



D26：特徵組合 - 數值與數值組合



PDF 下載

全螢幕



Sample Code & 作業內容

請參閱範例Day_026_Feature_Combination.ipynb中的計程車費率預估

作業1：參考今日教材，試著使用經緯度一圈的長度比這一概念，組合出一個新特徵，再觀察原特徵加上新特徵是否提升了正確率？

作業2：試著只使用新特徵估計目標值(忽略原特徵)，效果跟作業1的結果比較起來效果如何？

作業請提交Day_026_HW.ipynb

檢視範例

參考資料

特徵組合&特徵交叉 (Feature Crosses)

SegmentFalut [網頁連結](#)

這裡有一些延伸的特徵組合方式，例如講義中提到的運算組合方式，或者使用綜合特徵的離散化/分箱，或者將兩種獨熱編碼綜合...等有趣的特徵組合方式。這些方式提供同學在合成特徵時參考，但這些只是方法，合成特徵比較有效的方式還是參考領域知識(如果有的話)。

目錄

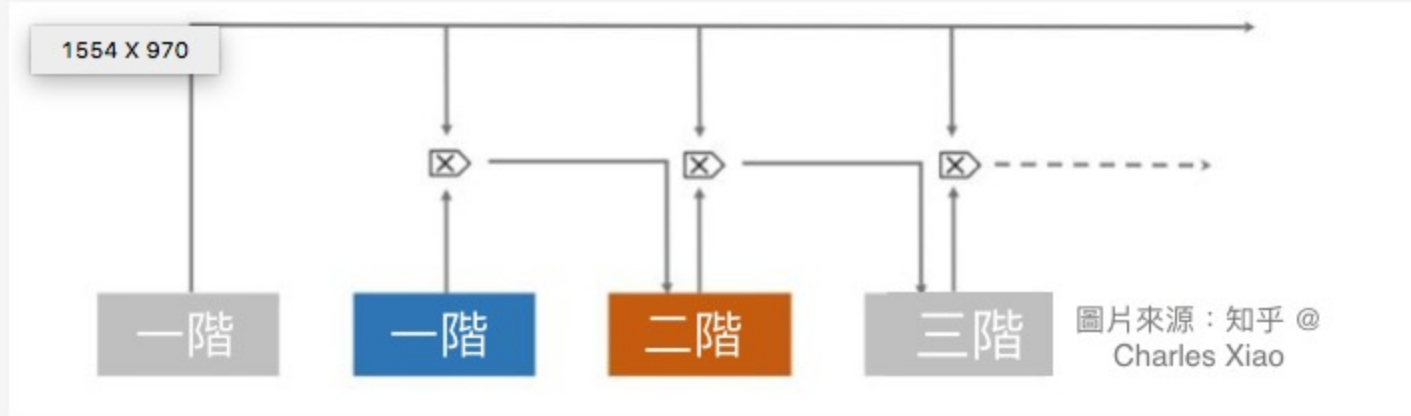
- [合成特徵\(synthetic Feature\)](#)
- 特徵組合(feature cross)：對非線性規律進行編碼
- 特徵組合的種類
- 特徵組合(Feature Crosses)：組合獨熱向量
- 代碼部分練習 學習目標：
- 使用分桶特徵列訓練模型
- 特徵組合
- 使用特徵組合訓練模型

簡單高效的組合特徵自動挖掘框架

知乎 [網頁連結](#)

這裡討論的「自動學習交叉特徵」，講的是當資料中類別種類多的類別特徵，這類問題我們一般的做法是使用LR(邏輯斯回歸) 或者FM(分解機器)...而本文則是討論FM進階的FFM。

不過這是在特徵都是屬於種類多的類別特徵時，才只好做的處理方式，如果這類特徵很少，或者在競賽中跳過這類特徵就可以有不錯準確度了，建議可以跳過這部分內容。



提交作業

請將你的作業上傳至 Github，並貼上該網址，完成作業提交

<https://github.com/>

確定提交

[如何提交](#) ▾

熱門問答

● Edwin · 2019.05.14

D26 Linear regression: cross validation score

經緯度比使用 $1 - \text{np.cos}((\text{d}[\text{'dropoff_latitude'}] + \text{d}[\text{'pickup_latitude'}])/2 * 2 * \text{np.pi}/360)$ 而不是 $2 - \text{np.cos}(40.75 * 2 * \text{np.pi}/360) = 0.75756$ 為什麼 Linear Regression 的 cross validation score 會從 0.025 提升到 0.48 從...

● 曾立嚴 · 2019.05.14 · 1 回答

D26 新增特徵組合後，線性迴歸 / 梯度提升樹比率變化問題

如D26作業，使用緯度差、經度差、座標距離特徵計算線性迴歸 / 梯度提升樹結果為：
Linear Reg Score : 0.02747902580681085 Gradient Boosting Reg Score : 0.8059786997870237 Linear...

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](#) ▾