# Homework 2 Report - Income Prediction

學號: R06942018, 姓名: 何適楷, 系級: 電信碩一

## 1 Problem 1

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

| | training | public | private |
|---|---|---|---|
| generative model | 0.8421 | 0.84533 | 0.84203 |
| logistic model | 0.8532 | 0.85712 | 0.84793 |

logistic 皆較佳，這也合理，因為 logistic model 的 function set 比較廣，generative model 則只考慮 normal distribution。

## 2 Problem 2

(1%) 請說明你實作的 best model，其訓練方式和準確率為何？
我最好的 model 是 logistic model，不過我另外有實做一個簡易的 Neural Network，形狀是 (6, 1)，training accuracy 是 0.86508，得到最好的 public score 是 0.85663，private score 是，感覺是 overfitting，但是加了一點 regularizationo($1 \times 10^{-5}$)，結果變差了 (0.85552)。也有做過 (12,1) 的網絡，training accuracy 是 0.87000，得到最好的 public score 是 0.85049，private score 是，感覺是 overfitting，但是加了一點 regularizationo($1 \times 10^{-6}$)，結果變好一點 (public=0.85479)，但是還不足以過 strong baseline，所以又加了一點 ($1 \times 10^{-5}$)，但是結果變差了 (public=0.85085)。另外還有試了很多 mdoel，像是 (2,1), (3,1), ... (13,1) ,(4,4,1), (6,6,1) (12,12,1) (4,2,1) (4,3,1) (4,4,4) (6,6,6,6,6,6) 等都沒過，常常都是 overfitting 而且加了 regularization 和 dropout 沒有好很多，而且跑出來的結果也具有隨機性，很難確定最好的 function set。再 late summistion 中，有一個有超過 strong baseline，不過用同樣 model 重新跑去測結果沒有比較好。以下列出部分試的結果。

| | public | private |
|---|---|---|
| (5, 1) | 0.85038 | 0.85491 |
| (6, 1) | 0.85235 | 0.85626 |
| (5, 1) | 0.85835 | 0.85493 |
| (4, 1) | 0.85577 | 0.85272 |
| (3, 1) | 0.85321 | 0.85614 |

# 3 Problem 3

(1%) 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。以 logistic model 為例，影響最大的是收斂的次數。

| | iteraion | training | public | private |
|---|---|---|---|---|
| normalized data | 42000 | 0.8532 | 0.85712 | 0.84793 |
| raw data | 超過 200 萬次 | 0.8532 | 0.85712 | 0.84842 |

Learning Rate 皆為 $4 \times 10^{-6}$，方法皆為 RMSProp，可以看出 normalized model 的疊代次數遠小於沒有 normalized 的，因為沒有 normalized 的話，參數會傾向於改變 range 比較大的參數，小的都沒改到，所以收斂會很慢，raw data 的部分是有手動一直追佳 iteration 的次數，跑了一整個晚上，所以沒有記錄到實際次數，不過可以證實，raw data 真的會收斂!

# 4 Problem 4

(1%) 請討論你認為哪個 attribute 對結果影響最大？
我覺得 native country 影響最大，他的係數範圍為 1.38 到-2.82，相較於其他的 feature 只有零點幾。係數大代表對於整體值的貢獻大，影響自然就大。