學號：B06901063 系級：電機二 姓名：黃士豪

1. 請比較你實作的 generative model、logistic regression 的準確率，何者較佳？
   (public / private)
   generative model: 0.84189 / 0.83847
   logistic regression: 0.85786 / 0.85714
   logistic regression 的準確率較佳，因為 generative model 是理論值，且將所有資料視為常態分佈，常發生 underfit 的現象。
2. 請說明你實作的 best model，其訓練方式和準確率為何？
   我的 best model 是用 keras 疊了 3 層 activation function 為 sigmoid 的 DNN 模型，並在每兩層中增加 0.15 的 dropout，取 0.15 的 validation split 得出。其中原始的 feature 資料我有先進行 feature normalization，增加準確率，最後的準確率為 (public/ private) 0.86093 / 0.85689，雖然不是到非常好但是已經比一般的 logistic regression 結果好。
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
   (public / private)
   有加標準化：0.85245 / 0.85321
   沒加標準化：0.78918 / 0.78356
   很明顯的，有加標準化出來的準確率高相當多，推測是因為標準化能讓每個 feature 的 weight 更加平均，且避免 train data 中有極端值的影響。
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。
   (public / private)
   有加正規化：0.85245 / 0.85333 (lumbda = 1)
   沒加正規化：0.85245 / 0.85321
   有加正規化的數據因為叫為平滑，因此有略為改善準確率，但著實有限。

5. 請討論你認為哪個 attribute 對結果影響最大？

去除 attribute	分數(public / private)
age	0.85147 / 0.85100
workclass	0.85221 / 0.85308
fnlwgt	0.85307 / 0.85087
education	0.85221 / 0.85296
Education num	0.85171 / 0.85173
marital status	0.85221 / 0.85308
occupation	0.85233 / 0.85321
relationship	0.85221 / 0.85321
race	0.85196 / 0.85370
sex	0.84041 / 0.83490
capital gain	0.83894 / 0.83208
capital loss	0.84975 / 0.84657
hours per week	0.85319 / 0.84903
native country	0.82383 / 0.82557

可以發現 capital gain 和 native country 影響都甚大，尤其是 native country 直接下降了 3%，推測是因為各國物價不同造成的差異甚大。