學號：B04507025系級：電機四 姓名：韓秉勳

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

   兩者都取所有feature的一次項：

	public score	private score
logistic	0.85184	0.85321
generative	0.84643	0.84105

logistic有較好準確率，因generative由高斯近似，彈性不夠，反之logistic有足夠彈性逼近真實函數。

2. 請說明你實作的best model，其訓練方式和準確率為何？

   我架了一個簡單的dnn，feature全取再加age, capital gain , white取平方項，同時將含？的特徵去除，並用64-32-2的hidden layer train54個epoch, 另外也取5個一樣方式train的model ensemble。最後public 0.85859 private 0.85758，並未達到private strong，不過已算是我最好的結果。

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

   以下取logistic model，並將age, capital gain , white取平方項，同時將含？的特徵去除，iterate 100000次來做比較(同best model的取法)：

	public score	private score
not normalized	0.82346	0.81930
normalized	0.85638	0.85530

可看出normalized 後的準確率有明顯提高，因為有許多連續特徵數值遠大於1，若沒有normalize 某些軸的gradient變化也較大，無法降到較好的結果。

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。

   與上題取一樣feature的model:

	public score	private score
no regularization	0.85638	0.85530
λ=0.001	0.85687	0.85566
λ=0.01	0.85773	0.84891
λ=0.1	0.78046	0.77361

λ=1	0.77616	0.76956

可以看到public 0.01的regularization 分數最好，但private的表現似乎呈現相反關係，整體來說並沒有比較好的結果，可能因logistic不夠複雜不易overfitting，用regularize太多反而是underfitting。

5. 請討論你認為哪個attribute 對結果影響最大？

age 跟capital gain都是weight佔比較大的attribute,此兩項在現實生活跟薪水應該也是正相關，另外觀察膚色可能也有影響，固有加入white的選項作為二次，但若單取此三feature效果也是不太好。