

學號：R06942077 系級：電信碩一 姓名：洪健鈞

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

答：

將所有的 **feature** 都納入考量的情況下，且加入 **normalization**，發現 **logistic regression** 的準確率有 **0.84975**，而 **generative model** 的準確率卻只有 **0.84557**，相比之下，**logistic regression** 的準確率比較高。

2.請說明你實作的 **best model**，其訓練方式和準確率為何？

答：

我使用 **logistic regression** 來當作我的 **model**，並有挑選 **feature**，把 **hours per week** 與 **native_country** 從我的 **model** 裡去除，其餘 **feature** 皆使用，並加入二次方項。另外我還使用了一個方法，是 **random forest** 裡其中一個類似投票多數決的方法，因為這次的學習有一點 **random** 的誤差，所以這個方法我覺得有效，我將 **model** 跑 20 遍，取較多數 **model** 預測出來的答案為我預測的結果。準確率如下：
Public Score : **0.85503**
Private Score : **0.85100**

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：

在考慮所有 **feature** 的情況下，使用 **logistic regression**，尚未特徵標準化的 **model** 準確率僅 **0.82936**，在特徵標準化之後準確率提升到了 **0.84975**，有明顯的提升，所以特徵標準化對我的模型準確率是非常有幫助的。

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：

在考慮所有 **feature** 的情形下，且僅考慮一次項作為模型，並使用 **normalization**，未做正規化的模型準確率有 **0.84975**，而再做 **regularization** 之後卻只提升至 **0.85036**，發現 **regularization** 作用不大，我覺得可能是因為這次訓練次數不多，模

型沒有 **overfitting** 的情形, 所以 **regularization** 對於我的模型準確率沒特別大的影響。

5.請討論你認為哪個 **attribute** 對結果影響最大？

我認為是職業, 不同職業的起薪本來就不同, 需要更多專業的職業, 在就業市場上肯定是特別搶手, 就算隨著年資上升, 各行各業的薪水成長幅度應該也相差不大, 所以我覺得職業是對薪水最有影響力的 **attribute**。