

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

答：

	public	private	average
Logistic regression	0.85798	0.85210	0.85504
Generative model	0.85454	0.85222	0.85338

Logistic regression 做出來的結果稍微好一點，不過差別並不大

2.請說明你實作的 best model，其訓練方式和準確率為何？

答：

Best model 是使用 logistic regression，參考 sample code 作 normalization、validation set，並將 fmlwgt 刪掉然後全部加上平方項，所以總共有 210 項，validation set 大概切全部的 20~25%左右會有比較好的結果，並且 train 的次數太多也容易造成 over fitting，這樣做出來的準確率大概是 85.5%

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：

第一題作的兩個都已經有作 normalization，沒做 normalization 的情況下 logistic regression 做出來的準確率大概只有 78%，而 generative model 大概在 80%，並且在 train 的時候沒做 normalization 的數值會變化很大。

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：

作 regularization 之後準確率大概是 85.2%，比原本還稍低一點(取 $\lambda = 0.00001$)， λ 如果調大準確率會下降很多

5.請討論你認為哪個 attribute 對結果影響最大？

age	0.851605
sex	0.85332
capital_gain	0.84239
capital_loss	0.84879
hours	0.852404
Federal-gov	0.854185
Local-gov	0.855045
Never-worked	0.855414
Private	0.855046
Self-emp-inc	0.855537

Self-emp-not-inc	0.856519
State-gov	0.852834
Without-pay	0.85695

上表將這些 **feature** 每次拿掉其中一項之後做出來的準確度，可以發現拿掉 `capital_gain` 這項的準確率會最低，所以推測 `capital_gain` 對結果的影響最大