

學號：R07943150 系級：電子所一 姓名：吳辰鉉

1. 請比較你實作的 generative model、logistic regression 的準確率，何者較佳？

	Public	Private
Generative model	0.84092	0.84729
Logistic regression	0.86107	0.86167

不論在 public set 或是 private set , logistic 的準確率都好很多

2. 請說明你實作的 best model，其訓練方式和準確率為何？

利用 sklearn 的 GradientBoostingClassifier 套件並參考尋找適當參數的方法

在 public set 有 0.87323 在 private set 有 0.87592 準確率

3. 請實作輸入特徵標準化(feature normalization)並討論其對於你的模型準確率的影響
若沒有標準化則

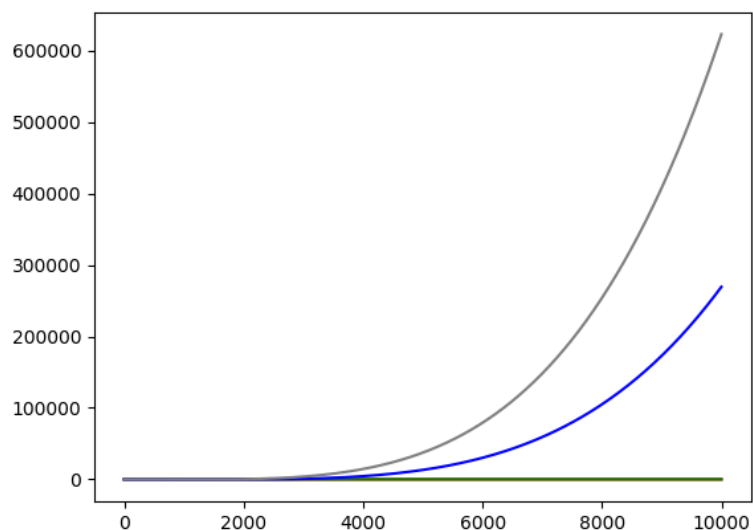
GradientBoostingClassifier 準確率分別為 0.87237 0.87469

Generative model 準確率分別為 0.84105 0.84631

Logistic model 準確率分別為 0.82667 0.83206

可見標準化對於 Logistic model 的影響最為顯著

4. 請實作 logistic regression 的正規化(regularization)，並討論其對於你的模型準確率的影響。



'green','blue','gray' 三條線分別對應 $\lambda = 0.001$ 0.003 0.005

若套用正規化則預測準確率較為不佳

5. 請討論你認為哪個 **attribute** 對結果影響最大？

有實際把訓練後每項特徵的權重拿出來比大小，其中最為顯著的是 **capital_gain**，這答案也合乎常理，畢竟通常是有一定多餘的錢或是能力才有辦法投資