

學號：R06922116 系級：資工碩一 姓名：賴柏恩

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

答： **Generative model**：準確率 84.58% (public) 84.24% (private)

Logistic model：準確率 79.92% (public) 79.52% (private)

在 data 不做任何處理下直接實作兩種 model，很明顯地看出 generative model 的表現較佳

2.請說明你實作的 **best model**，其訓練方式和準確率為何？

答：Best model 是使用 logistic regression，差別是在我將家庭的 attribute 去掉，因為我覺得應該其相關性較低，得到在 public set 85.57%的準確率，在 private set 上 84.90%的準確率

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。

答： **Generative model**：準確率 84.53% (public) 84.22% (private)

Logistic model：準確率 85.39% (public) 85.12% (private)

很明顯地看出 logistic model 在 data 經過 normalization 過後，表現得到很大的改善，一舉得超過了 generative model 的準確率，但在 generative model 上，data normalization 對於準確率並沒有什麼很明顯的改善。

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。

答： **with lambda = 0.1**

準確率: 0.83931 (public) 0.83319 (private)

with lambda = 0.01

準確率: 0.85393 (public) 0.85026 (private)

進行了正規化後，並沒有對於我的 model 有產生改善，可能是因為本身就沒有過度訓練，因此看不出有什麼效果

5.請討論你認為哪個 **attribute** 對結果影響最大？

我認為是 capital gain 對於結果影響最大，因為透過常理推斷，一個人投資淨收入越高，應該代表著其具有投資的本錢，因此收入應該不會太低。