學號：R06921084 系級： 電機碩一 姓名：陳治言

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

答：

在考慮所有的 feature 且不使用 normalization 的情況之下，我們可以發現 logistic regression 的準確率有 0.79153，但是 generative model 卻只會有 0.74936 的準確率，相較之下，logistic regression 感覺較佳。

2.請說明你實作的 best model，其訓練方式和準確率為何？

答：

使用 logistic regression 來實作，除了 hours_per_week 和 native_country 不考慮之外，其餘 feature 皆使用，先將數值化進行 normalization，並將 age 使用 2 次方與 3 次方，fnlwgt 使用 2 次方來進行訓練，最後的準確率如下：

- Public Score → 0.85798 (已達 Strong Baseline)
- Private Score → 0.85321 (已達 Strong Baseline)

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：

在考慮部分 feature 的情況下，且使用 logistic regression 來說，一開始未特徵標準化時，準確率只有 0.83243。使用特徵標準化之後，準確率提升至 0.84347，可以發現到 feature normalization 是非常有用的。

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：

在考慮部分 feature 的情況下，且使用 feature normalization 來說，一開始未正規化時，準確率只有 0.83243。使用特徵標準化，並將 λ 設為 0.1 與 1 來做比較，準確率分別為 0.83305 與 0.83277，可以發現到 regularization 作用並不大，有可能是因為訓練次數不多，推估模型上並沒有發生 overfitting。

5.請討論你認為哪個 attribute 對結果影響最大？

我認為是年齡與性別，主要是因為年齡較長者，其事業成就可能會較佳，自然而然，其收入可能會較高。性別是因為男女主要從事的職業也大不相同，加上我認為職場對於性別也有歧視，間接導致收入的分布。