100 道題目 > D23:類別型特徵-均值編碼

#### D23: 類別型特徵 - 均值編碼



## Sample Code & 作業内容

作業1:請仿照範例Day\_023\_Mean\_Encoder.ipynb,將鐵達尼範例中的類別型特徵改用均值編碼實作一次。作業2:觀察鐵達尼生存預測中,均值編碼與標籤編碼兩者比較,哪一個效果比較好?可能的原因是什麼?作業請提交Day\_023\_HW.ipynb

Q 檢視範例

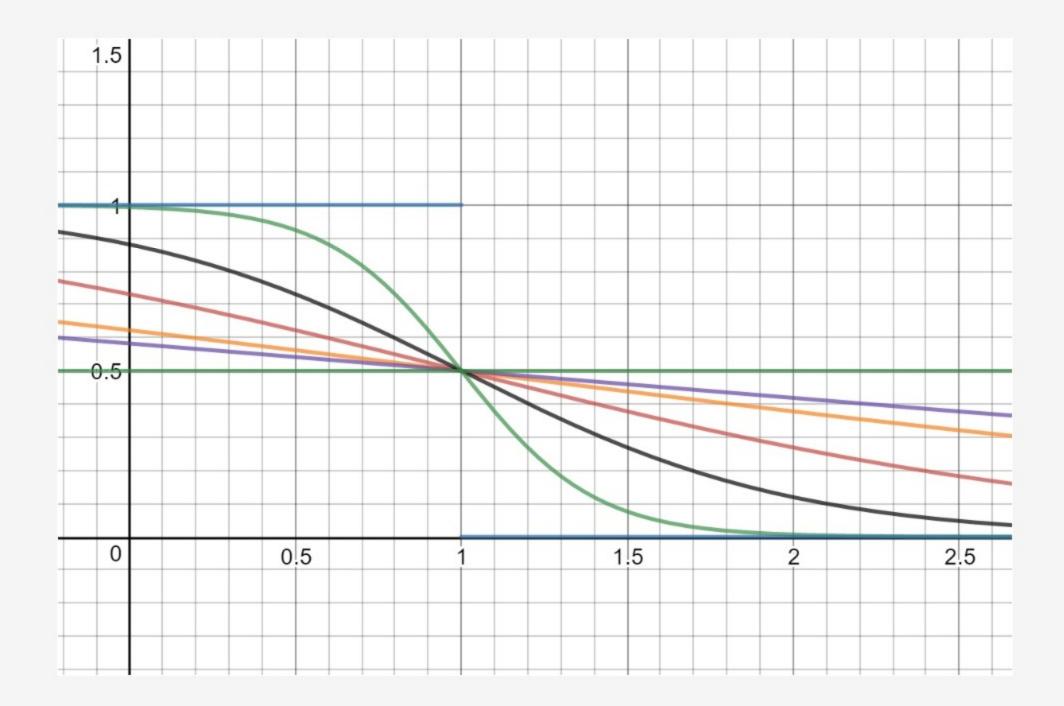
# 參考資料

### 平均數編碼: 針對高基數定性特徵(類別特徵)的數據處理/特徵工程

知乎 網頁連結

就實務上而言,均值編碼的意義在於當一個特徵有明顯意義,但是類別數量特別多(這裡說的"高基數")時可能有用,但最麻煩的點在於極度容易OverFitting,所以需要不同的平滑化方式。

在課程內使用平均因子的方法只是其一,這邊的內容也介紹了另一種較複雜的平滑化方式,提供同學參考。



# 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