學號：R05942017　系級：　電信碩二　姓名：黃梓鳴

1.請比較你實作的 generative model、logistic regression 的準確率，何者較佳？
答：
generative model 在 kaggle(public)上的準確率為 0.84533

logistic regression 的為 0.79213(只有 iteration 10 次，但有試過 1000 次，準確率增加不到一%) 明顯為 generative model 較高

2.請說明你實作的 best model，其訓練方式和準確率為何？
答：
在 kaggle(public)準確率為 0.86093，有用到 keras 和 tensorflow，有另外在 rawdata 發現 education_num 這個沒有整理在 X_train 的 feature，多這個 feature 讓我在 train 時的準確率高了 1%

Train 方法:先把資料做 **feature normalization(只有對連續的資料)**，丟到 5 層的 DNN 裡，在中間三層做 BatchNormalization、在每層間做 dropout、第一層用 RELU，二到四層用 LeakyReLU(alpha = 0.25)，第五層用 sigmoid。

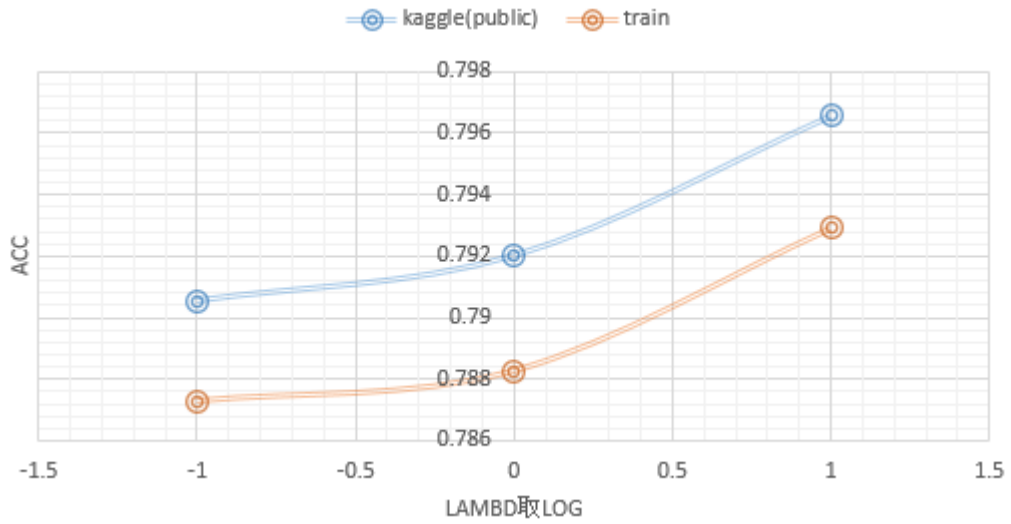
寬度分別是 107→32→50→50→50→1，optimizer='rmsprop'，
loss='binary_crossentropy'， metrics=['accuracy']

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：
對 logistic 作兩種 **normalization**，一種是只有對前六種(連續資料)，後一種是對所有 feature，對 train data 的 acc 分別為 81.8 和 80.7，由於前一種較高，把前一種丟到 kaggle(public)可得 acc=0.82063，而沒有 normalization 的為 0.79213，可見 normalization 對 logistic 實作可增加 acc

對 generative 也作同樣兩種 normalization 前一種的 acc 為 0.76475，後一種的為 0.78895，後一種較高，故也把它傳到 kaggle(public)得 acc=0.77051，明顯較沒有作 normalization 的低(0.84533) ，故得 normalization 會降低 generative model 的準確性

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：
分別以 lambda=0.1,1,10 帶入去訓練，得以下這張圖，沒有作 regularization 的 train data 的 Acc 為 0.78726，傳到 kaggle 上得到的 acc 為 0.79201，發現在 lambd=0.1 時 train data 的 acc 雖然上升一點，傳到 kaggle 得到的 acc 卻是下降

的，而 lambd=1,10 都是 train datac 和 kaggle 上的 acc 皆上升，跟老師的 PPT 有點不相符，但至少從這圖中可知，若 lambd 取得適當，是可以增加 acc 的



5.請討論你認為哪個 attribute 對結果影響最大？

我把每個 feature 都輪流刪掉一次(106 種變成 105 種 feature) 在用 generative model 下去跑，發現當沒有用 capital_gain 這個 feature 時，train data 的 acc 最低，故應該是 capital_gain 這個 feature 對結果影響最大