學號：R06922152 系級： 資工碩一 姓名：袁晟峻

1.請比較你實作的 generative model、logistic regression 的準確率，何者較佳？
答：
Generative model：kaggle public 0.84533、kaggle private 0.84227
Logistic model：kaggle public 0.84680、kaggle private 0.84485
Logistic 的準確率比 Generative 來的好
2.請說明你實作的 best model，其訓練方式和準確率為何？
答：
用 keras 的 sequential model 實作三層架構的 NN，第一層 106 個 neuron，activation 用 relu，第二層 64 個 neuron，activation 用 relu，第三層 1 個 neuron，activation 用 sigmoid，optimizer 用 adagrad，第一層到第二層中間 Dropout(0.5)，第二層到第三層時也 Dropout(0.5)，用這樣的 model 去做訓練，在 kaggle 上 public 分數為 0.85823、private 分數為 0.85284

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 generative model 的部分，特徵標準化對準確率的影響感覺不大，我測得的 training accuracy 都在 0.84 多
在 logistic model 的部分，若是不做特徵標準化，在算 sigmoid 使用 numpy.exp 的時候會 overflow，導致 model 訓練不起來，做完標準化之後訓練完的 model 可達到 0.84 多的準確率
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 logistic model 做了正規化之後，kaggle 上的 public 和 private 分數皆從 0.84 掉到了 0.83，可能原本的 model 並沒有很 overfit

5.請討論你認為哪個 attribute 對結果影響最大？
答：
根據自己猜測以及蒐集到的資訊，再加上測試的結果，種族是白人加上年紀比較大這些特徵容易是年薪>50K 的族群裡擁有的，所以我認為 Race 和 Age 對結果的影響最大