

D30：分類型特徵優化 - 葉編碼



PDF 下載

全螢幕

Sample Code & 作業內容

請參考檢視範例Day_030_Leaf_Encoding.ipynb

作業1：請對照範例，完成隨機森林的鐵達尼生存率預測，以及對應的葉編碼+邏輯斯迴歸

作業2：上述的結果，葉編碼是否有提高預測的正確性呢？

作業請提交Day_030_HW.ipynb

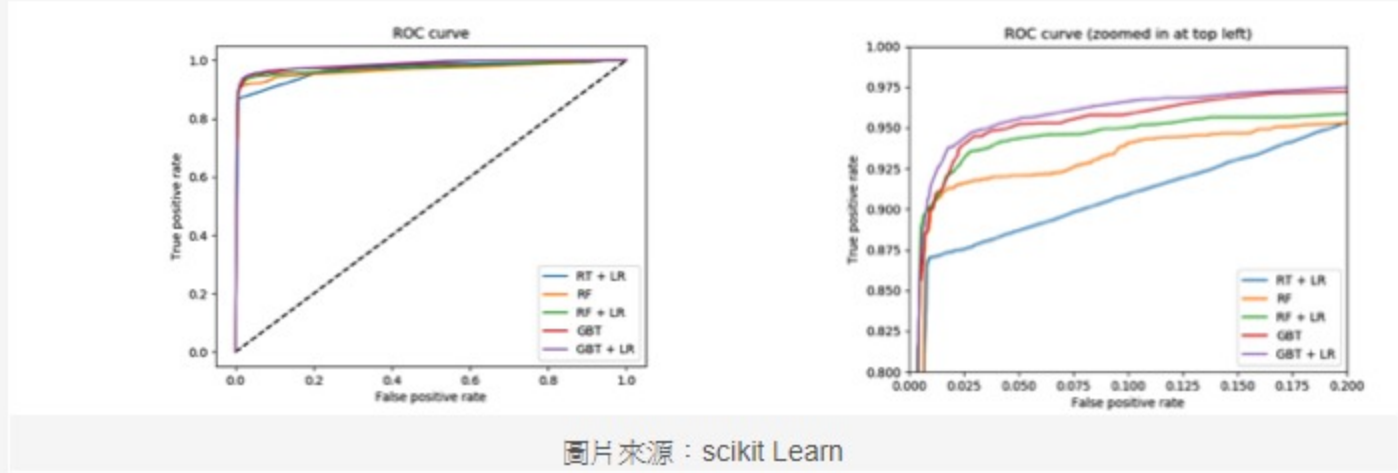
檢視範例

參考資料

Feature transformations with ensembles of trees

sklearn官網範例 [網頁連結](#)

這是講義內線面這張圖的出處，裡面有完整的葉編碼程式，因為裡面的一些細節需要多做解釋，因此我們加以註解改寫後放作為今天的範例當中，同學若是有興趣也可以直接參考這篇原文，裡面有一些原文解說。

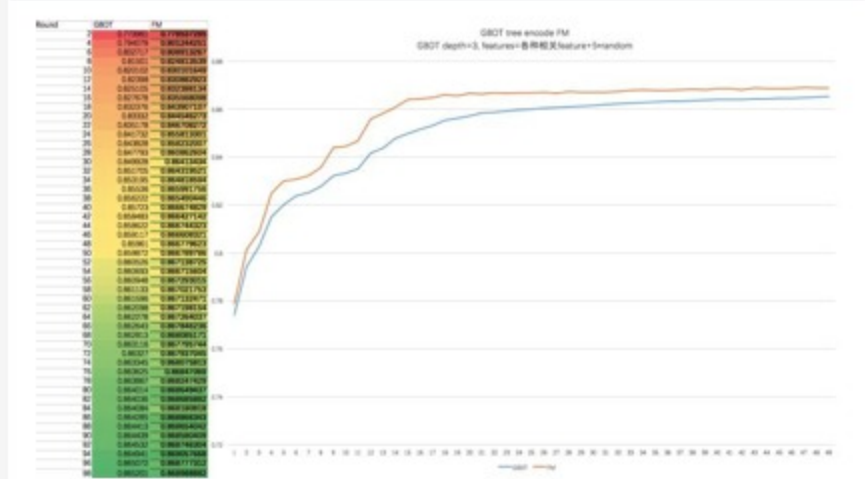


圖片來源：scikit Learn

CTR 預估[十一]：Algorithm-GBDT Encoder

知乎 算法那些事兒 [網頁連結](#)

這個網頁將葉編碼的應用，做了很詳盡的說明：包含使用的原因，包含的原理，以及葉編碼的戰史，如果很想弄清楚葉編碼，一定要看看這篇文章。



三分鐘了解推薦系統中的分解機方法 (Factorization Machine, FM)

每日頭條 [網頁連結](#)

最後是有關分解機的解說，因為這部分稍微有點複雜，需要先了解矩陣分解 (Matrix Factorization) 與推薦系統，如果對FM沒有興趣，可以跳過此連結，但很好奇FM到底是什麼的同學，可以由此入門。

Feature vector x															Target y								
x_1	1	0	0	...	1	0	0	...	0.3	0.3	0.3	0	...	13	0	0	0	...	5	y_1			
x_2	1	0	0	...	0	1	0	...	0.3	0.3	0.3	0	...	14	1	0	0	...	3	y_2			
x_3	1	0	0	...	0	0	1	...	0.3	0.3	0.3	0	...	16	0	1	0	...	1	y_3			
x_4	0	1	0	...	0	0	1	...	0	0	0.5	0.5	...	5	0	0	0	...	4	y_4			
x_5	0	1	0	...	0	0	0	...	0	0	0.5	0.5	...	8	0	0	1	...	5	y_5			
x_6	0	0	1	...	1	0	0	...	0.5	0	0.5	0	...	9	0	0	0	...	1	y_6			
x_7	0	0	1	...	0	0	1	...	0.5	0	0.5	0	...	12	1	0	0	...	5	y_7			
A	B	C	TV	NAH	SW	ST	...	T1	NAH	SW	ST	...	T1	NAH	SW	ST		
User				Movie				Other Movies rated				Last Movie rated											

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