站前 10分鐘 / 1小時 / 1天 的累計點擊量

以一筆 17:05 發生的網站瀏覽事件為例
同樣是1小時的統計，基礎分解會統計當日 17 時整個小時的點擊量
時段特徵則是會統計 16:05-17:04 的點擊量
兩者相比，後者較前者更為合理

重要知識點複習



- 時間型特徵最常用的是**特徵分解** - 拆解成年/月/日/時/分/秒的分類值
- **週期循環特徵**是將時間"循環"特性改成特徵方式, 設計關鍵在於**首尾相接**, 因此我們需要使用 sin /cos 等週期函數轉換
- 常見的週期循環特徵有 - 年週期(季節) / 周周期(例假日) / 日週期(日夜與生活作息), 要注意的是**最高與最低點的設置**

解題時間
It’s Your Turn

請跳出PDF至官網Sample Code＆作業
開始解題